INTERNATIONAL CONFERENCE ON

ADVANCED TECHNOLOGIES, COMPUTER ENGINEERING AND SCIENCE

Safranbolu, TURKEY

11-13 MAY 2018

2018 PROCEEDING BOOK

2018

PROCEEDING BOOK

ICATCES 2018

International Conference on Advanced Technologies, Computer Engineering and Science

11-13 May 2018 / Safranbolu, Turkey

Organized by

Karabuk University

Proceeding Book of the International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES 2018)

Editors

Prof. Dr. Mehmet AKBABA

Assoc. Prof. Dr. Oğuz FINDIK

Asst. Prof. Dr. Emrullah SONUÇ

Asst. Prof. Dr. Ümit ATİLA

Asst. Prof. Dr. Burhan SELÇUK

Res. Asst. Zeynep ÖZER

Res. Asst. Ayşe Nur ALTINTAŞ

Res. Asst. Yusuf Yargı BAYDİLLİ

Published, 2018

http://icatces.org/

icatces@karabuk.edu.tr

SCIENTIFIC COMMITTEE

Prof. Dr. Ahmet ARSLAN (Konya Gıda ve Tarım University)

Prof. Dr. Ali Karcı (İnönü University)

Prof. Dr. Basel Mahafzah (The University of Jordan)

Prof. Dr. Bilge DEMİR(Karabük University)

Prof. Dr. Cemil ÖZ (Sakarya Üniversitesi)

Prof. Dr. Derviş KARABOĞA (Erciyes University)

Prof. Dr. Erdal Çelik (Dokuz Eylül University)

Prof. Dr. Erkan ÜLKER (Selcuk University)

Prof. Dr. Fatih Vehbi ÇELEBİ (Yıldırım Beyazıt University)

Prof. Dr. Filiz ERSÖZ (Karabük University)

Prof. Dr. Ghulam Ali Mallah (Shah Abdul Latif University)

Prof. Dr. Haldun GÖKTAŞ (Yıldırım Beyazıt University)

Prof. Dr. Halil İbrahim BÜLBÜL (Gazi University)

Prof. Dr. Hamiyet Şahin KOL(Karabük University)

Prof. Dr. Harun UĞUZ (Selcuk University)

Prof. Dr. Hayrettin AHLATCI (Karabük University)

Prof. Dr. İdris KABALCI (Karabük University)

Prof. Dr. İhsan ULUER (Karabük University)

Prof. Dr. İsmail Rakıp KARAŞ (Karabük University)

Prof. Dr. Kerim ÇETİNKAYA (Karabük University)

Prof. Dr. Mario KOEPPEN (Kyushu Instutite Of Technology)

Prof. Dr. Mehmet AKBABA (Karabük University)

Prof. Dr. Mehmet ÖZALP (Karabük University)

Prof. Dr. Mehmet ÖZKAYMAK (Karabük University)

Prof. Dr. Mustafa YAŞAR (Karabük University)

Prof. Dr. Mykola S. Nikitchenko (Taras Shevchenko National University of Kyiv)

Prof. Dr. Nizamettin KAHRAMAN (Karabük University)

Prof. Dr. Nurhan KARABOĞA (Erciyes University)

Prof. Dr. Oleksandr I. Provotar (Taras Shevchenko National University of Kyiv)

Prof. Dr. Oleksandr O. Marchenko (Taras Shevchenko National University of Kyiv)

Prof. Dr. Raif BAYIR (Karabük University)

Prof. Dr. Sergiy D. Pogorilyy (Taras Shevchenko National University of Kyiv)

Prof. Dr. Serhii L. Kryvyi (Taras Shevchenko National University of Kyiv)

Prof. Dr. Vasyl M. Tereshchenko (Taras Shevchenko National University of Kyiv)

Prof. Dr. Valentina Emilia BALAS (University Aurel Vlaicu)

Prof. Dr. Yaşar BECERİKLİ (Kocaeli University)

Assoc. Prof. Dr. Abdrakhmanov RUSTAM (Ahmet Yesevi University)

Assoc. Prof. Dr. Amirtayev KANAT (Ahmet Yesevi University)

Assoc. Prof. Dr. Bilal Alataş (Fırat University)

Assoc. Prof. Dr. Ergin YILMAZ (Bülent Ecevit University)

Assoc. Prof. Dr. Ivan Izonin (Lviv Polytechnic National University)

Assoc. Prof. Dr. İsmail BABAOĞLU (Selcuk University)

Assoc. Prof. Dr. Kemal POLAT (Abant Izzet Baysal University)

Assoc. Prof. Dr. Mustafa Servet KIRAN (Selcuk University)

Assoc. Prof. Dr. Necaattin Barışçı (Gazi University)

Assoc. Prof. Dr. Oleksii I. Chentsov (Taras Shevchenko National University of Kyiv)

Assoc. Prof. Dr. Rabie A. RAMADAN (Cairo University)

Assoc. Prof. Dr. Tamila ANUTGAN (Karabük University)

Assoc. Prof. Dr. Taras V. Panchenko (Taras Shevchenko National University of Kyiv)

Assoc. Prof. Dr. Tufan TURACI (Karabük University)

Assoc. Prof. Dr. Tulep ABDIMUHAN (Ahmet Yesevi University)

Assoc. Prof. Dr. Yuliya Kozina (Odessa National Politechnic University)

Asst. Prof. Dr. Abdullah Cemil İLÇE (Abant Izzet Baysal University)

Asst. Prof. Dr. Adil HÜSEYİN (Karabük University)

Asst. Prof. Dr. Ahmet BABALIK (Selcuk University)

Asst. Prof. Dr. Berk Anbaroğlu (Hacettepe University)

Asst. Prof. Dr. Bilal Babayiğit (Erciyes University)

Asst. Prof. Dr. Burhan SELÇUK (Karabük University)

Asst. Prof. Dr. Caner ÖZCAN (Karabük University)

Asst. Prof. Dr. Çağrı SEL (Karabük University)

Asst. Prof. Dr. Ebubekir YAŞAR (Gaziosmanpaşa University)

Asst. Prof. Dr. Erkan DUMAN (Fırat University)

Asst. Prof. Dr. Fuat ŞİMŞİR (Karabük University)

Asst. Prof. Dr. Hakkı SOY (Necmeddin Erbakan University)

Asst. Prof. Dr. Hannah INBARAN (Periyar University)

Asst. Prof. Dr. İbrahim ÇAYIROĞLU (Karabük University)

Asst. Prof. Dr. İlhami Muharrem ORAK (Karabük University)

Asst. Prof. Dr. İlker TÜRKER (Karabük University)

Asst. Prof. Dr. İlker YILDIZ (Abant İzzet Baysal University)

Asst. Prof. Dr. İnan KESKİN (Karabük University)

Asst. Prof. Dr. Mehmet ŞİMŞEK (Duzce University)

Asst. Prof. Dr. Mesut GÜNDÜZ (Selcuk University)

Asst. Prof. Dr. Muharrem DÜĞENCİ (Karabük University)

Asst. Prof. Dr. Nesrin AYDIN ATASOY (Karabük University)

Asst. Prof. Dr. Nizar BANU (B S Abdur Rahman University)

Asst. Prof. Dr. Nursel YALÇIN(Gazi University)

Asst. Prof. Dr. Oktay AYTAR (Abant Izzet Baysal University)

Asst. Prof. Dr. Ömer Kaan BAYKAN (Selçuk University)

Asst. Prof. Dr. Ömer Muhammet Soysal (Louisiana State University)

Asst. Prof. Dr. Şafak BAYIR (Karabük University)

Asst. Prof. Dr. Şafak KAYIKÇI (Abant İzzet Baysal University)

Asst. Prof. Dr. Tuğba TUNACAN (Abant İzzet Baysal University)

Asst. Prof. Dr. Yasin ORTAKÇI (Karabük University)

Asst. Prof. Dr. Yüksel ÇELİK (Karabük University)

Asst. Prof. Dr. Zafer ALBAYRAK (Karabük University)

Dr. Adem DALCALI (Karabük University)

Dr. Ahmad Taher AZAR (Benha University)

Dr. Firdovsi A. Sharifov (V.M. Glushkov Institute of Cybernetics of NAS of Ukraine)

Dr. Kasım ÖZACAR (Karabük University)

Dr. Nilanjan DEY (Techno India College of Technology)

ORGANIZATION COMMITTEE

Honorary Committee
Prof. Dr. Refik Polat, Karabük University, Rector

Chair
Prof. Dr. Mehmet AKBABA, Karabük University

Assoc. Prof. Dr. Oğuz FINDIK, Karabük University

Co-Chair
Asst. Prof. Dr. Emrullah SONUÇ, Karabük University

Asst. Prof. Dr. Dmytro TERLETSKYI, Taras Shevchenko National University of Kyiv

Layout Editor
Res. Asst. Zeynep ÖZER

Organization Committee
Prof. Dr. Ali GÜNGÖR, Karabük University

Prof. Dr. İhsan ULUER, Karabük University

Prof. Dr. İsmail Rakıp KARAŞ, Karabük University

Prof. Dr. Mehmet Akbaba, Karabük University

Prof. Dr. Mehmet ÖZALP, Karabük University

Assoc. Prof. Dr. Oğuz FINDIK, Karabük University

Asst. Prof. Dr. Burhan SELÇUK, Karabük University

Asst. Prof. Dr. Hakan KUTUCU, Karabük University

Asst. Prof. Dr. Nesrin AYDIN ATASOY, Karabük University

Asst. Prof. Dr. Ümit ATİLA, Karabük University

Asst. Prof. Dr. Zafer ALBAYRAK, Karabük University

Secretary

Res. Asst. Dr. Ferhat ATASOY, Karabük University

Res. Asst. Ayşe Nur ALTINTAŞ, Karabük University

Res. Asst. Berna GÜNEŞ, Karabük University

Res. Asst. Elif KABULLAR, Karabük University

Res. Asst. Furkan SABAZ, Karabük University

Res. Asst. İdris KAHRAMAN, Karabük University

Res. Asst. Mehmet Zahid YILDIRIM, Karabük University

Res. Asst. M.Selman GÖKMEN, Karabük University

Res. Asst. Oğuzhan MENEMENCİOĞLU, Karabük University

Res. Asst. Rafet DURGUT, Karabük University

Res. Asst. Said DEMİR, Karabük University

Res. Asst. Yasemin SANDAL, Karabük University

Res. Asst. Yusuf Yargı BAYDİLLİ, Karabük University

Res. Asst. Zeynep ÖZER, Karabük University

Welcome to ICATCES 2018

It is a pleasure for us to offer you Abstracts Book for the 1st International Conference on Advanced Technologies, Computer Engineering and Science ICATCES'18. Our goal was to bring together leading academic scientists, researchers and research scholars to exchange and share their experiences and present their latest research results, ideas, developments, and applications about all aspects of advanced technologies, computer engineering and science. We decided to organize this event with the encouragement of our colleagues in the hope of transforming the event into a symposium series. Now, ICATCES'18 is honored by the presence of over 200 colleagues from various countries. Our warmest thanks go to all invited speakers, authors, and contributors of ICATCES'18 for accepting our invitation. We hope that you enjoy the symposium and look forward to meeting you again in one of the forthcoming ICATCES'19 event.

Best regards,

Chairman of Conference

CONTENTS

Measurement of Customer Satisfaction through Emotion Analysis in the Banking Sector 1
Video Object Tracking with PMHT 5
Designing Autonomous Landing System for Rotary Wing UAVs 10
Vegetation Extraction from Digital Orthophoto Maps Using Object-Based Segmentation and Decision Tree Classifier 14
A New Algorithm for Shape Estimation with a Low Cost Sensor 18
Fast Quadratic-Linear Approximated L1-norm for SAR Image Despeckling 23
MOTSA: A Multi-Objective Tree-Seed Algorithm 28
Solving of constrained problems via multi-objective vortex search algorithm 32
A Clustering Ranking Based Multiobjective Evolutionary Algorithm 36
A Multiobjective Evolutionary Algorithm Approach to Employee Bus Transportation Problem 38
Performance analysis of Galactic Swarm Optimization with Tree Seed Algorithm 40
The Binary Salp Swarm Algorithm with Using Transfer Function 45
Monogamous Crab Mating Optimization Algorithm for Solving Vehicle Routing Problem 49
A Kinect 2 Based Telerehabilitation Method for Shoulder Rehabilitation Exercises 53
Tournament Selection based Antlion Optimization Algorithm for Solving Quadratic Assignment Problem 57
Parallel Machine Scheduling using Improved Antlion Optimization Algorithm 63
Robot Path Planning using Gray Wolf Optimizer 69
Training Multi-Layer Perceptron using Opposition based Learning Spiral Optimization Algorithm 75
Kinect Based English Teaching Game for Preschool Children 80
WALRUS: A Retro Communication Gadget Based on Internet of Things Technologies 85
Performance Tuning in Database Systems in High Availability Architecture and Reducing Query Costs Strategies in Oracle Database Management System 89
Lifestyle Change Treatment with Cloud Based Mobile Tracking Application: VITAE 96
Research of the Traffic Flow on the Signal-controlled Intersections using Queuing Theory 100
Real-Time Monitoring and Control of The SDLC Process on a Single Automation in Core Banking Applications 104
Forensic Analysis of Persistent Data Storages Analyzing NTFS Formatted Drives 109
EDU-VOTING: An Educational Homomorphic e-Voting System 114
Design and Implementation of Greenhouse Automation System with Matlab GUI 120
Gray Image Enhancement with Regional Similarity Transformation Function (RSTF) 122
Trends in Cloud-Based Learning Management Systems 125
A Usability Analysis of Edmodo Learning Management System: The Case of a University 127
MultiMedia Application and IPv6 Addressing in Soft Switches 129
Trellis Combining Approach for Demodulate and Forward Based Multiple Access Relay Channels 135
A Review: Mobile Communication Past, Present Future 141
Effects of Sandstorms on Vehicular-to-Road Visible-Light Communication 146
Solving University Course Timetabling Problem Using Ant Colony Optimization: An Example of Mersin University Engineering Faculty 154
Determine the Effect of Genetic Algorithm Performance Parameters in Single-Machine Scheduling Problem 158
Artificial Atom Algorithm in Image Processing Applications 164
Improved Vortex Search Algorithm for Single Objective Optimization Problems 168
Car Model Categorization with Different Kind of Deep Learning Convolutional Neural Network Models 174
Increasing the Performance of SAR Image Despeckling Using Convolutional Neural Networks 179

Evaluation of Object Tracking Performance of ADNet Method with Different Data Sets and Color Spaces 182
A Study on Prediction Success of Machine Learning Algorithms for Wart Treatment....186
Turkish Sign Language Alphabet Recognition with Leap Motion....189
A Review on Web Crawlers and Ontology-Based Crawlers 193
Performance Comparison of Machine Learning Methods for Solving Handwriting Character Recognition Problem 197
Artificial Neural Networks Approach to Greenhouse Heating Requirement Estimation 203
Speech and Eye-Gaze-Based PC Control System for Disabled People....208
A Simple Heuristic Approach to Improve Performance of Extreme Learning Machine 212
PI and Self-Tuning PI Controller Design and Comparison for Speed Control of DC Motor....218
Cardiotocography Data Set Classification with Extreme Learning Machine 224
Movie Rating Prediction with Machine Learning Algorithms on IMDB Data Set....231
An Algorithm of Finding Extreme Points of Curves 236
Choosing optimal truncation level on estimation of the nonparametric regression with Padé approximation....239
Elongation of Partially-Thermally Insulated on the Lateral Surface of the Rod Under the Influence of Temperature, Heat Flow, Heat Transfer and Tensile Force....241
Impedance Control Of 2dof Serial Robot Manipulator....244
A Review on Measurement Methods of Non-Invasive Blood Glucose Level 250
A Survey on Predicting Survivability of Retinoblastoma on SEER Data....257
Design of A New Robot Manipulator for MALDI Sprayer....262
Investigation of Human Femoral Head by Wide Angle X-Ray Scattering (WAXS) Measurements....265
Design of A Triple Band Absorber Metamaterial for C-Band Radars with Golden Ratio Approaches....268
Antenna Selection Techniques for Digital Relaying based Cooperative MIMO Channels....271
Comparative Analysis of the Distributed Energy Resources Connections to Distribution Network....276
3D Visualization Thyroid CT Images Using Marching Cubes Algorithms....282
Design and Static Stress Analysis of Lifting Hook....287
A Voice Encryption Application Based on a Chaotic System with Single Parameter....290
A Pseudo Random Number Generator Design Based on a Four Dimension Chaotic System....293
An Automated Vulnerable Website Penetration 297
RC4 Stream Cipher Based Digital Color Image Encryption Using Chaotic Systems....302
Utilizing RFM Analysis and Apriori Methods on Turkey State Supply Office Data for Regaining Lost Customers....306
Blockchain: A Decentralized Approach to Big Data 310
Application of PageRank Algorithm in Linked Data....314
An Analysis of Rumor Spreading Fundamentals with a Case Study on Facebook 318
Design and Fabrication of High Gain Ultra-Wideband Antenna for Microwave Imaging and Radar Applications 325
VisionDetector: A Low-Cost Mobile Visual Impairment Detection System....331
Simulation Measurements of Electromagnetic Field Values for Detection of the Breast Cancer 336
A Study on BCI Speller Design and Analysis of Signal Window Length....340
Dynamic Programming Approach to Selling and Buying of Stocks in XU030 Index of BIST by Forecasting Stock Prices for Next Five Years....344
Software Development for Robotic Competition Eliminations Using Expert Systems....348
Technology Evaluation with HUP-Growth Mining Algorithm 352
A Real-time and Secure Patient Monitoring System (RSPMS)....355
Decreasing Failure and Turnover Rates In Double-Glazing Production With Failure Mode And Effect Analysis and An Application....360

A Literature Review on RFID Applications in Advanced Manufacturing Systems 365
A Simulation Approach of Revenue Maximization Strategies for Turkish Domestic Airlines 369
The Sustainability Indicators in Supplier Selection: The Furniture Industry 374
Evaluating Research Performance of the European Countries Through Social Network Analysis 377
Complex Network Analysis of Players in Tennis Tournaments 383
Complex Network Analysis of UEFA Europe League Competitions 389
Implementation of Clarke & Wright Savings Algorithm in Social Network Analysis: An Exemplary Event Planning System 394
Analysis of Transportability of Causal Effects in Wireless Sensor Networks 399
Three Axis Gimbal Design and Its Application 405
Training ANFIS using The Whale Optimization Algorithm 409
Usage of the A* Algorithm to Find the Shortest Path in Transportation Systems 415
Feature Selection for Gender Classification in TUIK Life Satisfaction Survey 418
Automatic Author Detection in Turkish Books Using N-Gram and Naïve Bayesian Approach 424
The Effect of Over-sampling and Under-sampling Techniques in Medical Datasets 429
Analysis of the Co-authorship Network of Turkish Engineering Research Society 433
The Effects of Ransom Software on IoT (Internet of Things) Systems 437
A Fast and Simple Computer Aided Lighting Simulator 442
A Survey of Uncertainties in MAPE-K Control Loop 447
Design and Control of Excavator Type Work Machine Simulator 453
Low Cost and Practical Data Acquisition System Using Labview: An Application 457
Implementation of NURBS curves on the LCD touch screen using FPGA 461
Intrusion Detection with Probabilistic Neural Network: Comparative Analysis 467
How to manage Software Architecture documentation in Scrum Framework 471
Software Architecture Documentation in Agile 476
Tracking the Architectural Quality: “W-Model of Software Architecture” 479
An Interpretation System from Turkish to Turkish Sign Language 483
Performance Evaluation of kNN, Support Vector Machines and Artificial Neural Network on Optical Character Recognition 485
Comparison of Classification Algorithms on NFC-Based Public Transport Data 487
An Efficient Human Action Recognition Framework with Pose-based Spatiotemporal Features 492
Real Time Drowsiness Driver Detection and Low Cost Warning System 498
Local Variance Switching Gaussian Filter 502
Image Compression Using SVD Method 508
Online Mine Detection Experiment with an Unmanned Underwater Vehicle 512
The Autonomous Yaw and Depth Controller Experiment of Unmanned Underwater Vehicle 517
Information Security Risk Assessment using Bayesian Network and Fuzzy Inference System: A Case Study 522
Increasing the security of Mobile Communication with Steganography 530
Rule-Based Performance Measurement in Open Source IDS Systems 535
Fusion of Full-Reference and No-Reference Anti-Spoofing Techniques for Ear Biometrics under Print Attacks 538
Data Security on Virtual Private Networks 543
Creating of Probability Maps of Earthquake Occurrences Using Kriging Method with the Geographic Information Systems (GIS): Estimates for 3 Section of the NAFZ (Western, Central, Eastern)-Part 2 547
A Study on Obtaining Rectified Photographs for Architectural Applications 550

Prediction of Aboveground Carbon Storage in Forest Areas Using Remote Sensing Data....556
Relationship with GIS of Neighborhood Features belong to Konya....560
Real Estate Valuation Using Artificial Neural Networks Method....566
The Effects of Urban Transformation on Real Estates and Land Valuation at City Plans....571
Application Artificial Neural Network in Mass Real Estate Appraisal for Centre Neighborhood of Konya....575
Comparison of Pixel Based and Object Based Classification Methods on Wetland Areas: Example of Aslantaş Dam Lake....579
Artificial Neural Network Model Design for Daily Demand Prediction in Mass Meal Production....583
Determination of Coastline Changes at Kozan Dam Lake by Using Artificial Neural Networks Method....587
Particle Swarm Optimization Based Determination of Learning Parameters in Artificial Neural Networks with Backpropagation Learning Algorithm....592
Bug Localization by Using Information Retrieval and Machine Learning Algorithms....598
An Overview for the National Cyber Security Strategy....603
Comparison of Turkey ad European Union Computer Engineering Programs....610
A Quality Model for Evaluating Maintainability of Object-Oriented Software Systems....614
An Approximation of The Voronoi Diagram for A Set of Arcs....620
The Distance Effect in the Dosimetry Analysis of a Rat Model at GSM-900 Frequency Band: A Simulation Study....626
Modeling and Simulation of Complex Mechanical Systems Using Electrical Circuit Analog....630
3-D Modeling and Analysis of Shaded Pole Motors Using Finite Elements Method....635
An Intelligent Material Placement for Electrical Installation Project....639
Exposure Analysis of a Human Body due to Underground Power Cables and Magnetic Field Mitigation....642
Solving the Traveling Salesman Problem Using Parallelized Artificial Bee Colony Algorithm....646
An Artificial Bee Colony Algorithm and Its Application to Travelling Salesman Problems: Reverse Logistics Optimisation for Accumulator Recycling Companies....650
Selection of Facial Features using Genetic Algorithm under Different Illumination Conditions and Occlusions....652
Constrained Optimization Problems Solution with Salp Swarm Algorithm and Ant Lion Optimization....656
Performance Evaluation of Various Binary Variants of ABC Algorithm for Solving Knapsack Problem....661
An In-Vivo Study of Human Tibiofemoral Joint Kinematics by Using Dual Fluoroscopy System....666
Detection of Light Sleep Periods Using an Accelerometer Based Alarm System....672
Splitting Tensile Strength of Concrete Containing Zeolite and Diatomite under the Effect of H2SO4 by ANN....677
Predicting the Compressive Strength of Concrete Containing Zeolite under the Effect of H2SO4 by ANFIS....685
Investigation of the Effect of Mesh Density and Element Type on Behavior of Biphasic Soft Tissues in Finite Element Analysis....693
Application of Artificial Intelligence Methods in Software Testing....698
An Interactive Learning Method Based on Agent Systems....700
Occupational Health and Safety in the Chemical Industry; Modeling and Software Developed on the Basis of Fuzzy Logic to Prevent Job Accidents....704
Occupational Health and Safety in the Metal Industry; Modeling and Software Developed on The Basis of Fuzzy Logic to Prevent Job Accidents....714
Detection of DDOS Attacks in Network Traffic Using Deep Learning....722
Data-Driven Estimation of Direction of Gravity from a Single Image....727
Live Target Detection with Deep Learning Neural Network and Unmanned Aerial Vehicle on Android Mobile Device....731
Deep Learning Based Vehicle Detection on Cross-Roads....737
An Empirical Comparison of Data Mining Tools and Migrating Birds Optimization Algorithm on Medical Diagnosis....741
A New IPv6 Addressing Strategy to Mitigate Reconnaissance Attacks....745
An Application of Temperature and Relative Humidity Data Obtaining by RF Communication....750

Optimization of Sensor Deployment for k-coverage in Wireless Sensor Networks....755
Cost-effective logging using SDN architecture....761
Optimizing Cellular Networks for Adaptive Video Streaming....765
Detection of EEG-Based Motor Imagery Tasks with 1D-Local Binary Pattern (LBP) Features....769
Development of A Multi-Objective Optimization Via Simulation Approach for Inventory Control Systems and Supplier Selection....773
Smart Traffic Signal and Routing System for Emergency Vehicles....779
Modeling of Voltage Sag/Swell Disturbances in Distributed Generation Systems....781
Efficient Design and Comparative Performance Analysis of PID Controller Applied to Automatic Voltage Regulator Employing Symbiotic Organisms Search Algorithm....785
Kinect Calibration and Data Optimization for Anthropometric Parameters....791
Some Experimental Studies of Segmentation of Different Medical Images using Color Difference on CIE L*a*b* Color Space....797
A Mobile Low-Cost Fire Detection System with Infrared Camera....802
Detecting Anomalies in Surveillance Videos with Spatio-Temporal Features....806
Automatic Segmentation and Labelling of 3D Human Activities....809
Understanding effects of hyper-parameters on learning: A comparative analysis....814
Artificial Bee Colony Algorithm for The Linear Ordering Problem....818
Systematic Literature Review on Security Vulnerabilities and Attack Methods in Web Services....821
Road Extraction Techniques from Remote Sensing Images: A Review....826
Delaunay Triangulation and Its Applications....830
Hand Gesture Recognition with One-Shot-Learning....833
Diacritic Restoration of Turkish Tweets with word2vec....839
Performance Evaluation of Bicycle Sharing System in Urban Transportation....843
Parallel Artificial Atom Algorithm for Large Scale Travelling Salesman Problem....844

Measurement of Customer Satisfaction through Emotion Analysis in the Banking Sector

B. KARADAG[1]

[1] Vakıf Participation Bank Inc. Istanbul/Turkey, bulut.karadag@vakifkatilim.com.tr

Abstract **- Customer satisfaction plays an essential role in financial institutions. Particularly in the banking sector, a series of customer satisfaction studies is carried out in order to ensure the continuity of the customers with the bank. The most important ones of these are; questionnaires, feedback from the branch staff, and the increase in the client's banking transactions. By analyzing these studies, the satisfaction level of the customers can be measured. However, these measurements do not have a complete mathematical accuracy and can be misleading. In this study; the facial image of the customers who come to the bank are perceived by the camera and evaluated by emotional analysis. In this way, it is determined that the customer is satisfied or non-satisfied when leaving the bank. Problematic banking applications are also identified with the help of the analytical result of emotional analysis. In addition to this, the branch personnel of the bank who behave well to the customers are determined and performance evaluation is made more accurately.**

Keywords - Emotion detection, customer satisfaction, emotional analysis, sentiment analysis, facial recognition

I. INTRODUCTION

Banks are financial institutions within the service industry in terms of their structures [1]. In today's competitive market, positive relationships between banks and customers are a dominant factor for business. In the banking sector it is generally adopted as a customer-centric approach. Therefore, customer satisfaction is considered as the unique factor on providing the best service [2]. Customer's satisfaction related to product or service affects the positive and negative perceptions of the product or service. For this reason, banks have to learn their feedback of their product or service satisfaction in order to make their customers more loyal [3]. Taking decisions in the direction of customer satisfaction is only possible by having information about customer behaviors. Customers' satisfaction depends on that they compare the service we give them with the expectations they have. The customer is satisfied when the customers' expectations are fulfilled. Even if the service is above the expectations, the customer is highly satisfied [4].

Banks try different methods to retain available customers and to gain new customers. In addition, banks attempt to increase the number of products available customers use and to provide in-house customers increase their usage of other services. The main purpose of these works is to provide customer satisfaction. Banks gain the loyalty of customers by ensuring the continuity of customer satisfaction. The most effective and correct way to increase customer loyalty is to provide quality service to customers [1]. The importance of customer satisfaction cannot be denied for any business. Because it is much more economical to keep existing clients in the long term than to find new customers [5].

Due to the effect of the increasing rivalry, a great intensify is experienced in the usage of technology related to our world's banking in recent years. Banks rapidly are trying to enrich their classical technological products such as ATM, POS, telephone and computer banking with new products and new service concepts. In this respect, the concept of electronic banking in the vision of all banks is in the top of the list. Banks are introducing new applications such as "Call Centers", "Internet Banking", "Mobile Banking" and "Customer Relationship Management" in order to provide better services to their customers and to provide 24-hour service. If the feedbacks from all these services are positive, we think that the customers are satisfied.

Table 1: Causes of participants' bank preferences

Causes	First Important Cause		Second Important Cause		Third Important Cause	
	frequency	%	frequency	%	frequency	%
Reliability	**81**	**61,4**	3	2,3	2	1,5
Physical properties	5	3,8	7	5,3	1	0,8
Compliance with credit conditions	9	6,8	17	12,9	8	6,1
Interest, welcome and respect	8	6,1	**28**	**21,2**	13	9,8
Easy access	4	3,0	27	20,5	10	7,6
Banking operations	4	3,0	13	9,8	5	3,8
Credit card	0	0	8	6,1	15	11,4
Internet banking	2	1,5	8	6,1	8	6,1
Call center	0	0	3	2,3	5	3,8
Service variety	1	0,8	5	3,8	21	15,9
Deposit high interest rate	1	0,8	2	1,5	5	3,8
Quick result of the process	7	5,3	1	0,8	**25**	**18,9**
Other	2	1,5	0	0	0	0
Total	124	93,9	122	92,4	118	89,4
No reply	8	6,1	10	7,6	14	10,6
Grand total	132	100,0	132	100	132	100,0

As seen in Table 1, participants indicated reliability (61.4%) as one of the three most important reasons for choosing a bank. The second important reason was interest, welcome and respect (21.2%) and the third one was the quick result of the process (18.9%) [6].

In this study, we will try to determine customer satisfaction realistically through emotional analysis in order to provide

customer satisfaction which is the main aim of banks. With the help of the Microsoft Cognitive Services Face API, the emotional state of the customer being at the bank is detected with certain periods. Then, by reporting, it will be determined which banking transactions have problems and which staff are more favorable to the customers. In this way, the 2nd and 3rd important causes mentioned in Table 1 have been ensured.

II. EMOTION TAXONOMY

According to the emotion theorists and psychologists, various of emotions can be categorized by starting from globally showed six fundamental emotions to complicated emotions which are originated from different culture with. The following theories are mainly used in emotion analysis studies; Ekman and Friesen's List of Basic Emotions (1971), Plutchik's Wheel of Emotions (1980) and Parrots' Classification of Emotions (2001). Ekman and Friesen in 1971[7] put forward six quintessential basic emotions like disgust, joy, sadness, fear, anger and surprise which are globally presented and identified from facial expressions. Microsoft Cognitive Emotion API results give us Ekman and Friesen's fundamental set of emotions. Therefore, customers will be evaluated through their list of basic emotions.

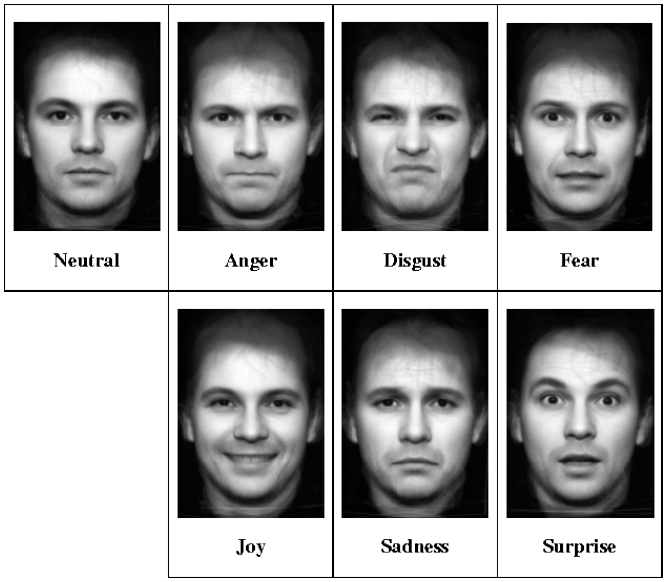


Figure 1: Ekman and Friesen's List of Basic Emotions.

III. IMPLEMENTATION

The emotional analysis of the customer is made with a structure that can be integrated into the banking application. The main aim for taking part in banking application is to get the transaction and employee information. In addition, there must be a camera connected to the branch staff's computer. The general model of emotion analysis integration is shown in Figure 2.

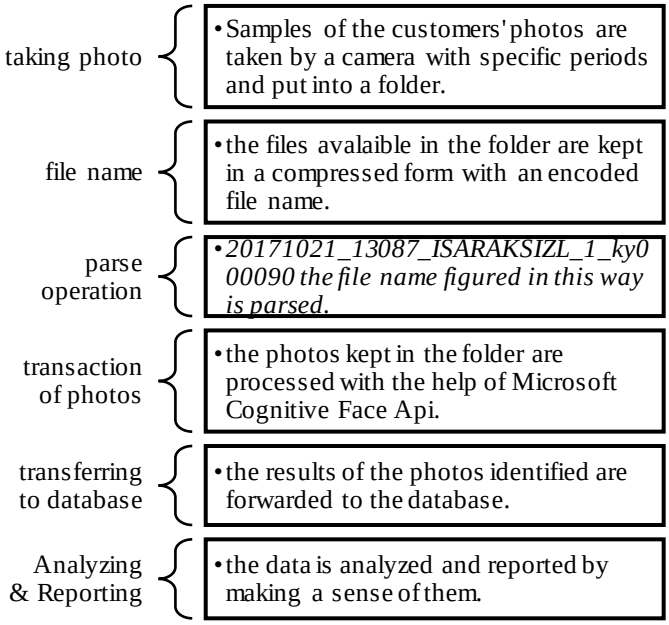


Figure 2: Generalized Model for Emotion Analysis.

A. *Taking Photo*

Samples of facial expressions are gathered in specific periods by a camera during the operation of the customer's request as seen in the Figure-3. The samples collected are put into the folder to be operated. The reason why the photos aren't processed as soon as they are taken is that the customer shouldn't lose any time while his operation is being done.

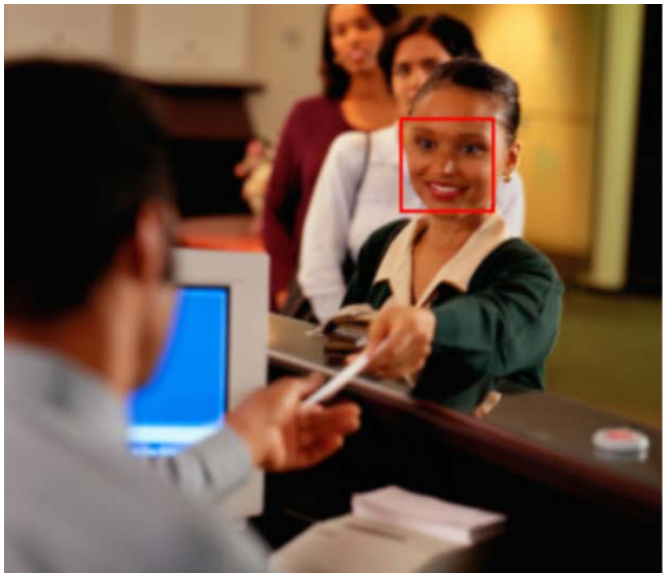


Figure 3: Taking customer's photo by a camera.

B. *File Name*

The names of the files are saved into the folders in a specific form as compressed. The example is seen in the Table-2.

Table 2: Sample of file name format

20171021_1400_ISARAKSIZL_1_bk000207.7z
20171021_13087_ISARAKSIZL_1_ky000090.7z

20171021_43679_ISARAKSIZL_1_gy000114.7z

C. Parse Operation

The samples of the photos saved are kept in zipped files. The result in the Figure-3 consists when we parse the file name. The areas of the file names are separated from each other with an underline (_). From the left, the date of the operation done is in the first tab, the client no processed in the bank is in the second tab, the screen information belonging to the process done in the banking application is in the third one, the information of the branch office in which the operation is done is in the fourth tab, the information of the staff who is carrying out the operation is in the fifth tab.

20171021_13087_ISARAKSIZL_1_ky000090.7z				
20171021 (Process Date)	*13087 (Customer Number)*	*ISARAKSIZL (Screen Code)*	*1 (Branch Code)*	*ky000090 (Personel Code)*

Figure 3: Parsing file name

D. Transaction of Photo

After the parsing, identifying the photos is started with the help of Microsoft Cognitive Face API. Firstly, to gain an API key by doing a recording on the platform of Microsoft AI + Machine Learning is needed for this operation. We use this key while processing the photo.

We take the output of the processed photo as a data set on the form of JSON. The finding of the operated photo in the Figure-1 can be observed in the Figure-4. We conclude after the process that the customer is really satisfied in the second when the sample of the customer's photo is taken and processed.

```
"emotion": {
  "anger": 0.0,
  "contempt": 0.0,
  "disgust": 0.0,
  "fear": 0.0,
  "happiness": 1.0,
  "neutral": 0.0,
  "sadness": 0.0,
  "surprise": 0.0
},
```

Figure 4: Result of the processed photo

E. Transferring the Database

We save both the result of the data processed in JSON form and the name of the zipped file in the database after parsing. A number of photos belonging to the customer are available since we take the samples of the customer's photos in specific periods.

F. Analyzing & Reporting

The outputs inset into the database are evaluated according to the processes of the branch office, the staff and the banking applications. When the records evaluated in the branch-based are analyzed, we learn that which branch is welcoming the customer in a better way for performance. In this way, we detect the behaviors' the branch staff. We also use this information in the evaluation of the staff performance. Finally, the problematic ones of the banking applications are found. For instance; imagine that it is detected some clients are not satisfied during the process of a telephone bill. In this situation, we decide there is a problem in the integration of the telephone bill process and all the actions are carried out by the Unit of Information Technology to resolve the problem.

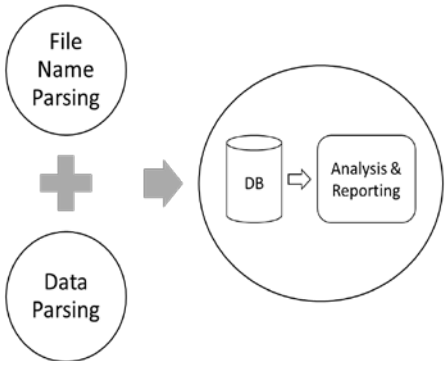


Figure 5: All transaction process of the data

IV. RESULT & ANALYSIS

After the images of customer are processed, the result in Table 3 occurs. In this way, the transaction-based results of the customer can be examined. When we review it in general, we find that the customer is satisfied according to the result. Any anger did not occur.

Table 3: Result of a customer's process

EmotionId	ProcessDate	BranchId	CustomerId	UserCode	ScreenCode	Anger	Contempt	Disgust	Fear	Happiness	Neutral	Sadness	Surprise
0	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	1	0	0	0
1	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	1	0	0	0
2	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0.81	0	0	0.19	0	0
3	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	0.86	0.14	0	0
4	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	0.78	0.22	0	0
5	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	1	0	0	0
6	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	0.87	0.13	0	0
7	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0.9	0.1	0	0	0	0
8	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	0	1	0	0
9	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	0	1	0	0
10	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	1	0	0	0
11	20170206	1	13087	ky000069	ISARAKSIZL	0	0	0	0	1	0	0	0

The most important outcomes of Ekman and Friesen's List of Basic Emotions are the anger and happiness. These two emotions, give us the most definitive result for the customer.

When we examine the Figure-6, we see that the customer is leaving the bank happily. In this way, we assume that the customer will be in connection with the bank again. It is a good progress for the bank's profit.

However, if we look at the linear line of happiness while the customer is in the bank, we can see that the happiness of the customer goes down and the neutral situation tends to increase. This situation is not a problem for the bank because the neutral situation does not give us an exact result. Figure 7 and Figure 8 show the linear line of happiness and neutral status. If the customer was angry instead of neutral, then we could say that

the customer came to the bank happily, but he was angry when he left. We think that this customer is not satisfied.

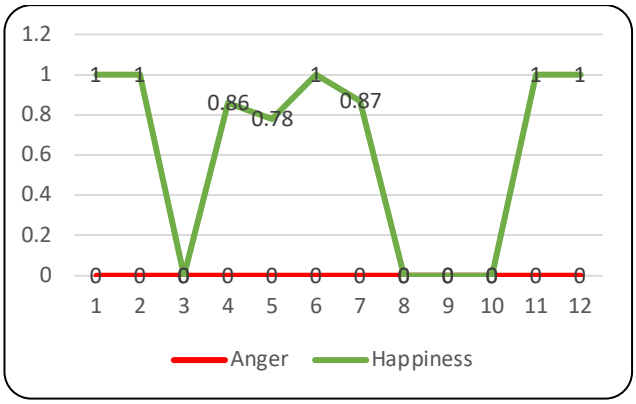


Figure 6: Anger and happiness status of a sample customer

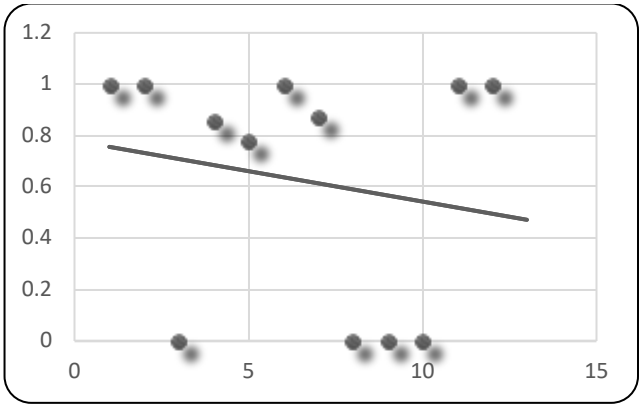


Figure 7: Linear result of the happiness status

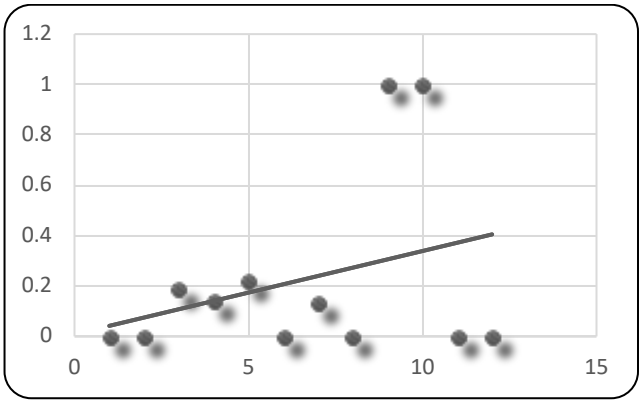


Figure 8: Linear result of the neutral status

V. CONCLUSION

In this study, we conducted emotional analysis with the aim of measuring customer satisfaction. In this way, it is aimed to increase the banking income by ensuring the continuity of the in-house customers. Additionally, problematic banking applications are identified and fixed by the Information Technology Unit. Besides, the staff who does not show good behavior towards the customer can be detected. Lastly, branches and branch staff behaving well are rewarded and rivalry can be ensured.

Today all of the existing vision systems for facial muscle action detection deal only with the frontal-view face images. It should be noted that, some people may not show their emotion and mental state by facial expression. It will be more accurate to examine their emotions through sound analysis as well as the image of these people.

REFERENCES

[1] *Bankacılık ve Sigortacılık Araştırmaları Dergisi Cilt 1 Sayı 3-4 (Ocak 2012), ss.4-15*.

[2] Tolon, M. (2004). *Ticari Bankalarda Pazarlama Stratejilerinin Uygulanması ve Türkiye'deki Ticari Bankalar Üzerine Bir Araştırma.* Verimlilik Dergisi (4), 63-88.

[3] Yılmaz, Veysel ve Çatalbaş Karpat, Gaye, (2007), "*Kredi Kartlarına İlişkin Algının Müşteri Memnuniyeti ve Sadakati Üzerine Etkisi*", Finans Politik & Ekonomik Yorumlar Dergisi, Cilt: 44 Sayı:513, s.83-94.

[4] Kotler, P., (2000) *Marketing management*. USA: Prentice Hall.

[5] MISHRA, Anubhav Anand, (2009), "*A Study on Customer Satisfaction in Indian Retail Banking*", The IUP Journal of Management Research, Vol. VIII, No. 11, p.45-61.

[6] *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi* Y.2011, C.16, S.2, s.59-77

[7] D. Anurag,S. Ashim, '*A Comparative Study on different approaches of Real Time Human Emotion Recognition based on Facial Expression Detection*', International Conference on Advances in Computer Engineering and Applications, 978-1-4673-6911-4/15/$31.00©2015 IEEE.

Video Object Tracking with PMHT

A. G. PAKFILIZ[1]

[1]Baskent University, Ankara/Turkey, apakfiliz@baskent.edu.tr

Abstract **- Target detection and tracking in video data is a complex problem. Especially for designing and developing surface to air tracking systems should be dealt with this problem strictly. In this study a combination of an image processing technique and Probabilistic Multi-Hypothesis Tracker (PMHT) is used. The target is discriminated from the video data for target initiation. In the subsequent steps the same discrimination method is used to obtain proper data for track continuation. For target identification from video data a transformation is used and target based pixels are discriminated from the background. Then image is transformed into point measurement data with amplitude information. Thus, video data is made suitable for tracking with PMHT. The resulting algorithm performs target detection and tracking operations automatically.**

Keywords **- Video target detection, tracking, PMHT.**

I. INTRODUCTION

TARGET detection and tracking automatically may be a critical problem for surface to air tracking systems. For moving object detection in video data three common methods are used. These methods are background subtraction, temporal differencing, and optical flow. Background subtraction is a commonly used technique for detecting objects of interest in static scenes [1]. Temporal differencing, or change detection based on frame difference, attempts to detect moving regions by making use of the difference of consecutive frames (two or three) in a video sequence. This method is highly adaptive to dynamic environments, but generally does a poor job of extracting the complete shapes of certain types of moving objects [2]. The other approach for moving object detection is optical flow, which is defined as a velocity field in the image which transforms one image into the next image in a sequence, and several different algorithms have been derivated by using optical flow [3]. Optical flow method is a complex algorithm, and not suitable for real-time processing.

These techniques have some disadvantages. Background subtraction method is sensitive to the changes of light. Optical flow method is a complex algorithm, and not suitable for real-time processing. Frame difference method is simple and easy to implement, but the results are not as accurate as the results of the other methods. In this study RST (reciprocal pixel intensity measurement (RPIM) transformation - Sobel- Thresholding) method [4] is applied. Using RST method target based pixels take higher intensity values, and a suitable measurement data is obtained for discriminating target from background. After performing RST transformation a new technique is used for obtaining point measurement data. For this purpose pixel intensity values are used.

For target tracking purposes, PMHT algorithm is used. Following its development by Luginbuhl & Streit [5] several years previous, there was a significant effort both to prove the PMHT's superiority in terms of traditional tracking performance, and in its modification to this end. Thus, there has been a considerable exploitation of the PMHT's flexibility: it has been extended to deal with amplitude information [6], to incorporate alternative models such as of "homothetic" measurements and to function as a natural overlay to a hidden Markov "maneuver" process [7]. In this study amplitude information (AI) added Homothetic PMHT (HPMHT) is used.

For evaluating the performance of the proposed method, surface to air video data for different scenarios is used. In these scenarios different maneuvers, flying vehicles and environmental conditions for one target are taken into account. And the results of the algorithm performance are given in the experimental study section.

II. OBTAINING MEASUREMENT DATA

In this study, video data is brought into suitable form to use in the PMHT algorithm. Initially the RST method is used, and then the obtained data is converted into point measurement. In the first step of the RST method, a color (RGB) video frame is converted to a gray scale image using the following equation [8]:

$$I(m,n) = 0.299R(m,n) + 0.587G(m,n) + 0.114B(m,n) \quad (1)$$

here; R (m; n), G (m; n) and B (m; n) are the values of the red, green and blue color bands at each pixel. Then reciprocal pixel intensity measurement (RPIM) technique [4] is applied to this data using following equation;

$$I_R(m,n) = 255 \times \frac{1}{I(m,n)} \quad (2)$$

At this point of the process classic Sobel edge detection operator given in [8] is applied in order to reduce clutter effects and increase the probability of target discrimination. Sobel operator is a discrete differentiation operator, and computes an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel operator is either the corresponding gradient vector or the norm of this vector. Sobel operator is the partial derivative of two-dimensional *f (x; y)* function as the central computing 3x3 neighborhood at x and y directions. In order to suppress noise, a certain weight is correspondingly increased on the center

point, and its digital gradient approximation equations with respect to x; y directions are given in (3) and (4), respectively.

$$G_x = \left(f\left(x+1, y-1\right) + 2f\left(x+1, y\right) + f\left(x+1, y+1\right)\right) - \ldots \\ \ldots\left(f\left(x-1, y-1\right) + 2f\left(x-1, y\right) + f\left(x-1, y+1\right)\right) \quad (3)$$

$$G_y = \left(f\left(x-1, y+1\right) + 2f\left(x, y+1\right) + f\left(x+1, y+1\right)\right) - \ldots \\ \ldots\left(f\left(x-1, y-1\right) + 2f\left(x, y-1\right) + f\left(x+1, y-1\right)\right) \quad (4)$$

The Sobel operator is the magnitude of the above gradients as follows:

$$g\left(x, y\right) = \sqrt{G_x^2 + G_y^2} \quad (5)$$

Sobel's convolution template operators are given as in equation (6):

$$T_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \qquad T_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (6)$$

Sobel operator is applied to detect the edge of the image obtained in the output of RPIM process I_R , given in equation (2). In this case the horizontal template T_x and vertical template T_y can be used to convolute with the image, without taking into account the border conditions. The same size of two gradient matrix M_1 and M_2 may be obtained as the original image. Then the total gradient value *g(x,y)* is obtained by adding the two gradient matrices. By using total gradient value Sobel operator returns the edges in higher intensity values than the noise floor. It can be concluded that there is an edge where the gradient of the image is maximum or where the intensity level changes. To prune the pixel intensities that most probably emerges from clutter rather than target edges a threshold value is used. As a result the applied process can be summarized as RPIM, Sobel operator, and Thresholding (RST) operation.

Video frames obtained using the RST method are still in the form of amplitude modulated images. However, the PMHT algorithm to be used works with point measurement. An additional process is applied to make the measurements a point measurement. To convert this data to point measurement, the RST applied video frame is divided into 5x5 pixel areas, and the average amplitude values (E{PA}) of each field is found. Then, the average amplitude value of all the pixels in the sensor area (E{SA}) is found, and the threshold value (AI_e) is selected as 10% higher than this value.

$$E\left\{SA\right\} = \sum\nolimits_{SA} I\left(x, y\right) \quad E\left\{PA\right\} = \frac{\sum_{5\times5} I\left(x, y\right)}{5\times5} \\ AI_e = 1.1\times E\left\{SA\right\} \quad (7)$$

The average amplitude value of each field is compared with the threshold value, and if it is equal to or greater than the threshold level ($E\{PA\} \geq AI_e$) a measurement point is established in the center of the field by deciding that there is a measurement. The average amplitude value of the field is taken as the amplitude value of the generated measurement point. Thus, the image data becomes point measurement with amplitude information (AI). AI is used to increase the accuracy of track continuation, especially in dense clutter environment.

III. HPMHT Algorithm with AI

A. *Standard PMHT Algorithm*

The PMHT is a true multi-target tracking algorithm derived from the application of Expectation-Maximization (EM) algorithm [5]. A fundamental difference between the PMHT and other standard tracking methods is that the PMHT assumes that the assignment indicates for each measurement are independent random variables. This presents an additional parameter for the problem, $\mathbf{\Pi}$, the probability mass function of the assignment [9].

The PMHT is an iterative algorithm: it asymptotically approaches a local maximum of the EM auxiliary function by refining estimates for the states, $\mathbf{X}$, and the parameters, $\mathbf{\Pi}$. At the ith iteration, denote the estimated states and parameters as $\mathbf{X^{(i)}}$ and $\mathbf{\Pi^{(i)}}$ respectively. The estimated states of model m at scan t are denoted $\mathbf{x_t^{m(i)}}$ and the estimated prior assignment probability for model m at scan t is $\boldsymbol{\pi}_\mathbf{t}^{\mathbf{m(i)}}$. The iterations are repeated until the auxiliary function converges (or some other halting criterion is met) at iteration i^* and the estimated track state and parameters at the last iteration comprise the PMHT state and parameter estimates:

$$\hat{x}_t^m = x_t^{m\left(i^*\right)} \qquad \hat{\pi}_t^m = \pi_t^{m\left(i^*\right)} \quad (8)$$

To start the iterative algorithm, the PMHT requires initial estimates $\mathbf{X^{(0)}}$, and $\mathbf{\Pi^{(0)}}$. These may be obtained from earlier data if it is available, or may be merely guesses. $\mathbf{\Pi^{(0)}}$ can be initialized as a uniform distribution, $\boldsymbol{\pi}_\mathbf{t}^{\mathbf{m(0)}}=M^{-1}$. The estimates $\mathbf{X^{(i+1)}}$ and $\mathbf{\Pi^{(i+1)}}$ are found by maximizing the EM auxiliary function

$$Q\left(\mathbf{X}, \mathbf{\Pi} \middle| \mathbf{X}^{(i)}, \mathbf{\Pi}^{(i)}\right) = \sum_{t=1}^{T} \sum_{r=1}^{n_t} \sum_{k_{tr}}^{M} \log P\left(\mathbf{X}, \mathbf{K}, \mathbf{Z}; \mathbf{\Pi}\right) \ldots \\ \ldots P\left(\mathbf{K} \middle| \mathbf{X}^{(i)}, \mathbf{Z}; \mathbf{\Pi}^{(i)}\right) \quad (9)$$

with respect to $\mathbf{X}$ and $\mathbf{\Pi}$. $\mathbf{K}$ represents measurement to track assignment vector. At the end of the maximizing process posterior probabilities are obtained. The derivation of the intermediate steps of the maximizing process can be found in many papers such as [5,9]. Here, only the posterior probability, synthetic measurement and its covariance formulations will be given, as these are what change in the HPMHT. $w^n{}_{kr(t),r}$ is defined as the posterior probability (conditioned on measurements and the target state or assignment weights) that the r^{th} measurement at time t arises from target $k_r(t)$. In the original PMHT it is given by;

$$w_{l,r}^{n}(t)=\frac{\pi_l N\left\{\mathbf{z}_r(t);\mathrm{H}_l(t)\mathbf{x}_l^{n}(t),\mathbf{R}_l(t)\right\}}{\sum_{p=1}^{M}\left[\pi_p N\left\{\mathbf{z}_r(t);\mathrm{H}_p(t)\mathbf{x}_p^{n}(t),\mathbf{R}_p(t)\right\}\right]} \quad (10)$$

where $N(z|Hx,R)$ is Gaussian Probability Density Function (PDF) with Hx mean and $\boldsymbol{R}$ covariance. With the posterior probability given as above the synthetic values of the measurements $\{\tilde{\mathbf{z}}\}$ and corresponding measurement covariance $\{\tilde{\mathbf{R}}\}$ are respectively

$$\tilde{\mathbf{z}}_s(t)=\frac{\sum_{r=1}^{n_t} w_{sr}^{n}(t)\mathbf{z}_r(t)}{\sum_{r=1}^{n_t} w_{sr}^{n}(t)} \quad (11)$$

$$\tilde{\mathbf{R}}_s(t)=\frac{\mathbf{R}_s(t)}{\sum_{r=1}^{n_t} w_{sr}^{n}(t)} \quad (12)$$

The structure of the PMHT algorithm can be summarized in 3 items;

1. Calculate the w's based upon the measurements and the current track estimate,
2. Form the "synthetic" measurements and covariances from these w's, and
3. Update the track using a Kalman smoother.

B. The Homothetic PMHT (HPMHT)

The homothetic measurement model replaces the simple Gaussian model for $N(z_{tr}|H_{tm}x_{tm},R_{tm})$ [7, 9]. The HPMHT is a modification on the basic PMHT model such that measurements at scan t can come from any one of P Gaussian measurement processes, each having the same mean $H_{tm}x_{tm}$ but with differing covariances $\{\kappa_p{}^2\boldsymbol{R}\}_{p=1}^{P}$. Here for the notational convenience, the derivation is given for the single target case. With the new assumptions, modification to (10) yields that

$$w_{p,r}(t)=\frac{\pi_p N\left\{\mathbf{z}_r(t);\mathrm{H}_p(t)\mathbf{x}_p(t),\kappa_p^2\mathbf{R}(t)\right\}}{\sum_{l=1}^{M}\left[\pi_l N\left\{\mathbf{z}_r(t);\mathrm{H}_l(t)\mathbf{x}_l(t),\kappa_l^2\mathbf{R}(t)\right\}\right]} \quad (13)$$

Related with the modification in posterior probability, modified synthetic values of the measurements and corresponding measurement covariance are obtained as follows;

$$\tilde{\mathbf{z}}(t)=\frac{\sum_{r=1}^{n_t}\sum_{p=1}^{P} w_{p,r}(t)\mathbf{z}_r(t)/\kappa_p^2}{\sum_{r=1}^{n_t}\sum_{p=1}^{P} w_{p,r}(t)/\kappa_p^2} \quad (14)$$

$$\tilde{\mathbf{R}}(t)=\frac{\mathbf{R}(t)}{\sum_{r=1}^{n_t}\sum_{p=1}^{P} w_{l,r}(t)/\kappa_l^2} \quad (15)$$

C. AI Adding to HPMHT

For adding AI information to the HPMHT algorithm a modification is made to posterior probability, given in (10). In order to complete this process, some preliminary definitions should be given first.

The PDF of clutter originated AI is represented with $P_0(a)$, and the PDF of target originated AI is given with $P_1(a)$. $P_0(a)$, and $P_1(a)$ are defined as follows;

$$p_0(a)=a\exp\left(-\frac{a^2}{2}\right) \qquad a\geq 0 \quad (16)$$

$$p_1(a)=\frac{a}{1+d}\exp\left(-\frac{a^2}{2(1+d)}\right) \qquad a\geq 0 \quad (17)$$

where "a" represents AI, and "d" represents Signal to Noise Ratio (SNR).

A pre-selected threshold level (τ) is used to accept the measurement as valid. In this case, the probability of exceeding the threshold level of a target originated measurement is defined as Probability of Detection (P_D), on the other hand the probability of exceeding the threshold level of a clutter originated measurement is defined as False Alarm Probability (P_{FA}). P_D , and P_{FA} are defined as follows,

$$P_D=\int_{\tau}^{\infty} p_1(a)da \qquad P_{FA}=\int_{\tau}^{\infty} p_0(a)da \quad (18)$$

Also, definitions of target and noise originated measurements exceeding the threshold level are given below, respectively.

$$p_0^{\tau}(a)=\frac{1}{P_{FA}}p_0(a)=\frac{1}{P_{FA}}a\exp\left(-\frac{a^2}{2}\right) \qquad a>\tau \quad (19)$$

$$p_1^{\tau}(a)=\frac{1}{P_D}p_1(a)=\frac{1}{P_D}\frac{a}{1+d}\exp\left(-\frac{a^2}{2(1+d)}\right) \qquad a>\tau \quad (20)$$

The ratio of above two identities are known as Amplitude Likelihood Ratio (ALR) and defined as follows;

$$\lambda_l=\frac{p_1^{\tau}(a_l)}{p_0^{\tau}(a_l)} \quad (21)$$

AI information is added to HPMHT algorithm from posterior probability using ALR;

$$w_{p,r}(t)=\frac{\left(\pi_p N\left\{\mathbf{z}_r(t);\mathrm{H}_p(t)\mathbf{x}_p(t),\kappa_p^2\mathbf{R}(t)\right\}\right)\lambda_p}{\sum_{l=1}^{M}\left[\left(\pi_l N\left\{\mathbf{z}_r(t);\mathrm{H}_l(t)\mathbf{x}_l(t),\kappa_l^2\mathbf{R}(t)\right\}\right)\lambda_l\right]} \quad (22)$$

After the AI is added in the above manner, the remaining process of the HPMHT algorithm continues in the same way. The resulting algorithm can be nominated as HPMHT-AI.

IV. EXPERIMENTAL RESULTS

In this section some real life surface to air video data is used to evaluate the performance of the proposed method. The assumptions for generating point measurement data from video data has been explained in section 2. All the experimental studies has been accomplished for one target situation, and all the video data has been taken from ground to air.

In the experimental studies the point measurement with amplitude information obtained from video data is accomplished using a 250x250 pixel sensor area. The scan rate (Δt) was selected as 2 frames and the tracking algorithm was performed for T=30 scan steps. The process and measurement noise standard deviations σ_p and σ_m are equal to 0.1 and 10, respectively. Two-point initialization is used and the detection probability P_D is selected as 97%, probability of false alarm P_{FA} is 1%. The homothetic parameters used in this study has been selected as; P = 3, with κ_1=1, κ_2=4 and κ_3=9.

One example for a real video frame is given as follows. First obtaining measurement data is shown. The scenario is for 2.36 dB SNR for RST data and 9.1 dB SNR after transforming point measurement with AI. In Figure-1 video frame for one scan is given in RGB format. This frame is transformed into RST format and presented in scale data and display as image in Figure-2. Also the transformation of the same scan into point measurement with AI format is given in Figure-3.

Figure 1: Video Frame of RGB data.

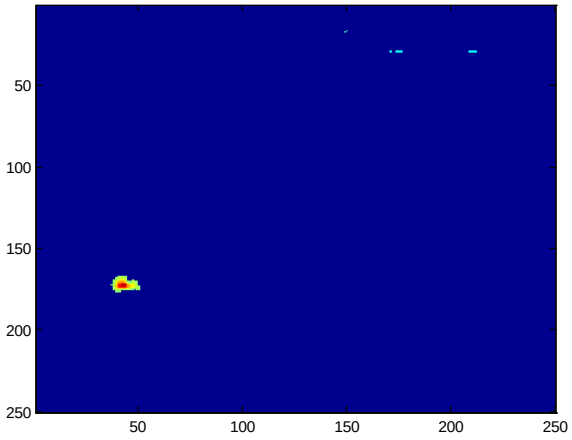


Figure 2: Video Frame after transformed to RST.

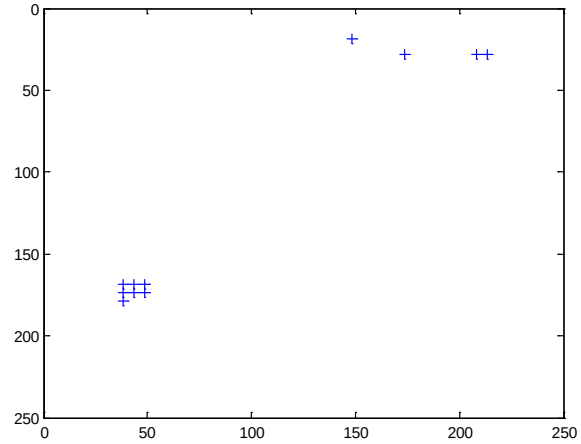


Figure 3: Point Measurement of the RST Data.

As one can see in the end of the process, the transformation of video data into point measurement data with AI is not only suitable for tracking with HPMHT-AI, but also the target detection is provided. Before giving the results for different scenarios the performance of HPMHT-AI algorithm for the above case is given in Figure-4. In the figure target centroids and corresponding HPMHT-AI estimations are given for throughout the tracking process.

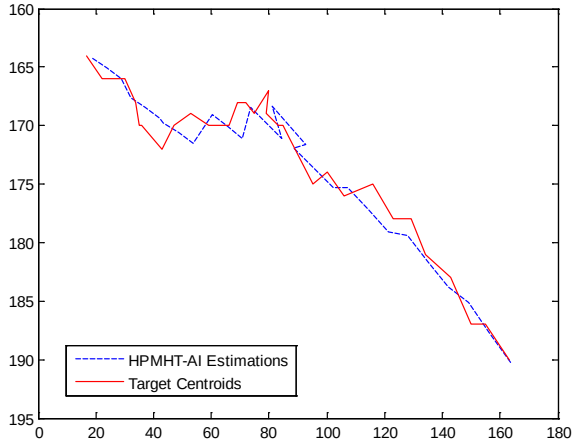


Figure 4: HPMHT-AI Estimations vs Target Centroids.

Experiments are conducted for 4 different conditions. These conditions are classified to their clutter and maneuvering conditions. Each condition takes place for a real life scenario. Some additional specifications should also be defined. These additional specs that exist in the scenarios are mean clutter number, maximum acceleration, and minimum SNR ratio of AI. It is also considered that the target is spread to the pixels due to its physical shape. The results are evaluated for mean estimation error and maximum estimation error. If any of them is greater than the square root of the target spreading pixel area then it can be concluded that the tracking process is unsuccessful. In Table-1, the results are given for four different scenarios corresponding to four different conditions.

Table 1: Experimental Results.

Scenario Type.	Mean Clutter No / Max acceleration (pxl/scan^2) / min. SNR of AI (dB)	Spread of Target (pxl^2)	Mean Estimation Error (pxl.)	Max. Estimation Error (pxl)
Low Clutter / Low Maneuver	8 / 0.318 pxl/scan^2/ 5.6 dB	150	2.4	7.1
High Clutter / Low Maneuver	75 / 1 pxl/scan^2/ 1 dB	54	1.97	4.7
Low Clutter / High Maneuver	11 / 3.25 pxl/scan^2/ 4.37 dB	121	3.8	8.5
High Clutter / High Maneuver	72 / 3.5 pxl/scan^2/ 2.3 dB	420	4.03	13.1

It can be seen from the table that the performance results of HPMHT-AI algorithm obtained using RST based point measurements gives highly satisfactory results for video object tracking.

V. CONCLUSION

In this study a novel approach for tracking air vehicles from ground using video sensors is proposed. The tracking is conducted via HPMHT which is a point measurement based technique. For this reason, an image processing method called RST is applied to the video sequences and target determination is provided. Then RST data is converted to point measurement with AI. Thus, the video data is converted a totally different form, but it still conveys the whole information. This point measurement data is used in HPMHT-AI algorithm for track continuation. The results obtained by using the proposed image processing technique and the HPMHT-AI algorithm are quite satisfactory.

REFERENCES

[1] A. Sobral, A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.

[2] D. Zhou, H. Zhang, "Modified GMM background modeling and optical flow for detection of moving objects," IEEE Int Conf on Systems, Man and Cybernetics (ICSMC 2005), 2005; p.2224-2229.

[3] B. Galvin, B. McCane, K. Novis, D. Mason, S. Mills, "Recovering Motion Fields: An Evaluation of Eight Optical Flow Algorithms," In BMVC, 1998, vol. 98, p. 195–204.

[4] A.G. Pakfiliz, " Automatic Detection of Aerial Vehicle in Cloudy Environment by UsingWavelet Enhancement Technique" Radioengineering, vol. 26, no. 4, pp.1169-1176, 2017.

[5] R. Streit and T. Luginbuhl, "Probabilistic MultiHypothesis Tracking", NUWC-NPT Technical Report 10,428, February 1995.

[6] Y. Ruan and P. Willett, "The Multiple Model PMHT and its Application to the Second Benchmark Radar Tracking Problem", IEEE Transactions on Aerospace and Electronic Systems, vol. 40, 2004, p. 1337-1347.

[7] M.Efe, Y. Ruan, P. Willett, "The Pedestrian PMHT", In Proceedings of the 5th ISIF, vol. 2, 2002, p. 838—845.

[8] R. C. Gonzalez, R. E. Woods, "Digital Image Processing", 2nd ed, Prentice Hall, 2002.

[9] S.J. Davey, "Extensions to the Probabilistic Multi-Hypothesis Tracker for Improved Data Association", PhD Thesis, The University of Adelaide, 2003.

Designing Autonomous Landing System for Rotary Wing UAVs

C. CICEKDEMIR[1], H. UCGUN[2], U. YUZGEC[3] and M. KESLER[4]

[1] Vocational High School, Bilecik Seyh Edebali University, Bilecik/Turkey
cagri.cicekdemir@bilecik.edu.tr
[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
hakan.ucgun@bilecik.edu.tr
[3] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
ugur.yuzgec@bilecik.edu.tr
[4] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
metin.kesler@bilecik.edu.tr

Abstract - **Unmanned Aerial Vehicles (UAVs), are air aircraft that can fly autonomously without human control or that can be controlled by a ground station. In parallel with the technological developments of the present day, it has been observed that the use of UAVs is also greatly increased. In this context, new working areas were needed together with the widespread use of UAVs. As a result of the researches, the autonomous control works was seen to become the foreground. Within the scope of this study, an autonomous landing system has been implemented for UAVs. The images taken with the help of the camera located at the bottom of the UAV are subjected to image processing techniques on the Linux based operating system to determine the landing track. In the scope of the study, the landing station detected is a circle with a "H" in the middle. The letter "H" on the landing station can be changed color according to the location. For example; the landing station can be red in football ground, can be blue in empty ground. Images taken from the camera are subject to the image processing technique. First of all, the image is removed from noises. Then, it is classified according to the "H" letter color on landing station. Pixels which are same with "H" letter are changed to white color, others changed to black color. So that, it is determined how many shaped has been found. The shapes are ordered according to their size and compared with the landing station. If the similarity rate is over 75%, the target is determined. Once the determination, the altitude of UAV will gradually decrease and the image acquisition and measurement process will be repeated until it reaches a desired position. After arriving at the desired position (Eg z<=20 cm), the UAV performs descent by stopping engines.**

Keywords - **UAV, Rotary Wing, Landing Station, Image processing, Autonomously**

I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs), one of the innovations brought by today's technologies, have reached massive amounts due to their low costs. The increase in users has also increased accident rates due to careless use. To eliminate such problems, autonomous control systems have been needed. These systems are able to provide UAVs to autonomously move from one point to another point and also fast response from the reactions in the environment. But the average flight time for UAVs can range from 10 to 30 minutes, although battery technology is progressing day by day [1, 2]. Flight times are further reduced by the size and weight of the UAVs. When the battery is at insufficient level, autonomous moving UAV may be in danger of falling over time.

In order to prevent such problems, with the additional enhancements integrated into the autonomous moving systems, the autonomous landing operation can be performed for the nearest ground station if the battery level reaches the critical point.

Autonomous landing systems can be done by GPS based algorithms. But, in GPS-based positioning, usually 2-3 meters of error can occur [3]. In this case, the UAV may land at a different point than the ground station and problems may occur. The problems of GPS-based autonomous landing systems can be solved with the help of image processing techniques.

In this study, the UAV is landed at the ground station which is in view by the image processing techniques. During landing operations, the images taken from the camera are subjected to various image processing methods to determine the ground station. In the scope of the study, the images which are taken with a wide-angle camera are send to a microcontroller in electronic environment. Images were processed on microcontroller with OpenCV libraries and C++ programming language.

II. LITERATURE REVIEW

Different studies have been conducted on UAVs for autonomous landing systems. Youeyun Jung and his colleagues have been working on the detection of the "H" lettering, which is yellow on the black ground plane. Height values were continuously analyzed with IMU after the target was detected. The deformations on the image at the near distance were prevented by marking the middle of the "H" [4].

Karl E. Wenze and his colleagues have been working on the detection of the "H" lettering, which is white and black in color. They used an infrared camera in that work. They found that the images taken from the camera using the infrared technology are clearer [5].

III. MATERIAL AND METHOD

In this study, firstly we designed the shape and size of the ground station. The used camera is connected to the UAV via usb cable. The analysis process begins when the ground station enters the view of the camera which has a wide angle of 150 degrees. The images from the camera is initially blurred to remove noises. Afterwards, undesired colors are extracted from the image by the color detection algorithm. The shapes of the images are determined by the contour detection algorithm. The shapes are filtered according to their size and small shapes are taken from the image. The remaining shapes are compared with the ground station image in order to find the target. The detected image must have a similarity rate of at least 75% of the target image. If the similarity condition is met, the landing command is sent to the autonomous moving system. The system comprises Raspberry Pi 3 microcontroller control card for processing images. Obtained images were subjected to image processing algorithms with OpenCV 3.2.0 library.

The ground station shown in the block diagram in Figure 1 is formed with the universal letter "H" used all over the world. The circle is used outside of the letter "H" so that the target can be detected successfully, regardless of the angle from which is viewed. Red color is preferred for circle and "H" because there are fewer in our living spaces and in nature. The ground station is designed in such a way that the background color is white.

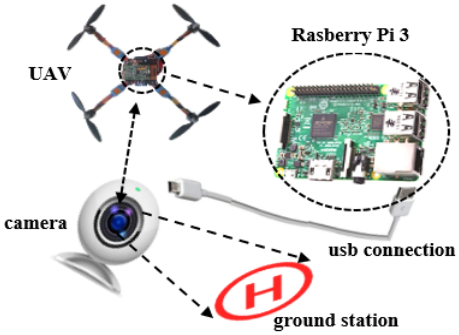


Figure 1: Block Diagram

A. *Raspberry Pi 3*

The Raspberry Pi microcontroller card, which was developed by Raspberry Pi Foundation company in 2009 to expand the use of computers in 3rd world countries, has its own place with its hardware features and size. The Raspberry Pi control card, which is a credit card sized in size, can be used for many different operations thanks to its input / output modules [6]. In the scope of this study, Raspberry Pi 3 shown in Figure 2 was used as a control card. Debian-based Raspbian version is preferred for Linux distributions as operating system.

Figure 2: Raspberry Pi 3

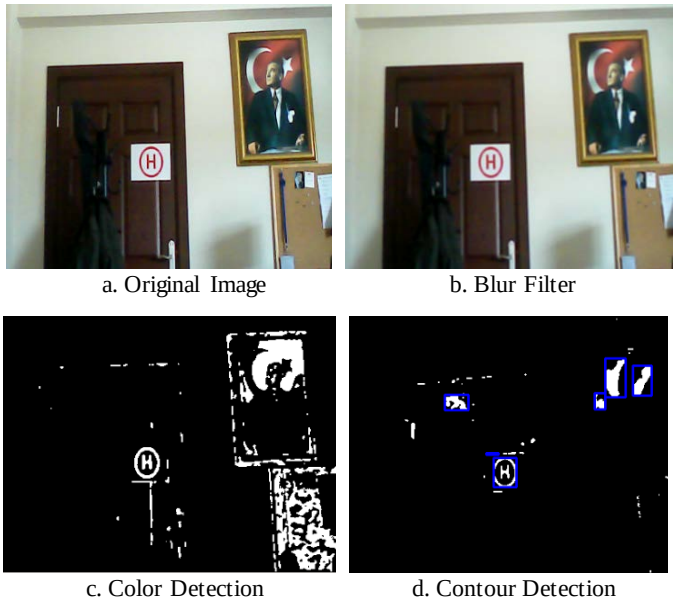
B. *OpenCV*

OpenCV, developed by Intel Corporation in 1999 and licensed under BSD license, is an open source "Computer Vision" library. It can be used to extract and process meaningful information contained in pictures or videos. Since OpenCV is a library developed independently from platforms, it can work in environments like Windows, Linux, MacOS, IOS, Android. OpenCV applications can be developed with very different programming languages such as C, C #, C ++, Python, Java, Perl, Ruby [7].

Within the OpenCV library, there are many algorithms for image processing and machine learning. With these algorithms, many processes such as plate recognition, face recognition, object classification, object discrimination, image comparison, human motion detection can be done easily [8].

IV. PROPOSED LANDING SYSTEM

The system is basically based on determining the location of the ground station using the OpenCV library and image processing algorithms of the images taken from the camera. The visual in Figure 3 shows how the ground station is detected, the results of the applied filters.



a. Original Image b. Blur Filter

c. Color Detection d. Contour Detection

Figure 3: Steps for Detection Ground Station

The images taken from the camcorder are processed in the method is called by the timer, not in the loop because of the mistakes that may occur if the microcontroller is stuck. The timer is set to 200 milliseconds so that 5 images in one second can be processed. The operations performed by the camera are explained below;

1) Images are taken in 640x480 format. The image is converted to a 340x240 size for faster processing. This is done using the resize() method of the OpenCV library [9, 10].
2) Figure 3.b shows the picture with average blur filter. This filter is used to remove the noises on the image. The filter takes the average of all pixels with a 3x3 kernel matrix moving around on the image, then equaling the pixel value in the middle. This method softens transitions between pixels.
3) Images taken from the camera are in RGB format as standard. Each pixel value is a 1x3 matrix structure. In the OpenCV library, other color spaces such as HSV, YCb Cr, LAB are also available. The HSV color space is more successful than the RGB color space when the primary colors are detected [11]. For this reason, the color space of the taken images is converted to HSV format. In the HSV color space, each pixel value is represented by a 1x3-dimensional matrix, as the same RGB color space. The first element of the matrix H represents the main color value, the second element S represents the saturation value, and the third element V represents the brightness value. As shown in Figure 4, the H value in the HSV color space is between 0 and 360, and the S and V values are between 0 and 255. But, in the OpenCV library, the H value is scaled from 0 to 180 for not overload the 8-bit area. The red color determined in the study is defined as 0 to 10 and 170 to 180 for the H value in the HSV color space.

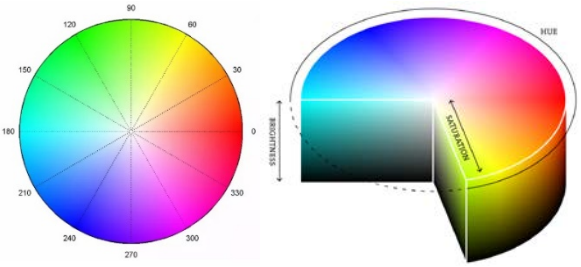


Figure 4: HSV Color Space

4) After changing the color space, the values of all the pixels on the image are checked according to whether they are in the red color range or not. Those that are out of the red color range are converted to white color, while those that are in the red color range are converted to black color. As a result of this operation, each pixel value is represented by an elementary matrix. On this account, you earn money from memory. Figure 3.c shows the image subjected to color filtering. In the OpenCV library, color space filtering is done using the inRange () method [12].
5) Interconnected pixels are analyzed in black-and-white formatted images. Each pixel is numbered according to whether it has the same value as its neighbor or not. So that, it is important that how many shapes are determined on the image. Figure 3.d shows the shapes detected on the image. In the OpenCV library, the findContours () method is used to detect shapes from black and white images [13]. This method is based on algorithm developed by Satoshi Suzuki et al. In 1985 on binary images [14].
6) The determined shapes are sorted in size from large to small. The detected shapes are compared with the ground station image recorded in the system. When the similarity rate reaches 75% and above, it is acceptable result that the ground station is detected.

The form screen designed by the C ++ programming language is shown in Figure 5. The image processing starts with "start cam" button on form. The similarity ratio of between shapes and ground station are shown on the right in Figure 5. At the same time, the position on the detected image that is also shown.

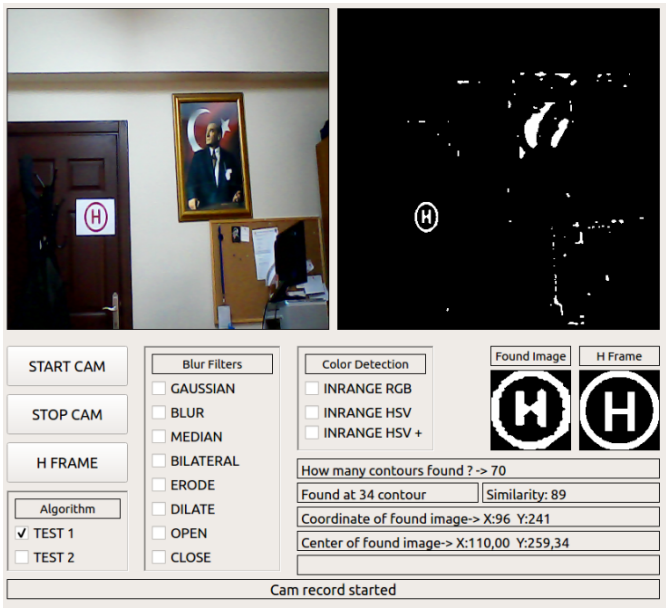


Figure 5: GUI of Ground Station Detection

V. CONCLUSION

In this study, it was aimed to detect the ground station, which is related to the autonomous landing procedures of an unmanned aerial vehicle (UAV). Approaching the ground station, the UAV identifies the location of the ground station with image processing techniques. After locating, the UAV initiates landing by sending commands to the autonomous moving system. At this time, if the ground station disappears, landing stops. Then, it rises to the original height which is before start landing. Within the scope of application, color identification algorithm is used from image processing techniques. For better results, different image processing techniques such as color space changing and contour detection algorithms have been used. All techniques were applied with the help of the OpenCV library in C++ language.

REFERENCES

[1] Saha, B., Koshimoto, E., Quach, C.C., Hogge, E.F., Strom, T.H., Hill, B.L., Vazquez, S.L., and Goebel, K. Battery health management system for electric uavs. In Aerospace Conference, 2011 IEEE.

[2] Sina Sharif Mansouri, Petros Karvelis, George Georgoulas, George Nikolakopoulos, Remaining Useful Battery Life Prediction for UAVs based on Machine Learning, The 20th World Congress of the International Federation of Automatic Control, Toulouse, France, 2017

[3] Ha Yoon Song and Jun Seok Lee, Detecting Positioning Errors and Estimating Correct Positions by Moving Window, PLoS One, 2015

[4] Youeyun Jungl, Robust Marker Tracking Algorithm for Precise VA V Vision-based Autonomous Landing, International Conference on Control, Automation and Systems, ICCAS 2015

[5] Karl E. Wenze, Low-Cost Visual Tracking of a Landing Place and Hovering Flight Control with a Microcontroller, J Intell Robot Syst (2010) 57:297–311, August 2009

[6] Aktaş, V., "What is Raspberry Pi?", Accessed February in 2018, http://www.volkanaktas.com/2014/05/raspberry-pi-nedir/

[7] Erişti, E., "Görüntü İşlemede Yeni Bir Soluk, OPENCV", Akademik Bilişim'10 - Academic Informatics Conference, 10 - 12 February 2010, Mugla, Turkey

[8] "About OpenCV", Accessed February in 2018, https://opencv.org/about.html

[9] "OpenCV Geometric Transformations", Accessed February in 2018, https://docs.opencv.org/3.0-beta/modules/imgproc/doc/geometric_transformations.html

[10] Ethan E. Danahy, Sos S. Agaian, Karen A. Panetta, Algorithms for the resizing of binary and grayscale images using a logical transform, Proc. SPIE 6497, Image Processing: Algorithms and Systems V, 2017

[11] Nursabillilah Mohd Ali, Nahrul Khair Alang Md Rashid and Yasir Mohd Mustafah, Performance Comparison between RGB and HSV Color Segmentations for Road Signs Detection, Scientific.net/AMM.393.550, September 2013

[12] "OpenCV Core Functions: InRange", Accessed February in 2018, https://docs.opencv.org/3.2.0/d2/de8/group__core__array.html#ga48af0ab51e36436c5d04340e036ce981

[13] "OpenCV Core Functions: findContours", Accessed February in 2018, https://docs.opencv.org/3.3.1/d3/dc0/group__imgproc__shape.html#ga95f5b48d01abc7c2e0732db24689837b

[14] Satoshi Suzuki and others. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing, 30(1):32–46, 1985.

Vegetation extraction from digital orthophoto maps using object-based segmentation and decision tree classifier

A. JAMIL[1] and B. BAYRAM[1]

[1] Yildiz Technical University, Istanbul/Turkey, akh.jamil@yahoo.com
[1] Yildiz Technical University, Istanbul/Turkey, bayram@yildiz.edu.tr

***Abstract* - The main objective of this study is to automatically extract tea gardens from large geographic areas using high-resolution digital orthophoto maps. To achieve this objective, object-based image analysis and decision tree (DT) classifier were integrated. For segmentation, multi-resolution image segmentation algorithm was used which is implemented in Definiens Developer commercial software. Both scale and compactness parameters were empirically calculated that produced optimal results. The segmented objects were selected manually for training the DT classifier from all used images. Spectral and textural features were extracted from each segment and to make the features robust against local variations, they were extracted at two image scales and final feature vector was formed by averaging the two feature vectors. The selected optimal features were used to train the DT classifier and then applied it on the test data to generate thematic maps for tea gardens. The performance of the proposed method was evaluated by comparing results with the reference data that produced promising results for mapping tea gardens (overall accuracy 88%) on our dataset.**

***Keywords* – Tree classification, decision trees, object-based segmentation, tea garden extraction.**

I. INTRODUCTION

THE advent of modern imaging sensors has made it possible to obtain images with very high spatial and spectral resolution, which makes it possible to extract objects even at fine scales. On the other hand, efficient and powerful image processing techniques are also proposed to take advantage of the data produced by such sensors. However, these techniques are devised keeping a specific dataset under consideration, therefore, they may not produce satisfactory results when applied on other datasets. Since, the process of deriving useful information from remotely sensed data is not trivial due to similarities in spectral information between various classes of interest [1], new methods are still required to investigate for remote sensing data classification.

Generally, machine learning based methods combined with object-based approaches have proven to be more effective for tree species information extraction from remotely sensed data [2]–[4]. Traditionally, pixel-based methods were used; however, they failed to produce optimal results for high-resolution images due to spectral similarity among classes. To overcome the problems with pixel-based approaches, object-based approaches are proposed that partition images into small homogeneous segments called objects, which are obtained by exploiting spectral and spatial relationships with neighboring pixels. In addition, object-based approach is useful as additional information can be obtained from image objects, such as textural and contextual information [2], [5], which is valuable for object detection and classification. Therefore, this study also focused on using an object-based approach for extraction of tea gardens.

Image segmentation is the key step of object-based image analysis. It subdivides the image into smaller regions using a predefined criteria of homogeneity and heterogeneity [6]. In this regard, all pixels belonging to same object are assigned same class during classification to overcome the problem of interclass similarities as in pixel-based approaches. A number of methods have been proposed in literature for image segmentation; for instance mean shift [7], region growing [8], multiresolution segmentation[9] etc.

Remote sensing data plays a vital role as it has been widely used to map tree species from large geographic areas. Obtaining this up to date information may be useful for making decisions at regional or even at national levels. For instance, the data can be used in various applications, such as environment monitoring, tree inventory forest management [3], forest mapping etc.[10], [11]. Therefore, there is need to develop new technique and continuous efforts are being made to derive useful information from remote sensing data.

The objective of this study is to automatically map tea gardens from large geographic areas using an object-based segmentation method and decision tree (DT) classifier. Object-based approach is useful to deal with high resolution data, DT have proved to be effective to deal with large-scale data, and their implementation is straightforward. The following sections describe the technical details of the proposed method.

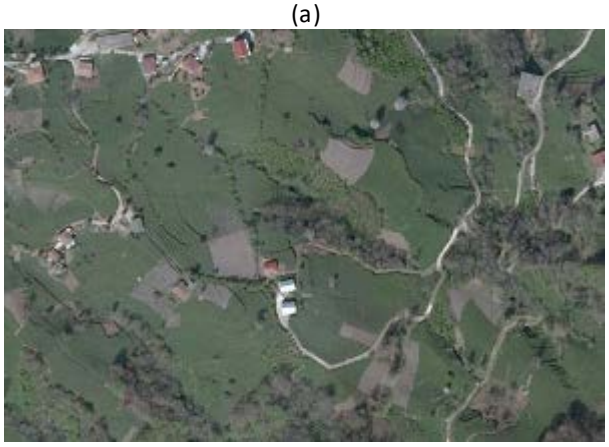
II. MATERIALS AND METHOD

A. Study Area

The study area selected for this study is known as Rize, which is a province of Turkey. It is located toward the Black-Sea region which is known for tea production. Rize and its neighbouring regions are dominated by thick forest and

agricultural land. Figure 1(a) shows map of Turkey and location of Rize province while (b) shows a sample image.

(a)



(b)

Figure 1: Study area (a) Map of Turkey (b) A sample image.

B. *Dataset*

Flights were made on 27th and 28th March in 2013 to acquire the images. The temperature of the region was ~15 °C and under bright sunny days. UltraCam-X digital aerial camera was used to capture the images by mounting it on an airplane. The flight altitude was about 4200 m with side and lateral laps to were 70% and 30% respectively and ground sample distance was set to 30 cm. Four bands namely red, green, blue, and near infrared were captured by the camera.

C. *Preprocessing*

Preprocessing step was primarily used to reduce salt and pepper effects and to remove non-vegetation areas. For the former one, Gaussian low-pass filter was applied on the input image to produce a slightly smoothened images using a kernel of size 5x5. For later, we used normalized difference vegetation index (NDVI) using red and near infrared bands and non-vegetation areas were removed.

D. *Image Segmentation*

Image segmentation is basis of object-based analysis methods. This study employed multiresolution image segmentation algorithm implemented in eCognition commercial software [12], which is extensively used for object-based analysis of high resolution images. The core algorithm implemented in the software is fractal net evolution approach (FNEA). FNEA is a bottom-up segmentation method, which starts with a single pixel and merges smaller segments with larger ones based on the homogeneity criteria. The segmentation process is controlled by selecting appropriate values for weighting of input data, scale, color and shape parameters. The scale parameter determines the maximum allowed heterogeneity for the resulting image objects. While the color determines the influence of spectral values and the shape (compactness and smoothness) determines the impact of the shape on the segmentation of objects[13]. Figure 2 shows the result of the FNEA applied on a sample image.

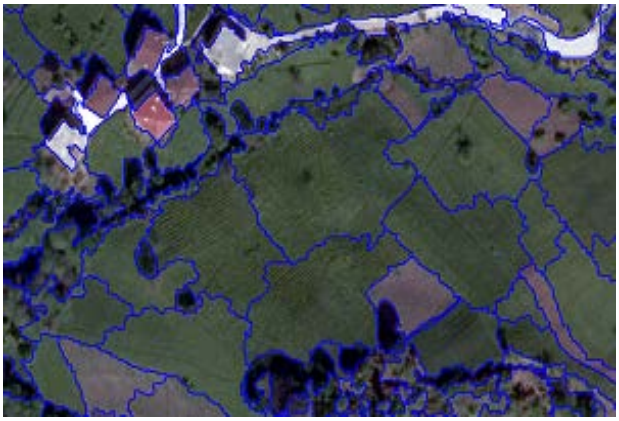


Figure 2: FNEA results obtained for a sample image.

The FNEA segmentation algorithm produced visually appealing results on the input image. However, there were some errors, especially towards the boundaries of the objects due to spectral similarity between different classes.

E. *Classification*

This study employed knowledge-based decision trees for tea garden extraction. DT is a reliable and robust approach used for both classification and regression. It applies a method for separating the data into homogeneous subsets using a threshold value of the input features in a recursive manner. The non-parametric nature of DT makes it more popular as it does not require prior knowledge about the distribution of data and it has the ability to process high-dimensional data fast [14].

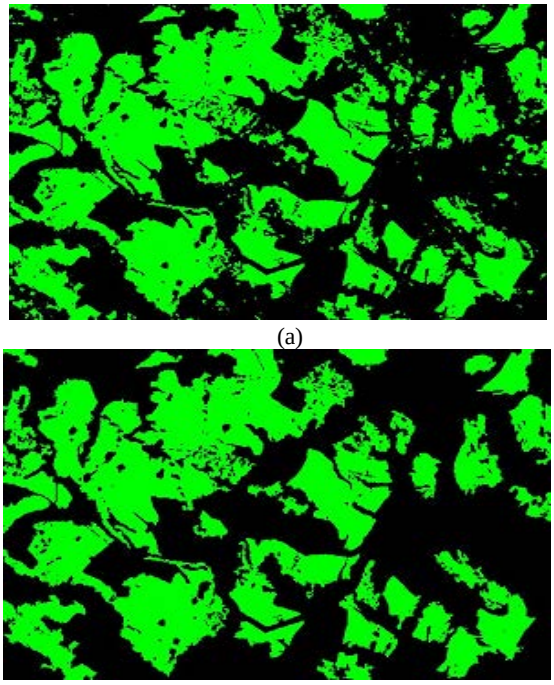
Both spectral and spatial features were used for training and testing the data. The classifier constructs a decision tree by performing a statistical analysis of the data. As the spectral and textural features of tea gardens are significantly different from other types of trees, it uses the criteria to split the data into desired classes.

The decision tree classifier analysis was performed using MATLAB 2017b © platform. The knowledge-based classification tree was constructed using the function fitctree of MATLAB. It requires two parameters: X (feature vectors) and Y(labels). After training the classifier, the features of the test data were passed and classifier predicted a label for each row of features set which contains samples of input images.

F. *Post-processing*

A post-processing step was included to remove some

unwanted objects and reduce noise from the output obtained from the DT classifier. Two image processing techniques were employed, morphological processing for filling small gaps within objects while geometrical constraints were applied to remove small objects which were not tea gardens, such as isolated trees and small grassy areas. Figure 3 shows the results before and after applying the post-processing step on a sample image.



(a)

(a)

Figure 3: (a) DT classifier result (b) Post-processing result after removing the noise

III. Results and Discussion

For object-based approaches, the results of the classification are highly influenced by the image segmentation. Therefore, selecting a good segmentation algorithm is highly desirable to obtain higher classification accuracies. Although, there is no general rule for selection of segmentation algorithm, however, one that produces good classification results is favored. We selected FNEA as it has been successfully used in various remote sensing applications.

FNEA requires tuning of various parameters, therefore by hit and trial method, optimal values for these segmentation parameters were obtained that produced visually suitable segments. After several trials, the scale parameter was set to 50, compactness to 0.8, shape to 0.2, and the weights for all the layers to 1.0 for all images.

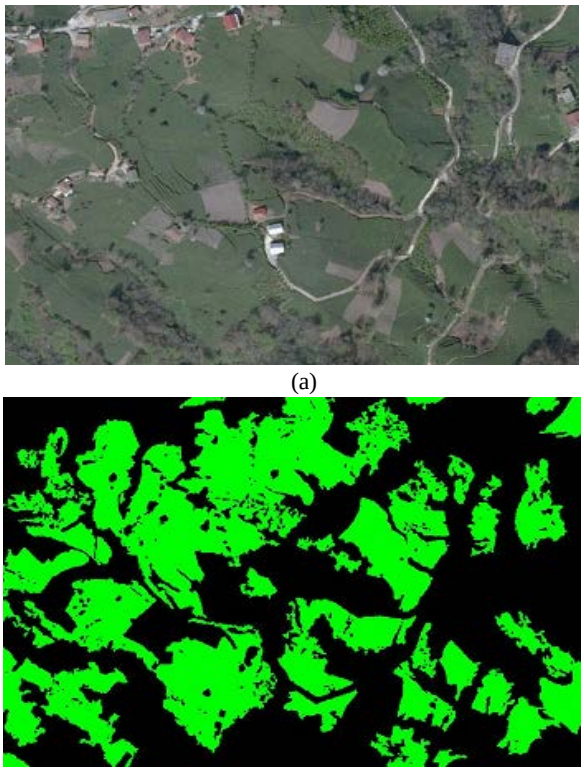
After image segmentation using tuned parameters, training samples were selected for two classes of interest (tea gardens and other types of species). An unbalanced training dataset was prepared by selecting 800 samples for tea gardens while 500 samples for other types of trees. While testing was performed on the whole dataset (all objects). The training samples were carefully selected for each type of classes by visually inspecting the objects and having good homogeneity. Since, the images were mostly dominated by tea gardens, therefore, relatively higher number of training data was selected for tea gardens compared to that of other types of tree species. Table 1 summarizes the amount of training and testing data used in this study.

Table 1: Summary of test and training data used

Class Type	Training Object	Testing Object
Tea Gardens	800	All Images
Other Trees	500	All Images

The performance of the proposed method was evaluated by overlaying the output results with the reference data. The results were then quantified in the form of overall accuracy and Cohen's Kappa Coefficient for tea gardens. It is worth mentioning here that evaluation was done only for tea gardens for two main reasons: first, the objective was to extract the tea garden areas; secondly, the reference data was only available for the tea gardens.

The overall accuracy (OA) was 88% and the Cohen's Kappa Coefficient is 0.72 for tea garden extraction using the proposed method. A sample output as a thematic map for tea gardens (green color) obtained for the proposed method is shown in figure 4. (a) is the sample input image while (b) is the final result of the proposed method. Table 2 summarizes the quantitative results obtained for whole dataset.



(a)

(b)

Figure 4: The results obtained for the proposed method. (a) A sample image (b) Thematic map of tea gardens by applying DT.

Table 2: Classification accuracy of proposed method (%)

Class	Overall Accuracy	Kappa Coefficient
Tea Gardens	88.01	0.72

It is interesting to note that lower value for Cohen's Kappa Coefficient was due to recently cultivated tea gardens. Even visually these gardens seem like bare land in the obtained digital images, however the photogrammetric experts evaluated them as tea gardens and during reference data creation they were marked with white pixels indicating them as tea gardens. From technical point of view, however, the proposed method is primarily designed to classify vegetation areas, therefore, recently cultivated tea gardens will be rejected considering them as non-vegetation areas. Before start of this work, it was assumed that only vegetation areas will be further processed for classification and rest of the areas will be rejected as non-vegetation areas. These cases affected the classification accuracy of the classifier. However, the accuracy can be improved by adding a manual editing process where users will be able to manually add the missing tea gardens to the classification output due to lack of vegetation information in those particular areas which is not part of this work. Few examples of such cases are highlighted in red colored polygons in Figure 5.

Figure 5: Some samples of non-vegetation areas but they are marked as tea areas since they have been recently cultivated.

IV. CONCLUSION

This work focused on extraction of tea gardens from high resolution digital orthophoto maps by combining object-based image analysis technique with decision tree classifier. After applying a preprocessing step to reduce the impact of noise, the region-growing algorithm was applied to transform individual pixels into an object-based representation. From each object several spectral and texture features were extracted and fed into the DT classifier for training. After training the classifier, the same set of features were extracted from the test images and sent to the DT classifier for classification which produced thematic maps for tea gardens. A post processing step was employed to remove small objects by applying morphological and geometrical constraints. Finally, the results were compared with the reference data for evaluation.

Our findings indicate that integration of object-based image segmentation and DT is effective for extraction of tea garden from high-resolution digital orthophoto maps. It is worth noting that the proposed method can be adopted for other types of vegetation classification. In such case, appropriate training data should be selected for each class and train the classifier. In future, we intend to work on different classifier combination to improve the accuracy of the classification.

ACKNOWLEDGMENT

The authors would like to thank the EMI Group Inc., Turkey for providing the important data for this study, which is a part of TEYDEP Project entitled "Development of Object Based Neural Network Image Processing System Determination of Vegetation and Forestry Boundaries" (Project No. 7140512). This work was supervised by EMI Group-Turkey, and consulted by Prof. Dr. B. Bayram.

REFERENCES

[1] Y. Chen, Y. Ge, G. B. M. Heuvelink, R. An, and Y. Chen, "Object-Based Superresolution Land-Cover Mapping From Remotely Sensed Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 1, pp. 328–340, Jan. 2018.

[2] T. Blaschke, G. J. Hay, M. Kelly, S. Lang, P. Hofmann, E. Addink, R. Queiroz Feitosa, F. van der Meer, H. van der Werff, F. van Coillie, and D. Tiede, "Geographic Object-Based Image Analysis – Towards a new paradigm," *ISPRS J. Photogramm. Remote Sens.*, vol. 87, pp. 180–191, Jan. 2014.

[3] A. Jamil and B. Bayram, "Tree Species Extraction and Land Use / Cover Classification From High-Resolution Digital Orthophoto Maps," *IEEE J. Sel. Top. Appl. EARTH Obs. Remote Sens.*, vol. 11, no. 1, pp. 89–94, 2017.

[4] Z. Zhong and J. Li, "Generative Adversarial Networks and Probabilistic Graph Models for Hyperspectral Image Classification," no. February, 2018.

[5] M. Åkerblom, P. Raumonen, R. Mäkipää, and M. Kaasalainen, "Automatic tree species recognition with quantitative structure models," *Remote Sens. Environ.*, vol. 191, pp. 1–12, 2017.

[6] Q. Li, C. Wang, B. Zhang, and L. Lu, "Object-Based Crop Classification with Landsat-MODIS Enhanced Time-Series Data," *Remote Sens.*, vol. 7, no. 12, pp. 16091–16107, 2015.

[7] D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, 2002.

[8] S. Kamdi and R. K. Krishna, "Image Segmentation and Region Growing Algorithm," *Int. J. Comput. Technol. Electron. Eng.*, vol. 2, no. 1, pp. 103–107, 2011.

[9] P. Happ, R. Ferreira, and C. Bentes, "Multiresolution segmentation: a parallel approach for high resolution image segmentation in multicore architectures,"… *Object-Based Image …*, 2010.

[10] U. Maulik and D. Chakraborty, "Remote Sensing Image Classification : A survey of support-vector-machine-based advanced techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 33–52, 2017.

[11] S. E. Franklin and O. S. Ahmed, "Deciduous tree species classification using object-based analysis and machine learning with unmanned aerial vehicle multispectral data," *Int. J. Remote Sens.*, vol. 0, no. 0, pp. 1–10, 2017.

[12] Definiens, *Definiens Professional 5: Reference Book. Munich: The Imaging Intelligence Company*. 2006.

[13] I. L. Castillejo-González, J. M. Peña-Barragán, M. Jurado-Expósito, F. J. Mesas-Carrascosa, and F. López-Granados, "Evaluation of pixel- and object-based approaches for mapping wild oat (Avena sterilis) weed patches in wheat fields using QuickBird imagery for site-specific management," *Eur. J. Agron.*, vol. 59, pp. 57–66, Sep. 2014.

[14] T. Sasaki, J. Imanishi, K. Ioki, Y. Morimoto, and K. Kitada, "Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data," *Landsc. Ecol. Eng.*, vol. 8, no. 2, pp. 157–171, Jul. 2012.

A New Algorithm for Shape Estimation with a Low Cost Sensor

K. ŞALOĞLU[1], A. HOŞAFÇI[2], M. BİRBİLEN[3], A. BAYBÖRÜ[4], E. DİNLER[5] and A. GÜNEŞ[6]

[1] Atilim University, Ankara/Turkey, saloglu.keziban@student.atilim.edu.tr
[2]Atilim University, Ankara/Turkey, hosafci.arda@student.atilim.edu.tr
[3]Atilim University, Ankara/Turkey, birbilen.merve@student.atilim.edu.tr
[4]Atilim University, Ankara/Turkey, bayboru.alpay@student.atilim.edu.tr
[5]Atilim University, Ankara/Turkey, dinler.ege@student.atilim.edu.tr
[6]Atilim University, Ankara/Turkey, ahmet.gunes@atilim.edu.tr

***Abstract* - In this paper, a measurement system composed of a single low-cost sensor and a shape estimation algorithm is proposed. Measurement mechanism consists of a low-cost infrared sensor, two servo motors and a microcontroller. This mechanism provides flexibility to scan at different heights and angle intervals. From the obtained scanning results, the features of the shapes are extracted and objects are classified accordingly. Classified objects are square prism, triangular prism, cylinder and pyramid. The extracted features are change in the width of the object depending on its height, having corner, corner angle and slope of the surface. The algorithm uses segmentation and fractured line fitting on the contour data of the shapes. The proposed algorithm is proved to be superior to our former method using RANSAC algorithm.**

***Keywords* - shape estimation, low-cost sensors**

I. INTRODUCTION

Shape estimation is one of the important research areas for robot vision, robotics and pattern recognition. Shape of the objects is an important feature which can eventually contribute to the object recognition and classification. Data can be collected using camera, infrared or ultrasonic sensors.

Most sensors collect surface data from the objects. Digital images captured by cameras, also contains the contour information of the objects. To obtain contour data from the digital images, algorithms like "Adaptive Discrete Curve Evolution" may be used [1]. This algorithm deletes the deformed parts of the boundary points of an image and keeps the points that contain visual information. Algorithm ends when all the significant points are included. The contour data can be used to estimate the shapes efficiently. Lines or curves are fitted to boundary points and feature vectors are obtained to classify the shape of objects [2]. Another study focuses on using centroid distance function curve to do shape recognition on two dimensional shapes. This function presents the distance of the contour points to the centroid of the shape [3]. To recognize shapes, shape context is also used [4]. All boundary points of the shape are defined with respect to a single boundary point. This feature is compared with the shape context value of the shape in the database and dissimilarity is measured. Thus, the shape recognition is done. This study is done on digital images having two dimensional shape contours of objects.

Another family of sensors to estimate shapes are infrared or ultrasonic sensors. Those sensors may be combined to form different sensor arrays. An ultrasonic sensor array containing four sensors is used to estimate the shape of cube, cylinder and rectangular prism [5]. However, measurements are taken at only one horizontal slice. Another sensor array containing 16 ultrasonic sensors is used to distinguish and classify a cylinder and a rectangular prism and distorted ultrasonic waves are utilized to identify the shape of the objects using genetic algorithm [6]. This algorithm detects the distortion created by different shapes. Another sensor array alternative is a pentagon sensor array which is applied using infrared sensors [7]. The object is placed inside of the pentagon sensor array. Hemisphere, ellipse and cylinder shaped objects' shape are reconstructed. In this study, shape of the object is not classified. In another study, data is collected with an infrared range finder in order to do mapping for a mobile robot [8]. Collected data is processed in three steps: cluster, segment and fit. Thus, classification of line, arc or corner is done [8].

All approaches discussed above, contain measurements performed at only a fixed height. Another common point of the former studies is to use multiple sensors. However, the cost increases with the increasing number of sensors. Approaches that use digital image of the object is not sufficient since camera is a relatively expensive sensor with respect to the infrared or ultrasonic sensors. Additionally, the sensor arrays are not a flexible alternative to use on mobile robots due to higher mounting space requirements. Therefore, a system that contains minimum number of low- cost sensors and provides flexible scanning to collect data is needed for mobile robots.

In this paper, a two Degrees of Freedom (DoF) mechanism is designed to collect data. The objects are scanned with the mechanism. Collected data is classified using proposed algorithm with respect to the determined features of the shapes. Proposed algorithm separates the collected points into segments and does fractured line fitting between segmented parts to evaluate features. Classification results are compared with the previous study that uses RANSAC (Random Sample Consensus) algorithm for line fitting [9].

II. SETUP

Measurement setup consists of two servo motors and one GP2Y0A02YK0F infrared proximity sensor. Setup is shown in

Fig. 1 together with the rotation and coordinates axes. The system is capable of rotating around z and x axis. This allows taking measurements from horizontal planes at different heights and angle intervals.

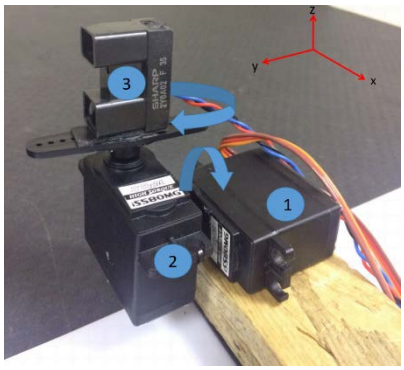


Figure 1: Designed mechanism; servo motor rotating around y axis is labeled as 1, servo motor rotating around z axis is labeled as 2. The infrared sensor is labeled as 3.

Infrared sensor gives analog voltage output depending on the distance of the object measured. However, the sensor should be calibrated in order to measure distance values. The calibration of the sensor is done fitting a fifth degree polynomial that relates voltage outputs by distance values as seen in Fig.2.

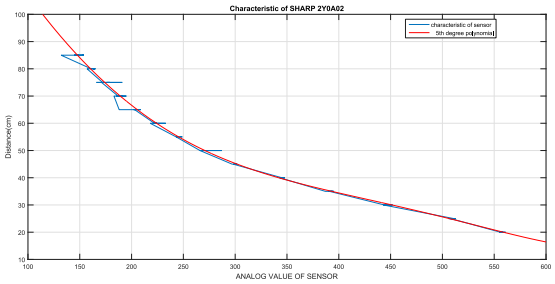


Figure 2: Calibration curve of the infrared sensor.

During the measurements, a distortion occurs at the output. Setup works similar to a radar while scanning the objects. Therefore, the data is collected at different angle values of the servo motor. This creates a distortion since changing angle of motor causes a change in the perspective of the sensor. This means the data should be rectified. This rectification is done using basic trigonometric functions. Sensor field of vision is divided into three regions and trigonometric functions are used to find the distance value independent from the perspective of the sensor. This is shown in Fig.3. The function used for rectification of measurements in region I is:

$$d = r_m \cos(a) \qquad (1)$$

Where d is the rectified distance, r_m is the measured distance, a is the angle of the motor.

Similar calculations are done for the case of objects standing on region II and III.

Additionally, impulsive noise is present on the scanning output of the sensor. In order to clean the noise from the output, tenth order median filter is applied. All the steps including rectification and median filtering are shown in Fig.4

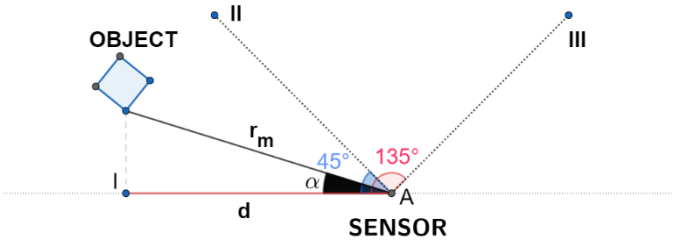


Figure 3: Coordinate transformation applied to region I: The sensor is at point A, the radial distance between the sensor and object is "r_m" and the obtained distance after rectification is "d".

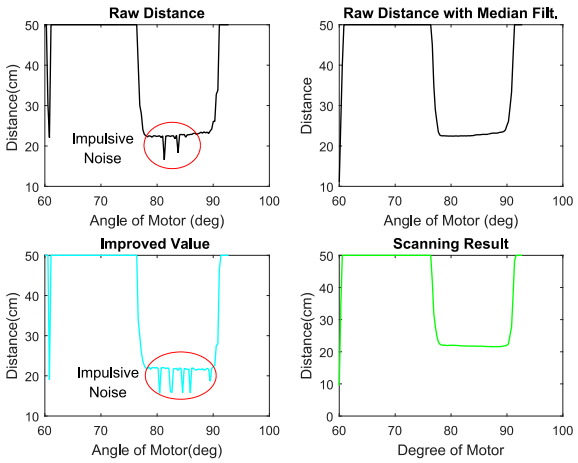


Figure 4: Rectification and median filtering steps on the raw scanned data.
sdsdsdds

III. MEASUREMENT AND CLASSIFICATION

Measurements are repeated for all the objects which are shown in Fig.5. Scanned data includes the contour information of the shapes. This information is evaluated in order to obtain features for classification.

Scanned shapes are square prism, cylinder, triangular prism and pyramid. Obtained scanning results of these shapes are shown in Fig.6.

Scanning results are also obtained using a single ultrasonic sensor. The measured contours are shown in Fig.7. All contours show similar curvatures and the shapes cannot be distinguished. Since the wavelength of the acoustic waves are comparable to dimensions of the fine details on the surfaces, the patterns of the contour plots do not give enough information to classify the shapes.

According to measurements using infrared sensor, surface curvature of cylinder is higher than the square prism. Also, corner angle of square prism is higher than the triangular prism. Width of the pyramid is decreasing with its height. Using these properties, four features are generated. These features are:

- change in width depending on height
- having a corner
- corner angle
- slope of the surface

The flow of the proposed algorithm is summarized in Fig.8. The shapes are firstly classified with respect to corner and height features. This principle of this step is shown in Fig. 9.

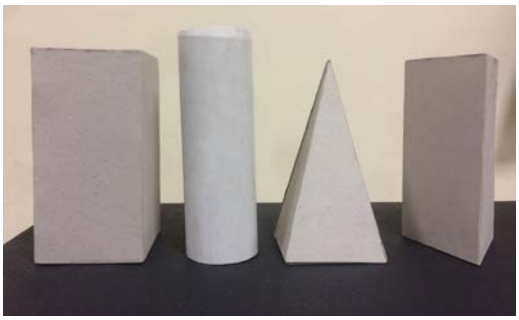


Figure 5: Scanned objects.

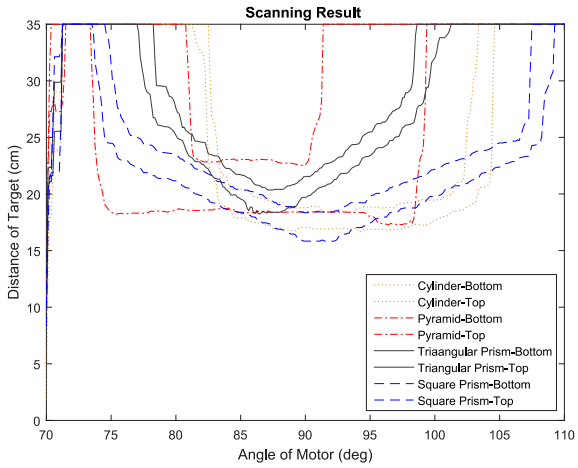


Figure 6: Contour data obtained using infrared sensor from upper and lower sections of square prism, cylinder, pyramid, triangular prism.

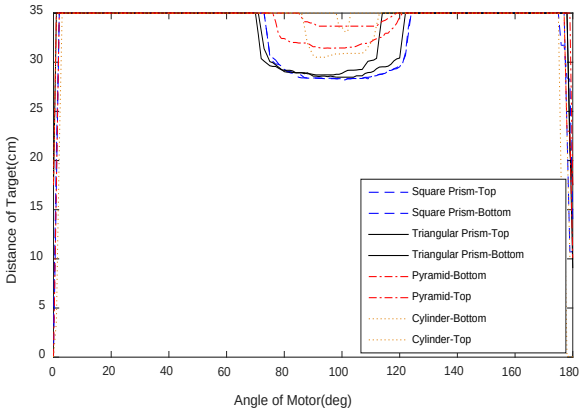


Figure 7: Contour data obtained using ultrasonic sensor from upper and lower sections of square prism, cylinder, pyramid, triangular prism.

The change in objects width with its height determined using two arrays containing surface information of the shape.
Those arrays are obtained from scanning the object at two different heights. Difference of the arrays is calculated. If the difference of the surfaces is greater than a threshold value for certain amount of value, it is considered the width of the object is getting decreased with its height. Defined groups depending on height feature is shown in Fig.9.

In order to determine whether the object has corner or not, the mean of the dataset is found. If the minimum data in the dataset is smaller than the mean value for a certain threshold, it is considered as the shape has corner.

While comparing surface slope or corner angle, segmentation is needed. In order to define surface slope, the surface of the shape is split into three parts. The lines are fitted to these segments. Finally, slope of these three lines are compared. Fig.10 shows an example of segmentation done on the shape. If the slopes of those lines are getting increased, then classification results in cylinder.

While calculating the corner angle, corner point is defined. The right hand side and left hand side points of this corner are used to fit lines. The angle between those lines is the corner angle of the shape. Fig.11 shows an example of fitted lines to the vertices of the shape. If the angle is greater than the value of 95, classification results in angled square prism. This threshold value is chosen depending on the measurements.

Finally, the algorithm is tested by manually moving each object 40 times. In these measurements, objects are placed randomly in front of the sensor at 20-30 cm distances and 60-120 degrees interval. The success of the algorithm is shown in Table 1.

Table 1: Shape Recognition Results with Proposed Algorithm

	Classification Result			
	Square Prism	Triangular Prism	Cylinder	Pyramid
Square Prism	40	0	0	0
Triangular Prism	6	34	0	0
Cylinder	11	0	29	0
Pyramid	0	0	0	40

Classification results show that if the number of segmented lines increased, the success of classification increases. Also, boundary values could be chosen as intervals to decrease the errors. For instance, triangular prism is classified wrongly as square prism since the corner angle differs from the specified value by ±1. If an interval would be specified, this type of errors would be decreased. Also, success of classifying pyramid shows that doing measurement from different heights is important for shape estimation.

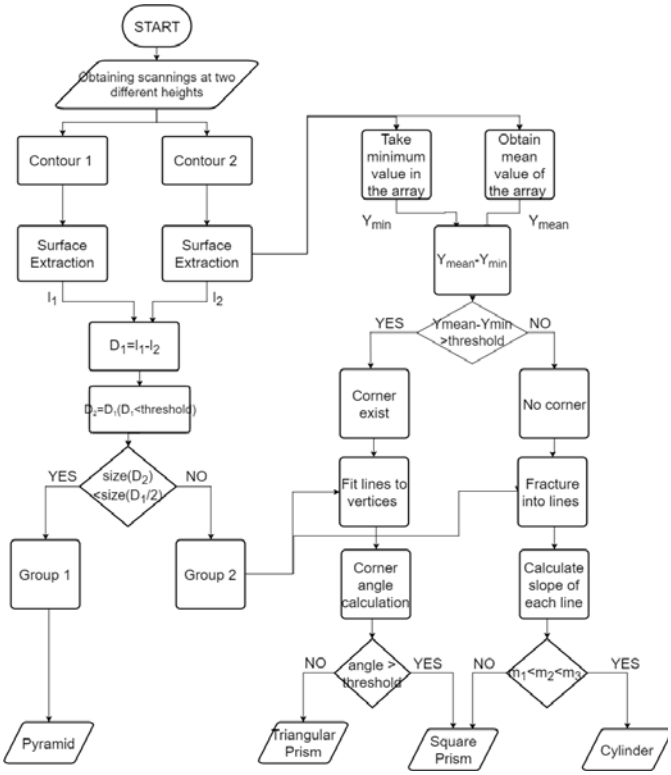


Figure 8: Shape estimation algorithm flowchart. Group 1 represents pyramid, Group 2 represents square prism, triangular prism and cylinder objects.

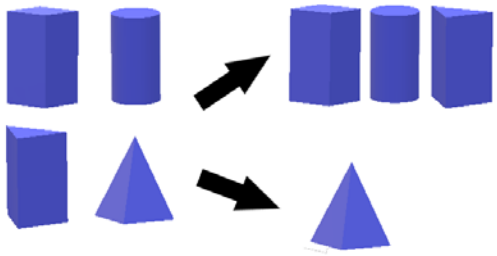


Figure 9: Shapes are divided into two groups according to the decrease on their width while the height is increasing.

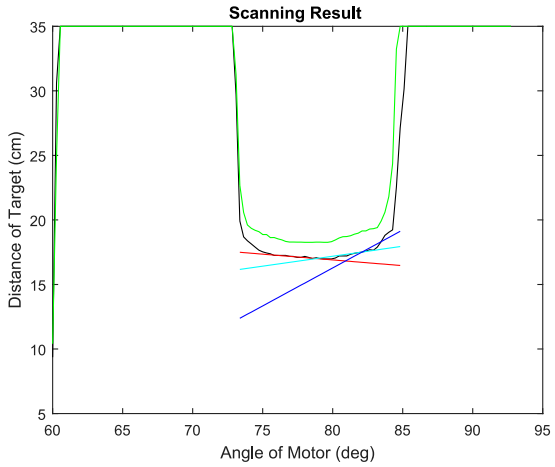


Figure 10: Segmentation and fitted lines on the contour of the cylinder

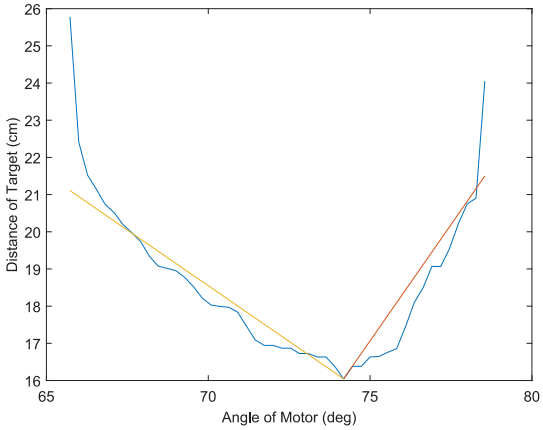


Figure 11: Fitted lines for the shapes that has corners.

RANSAC algorithm is applied for line fitting purposes in shape estimation algorithm in a previous study. In that study, there is no segmentation applied on the surface to calculate surface slope or corner angle [9]. The results of that algorithm, tested by the same 40 data, for classification of the shapes are given in Table 2.

In Table 1 and Table 2 it is seen that the algorithm using segmentation has more correct estimations than algorithm using RANSAC. Segmentation decreased the number of errors in cylinder from 19 to 11. Also, the triangular prism error is decreased from 11 to 6 since the corner angle is calculated more accurate than previous study. Therefore, the proposed algorithm is superior to the previous method.

Obtained errors in cylinder is due to the errors in segmentation and some measurement errors. When the classification is done, slope of the lines defining surface of the cylinder is described as their slope is getting higher. However, in some of the cases, this could not be obtained. Errors in the case of triangular prism are caused by the cases where the corner of the triangular prism is not detectable by the sensor. That is, when the rectangle lateral surface of triangular prism is visible by the sensor.

Table 2: Shape recognition results with RANSAC

	Classification Result			
	Square Prism	Triangular Prism	Cylinder	Pyramid
Square Prism	36	1	3	0
Triangular Prism	11	29	0	0
Cylinder	19	0	21	0
Pyramid	0	0	0	40

IV. CONCLUSIONS

This paper proposes a two DoF mechanism and an algorithm for shape estimation using a low-cost infrared sensor and an ultrasonic sensor. It is seen that the information obtained by ultrasonic sensor is not capable of distinguishing between shapes. Therefore, the algorithms are developed only for the infrared sensor.

Measurement data obtained using infrared sensors is rectified and median filtered. Finally, refined data is used to do classification depending on the features extracted from the contours of the shapes. Proposed algorithm includes segmentation. This algorithm is compared with the previous study that uses RANSAC. The classification results of both methods are presented in Table 1 and Table 2. Obtained errors while classifying the shapes are evaluated. It is concluded that segmentation is increased the success of the algorithm compared with the previous study.

Also, measurement mechanism is a better alternative than sensor arrays since it provides measurements at different angle and height intervals.

In the future studies, success of the algorithm will be increased. Algorithm will be combined with object detection and target tracking functions. Also, the measuring mechanism will be implemented on a mobile platform.

REFERENCES

[1] D. Hu, W. Huang, L. Shang, Z. K. Zhu, and J. Yang, "Adaptive discrete curve evolution for shape recognition," in 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014). IEEE, December 2014.

[2] T. Avci, G. Kokdemir, Z. Kurt, and K. Ozkan, "Shape features based conic arcs for leaf recognition," in 2014 22nd Signal Processing and Communications Applications Conference (SIU). IEEE, Apr 2014.

[3] M. Ozturk, "Geometrical shape recognition based on CDF extreme points analysis," in 2015 23nd Signal Processing and Communications Applications Conference (SIU). IEEE, May 2015.

[4] S. G. Salve and K. Jondhale, "Shape matching and object recognition using shape contexts," in 2010 3rd International Conference on Computer Science and Information Technology. IEEE, Jul 2010.

[5] A. R. Patkar and P. P. Tasgaonkar, "Object recognition using horizontal array of ultrasonic sensors," in 2016 International Conference on Communication and Signal Processing (ICCSP). IEEE, Apr 2016.

[6] K. Ohtani and M. Baba, "Ultrasonic robot eye for shape recognition employing a genetic algorithm," in IMEKO World Congress, 2003.

[7] S. Daud, S. M. Sobani, M. Ramiee, N. Mahmood, P. Leow, and F. C. Harun, "Application of infrared sensor for shape detection," in 2013 IEEE 4th International Conference on Photonics (ICP). IEEE, Oct 2013.

[8] J. Zhang, K. Pan, and K. Zhao, "Feature extraction from infrared scan data based on SM-MS algorithm for autonomous mobile robot," in 2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, Dec 2015.

[9] K. Şaloğlu, A. Hoşafçı, M. Birbilen, Y. A. Bulut, A. Gunes, "Düşük maliyetli sensörler ile şekil tanıma," in Proceedings of SİU İzmir, 2-5 Mayıs, 2018.

Fast Quadratic-Linear Approximated L1-norm for SAR Image Despeckling

F. NAR[1] and F. ATASOY[2]

[1] Konya Food and Agriculture University, Konya/Turkey, fatih.nar@gidatarim.edu.tr
[2]Karabuk University, Karabuk/Turkey, ferhatatasoy@karabuk.edu.tr

Abstract **- Speckle noise, inherent in synthetic aperture radar (SAR) images, degrades the performance of the various SAR image analysis tasks. Thus, speckle noise reduction is a critical preprocessing step for smoothing homogeneous regions while preserving details. Therefore, in a recent study, SDD-QL method was proposed which is a variational approach where l_1-norm total variation regularization term is approximated in a quadratic and linear manner to increase despeckling accuracy while decreasing the computation time. In this study, we propose a simple and efficient mechanism to increase the speed of SDD-QL method up to order of magnitude while obtaining the same despeckling result. Computational efficiency and despeckling performance of the proposed method are compared to the SDD-QL method on real-world SAR images.**

Keywords **- synthetic aperture radar, speckle reduction, fast noise filtering, l_1-norm, total variation, SDD, SDD-QL.**

I. INTRODUCTION

SYNTHETIC Aperture Radar (SAR) is an active sensor system that microwave pulses are transmitted by an antenna towards the earth surface and then images are generated from backscattered signals. SAR sensors can provide day-and-night imagery of Earth and images are acquired independent of atmospheric and sunlight condition such as rain, cloud, and fog [1]. The increase in the quality and quantity of SAR images from air and satellite platforms has led to an increase in the need for automatic and manual analysis. However, SAR images contains speckle noise which is a granular pattern that degrades the quality of SAR images [2].

The filtered reflectivity value is calculated by each pixel value within a sliding window that is weighted average in the speckle noise reduction approaches. Examples of these approaches are Lee [3], Frost [4], and improved versions of these filters [5], and also Kuan filter [6]. This type of despeckling methods are insufficient to protect details such as edges, point scatterers, and textures. Non-Local Mean Filtering (NLM) to overcome these failing obtains a very good performance especially for textured regions [7]. Probabilistic Patch-Based (PPB) [8] algorithm and SAR-BM3D [9] are the state-of-the-art NLM-based methods for despeckling of SAR images. As another approach, diffusion-based methods which are inspired from physics are also proposed. For example, the speckle-reducing anisotropic diffusion (SRAD) method provides speckle-reducing using Lee filter based on a partial differential equation model in anisotropic diffusion [10].

The Total Variation (TV) based noise reduction approach proposed by Rudin, Osher and Fatemi (ROF) [11] has led to many noise reduction studies in the literature. In the ROF model, a noise reduction approach that preserves the edges by applying the l_1-norm penalty to the derivatives of the reflectance values in the image is obtained. Inspired by ROF model, various TV based speckle reduction methods are proposed for SAR images which employs so called variational methods [12 – 16]. In a later study, an affective despeckling noise reduction method with a single parameter is proposed that makes it possible to make the TV regularizer term as l_0-norm, fractional norm, and l_1-norm [17], namely sparsity driven despeckling (SDD). In a recent study, quadratic approximation in [17] is improved with quadratic-linear (QL) approximation for l_1-norm case which leads to better despeckling with lower execution times [18], namely SDD with QL (SDD-QL).

In this study, accuracy parameter of TV regularization term in SDD-QL is adaptively changed in iterations which leads to same despeckling result up to order of magnitude faster execution times.

II. PROPOSED METHOD

In [18], speckle reduction for the SAR image is defined as the minimization of the following variational cost function with respect to f:

$$J(f) = \frac{1}{2N}\sum_{p=1}^{N}(f_p - g_p)^2 + \lambda|(\nabla f)_p| \qquad (1)$$

where g is observed speckled image, f is the desired despeckled image, N is the pixel count, p is the pixel index number, λ is a positive value determining smoothing level, and ∇ is the gradient operator. In the cost function, the data fidelity term ensures f stays similar to g in l_2-norm manner and TV regularization term implies penalty on the changes in image gradients in l_1-norm manner.

In [18], proposed approach for the minimization of non-differentiable cost function given in Equation 1 is named as SDD-QL. SDD-QL uses efficient quadratic-linear approximation to convert non-differentiable cost function in Equation 1 to below differentiable cost function.

$$J^{(n)}(f) = \frac{1}{2N}\sum_{p=1}^{N}(f_p - g_p)^2 + \gamma(f_p - \hat{f}_p)^2 + \lambda[(1-\alpha)(w_{x,p}(\partial_x f)_p^2 + w_{y,p}(\partial_y f)_p^2) + \alpha(s_{x,p}(\partial_x f)_p + s_{y,p}(\partial_y f)_p)] \quad (2)$$

where n is the iteration number, $\hat{f}_p$ is a proxy constant for f_p, $(f_p - \hat{f}_p)^2$ is a slow-step regularization (SSR) term for forcing f_p stays close to $\hat{f}_p$ since QL approximation is only accurate around $\hat{f}_p$, γ is a positive weight of SSR term, $w_{x,p} = \left(\left|(\partial_x \hat{f})_p\right| + \varepsilon\right)^{-1}$, $w_{y,p} = \left(\left|(\partial_y \hat{f})_p\right| + \varepsilon\right)^{-1}$, $s_{x,p} = sgn\left((\partial_x \hat{f})_p\right)$, $s_{y,p} = sgn\left((\partial_y \hat{f})_p\right)$, and α is QL approximation parameter $(0 \le \alpha < 1)$. In QL approximation, ε is approximation accuracy parameter $(\varepsilon > 0)$ where accuracy gets better as ε gets smaller. Note that, $\left|(\nabla f)_p\right|$ in Equation 1 is taken as sum of $\left|(\partial_x f)_p\right| + \left|(\partial_y f)_p\right|$ and then approximated using QL approximation in Equation 2 (see [18] for details). Superscript n in $J^{(n)}(f)$ shows that cost function must be minimized in an iterative manner due to employed QL approximation. Here, if γ is set as 0 and α is set as 0 then cost function of SDD-QL method become same with the cost function in SDD method. In SDD-QL study, γ is set as 1 and α is set as 0.5 [18].

Equation 2 can be represented in matrix-vector form as below:

$$J^{(n)}(v_f) = \frac{1}{2N}\Big((v_f - v_g)^\top(v_f - v_g) + \gamma(v_f - v_{\hat{f}})^\top(v_f - v_{\hat{f}}) + \lambda[(1-\alpha)(v_f^\top \mathbf{D_x}^\top \mathbf{W_x} \mathbf{D_x} v_f + v_f^\top \mathbf{D_y}^\top \mathbf{W_y} \mathbf{D_y} v_f) + \alpha(s_x^\top \mathbf{D_x} v_f + s_y^\top \mathbf{D_y} v_f)]\Big) \quad (3)$$

where v_g, v_f, $v_{\hat{f}}$, s_x, s_y are vector forms of g_p, f_p, $\hat{f}_p$, $s_{x,p}$, $s_{y,p}$, and $\mathbf{W_x}$, $\mathbf{W_y}$ are diagonal matrix form of $w_{x,p}$, $w_{y,p}$, and $\mathbf{D_x}$, $\mathbf{D_y}$ are the Toeplitz matrices as the forward difference gradient operators where derivatives are zero at the right and bottom boundaries respectively.

Equation 3 is differentiable; thus, one can take its derivative with respect to v_f and equalize it to zero to obtain its minimum which leads to a linear system as given below:

$$\mathbf{A} v_f^{(n+1)} = b$$
$$\mathbf{A} = (1+\gamma)\mathbf{I} + \lambda(1-\alpha)\mathbf{L} \quad (4)$$
$$b = v_g + \gamma v_{\hat{f}} - \frac{\lambda}{2}\alpha(\mathbf{D_x}^\top s_x + \mathbf{D_y}^\top s_y)$$

where $\mathbf{I}$ is identity matrix and $\mathbf{L} = \mathbf{D_x}^\top \mathbf{W_x} \mathbf{D_x} + \mathbf{D_y}^\top \mathbf{W_y} \mathbf{D_y}$. Iteration number is n for the $\mathbf{A}$, $\mathbf{L}$, $\mathbf{W_x}$, $\mathbf{W_y}$, b, v_f, $v_{\hat{f}}$, s_x and s_y unless it is explicitly stated. Here, $\mathbf{A}$ is a sparse, positive definite, and 5-point Laplacian matrix. Since $\mathbf{A}$ is positive definite, cost function in Equation 3 is strictly convex hence minimization leads to global minimum. Strict convexity of the cost function is not shown in SDD-QL paper; so, we show the proof of strict convexity of SDD-QL cost function (Equation 3) as follows:

Let $\mathbf{A}$ in the Equation 4 be the second derivative (i.e. Hessian) of the cost function $J^{(n)}(v_f)$ given in the Equation 3. A symmetric matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ is called positive definite (thus $J^{(n)}(v_f)$ is strictly convex), denoted by $\mathbf{A} \succ 0$, if $x^\top \mathbf{A} x > 0$ for every x $\in \mathbb{R}^N$ with $x \neq 0$.

$$\frac{\partial^2 J^{(n)}(v_f)}{\partial^2 v_f} = (1+\gamma)\mathbf{I} + \lambda(1-\alpha)\mathbf{L} = \mathbf{A} \quad (5)$$

Identity matrix $\mathbf{I}$ is positive definite and γ is positive. Also both λ and $(1-\alpha)$ are positive; so, we should only show that $x^\top \mathbf{L} x > 0$ to guarantee that $x^\top \mathbf{A} x > 0$ for all non-zero x. $x^\top \mathbf{L} x > 0$ can be expanded as below two inequalities after $x^\top$ and x are distributed from left and right onto $\mathbf{L}$:

$$x^\intercal(\mathbf{D_x}^\top \mathbf{W_x} \mathbf{D_x})x > 0, \quad x^\intercal(\mathbf{D_y}^\top \mathbf{W_y} \mathbf{D_y})x > 0$$
$$v_x^\top \mathbf{W_x} v_x > 0, \quad v_u^\top \mathbf{W_y} v_y > 0 \quad (6)$$

where $v_x = \mathbf{D_x} x$ and $v_y = \mathbf{D_y} x$. Here, v_x and v_y are non-zero vectors since x is a non-zero vector and $\mathbf{D_x}$ and $\mathbf{D_y}$ are Toeplitz matrices with non-zero diagonal elements. Both inequalities are satisfied since $\mathbf{W_x}$ and $\mathbf{W_y}$ are all diagonal matrices with positive entries. Thus, both first and second order optimality conditions are satisfied which shows that cost function $J^{(n)}(v_f)$ given in Equation 3 is strictly convex.

Pseudo-code of the proposed fast SDD-QL method is given in Algorithm 1. In Algorithm 1, implementation of all the steps are easy and computationally cheap, except for solving the linear system $\mathbf{A} v_f = b$. Since $\mathbf{A}$ is sparse and positive definite, preconditioned conjugate gradient (PCG) with incomplete Cholesky preconditioner (ICP) is used to obtain computational efficiency for solving the linear system. For PCG maximum iteration is set to 10^2 and convergence tolerance is set to 10^{-2}. Note that, all the matrices ($\mathbf{D_x}$, $\mathbf{D_y}$, $\mathbf{W_x}$, $\mathbf{W_y}$, $\mathbf{A}$, $\mathbf{L}$) in Algorithm 1 are sparse with regular structures.

In SDD and SDD-QL, solution becomes smoother as the iteration in Algorithm 1 proceeds where sharp edges appears in last iterations. Essentially in SDD and SDD-QL, in first iterations smoothing is happening similar to l_2-norm manner while in last iterations smoothing is happening in l_1-norm manner. This behavior can be controlled by changing ε such that $\varepsilon = 1$ will lead to a smoothing similar to using l_2-norm TV term while small values of ε will lead to a smoothing using l_1-norm TV term. In SDD-QL method, ε is a fixed value whereas in our study we propose to adaptively change the ε such that ε is 1 in first iteration and converges to desired value as iteration proceeds. In l_1-norm TV filtering, condition number (CN) of $\mathbf{A}$ gets higher as ε gets smaller. Higher CN means ill-posed linear system which requires more PCG iterations. Therefore, as proposed in this study, adaptively changing ε from 1 to a lower

value as iteration proceeds saves a considerable amount of computation, especially when small ε value is used. Proposed approach is also easy to implement and requires no extra computation since it does only simple scalar computations (2 lines in Algorithm 1 that contains ε_*) to calculate adaptive ε. Thus, only adding 2 lines of code to SDD-QL method is sufficient to obtain the proposed Fast SDD-QL method.

FastSDD-QL(g, λ, ε, α, γ, n_{max})
$v_f \leftarrow g$ // assign image data g to vector v_f as an initial solution
$v_g \leftarrow g$ // assign image data g to vector v_g
$\varepsilon_* = e^{-log(\varepsilon)/(n_{max}-1)}$ // initialization for the proposed adaptive ε
for $n \leftarrow 1$ **to** n_{max} **do**
  $\varepsilon = \varepsilon_*^{(1-n)}$ // assignment for the proposed adaptive ε
  $v_f \leftarrow v_f$ // assign v_f to proxy constant v_f
  $\mathbf{W_x} \leftarrow [diag(|\mathbf{D_x} v_f|+\varepsilon)]^{-1}$ // $\mathbf{D_x} v_f$ is x-derivative of v_f
  $\mathbf{W_y} \leftarrow [diag(|\mathbf{D_y} v_f|+\varepsilon)]^{-1}$ // $\mathbf{D_y} v_f$ is y-derivative of v_f
  $s_x \leftarrow sgn(\mathbf{D_x} v_f)$ // signum of x-derivative of v_f
  $s_y \leftarrow sgn(\mathbf{D_y} v_f)$ // signum of y-derivative of v_f
  $\mathbf{L} \leftarrow \mathbf{D_x}^\top \mathbf{W_x} \mathbf{D_x} + \mathbf{D_y}^\top \mathbf{W_y} \mathbf{D_y}$ // construct matrix $\mathbf{L}$
  $\mathbf{A} \leftarrow (1+\gamma)\mathbf{I} + \lambda(1-\alpha)L$ // construct matrix $\mathbf{A}$
  $b \leftarrow v_g + \gamma v_f - \lambda(\alpha/2)(\mathbf{D_x}^\top s_x + \mathbf{D_y}^\top s_y)$ // construct vector b
  solve $\mathbf{A} v_f = b$ // solve equation (4) to find v_f for the next iteration
end
$f \leftarrow v_f$ // converts vector v_f to an image f
return f // returns despeckled image f

Algorithm 1: Fast Quadratic-Linear Approximated l_1-norm Despeckling.

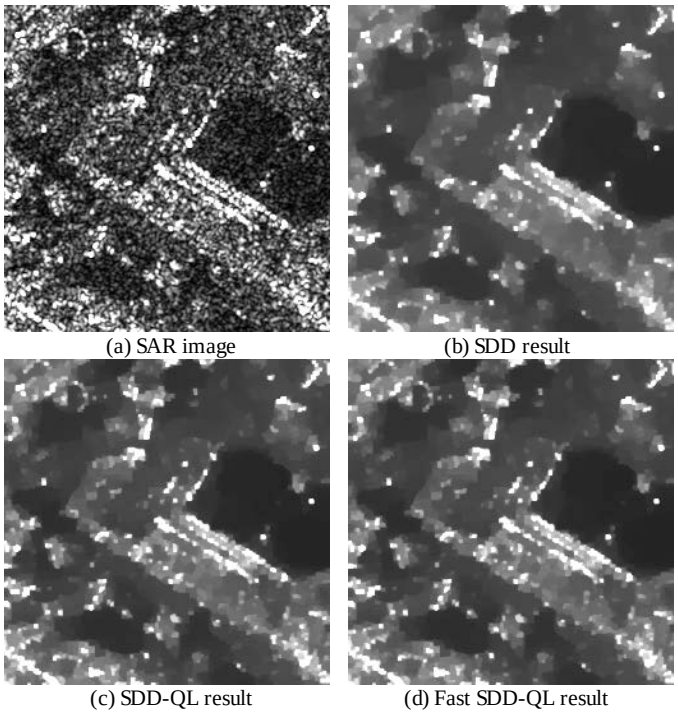
Fast SDD-QL method is the extension of the SDD-QL method where SDD-QL method is the extension of the SDD method. Therefore, Algorithm 1 can be executed as these three methods by setting the parameters properly as given in Table 1. As seen in Table 1, SDD method does not use SSR, QL, and adaptive ε and SDD-QL method does not use adaptive ε while Fast SDD-QL uses all of them. Essentially, for fixed ε, removing the two lines with ε_* symbol in the Algorithm 1 is sufficient.

Table 1: Parameter setting for SDD, SDD-QL and Fast SDD-QL in Algorithm 1.

Method	λ	γ (SSR)	α (QL)	ε
SDD	λ	0	0	*fixed*
SDD-QL	λ	1	0.5	*fixed*
Fast SDD-QL	λ	1	0.5	*adaptive*

III. Results and Analysis

In this section, Fast SDD-QL is analyzed to show its computational efficiency comparing to SDD method and SDD-QL method. SDD, SDD-QL, and Fast SDD-QL are all developed in C++ using the coding optimizations given in [17] and compiled as 64-bit executables. For all methods, 32-bit floating point arrays are used so all the calculations are realized with 32-bit floating point arithmetic. 32-bit floating point is preferred since its usage lead to same results with 64-bit floating point while 64-bit floating point requires twice memory and also slower. In all the experiments, Intel i7-6700K 4 GHz CPU is used as hardware, TerraSAR-X sample SAR image of India Visakhapatnam port (spot-mode, 16-bit, VV polarization, resolution ≈1 meter, number of looks ≈ 1) is used as test image, and $\lambda = 100$, $\varepsilon = 10^{-5}$, and $n_{max} = 5$ are used as default parameters unless otherwise stated. Despeckling results for a 256×256 SAR image are given in Figure 1.

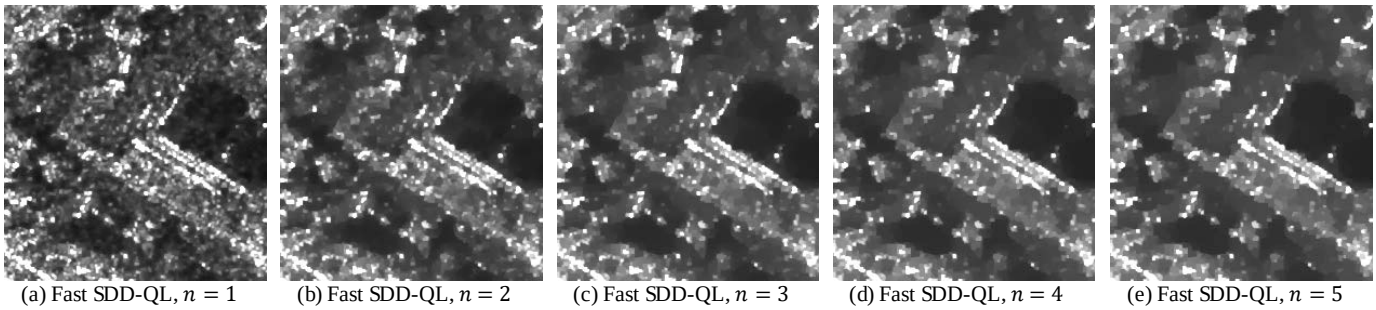


(a) SAR image (b) SDD result (c) SDD-QL result (d) Fast SDD-QL result

Figure 1: Despeckling results for a 256×256 SAR image.

A. Scale Space Construction

As mentioned before, solution becomes smoother as the iteration in Algorithm 1 proceeds where sharp edges appears in the last iterations. As level of smoothing increases in each iteration, these intermediate results can be used to construct a scale-space. Such scale spaces are used in various studies such as salient-region detection in SAR images [19] and change detection in multitemporal SAR images [20, 21]. Scale space constructed by intermediate results of Fast SDD-QL is given in Figure 2. Scale space constructed by SDD-QL is very similar to scale space constructed by Fast SDD-QL while scale space constructed by SDD is also similar.

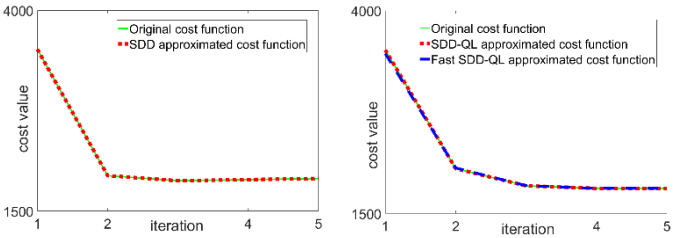


(a) Fast SDD-QL, $n = 1$ (b) Fast SDD-QL, $n = 2$ (c) Fast SDD-QL, $n = 3$ (d) Fast SDD-QL, $n = 4$ (e) Fast SDD-QL, $n = 5$

Figure 2: Scale spaces for the Figure 1.a (n is iteration number).

B. *Cost Function Graph*

In Figure 3 minimization of original cost function and approximated cost functions for SDD (Figure 3.a), SDD-QL, and Fast SDD-QL (Figure 3.b) methods are shown. Note that, cost function graph of SDD and SDD-QL is different since SDD-QL employs SSR term and QL approximation. For all methods, approximation accuracy is high so original cost functions and approximated cost functions are almost same since $\varepsilon = 10^{-5}$ is a small value. However, approximated cost function for Fast SDD-QL has small deviation from the original cost function in the first iterations since Fast SDD-QL uses $\varepsilon = 1$ in first iteration. In Fast SDD-QL, ε decreases as iteration proceeds and at last iteration ε becomes 10^{-5} and it converges to same global minimum as SDD-QL.

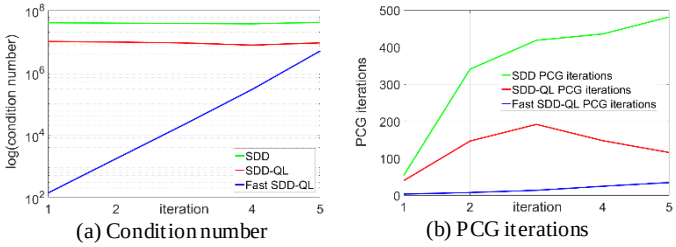


(a) SDD cost (b) SDD-QL and Fast SDD-QL cost

Figure 3: Cost function graph for SDD, SDD-QL, and Fast SDD-QL methods.

C. *Computational Analysis*

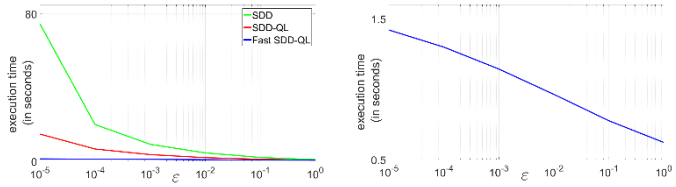
In Figure 4, condition numbers (CNs) and corresponding PCG iterations for the iteration of the three methods are shown. CNs are shown in log-scale since CNs of SDD and Fast SDD-QL have huge differences in their values (i.e. minimum CN is 138.5 and maximum CN is 4.1×10^7). As seen in Figure 4.a, Fast SDD-QL has very low CNs in the first iterations $\varepsilon = 1$ and similar CN with SDD-QL at the last iteration ($\varepsilon = 10^{-5}$). Lower CNs leads to lower PCG iterations since PCG iteration count is related to CN. Therefore, Fast SDD-QL has significantly lower PCG iterations (so lower execution times) comparing to SDD and SDD-QL.



(a) Condition number (b) PCG iterations

Figure 4: Condition number and corresponding PCG iterations (for the image in Figure 1.a)

In Figure 5, execution time for SDD, SDD-QL and Fast SDD-QL with respect to ε are shown. For $\varepsilon = 10^{-5}$ (high accuracy), Fast SDD-QL becomes 52.53 times faster than SDD and 10.45 times faster than SDD-QL. For $\varepsilon = 10^{-3}$, Fast SDD-QL becomes 8.15 times faster than SDD and 3.25 times faster than SDD-QL. For $\varepsilon = 10^{-1}$, Fast SDD-QL becomes 2.80 times faster than SDD and 1.55 times faster than SDD-QL.



(a) Execution times for 3 methods (b) Execution times for Fast SDD-QL

Figure 5: Execution time versus ε for 1024×1024 SAR image.

IV. CONCLUSION

In this study, we propose a simple and efficient mechanism to increase the speed of SDD-QL method, namely Fast SDD-QL. Proposed Fast SDD-QL method, SDD method, and SDD-QL method are all implemented using C++ and compiled as 64-bit executable while speed optimizations are turned on. Computational efficiency and despeckling performance of the proposed method are compared to the SDD method and SDD-QL method on real-world SAR images. Experiments show that Fast SDD-QL method can be order of magnitude faster comparing to SDD-QL method and much faster than SDD method especially when accurate despeckling is required, i.e. by using small ε value such as 10^{-5}. Although, Fast SDD-QL method brings significant speed ups comparing to SDD method and SDD-QL method, despeckling result is same with SDD-QL method and better than SDD method. Additionally, in this

study, we also show that Fast SDD-QL method can be used for generating nonlinear scale space without no extra computation.

REFERENCES

[1] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 1, pp. 6-43, 2013.

[2] F. Argenti, A. Lapini, T. Bianchi and L. Alparone, "A Tutorial on Speckle Reduction in Synthetic Aperture Radar Images," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 3, pp. 6-35, 2013.

[3] J. S. Lee, "Digital Image Enhancement and Noise Filtering by Use of Local Statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vols. PAMI-2, no. 2, pp. 165-168, 1980.

[4] V. S. Frost, J. A. Stiles, K. S. Shanmugan and J. C. Holtzman, "A Model for Radar Images and Its Application to Adaptive Digital Filtering of Multiplicative Noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vols. PAMI-4, no. 2, pp. 157-166, 1982.

[5] A. Lopes, R. Touzi and E. Nezry, "Adaptive speckle filters and scene heterogeneity," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 6, pp. 992-1000, 1990.

[6] D. T. Kuan, A. A. Sawchuk, T. C. Strand and P. Chavel, "Adaptive Noise Smoothing Filter for Images with Signal-Dependent Noise," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vols. PAMI-7, no. 2, pp. 165-177, 1985.

[7] A. Buades, B. Coll and J. M. Morel, "A Review of Image Denoising Algorithms, with a New One," *Multiscale Modeling \& Simulation*, vol. 4, no. 2, pp. 490-530, 2005.

[8] C. A. Deledalle, L. Denis and F. Tupin, "Iterative Weighted Maximum Likelihood Denoising With Probabilistic Patch-Based Weights," *IEEE Transactions on Image Processing*, vol. 18, no. 12, pp. 2661-2672, 2009.

[9] S. Parrilli, M. Poderico, C. V. Angelino and L. Verdoliva, "A Nonlocal SAR Image Denoising Algorithm Based on LLMMSE Wavelet Shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 606-616, 2012.

[10] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1260-1270, 2002.

[11] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259-268, 1992.

[12] G. Aubert and J.-F. Aujol, "A Variational Approach to Removing Multiplicative Noise," *SIAM Journal on Applied Mathematics*, vol. 68, no. 4, pp. 925-946, 2008.

[13] S. Durand, J. Fadili and M. Nikolova, "Multiplicative Noise Removal Using L1 Fidelity on Frame Coefficients," *Journal of Mathematical Imaging and Vision*, vol. 36, no. 3, pp. 201-226, Mar 2010.

[14] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Multiplicative Noise Removal Using Variable Splitting and Constrained Optimization," *IEEE Transactions on Image Processing*, vol. 19, no. 7, pp. 1720-1730, 2010.

[15] C. Liu, S. Zhu, "A convex relaxation method for computing exact global solutions for multiplicative noise removal," *Journal of Computational and Applied Mathematics*, vol. 238, pp. 144-155, 2013.

[16] C. Ozcan, B. Sen and F. Nar, *GPU efficient SAR image despeckling using mixed norms*, Proc. SPIE vol 9247, 2014, pp. 9247 - 9247 - 13.

[17] C. Ozcan, B. Sen and F. Nar, "Sparsity-Driven Despeckling for SAR Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 1, pp. 115-119, 2016.

[18] F. Nar, "SAR image despeckling using quadratic–linear approximated l_1-norm," *Electronics Letters*, vol. 54, pp. 387-389(2), 2018.

[19] Q. Zhang, Y. Wu, F. Wang, J. Fan, L. Zhang and L. Jiao, "Anisotropic-Scale-Space-Based Salient-Region Detection for {SAR} Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 3, pp. 457-461, 2016

[20] F. Bovolo and L. Bruzzone, "A detail-preserving scale-driven approach to change detection in multitemporal {SAR} images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2963-2972, 2005.

[21] B. Sevilmis, O. E. Okman, F. Nar, C. Demirkesen and M. Çetin, *A robust nonlinear scale space change detection approach for {SAR} images*, vol. 8892, 2013, pp. 8892 - 8892 - 13.

MOTSA: A Multi-Objective Tree-Seed Algorithm

G. OZCAN[1], A. OZKIS[1] and M.S. KIRAN[1]

[1]Selçuk University, Konya/Turkey, ozcanngul@gmail.com
[1]Selçuk University, Konya/Turkey, ahmetozkis@selcuk.edu.tr
[1]Selçuk University, Konya/Turkey, mskiran@selcuk.edu.tr

***Abstract* – The Tree-Seed Algorithm, TSA for short, has been proposed for solving single-objective optimization algorithm by inspiring the relation trees and seeds in nature. In this study, its multi-objective variant tree-seed algorithm, MOTSA, is proposed for solving multi-purpose optimization problems by motivating its performance on single-objective problems. In order to overcome selection issue in multi-objective problems, the well-known strategies, non-dominated sorting and crowding distance, of NSGA-II has been integrated with the proposed MOTSA. By doing this, the highest quality solutions from the combined populations of trees and seeds are selected and transferred to the next generation. The MOTSA are applied to solve well-known three multi-objective benchmark problems and obtained results show that MOTSA is an alternative multi-objective optimization algorithm.**

***Keywords* – Tree Seed Algorithm, multi-objective, optimization, non-dominated sorting, crowding distance.**

I. INTRODUCTION

THIS study proposes a new multi-objective tree seed algorithm, MOTSA, by using NSGA-II strategies and encourages researchers working on and improving the algorithm. The optimization can be addressed in two ways, single and multipurpose and this study focus on multi-objective optimization by modifying basic TSA to MOTSA by using the non-dominated sorting and crowding distance strategies of NSGA-II [1].

Tree-Seed Algorithm (TSA) is a recently proposed population-based metaheuristic algorithm to solve single objective continuous optimization problems. Real-world problems usually consist of simultaneous optimization of multiple goals that are contradictory [2]. While the MOTSA is based on the solution update rule of TSA, the better individuals are selected by applying Pareto-based elitism in determining the new parents and the individuals to be transferred to the next generation. In the applied elitism process, non-dominated sorting and crowding distance ordering methods in NSGA-II are used. Non-dominated sorting is used to determine the Pareto surface. With crowding distance, the diffusion of the algorithm is strengthened by identifying the most outstanding solutions found in the population.

In the present work, both algorithms were examined and the strongest features of the algorithms were determined. MOTSA has been introduced as a new method which is synthesized by NSGA-II and TSA strategies. In the second part of this article, the definition of optimization is made. In the third chapter, single-purpose optimization is explained by TSA. The fourth part deals with multi-objective optimization. While the proposed algorithm features and working principle are in the fifth section, the graphical results of the algorithm on the related benchmarks are in the last part.

II. OPTIMIZATION

Optimization term as mathematical definition; to minimize or maximize a real function is to achieve the solution of the problem by specifying the value in the range where the intent is intrinsic and calling the function. Optimization process consists mainly of two components as modeling and solving. Modeling is the mathematical expression of the problem encountered in real life, and the analysis includes obtaining the best solution that provides this model [3].

Optimization can be divided into two parts according to the number of goals in problems involved: multipurpose and single purpose. In this article, TSA of single-purpose optimization methods and NSGA-II of multi-objective optimization methods are explained in the next sections in order to describe MOTSA.

III. TSA (TREE-SEED ALGORITHM)

TSA [4] is a heuristic iterative optimization algorithm. This algorithm usually achieves quickly and easily to the optimal or near optimal solution. In the TSA, each tree and seed found in the population represents a solution. There are two equations in the algorithm for seed production, which is candidate solution and these are given as follows:

$$S_{ij} = T_{ij} + \alpha_{ij} \times (B_j - T_{rj}) \quad (1)$$

$$S_{ij} = T_{ij} + \alpha_{ij} \times (T_{ij} - T_{rj})$$

For seed production, one them is used and the most important point is which equation will be chosen to create a new seed. This selection is controlled by the control parameter of the method called the search tendency, ST, in the range [0, 1]. The

higher value of ST provides strong local search and speed convergence, and the lower value of ST causes slow convergence but strong global search. Briefly, diversity is ensured in the seed production by this parameter. In other words, the TSA's search abilities are controlled by the ST parameter [4]. In the basic TSA, another important point is how many seeds will be produced for each tree. This is analyzed in the first study [4], and it is proposed that this can be properly adjusted by the stand size (number of trees). The number of seeds, which will be produced for each tree, is between 10% and 25% of the stand size. The step by step definition of the basic TSA is given in Figure 1.

Step 1. The initialization of the algorithm.
Set the number of population size (N).
Set the ST parameter for the method.
Set the dimensionality of the problem (D).
Decide the termination condition
Generate *N* random tree location on the *D*-dimensional search space using Eq. 5 (T).
Evaluate the tree location using objective function specified for the problem.
Select the best solution (B) using Eq.6.
Step 2. Searching with Seeds
FOR all trees
Decide the number of seeds produced for this tree.
FOR all seeds
FOR all dimensions
IF(*rand*<*ST*)
Update this dimension using Eq. 3 (S)
ELSE
Update this dimension using Eq. 4 (S)
END IF
END FOR
END FOR
Select the best seed and compare it with the tree.
If the seed location is better than tree location, the seed substitutes for this tree.
END FOR
Step 3. Selection of Best Solution
Select the best solution of population using Eq. 7.
If new best solution is better than the previous best solution, new best solution substitutes for the previous best solution.
Step 4. Testing Termination condition
If termination condition is not met, go to Step 2.
Step 5. Reporting
Report the best solution.

Figure 1: TSA Working Principle [4].

IV. NSGA-II (NON-DOMINATED SORTING GENETIC ALGORITHM)

A. *Genetic Algorithm*

Evolutionary algorithms that optimize problems by modeling biological processes are called genetic algorithms [5]. Genetic algorithms are an intuitive algorithm like TSA. Unlike broad search algorithms, they do not produce all possible different states to choose the best. In genetic algorithms, chromosomes come together with individuals and individuals to form a population. Population density is maximized or minimized by certain rules. Each new generation is created by combining survivors in sequences created by random information exchange. The working diagram of genetic algorithms is given in Figure 2.

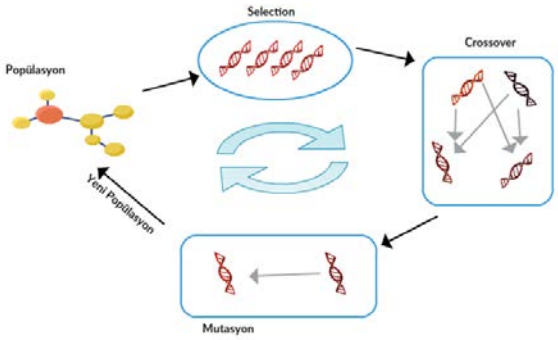


Figure 2: Working diagram of genetic algorithm.

B. *Pareto Optimal*

Almost all of the problems in the world are multi-purpose problems. These problems usually involve conflicting goals. Realize maximization and minimization in the concurrently for these conflicting purposes increases the complexity of the problem. In multi-objective optimization, because there are the objectives which are in competition with each other [6], the result is not a single output, but a set of optimal solutions. These solutions are known as Pareto-Optimal solutions. In the method of Pareto optimal, a vector containing all the objectives and the concept of dominance that allows you to choose between the solutions emerges.

The Pareto optimal method has two concepts:

1. The best solution is the solution that is not the worst in any of the goals and is better than the others in at least one goal.

2. The optimal solution is a solution that is not suppressed by any other solution in search space. Such an optimal solution is called the Pareto-Optimal Solution and the set of all these optimal solutions is called the Pareto-Optimal Set [7].

C. *Crowding Distance*

The crowding distance [8] value of any solution found in the optimization method refers to the density of the solutions around this solution.

Unlike other algorithms, the NSGA-II uses the crowding distance method in the determining rank levels process. To estimate the density of any point in the population; the average distance of two points located on both sides of the point being considered is calculated for all purposes. This amount is used to estimate the cubic perimeter length formed using the nearest neighbors [7]. The calculation of crowding distance is given in Figure 3.

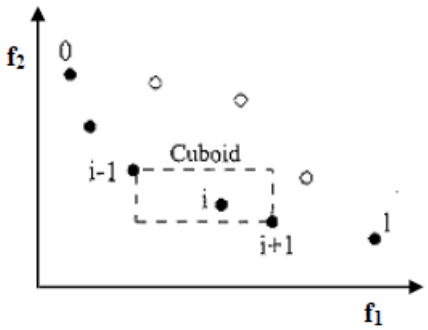


Figure 3: Calculate of Crowding Distance [8].

D. *Elitism*

In the selection process, a population of $R = P_0 + Q_0$ is created first. The population is sorted by the dominated levels. Then the crowding distance sequence is performed for each rank level in the population. A new population is created according to the order in the combined population [9].

The most important point of the elitism process is the sorting process. Population ranked according to rank levels, crowding distance ranking is performed again on each rank level members. After all the explanations, the working diagram of NSGA-II is given in Figure 4.

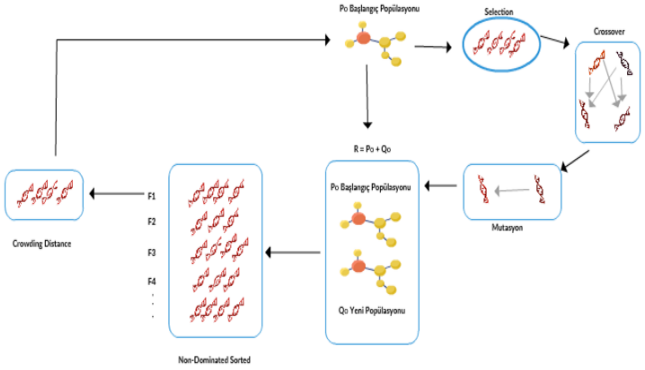


Figure 4: NSGA-II Working Principle.

V. MOTSA (MULTI-OBJECTIVE TREE SEED ALGORITHM)

In the proposed MOTSA, a solution similar to the NSGA-II is planned for multiple purposes of the TSA. In the algorithm, the function of producing seeds from trees according to the nature of TSA in seed production part is used, whereas the method of elitism of NSGA-II is used in comparison between seeds. The MOTSA's working diagram and algorithmic framework is given in Figure 5 and 6, respectively.

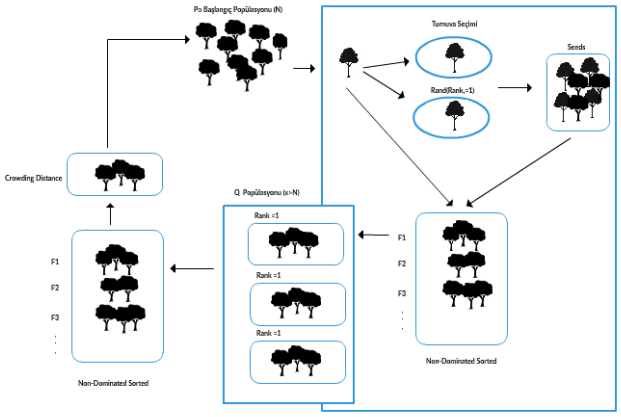


Figure 5: The working diagram of MOTSA

Firstly, the initial population size is determined and trees are randomly generated in the solution space. The fitness values of the generated trees are calculated. The dominated levels and rank values of the elements in the main population are determined.

The number of seeds to be formed from each tree in the population is determined randomly. Each seed is created using a randomly selected parent in the population with a Pareto-based tournament selection according to the ST parameter.

The elitism process is applied between the seeds and the trees. Determine rank levels and sorting operations are performed on the combined population. Crowding distance sorting is performed on the trees and seeds whose rank values are determined, and the most non-dominated solutions are determined. These operations are continued for the maximum iteration period and the best solution set is searched.

The MOTSA, which synthesizes the most powerful aspects of TSA and NSGA-II, provides fast and efficient results in search and sorting. In this respect, MOTSA is a method that can be developed for multi-purpose optimization researchers.

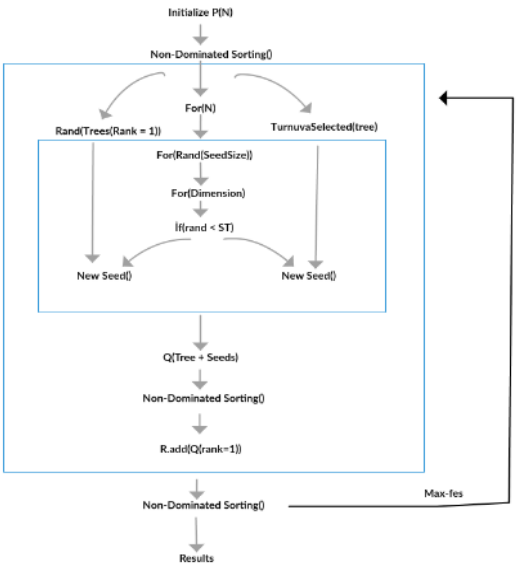


Figure 6: MOTSA Principle.

VI. RESULTS AND DISCUSSION

The proposed MOTSA designed for solving multi-objective problems has been applied to solve three benchmark problems, called Kursawe, Fonseca and Poloni. Under the different adjustment of control parameters, the MOTSA is run, and obtained results are presented in the following figures.

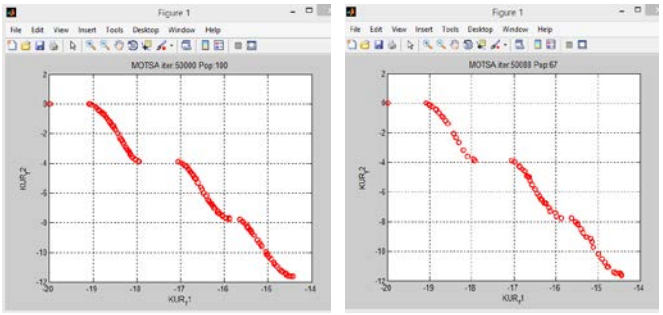


Figure 7. The results on Kursawe problem

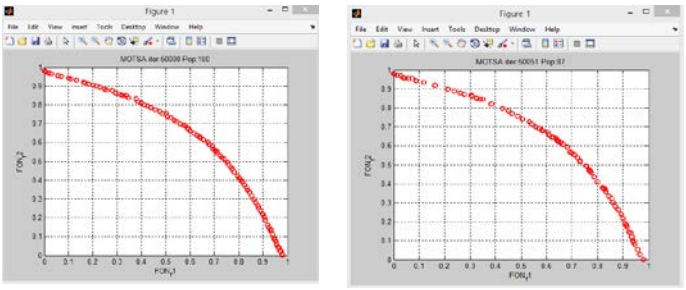


Figure 8. The results on Fonseca

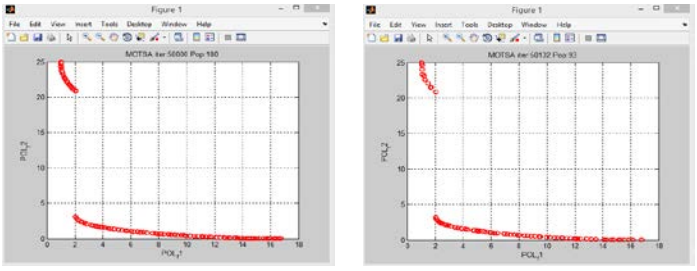


Figure 9. The results on Poloni problem

When we analyze the results given in the Figure 7, 8 and 9, the MOTSA produces the comparable and promising results. The stand size affects the performance of MOTSA on solving the problem.

VII. CONCLUSIONS AND FUTURE WORKS

In this study, a multi-objective optimization algorithm based on TSA has been proposed, named MOTSA. The crowding distance and non-dominated sorting strategies of NSGA-II has been integrated with MOTSA in the present work. The MOTSA has been applied to solve well-known three multi-objective benchmark problems, Kursawe, Fonseca and Poloni. The promising and comparable results are obtained by the MOTSA. In near future, we will apply it to solve a large set of benchmark problems and compare it with the state-of-art multi-objective optimization algorithms.

REFERENCES

[1] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, Ieee T Evolut Comput, 6 (2002) 182-197.

[2] Carlos A. Coello Coello, Gary B. Lamont and David A. Van Veldhuizen, 2007, Second Edition, Evolutionary Algorithms for Solving Multi-Objective Problems.

[3] Türkay, M. (2006), Optimizasyon Modelleri ve Çözüm Metotları, No;12, Ankara, 22-23.

[4] Kıran, M.S. (2015), TSA: Tree-seed algorithm for continuous optimization, Department of Computer Engineering, Faculty of Engineering, University of Selcuk, 42075 Konya, Turkey.

[5] Kubalik, J. Evolutionary Algorithms: Multi-Objective Optimization, Department of Cybernetics, CTU Prague

[6] Ergül, E.U. (2015), Çok Amaçlı Genetik Algoritma Yöntemlerinin Başarımının Belirlenmesi için İki Yeni Ölçüt Önerisi, Amasya, Türkiye.

[7] Sağ, T. ve Cunkaş, M. (2009),Çok Amaçlı Genetik Algoritmalar İçin Bir Çevrimdışı Performans Değerlendirmesi, Konya, Türkiye.

[8] Félix-Antoine Fortin ve Marc Parizeau, 2009, Revisiting the NSGA-II Crowding-Distance Computation, Canada.

[9] Open Source Engineering, Multı-objective optimization and genetic algorithm, Multiobjective optimization with NSGA-II

Solving of constrained problems via multi-objective vortex search algorithm

A.OZKIS[1], G. OZCAN[2], A. BABALIK[1], M.S. KIRAN[1]

[1] Selcuk University, Konya/Turkey, ahmetozkis@selcuk.edu.tr
[1] Selcuk University, Konya/Turkey, ababalik@selcuk.edu.tr
[1] Selcuk University, Konya/Turkey, mskiran@selcuk.edu.tr
[1]Selcuk University, Konya/Turkey, ozcanngull@gmail.com

Abstract **- In this study, multi-objective vortex search algorithm (MOVS), a new multi-objective metaheuristic optimizer, has been used to solve 3 different constrained benchmark problems. The performance of the MOVS algorithm is compared with NSGA-II, is a well-known multi-objective evolutionary algorithm, on different performance metrics. Obtained metric results show that the MOVS is a promising algorithm for solving multi-objective benchmarks. This study encourages researchers working on constrained mathematical or real-world problems to use the MOVS algorithm.**

Keywords **– multi-objective optimization problems, metaheuristic approach, constrained problems, multi-objective vortex search algorithm.**

I. INTRODUCTION

MOST of real-world problems involve multiple purposes and constraints. These problems are named as multi-objective optimization problems or shortly MOOPs. It is quite difficult to produce a satisfactory solution in a MOOP concurrently with each purpose and constraint. These difficulties arise from the following reasons: i) in a MOOP, objective functions often have an inverse relationship. So the solution that produces a good value for one objective can produce a bad value for another objective (See Figure 1). Therefore, there is no solution that produces the best value for all purpose functions. ii) A MOOP is usually a member of the class of NP-hard problems that cannot be solved at the polynomial time.

In 1950's, researchers first used classical mathematical methods (CMMs) to solve constrained NP-hard MOOPs. However, CMMs have trouble solving MOOPs because of two basic deficiencies: i) the success of CMMs is largely dependent on the MOOPs character. A CMM produces appropriate solutions for a MOOP, while may fail for another MOOP. ii) When solving a MOOP, the computational cost of CMMs is quite high. In order to overcome these two deficiencies, scientists have tried different approaches when solving MOOPs. One of these approaches is metaheuristic-based algorithms (MBAs). The MBAs have attracted the interest of researchers to produce successful results independently of the characteristics of the problems and at low computational cost. For nearly 20 years, many researchers have proposed multi-objective MBAs to solve MOOPs. The pioneering algorithms [1-5] proposed about 20 years ago, now contribute to the proposal of a large number of new algorithms [6-9]. In this work, the performances of the well-known NSGA-II [2] and recently proposed MOVS [9] algorithms are compared with 3 different constrained MOOPs using hypervolume (HV), spread, and epsilon metrics.

The remaining section of the study is organized as follows. Concept of multi-objective optimization is described in the next section. NSGA-II and MOVS algorithms are explained in section 3. Experimental results and discussion handled in Section 4. Findings and conclusions are presented in Section 5.

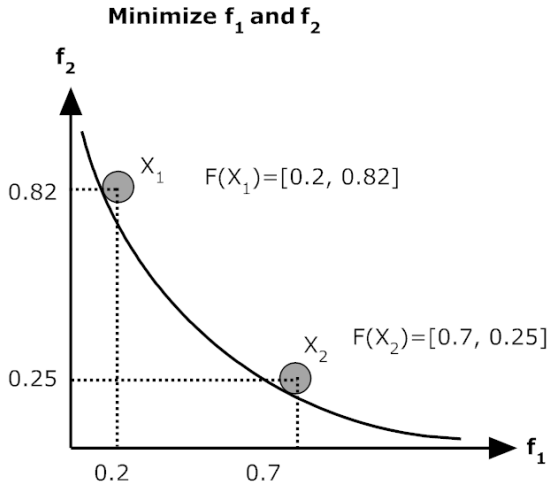


Figure 1: Inverse relationship of objective functions in a MOOP.

II. CONCEPT OF MULTI-OBJECTIVE OPTIMIZATION

A MOOP is mathematically expressed as follows [7]:

$$\begin{gathered} Minimize: F(\vec{z}) = [f_1(\vec{z}), f_2(\vec{z}), \dots, f_n(\vec{z})] \\ \text{Subject to:} \\ g_i(\vec{z}) \geq 0, i = 1,2, \dots, m \\ h_j(\vec{z}) = 0, i = 1,2, \dots, p \end{gathered} \quad (1)$$

Here $\vec{z} = [z_1, z_2, \dots, z_k]^T$ is vector of decision variables, k is the number of decision variable, m is the number of inequality constraints, p is the number of equality constraints, g_i is the i_{th} inequality constraint, h_j is the j_{th} equality constraint, n is the number of objective functions.

In Figure 1, there are two different solutions as X_1 and X_2. The X_1 solution has obtained 0.2 for f_1, while the X_2 solution has obtained 0.7. So, X_1 is better than X_2 for f_1 function. Conversely, X_2 is better than X_1 for f_2 function. In such cases a conflict arises and it is not known which solution is better. In this case, X_1 and X_2 are non-dominated solutions for each other. The following condition is required for a solution to dominate another solution:

$$if \begin{array}{l} \forall i \in \{1,2,\dots,n\} f_i(X_1) \le f_i(X_2) \\ \wedge \exists j \in \{1,2,\dots,n\} f_j(X_1) < f_j(X_2) \end{array}, \; X_1 \succ X_2 \qquad (2)$$

Here, if the solution X_1 is better than X_2 at least in one function, and not worse than X_2 in any objective function, X_1 dominates X_2 and it is denoted as $X_1 \succ X_2$. Non-dominated solutions (NDSs) refer to the most quality solutions in a population and objective vectors of NDSs is called as pareto front or shortly PF.

III. NSGA-II AND MOVS

A. *Non-dominated genetic algorithm–II (NSGA-II)*

NSGA-II [2] is one of the well-known stochastic algorithms proposed to solve MOOPs. NSGA-II ranks the solutions of a population according to their qualities using non-dominated sorting (NS) and crowding distance (CD) strategies and transfers the best solutions to the next generation. For more information about the NSGA-II, please review [2].

B. *Multi-objective Vortex Search (MOVS)*

VS [10] has an adaptively determined radius and a search is made around the determined center point using this radius. Adaptive radius balances exploration and exploitation processes during iterations. MOVS [9] is a modification of the single-objective vortex search (VS) algorithm to solve MOOPs. The modifications made in VS to suggest the MOVS algorithm are given below:

- NS and CD strategies are added,
- Center solution is determined using tournament selection strategy,
- Crossover operation is applied,
- Parameter "a" that determines the value of the search radius is randomly produced between 0 and 1.

For more information about the MOVS algorithm, please review [9].

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. *Constrained MOOPs*

In this study, three different constrained MOOPs have been used to compare NSGA-II and MOVS. These problems are given in Table 1.

Table 1: Constrained MOOPs used in this study.

Problem name	Number of decision variables	Number of objective function	Number of constraints
Speed Reducer	7	2	11
Welded Beam	4	2	4
Water	3	5	7

Speed Reducer Design Problem
Minimize:

$$f_1(x) = 0.7854 x_1 {x_2}^2 \left(\frac{10 {x_3}^2}{3} + 14.933 x_3 - 43.0934\right) - 1.508 x_1 ({x_6}^2 + {x_7}^2) + 7.477 ({x_6}^3 + {x_7}^3) + 0.7854 (x_4 {x_6}^2 + x_5 {x_7}^2)$$

$$f_2(x) = \sqrt{(745\, x_4 / x_2 x_3)^2 + 1.69 \times 10^7} / 0.1 {x_6}^2$$

$$\forall_i g_1 \le 0$$

$$g_1(x) = \frac{1.0}{(x_1 {x_2}^2 x_3)} - \frac{1}{27}$$
$$g_2(x) = \frac{1.0}{(x_1 {x_2}^2 {x_3}^2)} - \frac{1}{397.5}$$
$$g_3(x) = \frac{{x_4}^3}{(x_2 {x_3}^2 {x_6}^4)} - \frac{1}{1.93}$$
$$g_4(x) = \frac{{x_5}^3}{(x_2 x_3 {x_7}^4)} - \frac{1}{1.93}$$
$$g_5(x) = x_2 x_3 - 40$$
$$g_6(x) = x_1 / x_2 - 12$$
$$g_7(x) = 5 - x_1 / x_2$$
$$g_8(x) = 1.9 - x_4 + 1.5 x_6$$
$$g_9(x) = 1.9 - x_5 + 1.1 x_7$$
$$g_{10}(x) = f_2(x) \le 1300$$

$$a = 745 x_5 / x_2 x_3$$
$$b = 1.575 \times 10^8$$
$$g_{11}(x) = \frac{\sqrt{a^2 + b}}{0.1 x_7^3} \le 1100$$

Table 2: Lower and upper bounds of decision variables of speed reducer design problem

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
LB	2.6	0.7	17	7.3	7.3	2.9	5
UB	3.6	0.8	28	8.3	8.3	3.9	5.5

Welded Beam Design Problem
Minimize:

$$f_1(x) = 1.10471 * x(1)^2 * x(2) + 0.04811 * x(3) * x(4) * (14.0 + x(2))$$

$$f_2(x) = 65856000 / (30 * 10^6 * x(4) * x(3)^3)$$

$$\forall_i g_1 \le 0$$

$$g_1(x) = \tau - 13600$$
$$g_2(x) = \sigma - 30000$$
$$g_3(x) = x(1) - x(4)$$
$$g_4 = 6000 - P$$

$$q = 6000 * \left(14 + \frac{x(2)}{2}\right); \; D = sqrt\left(\frac{x(2)^2}{4} + \frac{(x(1) + x(3))^2}{4}\right)$$
$$j = 2 * (x(1) * x(2) * sqrt(2) * (\frac{x(2)^2}{12} + \frac{(x(1)+x(3))^2}{4}))$$

$$a = \frac{6000}{sqrt(2) * x(1) * x(2)} \qquad \beta = Q * \frac{D}{J}$$

$$\tau = sqrt(\alpha^2 + 2 * \alpha * \beta * \frac{x(2)}{2*D} + \beta^2)$$

$$\sigma = \frac{504000}{x(4) * x(3)^2} \qquad tmpf = 4.013 * \frac{30 * 10^6}{196}$$

$$P = tmpf * sqrt(x(3)^2 * \frac{x(4)^6}{36}) * (1 - x(3) * \frac{sqrt(\frac{30}{48})}{28})$$

Table 3: Lower and upper bounds of decision variables of welded beam design problem

	x_1	x_2	x_3	x_4
LB	0.125	0.1	0.1	0.125
UB	5	10	10	5

Water Problem

$$f_1 = 106780.37 * (x(2) + x(3)) + 61704.67$$
$$f_2 = 3000 * x(1)$$
$$f_3 = 305700 * 2289 * x(2)/(0.06 * 2289)^{0.65}$$
$$f_4 = 250 * 2289 * exp(-39.75 * x(2) + 9.9 * x(3) + 2.74)$$
$$f_5 = 25 * (1.39/(x(1) * x(2)) + 4940 * x(3) - 80)$$

$$g_1(x) = 1 - (0.00139/(x(1) * x(2)) + 4.94 * x(3) - 0.08)$$
$$g_2(x) = 1 - (0.000306/(x(1) * x(2)) + 1.082 * x(3) - 0.0986)$$
$$g_3(x) = 50000 - (12.307/(x(1) * x(2)) + 49408.24 * x(3) + 4051.02)$$
$$g_4(x) = 16000 - (\frac{2.098}{x(1) * x(2)} + 8046.33 * x(3) - 696.71)$$
$$g_5(x) = 10000 - (\frac{2.138}{x(1) * x(2)} + 7883.39 * x(3) - 705.04)$$
$$g_6(x) = 2000 - (0.417 * x(1) * x(2) + 1721.26 * x(3) - 136.54)$$
$$g_7(x) = 550 - \left(\frac{0.164}{(x(1)+x(2))} + 631.13 * x(3) - 54.48\right)$$

Table 4: Lower and upper bounds of decision variables of water problem

	x_1	x_2	x_3
LB	0.01	0.01	0.01
UB	0.45	0.1	0.1

B. Experimental Settings

In this study, experiments have been conducted using these parameters: agent numbers of both algorithms are set as 50, number of maximum evaluation is set as 10,000 and runtime is set as 30 for each case. Hypervolume (HV), spread, and epsilon metrics have been used to evaluate the performance of the algorithms. The spread metric evaluates the dispersion performance of the PFs generated by the algorithms while evaluating the EP convergence performance. HV metric evaluates both convergence and dispersion performance. It is desirable that the value obtained for the HV metric is large, while small for other metrics. The mean and standard deviation values obtained by metrics for 30 runs are given in Table 5. The above value in each cell is the mean, and the bottom value is the standard deviation. Results that have better values are written in bold for easy reading.

Table 5: Mean and standard deviation values of metrics for 30 runs.

		Speed Reducer	Welded Beam	Water
HV	NSGA-II	9.67e-01 4.7e-04	9.00e-01 1.5e-02	3.46e-01 1.8e-02
	MOVS	**9.68e01** **3.2e-04**	**9.01e-01** **5.4e-03**	**3.61e-01** **1.7e-02**
SPREAD	NSGA-II	4.77e-01 4.4e-02	**5.69e-01** **5.9e-02**	5.85e-01 6.6e-02
	MOVS	**4.58e-01** **6.1e-02**	6.00e-01 6.7e-02	**5.58e-01** **5.0e-02**
EPSILON	NSGA-II	1.74e+01 3.1e+00	**7.25e-01** **8.1e-01**	1.52e+05 2.8e+04
	MOVS	**1.61e+01** **3.0e+00**	7.34e-01 4.6e-01	**1.39e+05** **2.9e+04**

The following conclusions are made when the results given in the Table 5 are examined:

For HV metric, it is seen that MOVS is slightly better than NSGA-II on speed reducer and welded beam problems. In addition, MOVS is clearly better than NSGA-II on water problem.

For spread metric, MOVS is better than NSGA-II on speed reducer and water problems while NSGA-II is better than MOVS on welded beam problem.

Epsilon metric results are similar to spread metric. MOVS is better than NSGA-II on speed reducer and water problems while NSGA-II is better than MOVS on welded beam problem.

The results given in Table 5 show that the MOVS algorithm produces generally more successful results than NSGA2 on these three problems. Moreover, the standard deviations obtained indicate that the MOVS algorithm is a robust and stable algorithm.

V. CONCLUSION

In this work, the performances of the well-known NSGA-II and recently proposed MOVS algorithms are compared with 3 different constrained MOOPs using HV, spread, and epsilon metrics. Obtained results show that MOVS algorithm has achieved better results than NSGA-II almost all cases for used problems. It is also seen that MOVS is a stable algorithm with low standard deviations obtained. The results obtained in this study encourage scientists working in the field of multi-objective optimization to use the MOVS algorithm.

ACKNOWLEDGMENT

The authors of this study would like to thank Scientific Research Projects Coordinatorship at Selcuk University and The Scientific and Technological Research Council of Turkey for their institutional supports.

REFERENCES

[1] D.W. Corne, J.D. Knowles, M.J. Oates, The Pareto envelope-based selection algorithm for multiobjective optimization, in: International Conference on Parallel Problem Solving from Nature, Springer, 2000, pp. 839-848.

[2] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, Ieee T Evolut Comput, 6 (2002) 182-197.

[3] J. Knowles, D. Corne, The pareto archived evolution strategy: A new baseline algorithm for pareto multiobjective optimisation, in: Evolutionary

Computation, 1999. CEC 99. Proceedings of the 1999 Congress on, IEEE, 1999, pp. 98-105.
[4] A.J. Nebro, J.J. Durillo, J. Garcia-Nieto, C.A.C. Coello, F. Luna, E. Alba, SMPSO: A New PSO-based Metaheuristic for Multi-objective Optimization, Mcdm: 2009 Ieee Symposium on Computational Intelligence in Multi-Criteria Decision-Making, (2009) 66-+.
[5] E. Zitzler, M. Laumanns, L. Thiele, SPEA2: Improving the strength Pareto evolutionary algorithm, TIK-report, 103 (2001).
[6] C. Dai, Y.P. Wang, M. Ye, A new multi-objective particle swarm optimization algorithm based on decomposition, Inform Sciences, 325 (2015) 541-557.
[7] G. Dhiman, V. Kumar, Multi-objective spotted hyena optimizer: A Multi-objective optimization algorithm for engineering problems, Knowledge-Based Systems, (2018).
[8] S. Mirjalili, P. Jangir, S. Saremi, Multi-objective ant lion optimizer: a multi-objective optimization algorithm for solving engineering problems, Appl Intell, 46 (2017) 79-95.
[9] A. Ozkis, A. Babalik, A novel metaheuristic for multi-objective optimization problems: The multi-objective vortex search algorithm, Inform Sciences, 402 (2017) 124-148.
[10] B. Doğan, T. Ölmez, A new metaheuristic for numerical function optimization: Vortex Search algorithm, Inform Sciences, 293 (2015) 125-145.

A Clustering Ranking Based Multiobjective Evolutionary Algorithm

E. DASDEMIR[1], B. Y. OZCAN[2]

[1]Hacettepe University, Ankara/Turkey, edasdemir@hacettepe.edu.tr
[2]Middle East Technical University, Ankara/Turkey, begum.ozcan@metu.edu.tr

Abstract **- We propose a new clustering ranking based multiobjective evolutionary algorithm. The algorithm uses decision maker's preferences to reduce the search space and obtain a final set of preferred Pareto-optimal solutions. A new clustering ranking operator using Hierarchical Clustering on Principle Components (HCPC) and K-means methods is developed. We also develop a new crossover operator. The algorithm is implemented on several problems. The work is still in progress.**

Keywords **– Multiobjective evolutionary algorithms, heuristic search, clustering ranking, crossover.**

I. INTRODUCTION

THE field of multiobjective optimization deals with the optimization problems consisting of more than one objective. A final set of solutions, which is called as Pareto-optimal frontier or set in the literature, is found at the end. Classical optimization methods are often able to end up with a single solution at a time. However, Multiobjective Evolutionary Algorithms (MOEA) have different approaches than the classical optimization procedures. Instead of finding a single exact solution, they are able to approximate the well-converged and diversified unknown set of Pareto-optimal solutions Therefore, they are suitable for multiobjective optimization problems and they have been very popular in the literature [1].

Multiobjective evolutionary algorithms simulates the natural evolution. An initial population of solutions is generated. Genetic operators; selection, crossover and mutation is applied to generate new solutions. The solutions with better objective function values survives one generation to another. The two ultimate goal of classical MOEA are converging to the Pareto-optimal frontier and finding a set of solution as diverse as possible. Then decision maker makes his final decision among these final set of solutions. Usefulness of this approach in low dimensional problems have been shown many times; however, difficulties start to occur when the number of objectives increases. Since the number of solutions in Pareto-optimal set increases proportionally with problem complexity, it gets computationally costly to find all of them. Moreover, visualization problem occurs with increased dimensionality. For example, Fowler et al. [2] explained that many difficulties such as computational and visualization problems occur with complex problems. An alternative idea is using the preference of decision makers at the time that the algorithm runs to direct the search to the preferred regions of the Pareto-optimal set. For example, Deb and Himanshu [3] stated, existing EMO methodologies can be still very functional, if a desired region of the Pareto-optimal front can be emphasized. These algorithms are called as preference-based multiobjective evolutionary algorithms and have been gained popularity recently.

II. METHODOLOGY

In this research, we developed a preference-based MOEA. Our preference based heuristic approach reduces the search area of multiobjective optimization problems and ends up with decision maker's preferred solutions. Decision maker gives his preferences with reference points. Then the algorithm tries to approximate solutions close to the decision maker's reference points. For the general structure, we have used a similar approach to the well-known algorithm NSGA-II [4], which unfortunately does not consider the decision maker's preferences. To converge to the preferred region of the Pareto-optimal frontier, we developed a new clustering ranking operator. We replaced the crowding distance operator of NSGA-II with this new operator. We cluster population in each generation, find cluster centers, calculate the distance between each cluster's center and reference points, then we assign better ranks to the solutions in closer clusters. As another innovation, we developed a new crossover operator, Cheating Based Crossover (CBC). During the crossover of two parents, we allow them to cheat each other with a third parent, which affects the genes of new offspring. Instead of usual crossover operators, three parents are used and three offspring are generated. This increases the probability of generating new tours in earlier generations. CBC will help practitioners to explore different regions of the solution space in earlier generations.

III. DEVELOPED EVOLUTIONARY ALGORITHM

The steps of the developed clustering ranking based multiobjective evolutionary algorithm are presented as follows:

Step 1. Generate a parent population.
Step 2. Apply nondominated and clustering ranking to sort parent population.
Step 3. Apply genetic operators to create offspring population.
Step 4. Combine parent and offspring population.
Step 5. Apply nondominated ranking and clustering raking to sort combined population members.
Step 6. Select next parent population from combined population.

Step 7. Repeat the steps 3-6 until a predetermined termination condition is satisfied.

Nondominated and clustering ranks obtained in *Step 2*, used in *Step 3* during tournament selection. The solution with better nondominated and clustering rank defeat the other one. After *Step 5*, to proceed next generation, a new parent population need to be selected from the combined population. This selection starts from solutions with better nondominated ranks. Whenever the remaining space is not enough the cover all solutions from the same nondominated rank, solutions are selected using their clustering rank, i.e. solutions closer to the reference points are preferred

IV. CLUSTERING RANKING OPERATOR

Clustering ranking operator works as follows: Population members are clustered into different clusters and the centroids of each cluster is found. The euclidean distance between reference point(s) and cluster centroids are calculated. The solutions in clusters closer to the reference point(s) are prioritized in both tournament selection and population size reduction. Hierarchical Clustering on Principle Components (HCPC) and K-means clustering methods are applied using the objective function values as attributes. HCPC is used for identifying groups of similar observations in a data set. HCPC requires Ward's Criterion (WC) to perform hierarchical clustering on principle clustering [5]. In K-means clustering approach, data is partitioned into the clusters based on the Euclidean similarity metric [6]. Clustering process is applied onto the population at the beginning of each of the iterations prior to genetic operators and algorithm steps in order to maintain diversity and to supervise population toward the reference point proposed by the decision maker.

V. CONLUSIONS AND FUTURE WORK

Developed preference-based MOEA focuses to the preferred regions of the Pareto frontier instead of wasting time to compute all Pareto-optimal frontier. The algorithm is demonstrated on several test problems. Experiments have shown that the algorithms succeed in finding Pareto-optimal solutions closer to the reference points defined by decision maker. The results are promising with different clustering approaches. We compeared two different clustering approaches, i.e. K-means and HCPC, for our clustering ranking operator. Both methods work wells, however K-means is faster and easier than the HCPC to use for smaller number objectives. However, if the number of objectives increases, it will be better to use HCPC; because, it uses PCA and reduces the dimensionality of the data. We also compared the effect of number of clusters using different cluster numbers. If decision maker wants to a focus more to a specific region closer to the reference point, larger number of cluster can be used. On the other hand, if decision maker wants to obtain a larger set of solutions, low number of cluster should be used. Our algorithm is able to work with multiple reference point simultaneously. This is another beauty of the algorithm; because, the decision maker may not be sure about his preferences to provide a single reference point or he may be interested in the different regions of the Pareto-optimal frontier too. Hence, the developed algorithm helps decision makers to explore different regions simultaneously.

Note that this work is still in progress. The results and conclusions are preliminary because the studies are still ongoing. Therefore, numerical results are not presented in this paper. In future, the efficiency of the algorithm need to utilized and the algorithm needs to be tested on more problems. An additional study can be done to increase the efficiency of eliciting reference information from the decision maker.

REFERENCES

[1] K. Deb, *Multi-objective optimization using evolutionary algorithms*. Vol. 16. John Wiley and Sons, 2001.

[2] J.W. Fowler, E.S. Gel, M. Köksalan, P. Korhonen, J. L. Marquis, J. Wallenius, "Interactive evolutionary multi-objective optimization for quasi-concave preference functions," *European Journal of Operational Research*, vol. 206-2, pp. 417-425, 2010.

[3] K. Deb, J. Himanshu, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints," *IEEE Trans. Evolutionary Computation*, vol. 18-4, pp. 577-601, 2014.

[4] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6-2, pp. 182–197, 2002.

[5] F. Husson, J. Josse and J. Pages, "Principal component methods-hierarchical clustering-partitional clustering: why would we need to choose for visualizing data?" *Technichal Report – Agrocampus*, Applied Mathematics Department. 2010.

[6] J. MacQueen, "Some methods for classification and analysis of multivariate observations," *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1-14, pp. 281-29, 1967.

A Multiobjective Evolutionary Algorithm Approach to Employee Bus Transportation Problem

E. DASDEMIR[1], M. C. TESTIK[1], O. M. TESTIK[1], C. TUNCER SAKAR[1]

[1]Hacettepe University, Ankara/Turkey,
edasdemir@hacettepe.edu.tr, mtestik@hacettepe.edu.tr, ozlemaydin@hacettepe.edu.tr, cerents@hacettepe.edu.tr

***Abstract* - A multiobjective bus transportation problem is studied in this research. Due to the complexity of the problem, a multiobjective evolutionary algorithm based solution approach is developed. The solution approach is implemented on a real world problem. The preliminary results are promising. The research is still in progress.**

***Keywords* - employee bus transportation, routing, multiobjective evolutionary algorithms, heuristic search.**

I. INTRODUCTION

EMPLOYEE bus transportation can be defined as providing transportation services to spatially distributed employees between their residence and the workplace. This problem is usually known as school bus routing problem in the literature and belongs to the NP hard class of problems. It shares similar properties with traditional vehicle routing problems; however, it involves additional objectives and restrictions. In this study, a solution approach is developed for the employee bus transportation problem with multiple objectives and complicating constraints. Multiobjective Evolutionary Algorithms (MOEA) is employed for the solution. Please note that this work is still in progress. Although the preliminary results obtained are promising, more study on the efficiency of the evolutionary algorithm and on the quality of the results need to be done.

II. PROBLEM DEFINITION AND LITERATURE REVIEW

Solution of this problem involves assignments of the busses to the pickup points and finding the shortest routes for each bus to collect passengers and transport them to the workplace. Hence, the assignment and shortest path problems are required to be solved simultaneously. In our study, there are three objectives that need to be considered simultaneously, which make the problem more complicated. First objective aims to find minimum number of required buses. There is also an upper bound for the number of available busses due to budget constraints. Therefore, the utilized number of buses is required to be smaller than this upper bound. Second objective aims to minimize the distance traveled by a bus, which involves picking up the employees from pick up points and transporting them to the workplace by following the shortest route between these points. Third objective aims to serve maximum number of employees. Due to the available upper bound on number of busses, some employees cannot be transported. The aim is to transport as many passengers as possible.

There are also several parameters that need to be determined and some assumptions need to be considered. Coordinates of the pickup points need to be predetermined. The distance required to travel until first pickup point by a bus is ignored, i.e. a bus starts its trip from first pickup point. The capacity of the busses used to pick up employees are limited and the number of employees to be transported by a bus should not exceed its total capacity. A bus can be assigned to several pickup points to collect passengers during its travel from a starting point to the workplace. However, a pickup point can only be assigned to single bus, i.e. all employees waiting in the same pickup point need to get on the same bus.

In multiobjective optimization, instead of a single solution as in a single objective case, a set of "best" solutions is found at the end. This final solution set is called the Pareto-optimal frontier. In this set, each solution is better than another one in at least one objective [1]. This final set is presented to a decision maker for a final decision. Although optimization problems are usually modeled and solved with mathematical methods, due to difficulties in complex problems solution time of mathematical approaches increases exponentially with problem size. Therefore, heuristic approaches have been developed, where the aim is to approximate optimal solutions by searching the solution space with effective and quick methods. Hence, a good solution is found in a reasonable time. Some of the well-known heuristic approaches are evolutionary algorithms, simulated annealing, particle swarm optimization, etc. [2].

Many researchers addressed similar bus routing problems in the literature. For example, Bowerman et al. [3] presented a multiobjective mathematical formulation for the school bus routing problem end employed heuristics for the solution. Schittekat et al. [4] developed a metaheuristic for bus routing problem with bus stop selection.

III. SOLUTION APPROACH

Multiobjective Evolutionary Algorithms (MOEA) have been developed for solving optimization problems with multiple objectives. These methods are able to approximate the well-converged and diversified set of Pareto-optimal solutions. MOEA are population based algorithms mimicking natural evolution. Starting with an initial population of solutions, population members with better objective function values survive through generations.

MOEA algorithms are well suited to the employee bus transportation problem considered here due to its multiobjective structure and complexity. An initial population of solutions are generated. Each solution is represented by its fitness value

(objective function values). Genetic operators, selection, crossover and mutation are applied to the population members in each generation. Solutions with better objective function values survive through the generations. In each generation, new solutions (offspring) are obtained by crossover, and structure of some solutions are changed using mutation. These processes repeated until the termination condition is satisfied. Final solution set constitutes Pareto-optimal frontier.

The NSGA-II, an MOEA algorithm developed by Deb [5], is employed here for a solution procedure. NSGA-II prefers nondominated solutions and keep diversity among the solutions from one generation to another. Representation, crossover and mutation operators of the original study are modified here to fit the special characteristics of this problem. Each solution in the population corresponds to an assignment. For example, for a problem with five pickup points and three busses, 1-3-2-1-3 is a solution, where the first bus is assigned to the 1st and 4th, the second bus is assigned to the 3rd, and the third bus is assigned to the 2nd and 5th pickup points. A solution is represented with its objective function values in the population. Initial population is randomly generated and then selection, crossover and mutation operators are applied during the generations to eliminate poor solutions and increase diversity. For selection, common tournament selection operator is used. Two solutions from the population are selected, tournaments are played between them, and good ones win and weak ones are eliminated. Solutions are first evaluated first based on their nondominated ranks and then based on their crowding distance rank. Details of the nondominated and crowding distance ranking can be found in the original paper. At the end, all winner solutions constitute a mating pool. Crossover operator is applied to the solutions in the mating pool that consist of tournament winners. The main purpose is to create new solutions to explore the solution space. Position Based Crossover is used in our algorithm. We also use mutation operator to increase the diversity in the population.

IV. IMPLEMENTATION

The approach described is implemented on a real world problem. The algorithm is programmed in R programing language. Coordinates of pickup points and the number of passengers waiting to be served in these points were given. The distances between these points are found by using the Google Maps Distance Matrix API. We run the API using R's "gmapsdistance" package. It is assumed that each bus can transport at most 45 passengers. Different upper bounds on the number of busses are considered. After running the evolutionary algorithm and Using the available information from the decision maker, several solutions are found.

The real world problem has 103 pickup points. There is an upper bound for the number of available busses. Each bus can carry 45 passengers. The number of employee need to be carried is much more than the total available seats. The results of the experiments on this problem are promising. However, more studies on the efficiency of the algorithm need to be done. Moreover, quality of the solutions need to be improved.

V. CONLUSION AND FUTURE STUDIES

The work is still in progress. Studies are ongoing on the evolutionary algorithm. After heuristic approach is completed, e.g. in future studies, the problem will be solved by using exact mathematical methods and the results will be compared with heuristic approaches. The efficiency of the solution approach also be improved and the approach may be tested on problems with different parameter settings.

REFERENCES

[1] K. Deb, *Multi-objective optimization using evolutionary algorithms, vol 16*. John Wiley & Sons, 2001.

[2] E. A. Silver, "An overview of heuristic solution methods," *Journal of The Operational Research Society*, vol. 55-9, pp. 936-956, 2004.

[3] R. Bowerman, B. Hall, and P. Calamai, "A multi-objective optimization approach to urban school bus routing: Formulation and solution method," *Transportation Research Part A: Policy and Practice*, vol. 29-2, pp. 107-123, 1995.

[4] P. Schittekat, J. Kinable, K. Sörensen, M. Sevaux, F. Spieksma, and J. Springael, "A metaheuristic for the school bus routing problem with bus stop selection,". *European Journal of Operational Research*, vol. 229-2, pp. 518-528, 2013.

[5] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6-2, pp. 182–197, 2002

Performance analysis of Galactic Swarm Optimization with Tree Seed Algorithm

E.KAYA[1], O.UYMAZ[2], S.KORKMAZ[3], E.SIRAMKAYA[4], M.S. KIRAN[5]

1 Selcuk University, Konya/Turkey, ersinkaya@selcuk.edu.tr
2 Selcuk University, Konya/Turkey, ouymaz@ selcuk.edu.tr
3 Selcuk University, Konya/Turkey, sedatkorkmaz@ selcuk.edu.tr
4 Selcuk University, Konya/Turkey, esiramkayan@ selcuk.edu.tr
5 Selcuk University, Konya/Turkey, mskiran@ selcuk.edu.tr

***Abstract* - The Galactic Swarm Optimization (GSO) is a novel method inspired by the movements of stars and star clusters. The GSO is a framework that uses the optimization methods known in the literature. GSO has a two-stage structure. In the first stage, the optimization method identifies possible good solutions by scanning the search space. In the second stage, the best solution is tried to be found by using possible good solutions. In the original GSO study, Particle Swarm Optimization (PSO) was used as an optimization method in both stage. In this study, Tree Seed Algorithm (TSA), a new optimization method in the literature, is used in GSO framework instead of PSO. In the experimental study, the performance of the GSO_TSA model has been investigated on numeric benchmark functions and obtained results are compared with GSO_PSO model.**

***Keywords* - Galactic Swarm Optimization, Particle Swarm Optimization, Tree Seed Algorithm, Benchmark Function.**

I. INTRODUCTION

Optimization refers to the process of minimizing or maximizing a function, examining or solving a problem systematically by selecting the objective and integer or integer values in a defined range and placing in the function. There are many algorithms in the literature related to optimization. Some of those; Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Artificial Bee Colony (ABC), Tree Seed Algorithm (TSA), Whale Optimization Algorithm (WOA), Gravitational Search Algorithm (GSA)[1-6].

Galactic Swarm Optimization (GSO) algorithm is a framework that is inspired by stars and planets. It is basically an algorithm running on two levels. In the first level, it is divided the population into a specified number of separate pieces. These pieces are called subpops. This subpopulation is implemented by a determined optimization algorithm. The best results from these subpopulations constitute the population of the second level. At the second level, the specified optimization algorithm is applied to this population. This process is repeated at the specified number of times. In the original GSO method, the particle swarm optimization method was used as the optimization method in the first and second stages.

Tree Seed Algorithm is a population-based heuristic search algorithm that has been recently proposed to solve continuous optimization problems. TSA has achieved considerable success in continuous optimization problems. This algorithm basically models trees and seeds. One or more seeds are produced from the trees. If the seed produced is better, it is replaced by the location of the tree. In determining the new locations for the seeds, the best solution produced or another tree location is taken into account.

In this study, a new approach is presented using the TSA method within the GSO framework. The proposed approach and traditional TSA have been tested with continuous optimization problems frequently used in the literature. The results have shown that the TSA used in the GSO framework gives better results than the traditional TSA.

II. METHODS

A. Galactic Swarm Optimization

Galactic Swarm Optimization has been developed by Muthiah-Nakaraj et al in 2016[7]. GSO basically accepts the movements of galaxies and stars. The GSO algorithm is a two-level framework that uses optimization algorithms. It basically consists of two levels. In the first level, the subpopulation is divided into a certain number of subpopes, and the optimization algorithm is run on the subpopulations. In the first level, the optimization algorithm for each subpopulation is applied and the best solutions are obtained for each subpopulation. The best results from the first level are used to construct the initial population of the second level. In the second step, the optimization algorithm determined on this population is applied. This process is repeated for the number of iterations. The best individual obtained is presented as a solution.

In the original GSO, the PSO algorithm is used. PSO is a population-based optimization algorithm. The PSO was introduced by Kenedy and Eberhard in 1995, inspired by the social behavior of bird flocks and fish schools[8]. In this study, we use the TSA method instead of the PSO as the optimization method. Flowchart of GSO framework that implemented by TSA is given in figure 1.

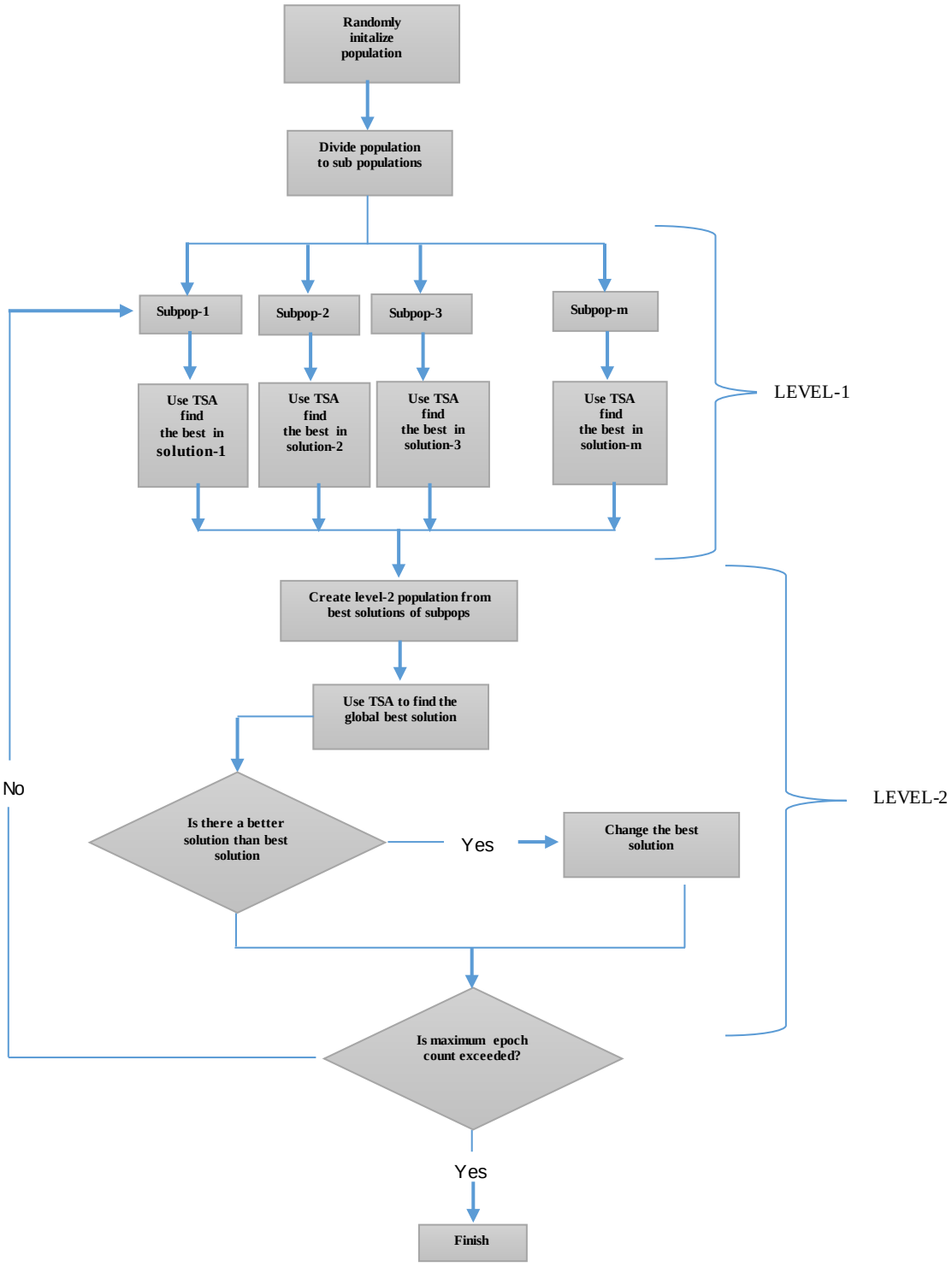


Figure 1: Flowchart of GSO framework that implemented by TSA

B. Tree Seed Algorithm

Tree-Seed Algorithm, TSA for short, is a novel population-based iterative optimization algorithm proposed for solving continuously-structure solution space by inspiring the relation between and their seeds in nature. In TSA, the location of trees and theirs seeds represent the

possible solution of the optimization problem and at the each iteration, a number of seeds is produced in order to search the solution space effectively. In the initialization of the algorithm, a certain number of trees (stand) is generated as follows:

$$T_{i,j} = T_j^{min} + r_{i,j}(T_j^{max} - T_j^{min}) \quad (1)$$

where, $T_{i,j}$ is the jth dimension of the ith tree, T_j^{min} and T_j^{max} are the lower and upper bound defined for the jth dimension, respectively, and $r_{i,j}$ is a uniform random number in range of [0,1]. After the stand is created, the fitness of each tree to the nature is calculated by using objective function specific for the optimization problem.

In TSA, there are two different solution update rule or seed creation mechanism, and which of them is selected for creating a seed is depended on a control parameter of the algorithm named as search tendency-ST. ST is in range of [0,1] and its high value provides fast convergence and local search, and lower value provides global search but causes slow convergence. In order to create a seed, the equations given below is used.

$$S_{i,j} = \begin{cases} T_{i,j} + \alpha_{i,j} \times (B_j - T_{r,j}) & if\ rand < ST \\ T_{i,j} + \alpha_{i,j} \times (T_{i,j} - T_{r,j}) & otherwise \end{cases} \quad (2)$$

where, S_i is the seed created for T_i, $T_{r,j}$ is a neighbor tree, B is the best tree based on the fitness, $\alpha_{i,j}$ is a random number in range of [-1,1] and rand is a random number produced in range of [0,1].

By using Eq. 2, a number of seeds are created for each tree and the best seed is fixed and compared with the current tree. If the fitness of the best seed is better than the fitness of the current tree, the current tree is withered and the new seed is located to the stand as a new tree. The number of seeds in TSA is an important point because search or optimization is done by these seeds. In the referenced study[4], this is controlled properly by depending on the stand size and it is randomly determined for each tree. The minimum and maximum number of seeds are selected as 10% and 25% of the stand size (N), respectively. The TSA is an iterative algorithm and there must be a termination condition for the algorithm. In the basic study of TSA, the maximum number of function evaluations (Max_FEs) is selected as termination condition of the algorithm.

The pseudo code of the algorithm is briefly presented as follows:

1. Determine the control parameters, N, ST, Max_Fes
2. Create Trees using Eq. 1.
3. Calculate Fitness of the Trees and save them.
4. Set FEs to N.
5. For each tree
    a. Determine the number of seeds for this tree.
    b. Produce seeds for this tree by using Eq. 2
    c. Select the best seed and compare it with the current tree
    d. If the best seed is better than the current tree, wither the current tree and locate the seed to the stand as a tree.
    e. Increase FEs by the number of produced seeds.
6. If FEs is equal or higher than Max_FEs, report the best solution; otherwise go to Step5.

III. EXPERIMENTAL RESULTS

In this study, GSO framework with TSA has been represented. The proposed method is compared with conventional TSA. 12 benchmark functions are used for comparison. The benchmark functions are detailed in Appendix [9, 10]. The benchmark functions are run in 30, 60 and 100 dimensions in the experimental study. In both methods, the maximum number of fitness calculation for 30, 60 and 110 dimensions was determined as 150,000, 300,000 and 450,000 respectively. The population size for the methods was set at 50. In TSA, search tendency (ST) parameter is taken as 0.1. The number of subpops for GSO is 10 and the number of individuals for each subpop is 5. The number of repetitions of both steps in the GSO method was set at 5.

All experimental studies were carried out independent 30 runs. Experimental results in 30, 60 and 100 dimensions are shown in Table 1, Table 2 and Table 3, respectively. In the tables, mean and standard deviations of 30 run are shown.

Table 1: Experimental Results for 30 Dimension

	GSO with TSA		**TSA**	
Function	**Mean**	**Std.Dev**	**Mean**	**Std. Dev.**
F1	**1,25E-08**	**1,26E-08**	0,055760013	0,032569647
F2	**3,41E-06**	**4,03E-06**	129936,239	79676,61338
F3	**9,68E-10**	**1,14E-09**	0,00647672	0,002625197
F4	**2,57E-06**	**4,43E-06**	331498,6658	306391,2741
F5	**1,52E-07**	**1,06E-07**	0,2826322	0,106916884
F6	137065,6667	61311,64533	**134605,2333**	**43531,14116**
F7	**0**	**0**	**0**	**0**
F8	**191251,8333**	**96567,10629**	638187,2667	199227,217
F9	**75,46737787**	**404,9161774**	406434,2	275459,7163
F10	**5678,195208**	**30573,40137**	335402,2333	231967,2558
F11	**0,000339058**	**0,000664279**	346498,4667	90155,06941
F12	**8,32E-09**	**1,60E-08**	0,003956195	0,001538774

Table 2: Experimental Results for 60 Dimension

	GSO with TSA		**TSA**	
Function	**Mean**	**Std.Dev**	**Mean**	**Std. Dev.**
F1	**0,036651919**	**0,026774993**	190965,1333	51774,35123
F2	**66122,51246**	**80142,90077**	286111,7333	325971,1452
F3	**0,00919557**	**0,006471114**	303341,1667	105265,899
F4	**3008978694 3**	**1,23778E+1 1**	1,07E+18	1,96E+18
F5	**0,00337202**	**0,00152403**	139588,9	55671,51479
F6	**359069,7**	**127279,8183**	425018,9	127458,1141
F7	**4,1**	**2,038790491**	38,83333333	6,132336876
F8	676420,6667	167854,3849	**166435,9333**	**11610,71613**

F9	**293647,7333**	**187220,274**	27824793506	506068042 68
F10	**311677,7667**	**244402,6638**	3,03926E+11	1,36527E+11
F11	**0,83770253**	**0,993606341**	279159,4	97107,62956
F12	**0,063678421**	**0,127654734**	143131,3333	80242,88452

Table 3: Experimental Results for 100 Dimension

	GSO with TSA		**TSA**	
Function	**Mean**	**Std.Dev**	**Mean**	**Std. Dev.**
F1	**391621,9333**	**288968,6826**	467683,2667	396967,1567
F2	**345199,2667**	234598,952	407476,6333	**40482,22908**
F3	424612	207022,7595	**211149,3333**	**57158,16892**
F4	**3,27E+32**	**9,51E+32**	3,45E+43	7,00E+43
F5	**31391,41657**	58585,10656	253575,9333	**47076,90186**
F6	563926,3667	**19277,70782**	**545222,7**	215847,5605
F7	**19,73333333**	**5,955016559**	9999,766667	775,2295868
F8	**132307,2667**	**8602,383201**	276117,8333	84093,41914
F9	**350311,9333**	**245037,9276**	4,23971E+12	1,45777E+12
F10	**358172,5**	**295986,3986**	1,16394E+13	3,4136E+12
F11	**332580,1333**	262470,4233	718887,2667	**233980,9837**
F12	326880,8	214501,0156	**167490,7**	**77205,9092**

When the results obtained for 30 dimensions were analyzed, the GSO with TSA method could not obtain better results only in F6 function. Similarly, in the 60-dimension results, GSO with TSA method gives better results than conventional TSA in benchmark functions except F8. GSO with TSA is successful in benchmark functions except F3, F6 and F12 in 100 dimension results. As a total, GSO with TSA success rates in the 30, 60 and 100-dimension results are %91.67, %91.67 and %75.00 respectively.

IV. CONCLUSION

GSO is a framework that can use other optimization methods. There are many successful optimization methods with different characteristics in the literature. The original GSO method uses the PSO method as the search method. In this work we used the TSA as a search method within the GSO. In our experimental study on 12 benchmark functions, the GSO with TSA method was compared with the traditional TSA method. GSO with TSA method is more successful than traditional TSA method.

REFERENCES

1. Holland, J.H., *Genetic Algorithms*. Scientific American, 1992. **267**(1): p. 66-72.
2. Kennedy, J. and R. Eberhart, *Particle swarm optimization*. 1995 Ieee International Conference on Neural Networks Proceedings, Vols 1-6, 1995: p. 1942-1948.
3. Karaboga, D. and B. Basturk, *A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm*. Journal of Global Optimization, 2007. **39**(3): p. 459-471.
4. Kiran, M.S., *TSA: Tree-seed algorithm for continuous optimization*. Expert Systems with Applications, 2015. **42**(19): p. 6686-6698.
5. Mirjalili, S. and A. Lewis, *The Whale Optimization Algorithm*. Advances in Engineering Software, 2016. **95**: p. 51-67.
6. Rashedi, E., H. Nezamabadi-Pour, and S. Saryazdi, *GSA: A Gravitational Search Algorithm*. Information Sciences, 2009. **179**(13): p. 2232-2248.
7. Muthiah-Nakarajan, V. and M.M. Noel, *Galactic Swarm Optimization: A new global optimization metaheuristic inspired by galactic motion*. Applied Soft Computing, 2016. **38**: p. 771-787.
8. Parsopoulos, K.E. and M.N. Vrahatis, *Particle Swarm Optimization method for Constrained Optimization problems*. Intelligent Technologies - Theory and Applications, 2002. **76**: p. 214-220.
9. Derrac, J., et al., *A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms*. Swarm and Evolutionary Computation, 2011. **1**(1): p. 3-18.
10. Gao, W.F., S.Y. Liu, and L.L. Huang, *A Novel Artificial Bee Colony Algorithm Based on Modified Search Equation and Orthogonal Learning*. Ieee Transactions on Cybernetics, 2013. **43**(3): p. 1011-1024.

APPENDIX

No of Funct.	Name	Search Range	Function
F1	Sphere	[100,100]	$f_1(\vec{X}) = \sum_{i=1}^{n} x_i^2$
F2	Elliptic	[100,100]	$f_2(\vec{X}) = \sum_{i=1}^{n} (10^6)^{(i-1)/(n-1)} x_i^2$
F3	SumSquares	[10,10]	$f_3(\vec{X}) = \sum_{i=1}^{n} i x_i^2$
F4	SumPower	[10,10]	$f_4(\vec{X}) = \sum_{i=1}^{n} \lvert x_i \rvert^{(i+1)}$
F5	Schwefel2.22	[10,10]	$f_5(\vec{X}) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$
F6	Schwefel2.21	[100,100]	$f_6(\vec{X}) = \max_i \{ \lvert x_i \rvert, 1 \le i \le n \}$
F7	Step	[100,100]	$f_7(\vec{X}) = \sum_{i=1}^{n} (\lfloor x_i + 0.5 \rfloor)^2$
F8	Schwefel2.26	[500,500]	$f_{14}(\vec{X}) = 418.98*\mathrm{n} - \sum_{i=1}^{n} x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$
F9	Penalized1	[50,50]	$f_{16}(\vec{X}) = \frac{\pi}{n}\{10\sin^2(\pi y_1) + \sum_{i=1}^{n-1}(y_i-1)^2[1+10\sin^2(\pi y_{i+1})] + (y_n-1)^2\} + \sum_{i=1}^{n} u(x_i,10,100,4)$ $y_i = 1+\frac{1}{4}(x_i+1)\ \ u_{x_i,a,k,m} = \begin{cases} k(x_i-a)^m & x_i > a \\ 0 & -a \le x_i \le a \\ k(x_i-a)^m & x_i < -a \end{cases}$
F10	Penalized2	[50,50]	$f_{17}(\vec{X}) = \frac{1}{10}\{\sin^2(\pi x_1) + \sum_{i=1}^{n-1}(x_i-1)^2[1+\sin^2(3\pi x_{i+1})] + (x_n-1)^2[1+\sin^2(2\pi x_{i+1})]\} + \sum_{i=1}^{n} u(x_i,5,100,4)$
F11	Alpine	[10,10]	$f_{18}(\vec{X}) = \sum_{i=1}^{n} \lvert x_i \cdot \sin(x_i) + 0.1 \cdot x_i \rvert$
F12	Levy	[10,10]	$f_{19}(\vec{X}) = \sum_{i=1}^{n-1}(x_i-1)^2[1+\sin^2(3\pi x_{i+1})] + \sin^2(3\pi x_1) + \lvert x_n - 1 \rvert [1+\sin^2(3\pi x_n)]$

The binary salp swarm algorithm with using transfer functions

E.KAYA[1], A.C.CINAR[2], O.UYMAZ[3], S.KORKMAZ[4], M.S.KIRAN[5]

[1] Dept. of Comp. Eng., Fac. of Eng. Selcuk University, Konya/Turkey, ersinkaya@selcuk.edu.tr
[2] Turkish State Meteorological Service Konya Division, Konya/Turkey, ahmetcevahircinar@gmail.com
[3] Dept. of Comp. Eng., Fac. of Eng. Selcuk University, Konya/Turkey, ouymaz@selcuk.edu.tr
[4] Dept. of Comp. Eng., Fac. of Eng. Selcuk University, Konya/Turkey, sedatkorkmaz@selcuk.edu.tr
[5] Dept. of Comp. Eng., Fac. of Eng. Selcuk University, Konya/Turkey, mskiran@selcuk.edu.tr

***Abstract* - The Salp Swarm Algorithm (SSA) is one of the recently proposed nature-inspired metaheuristic algorithms. SSA mimics the life cycle of salp swarms. Salp swarm is an animal group which lived in oceans. The navigating and foraging behaviors are the characteristic properties of the salp swarms. These behaviors are modeled as an optimization algorithm in SSA and it is firstly proposed for solving continuous optimization problems. In literature, there is no binary version of this algorithm which uses transfer functions. In this work, SSA is modified for solving binary optimization problems by using transfer functions. Transfer functions are used to convert the continuous decision variables to the binary decision variables. With this modification, the structure of SSA has not been changed, but only the Sigmoid and the Tangent Hyperbolic transfer functions are adapted. In order to validate the performance of the proposed binary SSA, a well-known pure binary optimization problem, uncapacitated facility location problems (UFLP), set is considered. UFLPs are used for a benchmarking of many metaheuristic algorithms such as; artificial bee colony, tree-seed algorithm, particle swarm optimization, differential evolution and artificial algae algorithm. The experimental results of 12 UFLPs are compared with each other and state-of-art algorithms. Experimental results demonstrate that the SSA is a promising solver for lower dimensional problems, but its performance should be improved on higher dimensional problems.**

***Keywords* – binary optimization, salp swarm algorithm, swarm intelligence, transfer functions, logistic functions.**

I. INTRODUCTION

Metaheuristic optimization algorithms are trending topic in the numeric optimization research field. Almost every month a new nature-inspired metaheuristic algorithm is proposed. The Salp Swarm Algorithm (SSA) is a population-based metaheuristic algorithm which mimics the navigating and foraging behaviors of salps [1]. This algorithm is in the swarm intelligence algorithms group. Mirjalili et al. [1] proposed the SSA for solving the single objective continuous optimization problems. Optimization problems are divided into two main groups according to the type of decision variables. The names of these groups are continuous and discrete. Binary optimization is a subfield of discrete optimization. In literature, there are many binary optimization algorithms [2-8] in order to effectively solve this type of problems. According to the literature review, there is no binary version of the SSA. In this work, we propose a new binary SSA for solving binary optimization problems by using transfer functions. Transfer functions convert the continuous decision variables to binary form. Thus, any modification is not required the structure of the algorithm, but only decision variables are transformed by transfer functions to binary solution space. In this work, we use sigmoid and tangent hyperbolic transfer functions in order to achieve this purpose. The log-sigmoid and tangent hyperbolic transfer functions produce values in range of [0, 1] and [-1,1] by depending on the input, respectively. The input range for these functions [-6, 6] is considered and the values in this range are converted to the binary variables by using the transfer functions. The uncapacitated facility location problem (UFLP) is a pure binary optimization problem and there is no integer or real number in this problem. A lot of combinatorial optimization problems such as set covering, set partitioning, airline crew scheduling [3-8] may be modelled by way of UFLP and thus, the solving the UFLP is an important attempt in binary optimization field.
The rest of the paper is organized as follows. Section 2 gives information about SSA. Binary SSA is explained in the Section 3. Experimental study is conducted in Section 4 and Section 5, respectively. Finally, the obtained results are discussed in Section 6.

II. SALP SWARM ALGORITHM (SSA)

A salp is an animal which lives in water and similar to jellyfishes. The swarming behavior of salps is modeled by Mirjalili et al. in 2017 [1] for an optimization algorithm. Salp Swarm Algorithm (SSA) is a population-based metaheuristic algorithm which mimics the navigating and foraging behaviors of salps. The salp swarm is named as salp chain in this algorithm. The single salp is shown in Figure 1(a) and the swarm of salps is shown in Figure 1(b).

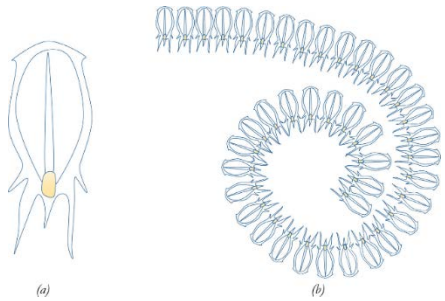


Figure 1: (a) The single salp, (b) The swarm of salps [1]

The salp population is divided into two main groups: leader and followers. The salps search the food source in the search space by using these behaviors. The position of the leader salp updates as follows:

$$x_j^1 = \begin{cases} F_j + c_1\left(\left(UB_j - LB_j\right) \times c_2\right) + LB_j & c_3 \geq 0 \\ F_j - c_1\left(\left(UB_j - LB_j\right) \times c_2\right) + LB_j & c_3 < 0 \end{cases} \quad (1)$$

where x_j^1 shows the position of leader salp in the jth dimension, F_j is the position of the food source in the jth dimension, UB_j states the upper bound of jth dimension, LB_j states the lower bound of jth dimension, c_1, c_2, and c_3 are random numbers. The parameters c_2 and c_3 are uniform random numbers which are generated between 0 and 1.

The parameter c_1 balances exploration and exploitation as follows:

$$c_1 = 2e^{-\left(\frac{4 \times iter}{Maxiter}\right)^2} \quad (2)$$

where $iter$ is the current iteration number and $Maxiter$ is the maximum number of iterations. The real salps use the following the model for movement.

$$x_j^i = \frac{1}{2}at^2 + v_0 t \quad (3)$$

where $i \geq 2$, x_j^i shows the position of ith follower salp in jth dimension, t is time, v_0 is the initial speed, and $a = \frac{v_{final}}{v_0}$ where $v = \frac{x - x_0}{t}$. For simplicity, this formula was converted to Eq.4. The position of the follower salps updated with Eq.4.

$$x_j^i = \frac{1}{2}\left(x_j^i - x_j^{i-1}\right) \quad (4)$$

The pseudo code of the SSA is given in Figure 2.

```
The salp population are randomly created between upper and lower bounds
x_i (i = 1, 2, ..., n)
while (termination criterion is not met)
  Calculate the objective function value of each salp
  LeaderSalp=the best salp
  Update c1 by Eq.2
    for each salp (x_i)
      if (i==1)
        Update the position of the leader salp by Eq.1
      else
        Update the position of the follower salps by Eq.4
      end
    end
  Salps outside the search space are restricted.
end
return LeaderSalp
```

Figure 2: Pseudo code of the Salp Swarm Algorithm

III. BINARY SALP SWARM ALGORITHM (BSSA)

The sigmoid and tangent hyperbolic transfer functions are used for binarization of the SSA. Continuous decision variables converted to the binary form with the transfer functions. In this way, no structural change is made to the algorithm. Before the calculating objective function values, continuous variables converted to binary variables. The continuous variables are created between -6 and 6, which are the upper and lower bound of the search space of the problem. Thus, the output of sigmoid function is between 0.0025 and 0.9975. Eq.5 shows how sigmoid transfer function (S) works in BSSA.

$$transfer(x_{i,j}) = \begin{cases} binX_{i,j} = 1, \ if \ rand(0,1) < S(X_{i,j}) \\ binX_{i,j} = 0, \ otherwise \end{cases} \quad (5)$$

The tangent hyperbolic transfer function is create values between -1 and 1. Eq.6 shows how tangent hyperbolic transfer function (TH) is used for creating 0 or 1 in BSSA.

$$transfer(x_{i,j}) = \begin{cases} binX_{i,j} = 1, \ if \ rand(0,1) < TH(X_{i,j}) \\ binX_{i,j} = 0, \ otherwise \end{cases} \quad (5)$$

The mathematical formula of sigmoid logistic function is shown in Eq.7 and the mathematical formula of the tangent hyperbolic logistic function is shown in Eq. 8.

$$S(x) = \frac{1}{1 + e^{-x}} \quad (7)$$

$$TH(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (8)$$

IV. UNCAPACITATED FACILITY LOCATION PROBLEM

The uncapacitated facility location problem (UFLP) is a pure binary optimization problem. In this problem, there are facilities, the customers, the logistic operations cost and the setting up cost. The main objective is to minimize the sum of these costs. For achieving this aim, a minimum number of facilities should be opened in the direction of customer demands.

$F = \{f_1, f_2, \ldots, f_m\}$ is the set of the facilities and m is the number of facilities, $C = \{c_1, c_2, \ldots, c_n\}$ is the set of customers and n is the number of customers, $S = \{s_f\}$ is the setting up cost of the facility f, $T = \left[t_{fc}\right]$ is the transportation costs from facility f to customer c. The objective function is as follows in Eq.9:

$$argmin\left\{O(X) = \sum_{f \in X} s_f + \sum_{c \in C} min\{t_{fc} | f \in X\} : \emptyset \subset X \subseteq F\right\} \quad (9)$$

where X is a solution, O() is the objective function. The several combinatorial optimization problems can be modelled by way of the UFLP form. For instance, set covering, set partitioning, airline crew scheduling [3-8]. Therefore, solving the UFLP means solving these problems immediately.

V. BENCHMARK SET

For validating the performance of the BSSA, we use 12 UFLPs which are directly taken from OR Library [9]. The problem names, sizes and optimum values are given in Table 1. The small size problems (cap71, cap72, cap73 and cap74) have

16 facilities and 50 customers. The medium size problems (cap101, cap102, cap103 and cap 104) have 25 facilities and 50 customers. The large size problems (cap131, cap132, cap133 and cap 134) have 50 facilities and 50 customers. The cost of optimal solutions are given in Table 1.

Table 1: The benchmark set

Problem name	Problem size	Cost of optimal solution
Cap71	16x50	932,615.75
Cap72	16x50	977,799.40
Cap73	16x50	1,010,641.45
Cap74	16x50	1,034,976.98
Cap101	25x50	796,648.44
Cap102	25x50	854,704.20
Cap103	25x50	893,782.11
Cap104	25x50	928,941.75
Cap131	50x50	793,439.56
Cap132	50x50	851,495.33
Cap133	50x50	893,076.71
Cap134	50x50	928,941.75

VI. RESULTS AND DISCUSSIONS

The relative error (RE) is used for the determining the performance of transfer functions and population sizes. The mathematical formula of the RE is shown in Eq.10.

$$RE(\%) = \frac{Best - Optimum}{Optimum} \times 100 \qquad (10)$$

where, Best is the obtained solution by the method, Optimum is the cost of optimum solution for the problem. The population sizes are taken as 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100. For making a fair comparison, we take 8E+4 for a maximum number of function evolutions (Max_FEs) as termination condition. The results of the compared algorithms are directly taken from [3, 6, 10]. Table 2 shows the RE values of Sigmoid BSSA.

Table 2: The RE values of Sigmoid BSSA

Problem/N	10	20	30	40	50	60	70	80	90	100
cap71	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00
cap72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
cap73	0.01	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.02	0.00
cap74	0.00	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.01	0.01
cap101	0.10	0.10	0.10	0.12	0.13	0.07	0.15	0.13	0.10	0.16
cap102	0.08	0.07	0.09	0.11	0.08	0.13	0.17	0.19	0.14	0.13
cap103	0.07	0.07	0.09	0.17	0.09	0.09	0.10	0.12	0.09	0.17
cap104	0.13	0.15	0.26	0.26	0.14	0.19	0.28	0.40	0.18	0.29
cap131	1.74	1.81	2.25	2.42	1.82	2.49	2.41	2.26	2.20	2.16
cap132	1.51	2.63	2.83	2.84	2.89	2.86	2.17	2.93	2.40	2.50
cap133	2.99	2.37	4.31	3.71	3.21	3.58	3.65	3.25	3.43	2.84
cap134	3.50	4.18	4.84	6.31	6.10	6.66	5.83	6.51	6.68	5.94
Total RE	10.14	11.38	14.79	15.95	14.46	16.08	14.79	15.78	15.26	14.20

According to the Table 2, the small-sized problems (cap71, cap72, cap73 and cap74) are solved optimally, the medium-sized problems (cap101, cap102, cap103 and cap104) are solved with about one percent less error than optimums and the large-sized problems (cap131, cap132, cap133 and cap134) are solved about five percent less error than optimums. The best population size is 10 according to the total RE. The Table 3 shows the RE values of BSAA based on Tangent Hyperbolic.

Table 3: The RE values of Tangent Hyperbolic BSSA

Problem/N	10	20	30	40	50	60	70	80	90	100
cap71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
cap72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
cap73	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.01
cap74	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.00	0.02	0.00
cap101	0.09	0.11	0.11	0.10	0.09	0.14	0.14	0.13	0.11	0.13
cap102	0.08	0.11	0.12	0.11	0.12	0.10	0.09	0.10	0.15	0.12
cap103	0.07	0.15	0.12	0.13	0.13	0.13	0.08	0.14	0.09	0.10
cap104	0.18	0.15	0.18	0.26	0.26	0.22	0.17	0.48	0.23	0.14
cap131	1.33	1.63	2.04	1.68	1.47	1.51	1.31	1.50	1.56	1.41
cap132	1.70	1.63	1.89	1.73	1.98	1.81	1.91	2.03	1.83	1.62
cap133	1.55	2.27	2.35	2.10	2.19	2.03	2.42	2.34	2.05	1.85
cap134	2.42	3.34	3.41	3.96	3.97	4.10	3.76	3.63	3.25	3.49
Total RE	7.43	9.38	10.24	10.07	10.21	10.04	9.89	10.36	9.30	8.88

According to the Table 3, the small and medium-sized problems are solved about optimally and the large size problems are solved about three percent less error than optimums. The best population size is 10 according to the total RE in Tangent Hyperbolic BSSA same as the Sigmoid BSSA. Additionally, according to the total RE values, the usage of tangent hyperbolic is better than the usage of sigmoid as a transfer function in BSSA.

The results of BSSA are compared with the results of binary variants of ABC, PSO, TSA and DE algorithms and the comparison results are given in Table 4.

Table 4: The comparison of state-of-art algorithms

Problem/N	BSSA-TanH	BPSO	SimLogicTSA	IBPSO	binABC	disABC	disDE/rand	binDE
cap71	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00
cap72	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00
cap73	0.00	0.02	0.00	0.20	0.00	0.00	0.00	0.00
cap74	0.00	0.01	0.00	0.40	0.00	0.00	0.00	0.00
cap101	0.09	0.05	0.00	0.60	0.00	0.00	0.00	0.00
cap102	0.08	0.01	0.00	0.73	0.00	0.00	0.00	0.00
cap103	0.07	0.04	0.00	0.64	0.00	0.00	0.01	0.00
cap104	0.18	0.08	0.00	1.00	0.00	0.00	0.00	0.00
cap131	1.33	0.13	0.00	2.42	0.00	0.62	0.00	0.00
cap132	1.70	0.09	0.00	3.60	0.00	0.09	0.00	0.01
cap133	1.55	0.11	0.00	5.26	0.12	0.03	0.01	0.01
cap134	2.42	0.13	0.00	7.63	0.00	0.00	0.00	0.00
Total RE	7.43	0.69	0.00	22.80	0.12	0.75	0.03	0.02

In Table 4, binary salp swarm algorithm with tangent hyperbolic transfer function (BSSA-TanH), binary particle swarm optimization (BPSO) [11], similarity and logic gate-based tree-seed algorithm (SimLogicTSA) [3], improved binary particle swarm optimization (IBPSO) [12], binary artificial bee colony (binABC) [6], similarity-based artificial bee colony (disABC) [4], similarity-based differential evolution with random position update rule (disDE/rand) [10], binary differential evolution (binDE) [10] are compared. The best algorithm is SimLogicTSA and the worst is the IBPSO for solving these UFLPS. The BSSA is a promising solver for

lower dimensional problems, but its performance should be improved on higher dimensional problems.

The convergences of population sizes on cap71, cap101 and cap 131 problems are shown in Figure 3-5, respectively. Due to the page limitation, the convergence graphs of the other problems are not included in this paper.

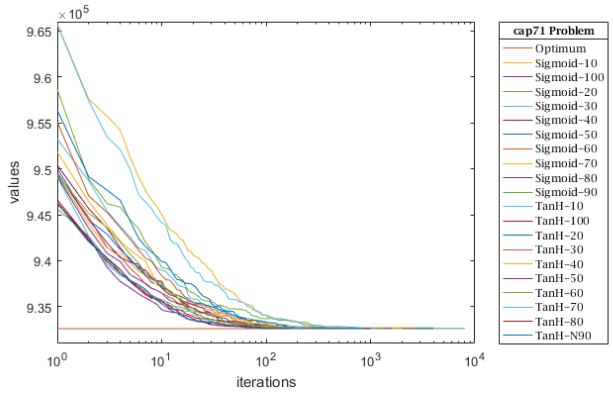


Figure 3: The convergences of population sizes on cap71 problem

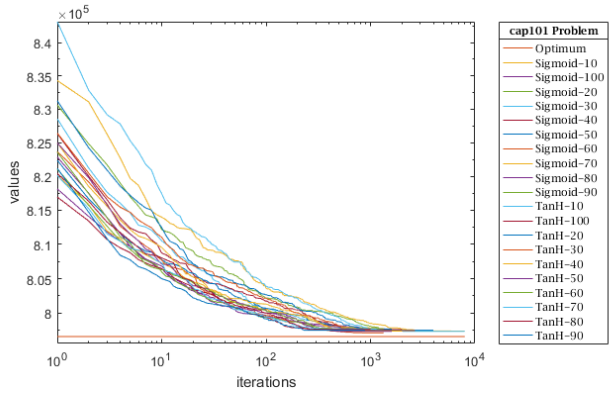


Figure 4: The convergences of population sizes on cap101 problem

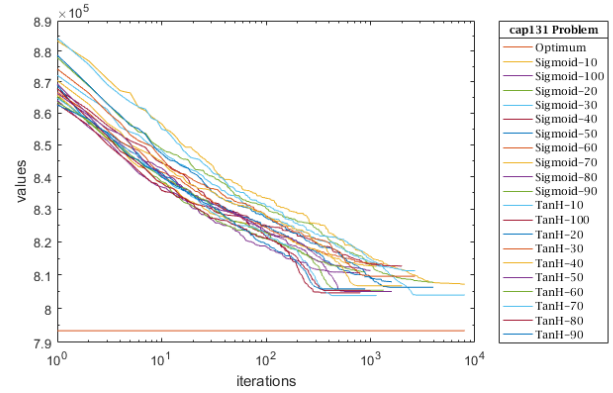


Figure 5: The convergences of population sizes on cap131 problem

The BSSA produces comparable solutions on small and medium-sized UFLPs, but for large instances BSSA produces non-optimal solutions and show lower performance when compared to state-of-art algorithms. To overcome this problem, different binarization mechanisms should be integrated with this algorithm.

REFERENCES

1. Mirjalili, S., et al., *Salp swarm algorithm: a bio-inspired optimizer for engineering design problems*. Advances in Engineering Software, 2017. **114**: p. 163-191.
2. Banitalebi, A., M.I.A. Aziz, and Z.A. Aziz, *A self-adaptive binary differential evolution algorithm for large scale binary optimization problems*. Information Sciences, 2016. **367**: p. 487-511.
3. Cinar, A.C. and M.S. Kiran, *Similarity and logic gate-based tree-seed algorithms for binary optimization*. Computers & Industrial Engineering, 2018. **115**: p. 631-646.
4. Kashan, M.H., N. Nahavandi, and A.H. Kashan, *DisABC: a new artificial bee colony algorithm for binary optimization*. Applied Soft Computing, 2012. **12**(1): p. 342-352.
5. Kiran, M.S., *The continuous artificial bee colony algorithm for binary optimization*. Applied Soft Computing, 2015. **33**: p. 15-23.
6. Kiran, M.S. and M. GÜNDÜZ, *XOR-based artificial bee colony algorithm for binary optimization*. Turkish Journal of Electrical Engineering & Computer Sciences, 2013. **21**(Sup. 2): p. 2307-2328.
7. Korkmaz, S., A. Babalik, and M.S. Kiran, *An artificial algae algorithm for solving binary optimization problems*. International Journal of Machine Learning and Cybernetics, 2017: p. 1-15.
8. Korkmaz, S. and M.S. Kiran, *An artificial algae algorithm with stigmergic behavior for binary optimization*. Applied Soft Computing, 2018.
9. Beasley, J.E., *OR-Library: distributing test problems by electronic mail*. Journal of the operational research society, 1990. **41**(11): p. 1069-1072.
10. Kashan, M.H., A.H. Kashan, and N. Nahavandi, *A novel differential evolution algorithm for binary optimization*. Computational Optimization and Applications, 2013. **55**(2): p. 481-513.
11. Kennedy, J. and R.C. Eberhart. *A discrete binary version of the particle swarm algorithm*. in *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*. 1997. IEEE.
12. Yuan, X., et al., *An improved binary particle swarm optimization for unit commitment problem*. Expert Systems with applications, 2009. **36**(4): p. 8049-8055.

Monogamous Crab Mating Optimization Algorithm for Solving Vehicle Routing Problem

B.ÇUBUKÇU[1], U. YÜZGEÇ[2]

[1] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
burakhan.cubukcu@bilecik.edu.tr

[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
ugur.yuzgec@bilecik.edu.tr

***Abstract* - Today, the use of meta-heuristic algorithms in solving non-linear and complex problems is becoming popular. These algorithms are inspired by natural selection, swarm intelligence, physical events and nature. Crab mating optimization algorithm is one of the meta-heuristic algorithms which was developed by V.R. Chifu in 2014. The crab mating optimization algorithm imitates the mating behavior of crabs in nature. One of the biggest shortcomings of this algorithm is that it has got the quite long running times for optimization problems. In this study, to increase the running speed of the original crab mating optimization algorithm, the mating process of the crabs in population was modified. A male crab mates with only a female crab in the new mating procedure. Therefore, the name of the proposed meta-heuristic algorithm comes from this new mating process as a Monogamous Crab (MC) mating algorithm. Vehicle routing problem (VRP) is the popular combinatorial optimization problem. VRP relates to the most appropriate route design to be delivered to a range of customer service by a number of fleets. In this study, the developed MC algorithm is adapted for the vehicle routing problem. The performance of MC algorithm was compared with those of metaheuristic algorithms. The results show that MC algorithm provides promising and competitive performance.**

***Keywords* - Metaheuristic algorithms, Crab mating algorithm, Vehicle routing problem.**

I. INTRODUCTION

In metaheuristic algorithms are alternative methods in place of the mathematical based exact optimization methods for solving optimization problems. In the optimization algorithms, the mathematical based methods are used such as backtracking, branch-and-bound and dynamic programming methods [1]. Metaheuristic algorithms are easy to understand and implementation to real-world optimization problem. These algorithms show a stochastic behavior and imitate biological or physical processes. The majority of metaheuristic algorithms are inspired events in nature, such as Ant Colony Optimization (ACO) algorithm [2][3], Particle Swarm Optimization (PSO) algorithm [4][5], Artificial Bee Colony (ABC) algorithm [6]–[9], Differential Evolution (DE) algorithm [10]–[12] and Genetic Algorithm (GA) [13]–[15].

Crab Mating Optimization Algorithm (COA) which is one of the metaheuristic algorithms, was developed by V.R. Chifu in 2014 [16]. This algorithm mimics the crab's mating behavior. COA includes two populations of male and female crabs. During optimization process, child crabs are generated using male and female crabs according to the mating probability of the crabs.

The crab mating optimization algorithm runs very slowly in comparison with other metaheuristic algorithms, although it finds global solutions in benchmark functions. In the COA, there are unused parameters such as the number of eggs, the number of fertilized eggs. For test functions, the probability values of mating calculated within the iteration loop of the algorithm are always close to zero or to one. This causes the loop to be entered all the time in the algorithm or not to be entered anytime. To increase the speed of COA, we have modified the mating process of the crabs in the population. A male crab mates with only a female crab in the new mating procedure. As a result, the name of the proposed new algorithm comes from this modified mating process as a Monogamous Crab (MC) mating algorithm. In this study, MC algorithm was adapted to the vehicle routing problem (VRP) that is the popular combinatorial optimization problem. The results show that the proposed MC algorithm provides competitive performance.

II. CRAB MATING OPTIMIZATION ALGORITHM (COA)

In 2014, Chifu et al. proposed Crab Mating Optimization Algorithm (COA) by examining mating and adult turnover behaviors of crabs [16]. Table 1 gives the concepts of the mating behavior of crabs and equivalent of these concepts in optimization process.

Table 1: COA concept.

Mating Behavior of Crabs	**Optimization Problem (OP)**
Female population	Candidate solutions of the OP
Male population	Candidate solutions of the OP
Probability of mating	Probability calculated by male crab
Mating process	Crossover between male-female
Eggs	Best individual's children after crossover operation
Number of eggs produced by each female	The number varies depending on the number of female crabs and the cost of female crabs.
Number of fertilized eggs	It depends on the male's fitness and his mating history.

The parameters used in COA are the algorithm convergence factor (β), the mutation points (*Mut*), number of male population (*NM*), the number of female population (*NF*), the crossover points (*Cr*), female receptivity degree (α). In the original COA, the values of *NF* and *NM* are initialized randomly. α, β, *Cr* and *Mut* values are fixed parameters. The pseudo code of COA is given below:

Algorithm 1: Pseudo Code of COA Algorithm.

Input: *PopSize*, *NF* , *NM*, α, β, *Cr*, *Mut*
Output: Best crab position, Best cost value.
Start
  Mating_counter ← 0
  PopFemale ← *Random_Initiliaze_PopFemale* (*NF*, *PopSize*)
  PopMale ← *Random_ Initiliaze_PopMale*(*NM*, *PopSize*)
  Rank_PopFemale_Fitness(PopFemale)
  Rank_PopMale_Fitness(PopMale)
  while(*iteration* < *maximum iteration*)
  for each male in PopMale
    for each female in PopFemale
      prob ← *MatingProb*(Male, Female, α)
      if(prob > Threshold)
        bestChild ← *Crossover*(Male, Female, *Cr*)
        Mating_counter ← Mating_counter + 1
        nc ← *NumberofClones*(female)
        Clones ← *Make_Clones*(bestChild, nc, β)
        Clones# ← *Fertilize_Clones*(Clones, Mating_counter)
        Clones# ← *Mutate*(Clones#, *Mut*)
      endif
    endfor
  endfor
  PopFemale ← PopFemale U *Half*(Clones#)
  PopFemale ← *Sort*(PopFemale)
  PopMale ← MalePop U *Half*(Clones#)
  PopMale ← *Sort*(MalePop)
  PopFemale ← *Eliminate_Worst_Fitness*(PopFemale)
  PopMale ← *Eliminate_Worst_Fitness*(PopMale)
  end while
  return *Get_Best_Fitness*(PopFemale, PopMale)
End

In COA's concept, the population is divided into male crab population and female crab population. Firstly, the mating probability with each female crab and each male crab is calculated by equation given below:

$$P(f,m) = \frac{Fitness(Male) * Mating\_counter}{\alpha} \quad (1)$$

Here, *Fitness(Male)* is the fitness/cost value of male crabs, *Mating_counter* is the number of mating male, and α is the receptivity degree of females. If the calculated mating probability is greater than the predetermined value then, crossing between the male and female and the best child is kept. Then the number of clones (eggs) and the number of the fertilized eggs are calculated. Later, these fertilized eggs are changed depending on the mutation coefficient.

Some of the children are added to the female population and the rest are added to the male population. As a final step, the newly formed population is sorted according to their fitness values, and the individuals with the worst fitness value from female and male populations are thrown out as many eggs as the number of eggs added for the population to expand.

III. Monogamous Crab Mating Optimization Algorithms (MC)

In solving optimization problems, COA works very slowly compared to other algorithms. One of the reasons there are unused parameters such as the number of eggs, the number of fertilized eggs in COA code. For benchmark functions, the mating probability calculated within the main loop of COA is always close to zero or close to one. This causes the loop to be entered all the time in the algorithm or not to be entered anytime. To overcome this drawback of COA, some modifications were added into the COA code. By mixing the female population before entering the loop, each male crab in the population was randomly mated with only a female crab using the following equations:

$$R_i = rndperm(PopSize) \quad (2)$$
$$Fp_m = Pf(R_i) \quad (3)$$

where, R_i represents the random index, *rndperm* generates a non-repeat random numbers, Fp_m denotes new female population order. We achieved to reduce the number of step in loop by about *1/PopSize*. The number of eggs and the number of fertilized eggs were removed from COA code to not change the population size. By these two modifications made on the code of COA, it is possible to accelerate COA. By updating the function regarding the mating probability, following equation was proposed:

$$P(f,m) = e^{-\left|\frac{Fitness(Male) \times Mating\_counter}{\alpha}\right|} \quad (4)$$

IV. Vehicle Routing Problem (VRP)

The Vehicle Routing Problem (VRP) is one of the combinatorial optimization problems. This problem comprises designing the optimal set of routes for groups of vehicles for the purpose of serving a given set of customers. In classic VRP, Each node is visited only once, by only one vehicle, and each vehicle has a limited capacity [17]–[19]. Figure 1 illustrates an example of VRP with one depot and 9 customers.

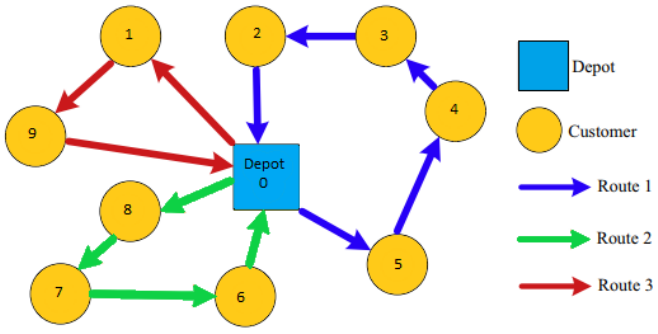


Figure 1: An example of VRP [20].

In solving such problem, main objective is to minimize the time and distance traveled. The objective function and model constraints are given as follows [18]:

$$
\begin{aligned}
&min \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=1}^{m} d_{ij}\, x_{ij}^{k} \\
&subject\ to \sum_{i=1}^{n} x_{0j}^{k} = 1, k = [1,m] \\
&\sum_{i=0}^{n} x_{ip}^{k} - \sum_{j=0}^{n} x_{pj}^{k} = 0, p = [0,n]; k = [1,m] \\
&\sum_{k=1}^{m} y_{i}^{k} = 1, i = [1,n] \\
&\sum_{i=0}^{n} q_i y_{i}^{k} \le a_k, k = [1,m] \\
&y_{i}^{k} \le \sum_{i=0}^{n} x_{ji}^{k}, i = 1,\cdots,n;\ k = [1,m] \\
&b_{i}^{k} + s_i + t_{ij} - M_{ij}\left(1 - x_{ij}^{k}\right) \le b_{j}^{k}, \\
&i,j = [1,n];\ k = [1,m] \\
&M_{ij} = l_i + t_{ij} - e_j, e_i \le b_{i}^{k} \le l_i, i = [1,n] \\
&y_{i}^{k} \ge 0, i = [1,n];\ k = [1,m] \\
&b_{i}^{k} \ge 0, i = [1,n];\ k = [1,m] \\
&x_{ij}^{k} \in \{0,1\}, i,j = [0,n];\ k = [1,m]
\end{aligned}
\tag{5}
$$

The model's decision variables are given below:

x_{ij}^{k}: 1, if j is supplied after i by vehicle k; 0, otherwise

b_{i}^{k}: moment at which service begins at customer i by vehicle k

y_{i}^{k}: fraction of customer's demand i delivered by vehicle k.

s_i : the service time at customer i

t_{ij}: travel time $(i \neq j)$

d_{ij}: distance traveled

q_i: the demand at point i

a_k: capacity of vehicle k

M_{ij}: constant with a large enough number

e_i: the earliest to start to service customer i

l_i: and latest time to start to service customer i

V. Experimental Results

The proposed MC algorithm has been implemented to vehicle routing problems (VRPs) with different sizes. In this study, VRP instance has been taken from www.yarpiz.com web site [21]. These instances have been solved by the proposed MC algorithm and it's result has been compared with the Simulated Annealing (SA) algorithm which is one of the well-known metaheuristic algorithms. The codes of MC and SA algorithms have been run on PC with Intel(R) Core(TM) i5-3230M CPU@2.60GHz RAM/8. Population size is 80, maximum number of iterations is 200. Fig.2 shows the results obtained at the end of one-time run by both of the metaheuristic algorithms.

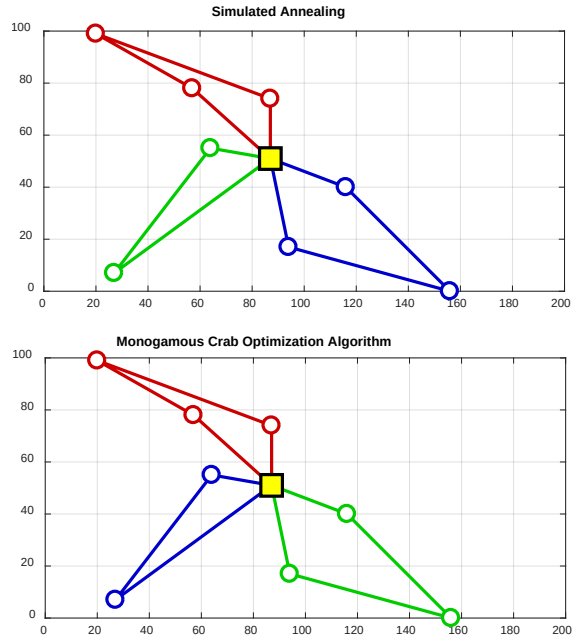


Figure 2: Results of SA and MC algorithms for VRP (8x3).

First VRP example consists of 8 customers, 1 depot, 3 routes. As can be seen from this figure, both algorithms provide the global solution for 8x3 VRP. In Fig. 3, the comparison of convergence curves of SA and MC algorithms is presented. According to this figure, MC algorithm has reached the global minimum earlier than the SA algorithm with a little difference. In second example, there are 10 customers, 1 depot, 3 routes. Figure 4 show the best solutions of VRP with 10x3 for both metaheuristic algorithms. The solutions for this instance provide the same solution for this instance again. In Fig. 5, Sa and MC algorithm's convergence curves are presented for 10x3 VRP.

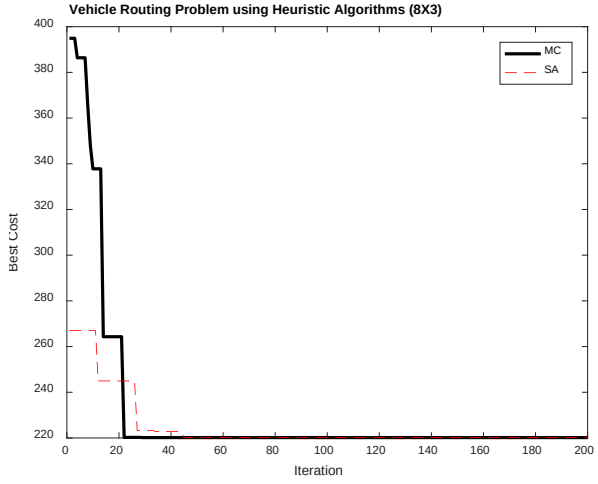


Figure 3: Convergence curves for VRP (8x3).

VI. Conclusion

In this study, Monogamous Crab (MC) mating algorithm was presented and this algorithm was implemented to the a combinatorial optimization problem known as vehicle routing problem (VRP). To evaluate the algorithm's performance on VRP, The proposed MC algorithm was compared with Simulated Annealing (SA) algorithm for VRP instances.

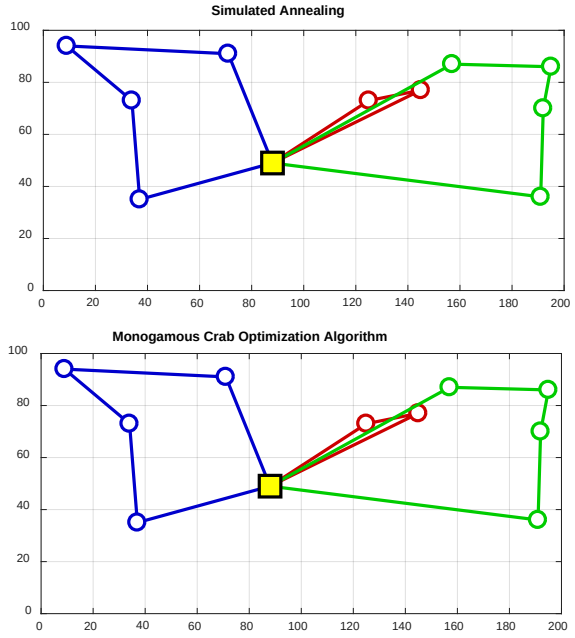


Figure 4: Results of SA and MC algorithms for VRP (10x3).

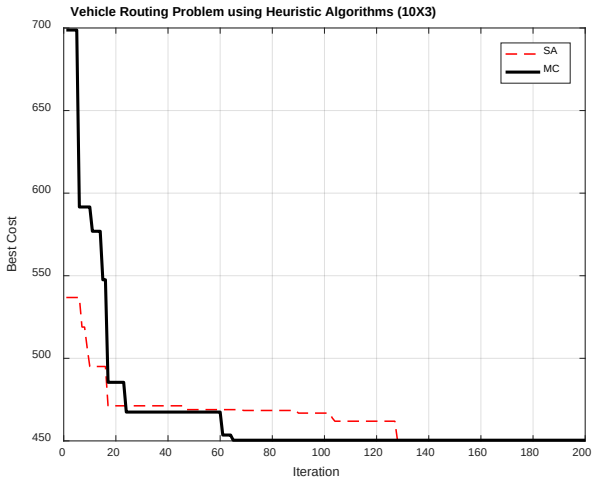


Figure 5: Convergence curves for VRP (10x3).

REFERENCES

[1] Z. Beheshti and S. M. H. Shamsuddin, "A review of population-based metaheuristic algorithm," *International Journal of Advances in Soft Computing and its Applications*, vol. 5, no. 1. 2013.

[2] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.

[3] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 26, no. 1, pp. 29–41, 1996.

[4] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Neural Networks, 1995. Proceedings., IEEE International Conference on*, 1995, vol. 4, pp. 1942–1948 vol.4.

[5] G. Moslehi and M. Mahnam, "A Pareto approach to multi-objective flexible job-shop scheduling problem using particle swarm optimization and local search," *Int. J. Prod. Econ.*, vol. 129, no. 1, pp. 14–22, 2011.

[6] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: Artificial bee colony (ABC) algorithm and applications," *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 21–57, 2014.

[7] E. . Caniyilmaz, B. . Benli, and M. S. . Ilkay, "An artificial bee colony algorithm approach for unrelated parallel machine scheduling with processing set restrictions, job sequence-dependent setup times, and due date," *Int. J. Adv. Manuf. Technol.*, vol. 77, no. 9–12, pp. 2105–2115, 2015.

[8] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput. J.*, vol. 8, no. 1, pp. 687–697, 2008.

[9] D. Karaboga, "An idea based on Honey Bee Swarm for Numerical Optimization," *Tech. Rep. TR06, Erciyes Univ.*, no. TR06, p. 10, 2005.

[10] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[11] K. V Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, vol. 28. 2005.

[12] U. Yüzgeç, "Performance comparison of differential evolution techniques on optimization of feeding profile for an industrial scale baker's yeast fermentation process," *ISA Trans.*, vol. 49, no. 1, pp. 167–176, 2010.

[13] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. Addison-We. 1989.

[14] N. M. Razali and J. Geraghty, "Genetic Algorithm Performance with Different Selection Strategies in Solving TSP," *Int. Conf. Comput. Intell. Intell. Syst.*, 2011.

[15] J. S. Kochhar, B. T. Foster, and S. S. Heragu, "HOPE: A genetic algorithm for the unequal area facility layout problem," *Comput. Oper. Res.*, vol. 25, no. 7–8, pp. 583–594, 1998.

[16] V. R. Chifu, I. Salomie, E. S. Chifu, A. Negrean, H. Jeflea, and M. Antal, "Crab mating optimization algorithm," in *2014 18th International Conference on System Theory, Control and Computing, ICSTCC 2014*, 2014, pp. 353–358.

[17] A. Goel and V. Gruhn, "A General Vehicle Routing Problem," *Eur. J. Oper. Res.*, vol. 191, no. 3, pp. 650–660, 2008.

[18] T. Caric and H. Gold, "Vehicle Routing Problem," *RAIRO-Operations Res. ...*, p. 152, 2008.

[19] P. Toth and D. Vigo, *The vehicle routing problem*, vol. 9. 2002.

[20] J. Luo, X. Li, M.-R. Chen, and H. Liu, "A novel hybrid shuffled frog leaping algorithm for vehicle routing problem with time windows," *Inf. Sci. (Ny).*, vol. 316, pp. 266–292, 2015.

[21] Yarpiz, Capacitated Vehicle Routing Problem (VRP) using SA, http://yarpiz.com/372/ypap108-vehicle-routing-problem/ Accessed 21 March 2018.

A Kinect 2 Based Telerehabilitation Method for Shoulder Rehabilitation Exercises

B. ÇUBUKÇU[1], U. YÜZGEÇ[2], R. ZİLELİ[3], A. ZİLELİ[4]

[1] Bilecik Şeyh Edebali University, Bilecik/Turkey, burakhan.cubukcu@bilecik.edu.tr
[2] Bilecik Şeyh Edebali University, Bilecik/Turkey, ugur.yuzgec@bilecik.edu.tr
[3] Bilecik Şeyh Edebali University, Bilecik/Turkey, raif.zileli@bilecik.edu.tr
[4] Bilecik State Hospital, Bilecik/Turkey, ahukaratas@hotmail.com

Abstract **- The number of people with some kind of physical disabilities in the world is around 1 billion. The rate of people who need physiotherapy increases with the aging world population every passing day. Physiotherapy may not be completed with exercises made only in hospital. Therefore, patients should do the exercises given by the physiotherapist at home. The main problem here is that only 31% of the exercises performed by the patients are done correctly. Nowadays, in addition to traditional treatment methods there are studies on telerehabilitation to solve such problems. The aim of this study is to ensure that the shoulder rehabilitation exercises are performed at home by the patients and to provide physiotherapists with meaningful data about the exercises. In this study, Improved Shoulder Physiotherapy Application (ISPA) was presented using Microsoft Kinect 2 for shoulder rehabilitation exercises. ISPA is a hardware and software product that uses the joints on the patient's skeletal system. In the proposed system, the angular values are calculated using the joint points taken from Kinect 2 and the patients are simultaneously guided to do shoulder rehabilitation exercises correctly.**

Keywords **- Telerehabilitation, Kinect 2, Physiotherapy, Shoulder rehabilitation**

I. INTRODUCTION

People with physical disabilities need to be physically treated in part or in their lifetime. According to studies conducted, the proportion of people who need physical therapy together with the duration of life is increasing every day. Today, the number of people with physical disabilities is over 1 billion in the world [1].

Patients who have to do physiotherapy exercises encounter some problems. One of the most important problems is whether be made the exercises correctly at home or not. The physiotherapy process is completed only by the exercises the patient doing at home, not ending at the hospital. In order to the physiotherapy to succeed, the exercises given by a specialist physiotherapist need to be repeated thousands of times by the patient. One way to do this is to have the daily home exercises done with a physiotherapist. But there are times when it is not financially possible. Patients do not have the opportunity to exercise with physiotherapist every day. So they can encounter problems making exercises correctly. [2]. The study shows that the patients who make exercises alone at home can only finished 31% of all exercises correctly [3]. These exercises done incorrectly can affect the patient's health in a negative way. [4]. Yet another study shows that; 65% of the patients do not understand the exercises being done at home, whether they are doing it correctly [5].

Telerehabilitation is often used to solve this problem in physiotherapy. Telerehabilitation is a term used to describe the provision of rehabilitation services at a distance using telecommunications technology as the service delivery medium [6]. At the same time, patients can do exercises at home with this technology as doing exercises with expert physiotherapists [7]. Studies show that exercises done by visual technology is effective as done with the expert physiotherapist by the conventional method [8-10].

Microsoft Kinect Sensor (MS Kinect) is the one of the popular and useful camera technologies for telerehabilitation. Stable and successful results with MS Kinect were observed in clinical trials [11-12]. In the literature, there are many studies on rehabilitation with MS Kinect. For example, F Cary et al. studies kinect based system and artificial neural networks classifiers for physiotherapy assessment and K. LaBelle's studies evaluating of kinect joint tracking for clinical and in-home stroke rehabilitation tools can be shown as an example on partial paralysis [13-14]. There are lots of studies about elderly people who are able to do daily exercises at home with MS Kinect [15]. There are also studies like pilot study of a kinect-based video game to improve physical therapy treatment to make exercises in the game environment with MS Kinect [16].

In the second part of this study, MS Kinect camera is summarized and in section 3 describes **I**mproved **S**houlder **P**hysiotherapy **A**pplication (ISPA) developed using Kinect 2 for patients with shoulder joint, muscle and tendon injury.

II. MS KINECT

Kinect has 2 models: Kinect 360 (Kinect 1) released in 2010 and Kinect ONE (Figure 1) released in 2014. Production of Kinect ONE for windows computers has been discontinued. However, Kinect ONE's (Kinect 2) version for X-Box can be connected to the computer with an adapter.

Figure 1: Kinect ONE [17]

Kinect 2 is based on TOF (time of flight) principle. The reflection of the signal is recorded by sending infrared lights and the delay between the emitted signal and the received signal is calculated [18]. The differences between Kinect 1 and Kinect 2 are summarized in Table 1.

Table 1: Kinect 1 and Kinect 2 Specifications [18]

Component	Kinect 1	Kinect 2
Technique	Structured-light	Time-of-flight
Depth Sensor	1.8 to 3.5 m	1.3 to 3.5 m
IR Depth Image	320 x 240	512 x 424
Color Image	640 x 480	1920 x 1080
Infrared Image	no IR	512 x 424
Audio Stream	16 kHz, 16-bit	48 kHz, 16-bit
Field of View hor.	57 degrees	70 degrees
Field of View ver.	43 degrees	60 degrees
Minimum Latency	102 ms	20-60 ms

From Table 1, it can be observed that Kinect 2 is a depth camera with better resolution than Kinect 1. In the Skeleton Tracking feature used in this study, Kinect 1 tracks 20 points while Kinect 2 is tracks 25 points. The number of skeletons they can see at the same time is 2 in Kinect 1 and 6 in Kinect 2. These features can also be listed as the reasons for using Kinect 2 in ISPA.

III. Improved Shoulder Physiotherapy Application (ISPA)

ISPA developed by using Kinect 2 and Kinect SDK 2.0. ISPA was developed using Visual Studio environment and is designed as Windows Presentation Form (WPF). The software is written using C # programming.

In ISPA, the conditions were designed for a total of 10 positions for left shoulder muscle tendon patients, four of which were starting position and six were exercising. With these conditions, it is aimed to perform shoulder physiotherapy exercises such as abduction, flexion, external rotation, internal rotation, extension to the users. Physiotherapist ideas were taken so that these exercises can be done in the simplest way by users. Than ISPA interface was created. As shown in Figure 2, ISPA interface consists of 4 main sections.

In the first part, there is the screen showing the model of the user. In the second part, the gif animation is presented to show the correct state of the exercises. In the third part, it comprises the exercise and time informing area. In the fourth part, there are warning messages for the users.

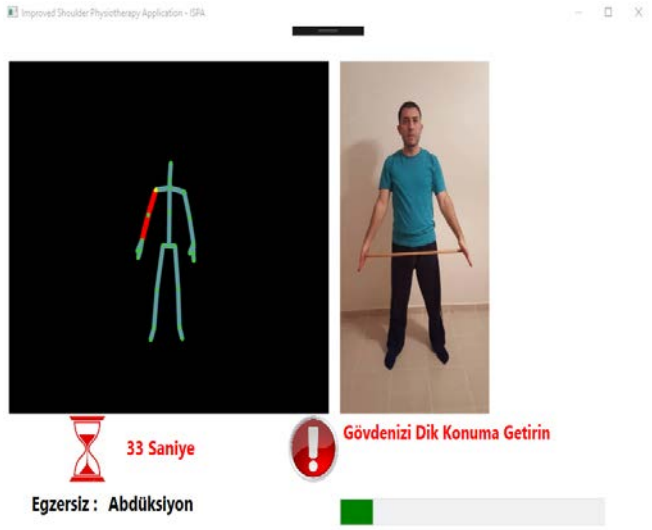


Figure 2: ISPA interface.

The requirements that emerged in the tests of the previously developed Shoulder Physiotherapy Application (SPA) have been added to the ISPA. Thus ISPA has been made more useful than first SPA version for both users and physiotherapists. In ISPA, left and right shoulder exercises are separated from each other. In addition, the exercise arm is shown in a different color. Thus, the confusion of the users is eliminated. It has been observed that the angle values set as the passing conditions of the exercises are not sufficient. Because some users have not been able to reach these angle values. As a solution to this, we have also added a time requirement to ISPA. Thus, even if the user does not make an exercise, this exercise is over after a pre-defined time. The best angle value the user performed in the exercise is recorded. By taking the opinions of the physiotherapist, every element that could distract the user from the interface is removed. One of the major developments in ISPA is the ability to do angular calculations instead of distance calculations.

In all exercises, skeletal points are taken from the Kinect sensors. As shown Figure 3, for the exercises to be performed, a triangle is formed with three corresponding positions A, B, C where $A = (x_1, y_1, z_1)$, $B = (x_1, y_1, z_1)$, $C = (x_3, y_3, z_3) \in R^3$.

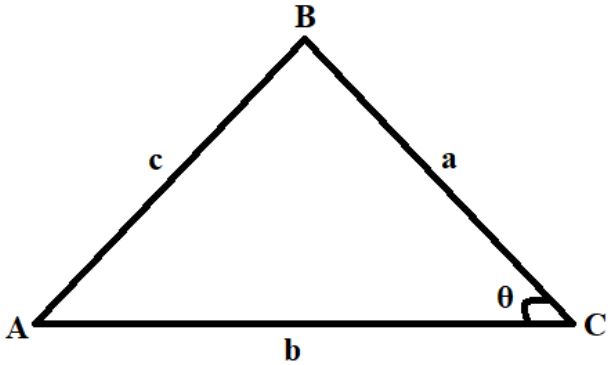


Figure 3: Formed Triangle.

Each edge length of this triangle is calculated using Equations 1-3.

$$a = \sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2 + (z_2 - z_3)^2} \quad (1)$$

$$b = \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2 + (z_1 - z_3)^2} \quad (2)$$

$$c = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (3)$$

Thanks to the Cosine Theorem (Eq. 4), the angle values required for the exercises are calculated.

$$\theta = \hat{C} = \arccos\left(\frac{a^2 + b^2 - c^2}{2ab}\right) \quad (4)$$

As an example, the angle used in extension exercise given in Figure 4. This exercise requires the angle A to be 45 degrees in order to complete it.

Figure 4: Shoulder extension.

For some exercises, 25 points provided by Kinect v2 shown in Figure 5 did not suffice. In these cases, the projections of the points in R^3 are taken and the distances are calculated by Equations 1-3. Similarly, these distances and Eq. 4 are used to obtain angles. These angles are rounded to integer values to provide the conditions physiotherapists want.

The most important feature that makes ISPA an intelligent system is that ISPA can instantly control the user's body position and make the necessary guidance to the user. While attempting to achieve the desired angle values in physical therapy exercises, the body position must also be kept within certain limits. ISPA is able to determine the user's posture, the user's standing, the distance between the two hands, the position of the hands relative to the body, etc. In the case of any faults in the user's body position, the ISPA informs the user of the warning and allows the body position to be corrected. At the same time, ISPA can ignore exercises with the non-correct body position.

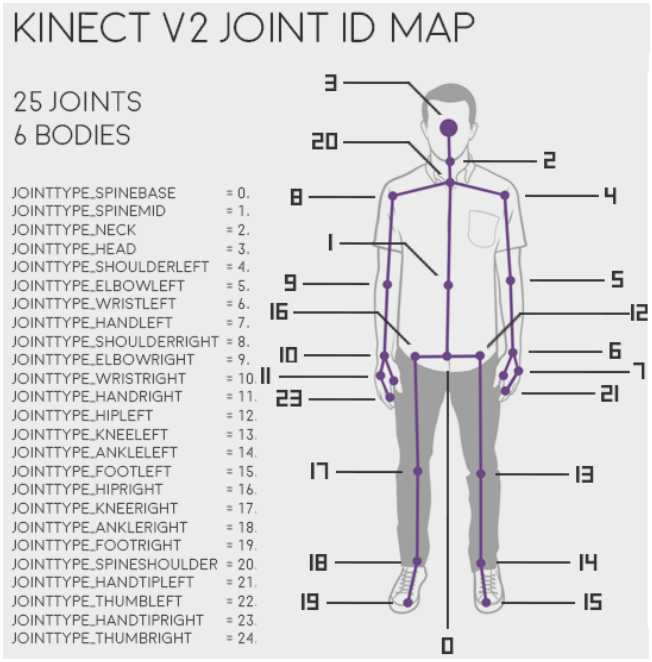


Figure 5: MS Kinect v2 skeleton joints [19].

An image from the tests with ISPA is shown in Figure 6. ISPA was first developed patients with shoulder muscle, tendon injury. Also ISPA software can be developed and used for many diseases.

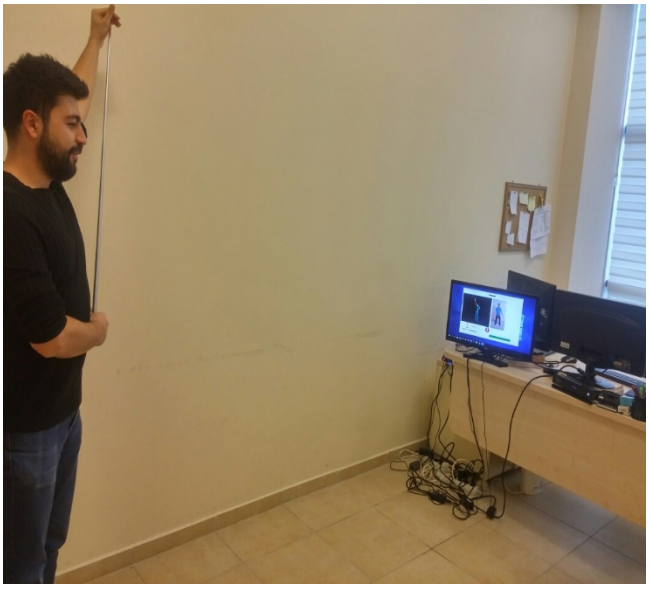


Figure 6: A photo taken during the ISPA tests.

IV. CONCLUSION

Studies show that telerehabilitation is gaining in importance day by day and it is at least as effective as traditional rehabilitation methods. Rehabilitation by image processing is an effective method in the field of telerehabilitation. Within this

scope ISPA gives promising result for shoulder rehabilitation. It has been observed that angular measurements in the ISPA provide better results than distance measurements and give more useful data for physiotherapists. ISPA's plain interface allows users to focus on the exercises without distracting them. Kinect2, used as an image processing camera in ISPA, also produces satisfactory results.

It has been observed that Kinect 2 for the upper body exercises used in ISPA, achieved better results than the lower body exercises. The skeleton tracking ability of Kinect v2 is better than that of Kinect v1.

In future studies, it is considered useful to perform validation reliability tests on the abduction, extension, flexion and rotation exercises performed by ISPA in Kinect2. SPA's software can be optimized to meet the needs of more physical disease. With a web platform to be developed, the exercise data of the patients can be presented to the physiotherapist instantly, and this platform can provide momentary interaction between patients and physiotherapists.

ACKNOWLEDGMENT

Authors would like to acknowledge the support provided by Bilecik Seyh Edebali University under grant 2017-01.BŞEÜ.03-04.

REFERENCES

[1] http://www.who.int/mediacentre/factsheets/fs352/en/ (20.03.2018).
[2] B. Çubukçu and U. Yüzgeç, "A Physiotherapy Application with MS Kinect for Patients with Shoulder Joint , Muscle and Tendon Damage," pp. 225–228, 2017.
[3] M. Shaughnessy, B. M. Resnick, and R. F. Macko, "Testing a model of post-stroke exercise behavior," *Rehabil. Nurs.*, vol. 31, no. 1, pp. 15–21, 2006.
[4] D. Tino and C. Hillis, "The full can exercise as the recommended exercise for strengthening the supraspinatus while minimizing impingement," *Strength Cond. J.*, vol. 32, no. 5, pp. 33–35, 2010.
[5] N. Duarte, O. Postolache, and J. Sharcanski, "KSGphysio – Kinect Serious Game for Physiotherapy," *Int. Conf. Expo. Electr. Power Eng.*, no. Epe, pp. 16–18, 2014.
[6] T. G. Russell, Physical rehabilitation using telemedicine. Journal of telemedicine and telecare, 13(5), 217-220, 2007.
[7] N. P. Bidargaddi and A. Sarela, Activity and heart rate-based measures for outpatient cardiac rehabilitation. Methods of information in medicine, 47(03), 208-216, 2008.
[8] A. S. Rizzo and G. J. Kim, A SWOT analysis of the field of virtual reality rehabilitation and therapy. Presence: Teleoperators & Virtual Environments, 14(2), 119-146, 2005.
[9] P. L. Weiss, D. Rand, N. Katz, and R. Kizony, "Video capture virtual reality as a flexible and effective rehabilitation tool," *J. Neuroeng. Rehabil.*, vol. 1, pp. 1–12, 2004.
[10] B. Çubukçu and A. Çetin, Ms kinect applications and opportunities for people with physical disabilities. Imuco, 459, 2016.
[11] P. Rego, P. M. Moreira, and L. P. Reis, "Serious Games for Rehabilitation A Survey and a Classification Towards a Taxonomy," *5th Iber. Conf. Inf. Syst. Technol.*, no. November 2015, pp. 1–6, 2010.
[12] P. Fikar, C. Schoenauer and H. Kaufmann, "The Sorcerer's Apprentice A serious game aiding rehabilitation in the context of Subacromial Impingement Syndrome," *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, Venice, 2013, pp. 327-330.
[13] F. Cary, O. Postolache, and P. S. Girão, "Kinect based system and Artificial Neural Networks classifiers for physiotherapy assessment," *IEEE MeMeA 2014 - IEEE Int. Symp. Med. Meas. Appl. Proc.*, 2014.
[14] K. Labelle, "Evaluation of Kinect Joint Tracking for Clinical and in-Home Stroke Rehabilitation Tools," *Readings*, p. 67, 2011.
[15] S. Bragaglia, S. D. Monte and P. Mello, "A Distributed System Using MS Kinect and Event Calculus for Adaptive Physiotherapist Rehabilitation," *2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems*, Birmingham, 2014, pp. 531-538.
[16] J. S. Brown, "Pilot Study of a Kinect-Based Video Game to Improve Physical Therapy Treatment," 2013.
[17] https://blogs.msdn.microsoft.com/kinectforwindows/2014/06/05/pre-order-your-kinect-for-windows-v2-sensor-starting-today/ (18.03.2018)
[18] C. Amon and F. Fuhrmann, "Evaluation of the Spatial Resolution Accuracy of the Face Tracking System for Kinect for Windows V1 and V2," *6th Congr. Alps-Adria Acoust. Assos.*, no. October, pp. 9–12, 2014.
[19] https://vvvv.org/documentation/kinect (03.02.2018)

Tournament Selection based Antlion Optimization Algorithm for Solving Quadratic Assignment Problem

H.KILIÇ[1], U. YÜZGEÇ[2]

[1] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
haydar.kilic@bilecik.edu.tr
[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
ugur.yuzgec@bilecik.edu.tr

Abstract **- Quadratic Assignment Problem (QAP) is based the facilities allocation, and it is a difficult combinatorial optimization problem. The objective of this problem is to make total assignment cost minimum while being assigned facilities to locations that are already known. To solve this problem, the different methods are used in the literature. Meta-heuristic algorithms are ones from these methods and in this study, we present a new version of recent antlion optimization algorithm for QAP. AntLion Optimization (ALO) algorithm was developed by Mirjalili in 2015. ALO algorithm is based on the hunting behaviour of the antlion. This algorithm comprises five stages: random walking mechanism, constructing trap, trapping in the antlion's pit, sliding ants in the pit, catching the prey and reconstructing the pit. Although ALO algorithm is successful in benchmark functions of multi dimensions, it has got some drawbacks. The most notable improvement is the use of tournament method instead of roulette wheel method. In ALO algorithm, the antlion is chosen from the population by roulette wheel method for using in each ant's random walking model. The roulette wheel method is more successful in maximization problems. In the minimization problems, the tournament selection method is more efficient method than the other selection methods Therefore, we used the tournament selection method in this study instead of the roulette wheel method on random walking mechanism. This proposed algorithm has been called the tournament selection based antlion optimization algorithm (TALO). To evaluate the performance of TALO algorithm, we used well-known meta-heuristic algorithms. The results provide the proposed TALO algorithm has the best performance in comparison with those of the other algorithms.**

Keywords **- Tournament Selection, Antlion, Quadratic Assignment Problem.**

I. INTRODUCTION

In mathematics and computer science, meta-heuristic algorithms are becoming an important role on solving to the optimization problems. A wide usage areas of meta-heuristic algorithms have come up over the last three decades, and many meta-heuristic algorithms such as differential evolution, particle swarm optimization, artificial bee colony, ant colony, have become popular for real-world optimization problems. In general, the meta-heuristic algorithms imitate some mechanisms as animal feeding habits, mating motivation or hunting techniques in nature [1][2]. Some of the most popular algorithms are Genetic Algorithm (GA) [3], Particle Swarm Optimization (PSO) algorithm [4][5], Artificial Bee Colony (ABC) algorithm [6]–[8], Differential Evolution (DE) algorithm [9][10], Ant Colony (ACO) algorithm [11].

The quadratic assignment problem (QAP) which is one of the most difficult combinatorial optimization problem was presented in 1957 by Koopmans and Beckmann [12]. It is defined as a facilities allocation problem. These facilities are located in many places that are already known and at the least costly ones. In this problem, the sum of the costs for each facility is the general cost function. In literature, there are several works about solving QAP using the meta-heuristic algorithms. Some of these are: solving QAP by simulated annealing algorithm [13][14], by genetic algorithm [15], by ant colony algorithm [16], by tabu search algorithm [17] and by particle swarm optimization algorithm [18].

Antlion optimization algorithm (ALO) which was presented by Seyedali Mirjalili [19] in 2015 is a meta-heuristic algorithm. The antlions come from the Myrmeleontidae family of predatory insect species, which take their name from extremely interesting nutritional behavior in the larval period. ALO algorithm imitates the hunting mechanism of the antlion in larvae phase. The ALO algorithm is basically derived from the hunting strategy of antlions. There are five main steps: the random walking mechanism of ants, building a trap, trapping in the antlion's pits, sliding ants towards antlion, catching the prey and rebuilding the pit [19][20]. There are some studies reported in the literature about real-world optimization applications or increasing performance of the ALO algorithm. These are optimal non-convex and dynamic economic load [21], PID controller parameters design [22], optimal flexible process planning [23], automatic generation control of interconnected power system [24], optimal route planning for unmanned aerial vehicle [25], multi objective optimal generation scheduling [26], determining the optimal coefficients of IIR filters [27].

ALO algorithm presents the effective results for different benchmark optimization problems, but it has got some disadvantages regarding algorithm's mechanism. One of the handicaps of ALO algorithm is that it has got the long runtime

during optimization. The reason of this is the random walking model used in its code structure. In this study, ALO algorithm is developed by eliminating some deficiencies of the original algorithm and a modified ALO algorithm is presented. We have changed the distance of random walking model as 20% of maximum iteration number. In Mirjalili's work [19], the antlion is selected from the population by roulette wheel method for random walking procedure. The tournament selection method is preferred rather than roulette wheel method in minimization problems. In the minimization problems, the tournament selection method is more efficient method[28][29]. For that reason, we preferred the tournament selection method to the roulette wheel method for the random walking mechanism. Furthermore, some new movements were defined between lower and upper boundaries around the antlion in the phase of trapping on antlion pits. These movements provide that ants walk more effectively around the selected antlion in the search space.

II. ANTLION OPTIMIZATION ALGORITHM (ALO)

The mathematical model of the original ALO algorithm is given briefly in this section. The hunting strategy of antlions shows their unique hunting behaviour. The hunting technique consists of setting up a trap, hiding in this trap, waiting for its prey, preventing the prey escape from trap, catching its prey and rebuilding the trap. The ALO algorithm imitates hunter-prey relationship between antlions and ants in the trap. After initializing the antlion population, the random walk model of ants is started. The mathematical model of these walks is given below:

$$X(t) = \begin{bmatrix} 0 \\ cumsum(2r(t_1) - 1) \\ cumsum(2r(t_2) - 1) \\ \vdots \\ cumsum(2r(t_n) - 1) \end{bmatrix} \quad (1)$$

where n stands for the maximum number of iteration, *cumsum* is the cumulative sum, t denotes the step of random walk, and $r(t)$ is the stochastic function as defined:

$$r(t) = \begin{cases} 1, & if rand > 0.5 \\ 0, & if\ rand \leq 0.5 \end{cases} \quad (2)$$

In order to keep ant's random walks in the search space, it is normalized by the equation given below:

$$X_i^t = (X_i^t - a_i)(d_i^t - c_i^t)(b_i - a_i)^{-1} + c_i^t \quad (3)$$

where i denotes the variable number's value, t is the iteration number, a stands for the minimum value of the random walk ($a = min(X)$), b is the maximum value of the random walk ($b = max(X)$), c denotes the lower value of the dynamic limit around the antlion, d is the upper value of the dynamic limit around the antlion.

Once the ants begin to fall into the trap, the antlions throw sand to hold in the trap and slide them down towards the center of the trap. The mathematical model regarding this behaviour is given below:

$$c_i^t = Antlion_i^t + c^t \quad (4)$$

$$d_i^t = Antlion_i^t + d^t \quad (5)$$

$$c^t = c^t . I^{-1} \quad (6)$$

$$d^t = d^t . I^{-1} \quad (7)$$

where $Antlion_i^t$ denotes the position of the selected *i-th* antlion at *t-th* iteration, and I stands for the sliding ratio. The ants are moved around the elite antlion and the antlion selected by roulette wheel method using the model given in Eq. (8).

$$Ant_i^t = 0.5(R_A^t + R_E^t) \quad (8)$$

where R_A^t denotes the antlion selected by roulette wheel method and R_E^t denotes the elite antlion obtained by Eq.(3) at each iteration. After catching the prey, antlions update their positions according to the eaten ants' fitness values. Eq.(9) gives the update mechanism of antlion for minimization problems:

$$if\ f(Ant_i^t) < f(Antlion_i^t), Antlion_i^t = Ant_i^t \quad (9)$$

ALO algorithm's pseudo code is given follows:

Algorithm 1: Pseudo Code of ALO Algorithm.

```
Initialize the positions of antlions
Calculate the cost values of antlions
Save the best antlion and its position (elite antlion)
  while (iteration < maximum iteration)
    for (each antlion)
        Select antlion using roulette wheel method
        Slide randomly walking ants in a trap
        Generate ant's random walk route around elite antlion
        Generate ant's random walk route around selected antlion
        Normalize random walks
        Calculate the position of ant using Eq. (8)
    end for
    Calculate the cost values of ants
    Combine ants and antlions
    Sort according to their costs and take first population size
    Update the elite antlion
  end while
```

III. TOURNAMENT SELECTION BASED ANTLION OPTIMIZATION ALGORITHM (TALO)

In this study, we developed ALO algorithm by improvements on the random walk model, hunting mechanism, selection procedure and etc. In the original ALO code, the random walking mechanism uses the maximum iteration number to generate the ant's random walking route. This is not effective method for run time of the algorithm. For that reason, first innovation on the ALO algorithm was realized by reducing the random walk size. We used n value as 20% of maximum iteration number in Eq. (1). Therefore, the long running time of the ALO algorithm has been shortened significantly.

In the stage of sliding ants towards the antlion's trap, the ants are shifted toward the antlion by a certain rate of slippage.

We made some improvements on the antlion's pit and on the shifting of the ants by throwing sand. It is given by the following mathematical model.

$$\left.\begin{aligned} c_i^t &= Antlion_i^t + c^t \\ d_i^t &= Antlion_i^t + d^t \end{aligned}\right\} \; if\ 0.75 < opt < 1 \tag{10}$$

$$\left.\begin{aligned} c_i^t &= Antlion_i^t - c^t \\ d_i^t &= Antlion_i^t - d^t \end{aligned}\right\} \; if\ 0.5 < opt < 0.75 \tag{11}$$

$$\left.\begin{aligned} c_i^t &= -Antlion_i^t + c^t \\ d_i^t &= -Antlion_i^t + d^t \end{aligned}\right\} \; if\ 0.25 < opt < 0.5 \tag{12}$$

$$\left.\begin{aligned} c_i^t &= -Antlion_i^t - c^t \\ d_i^t &= -Antlion_i^t - d^t \end{aligned}\right\} \; if\ opt < 0.25 \tag{13}$$

where *option* denotes a variable chosen randomly. Thanks to the updating the shift rates, we have provided a more accurate and faster hunting mechanism. Unlike antlion updating mechanism used at the end of iterations in the ALO algorithm, in the new updating mechanism, cost values of the ants and antlions are compared for each pair of ant and antlion. If the ant's cost value is better than antlion's cost, antlion's position is updated as ant position. Another innovation is about ants which go out of the search area. When the ant's position is out of the search space, they come back the search space again unlike the original ALO algorithm. This idea ensures that the ants take the random positions in the search space.

$$Ant_i^t = b_{low} + rand \times \left(b_{up} - b_{low}\right),$$
$$if\ Ant_i^t > b_{up}\ or\ Ant_i^t < b_{low}$$

where *rand* stands for a random number in interval [0,1], b_{low} denotes the lower limit and b_{up} is the upper limit of the search space.

In meta-heuristic algorithms, the selection procedure is used for selecting better individuals from the population. Some of examples are roulette wheel method, tournament selection method, truncation selection, linear ranking selection and exponential ranking selection. For especially minimization problems, tournament selection is the most efficient method [29][30]. In this method, a tournament is realized between individuals selected randomly and the individual with the best cost value becomes the winner of the tournament. The parameter of this method is the tournament size, known as tour. This parameter may be become a value ranging from 2 to number of population.

In this study, we focused a minimization problem known as QAP, so we preferred the tournament selection method instead of the roulette wheel selection method used in the original ALO algorithm. The tour size was selected as 2. In the tournament method, two groups from the population are randomly selected and the size of each group is found by division of the population size to tournament size. Tournament Selection based Antlion Optimization (TALO) algorithm's pseudo code is given below:

Algorithm 2: Pseudo Code of TALO Algorithm.

```
Initialize the positions of antlions
Calculate the cost values of antlions
Save the best antlion and its position (elite antlion)
  while (iteration < maximum iteration)
    for (each antlion)
      Select antlion using tournament selection method
      Slide randomly walking ants in a trap as Eqs. 10-13
      Generate ant's random walk route around elite antlion
      Generate ant's random walk route around selected antlion
      Normalize random walks
      Calculate the position of ant
        if Ant is out of the search space
          Relocate the ant in search space
        end if
    end for
    Calculate the cost values of ants
    for (each antlion)
      if Cost value of ant is better than that of antlion
        antlion eats ant (update the position of antlion)
      end if
    end for
    Update the elite antlion
  end while
```

IV. Quadratic Assignment Problem (QAP)

The Quadratic Assignment Problem (QAP) was presented for the first time by Koopmans and Beckman [12]. In solving this problem, main aim is to find minimum cost for total assignment while assigning facilities to locations. The mathematical model objective function of QAP is given below:

$$min \sum_{i,j=1}^{n} \sum_{p,q=1}^{n} w_{ij}\ d_{pq} x_{ip} x_{jq}$$
$$subject\ to \sum_{i=1}^{n} x_{ij} = 1, \tag{15}$$
$$\sum_{j=1}^{n} x_{ij} = 1,$$
$$x_{ij} \in \{0,1\}, 1 \le i,j \le n$$

where w_{ij} stands for the weight/flow coefficients in range i_{th} and j_{th} facilities and d_{pq} denotes the distance in range p_{th} and q_{th} locations.

V. Experimental Results

The proposed TALO algorithm has been implemented to quadratic assignment problem. In this study, QAP instance has been taken from www.yarpiz.com web site [31]. This instance has been solved by TALO algorithm and it's result has been compared with the several well-known meta-heuristic algorithms, such as ALO algorithm, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Firefly Optimization Algorithm (FOA). This problem includes the $W[20 \times 20]$ weight matrix and $D[20 \times 20]$ distance matrix. This problem comprises three different special situations. First of all, the 19th and 20th facilities must be as close as possible, then, the

11th and 16th facilities must be as close as possible. Finally, the 1st and 13th facilities must be as far as possible. These three critical states are indicated in the weight matrix as follows:

$$w(19,20) = w(20,19) = 10000$$
$$w(11,16) = w(16,11) = 10000$$
$$w(1,13) = w(13,1) = -10000$$

The codes of TALO and other algorithms have been run on PC with Intel(R) Core(TM) i5-3230M CPU@2.60GHz RAM/8. Each algorithm has been run 10 times. Population size is 20, maximum number of iterations is 1000. The parameters of meta-heuristic algorithms used for QAP performance tests are given in Table 1. Fig.1 shows the results obtained at the end of one-time run by the TALO and other meta-heuristic algorithms.

Table 1: Parameters of meta-heuristic algorithms.

Algorithm	Parameters
GA	Crossover Coefficient : 0.4 Mutation Coefficient : 0.8 Selection Pressure Coefficient : 5
PSO	Inertia Weight : 1.0 Inertia Weight Damping Ratio : 0.99 Personal Learning Coefficient : 1.5 Global Learning Coefficient : 2.0
FOA	Light Absorption Coefficient : 1.0 Initial Attraction Coefficient: 2.0 Mutation Coefficient : 0.2 Mutation Coefficient Damping R. : 0.98
ALO	Number of Antlions : 20
TALO	Number of Antlions : 20 Tournament Size : 2

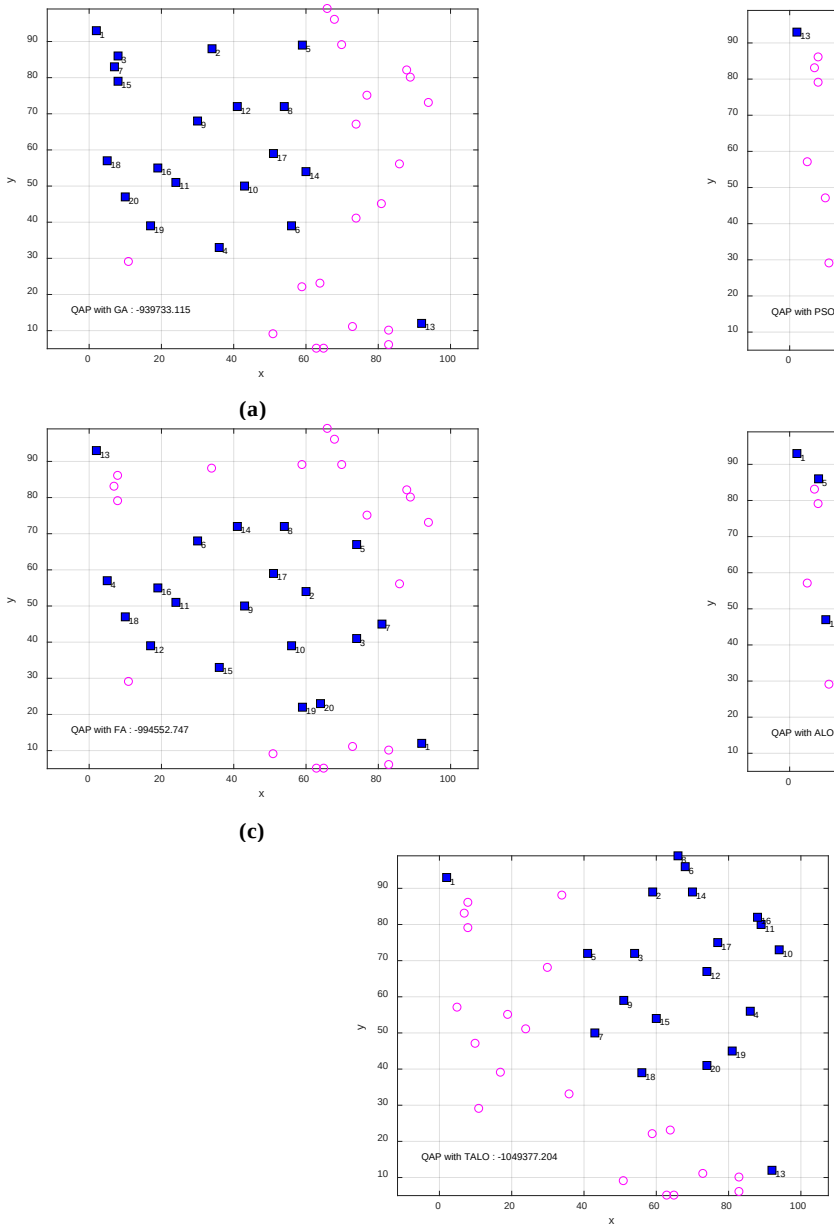


Figure 1: QAP Results obtained by meta-heuristic algorithms, (a) GA, (b) PSO, (c) FA, (d) ALO, (e) TALO.

In these figures, blue squares denote assigned facilities and circles denote empty location (not assigned facility). TALO result has got the best cost value (-1049377.204) among all results. According to the results of all algorithms, facility pairs (19-20), (11-16) are shown to be at close locations and facility pairs (1,13) be at far locations from each other. In Fig. 2, the convergence curves are shown for TALO and meta-heuristic algorithms used in QAP test.

For the comparison results with 10 independent runs of TALO and the other meta-heuristic algorithms, we used four metrics: mean cost, standard deviation, best cost and worst cost. Fig. 3 presents the best cost curves obtained by the TALO algorithms for each runs. In Table 2, QAP results with 10 independent runs summarized.

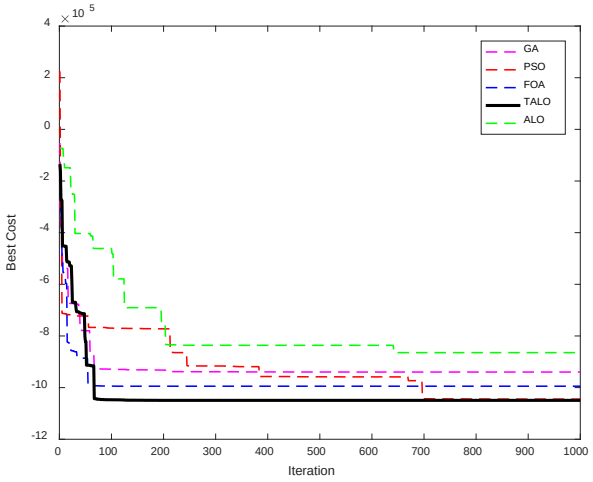


Figure 2: Comparison results of convergence curves.

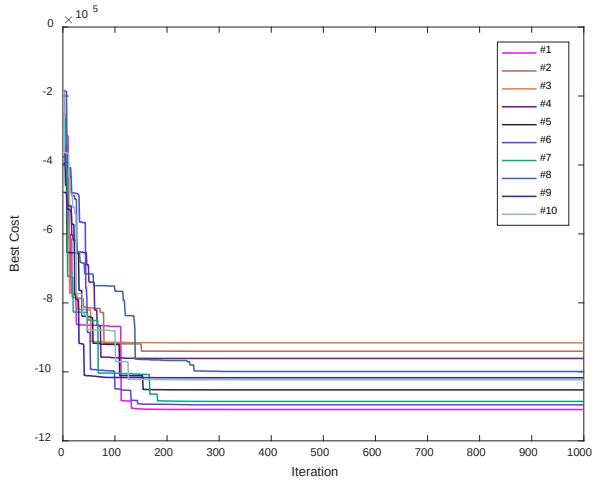


Figure 3: Convergence curves of TALO algorithm with 10 runs.

Table 2: The QAP results with 10 runs of TALO and other meta-heuristic algorithms.

	Mean Cost	**Std. Dev.**	**Best Cost**	**Worst Cost**
GA	-1040715.59	**32096.99**	-1078899.84	**-998807.47**
PSO	**-1050518.95**	48744.72	-1094306.47	-972582.99
FOA	-1039310.94	47073.92	-1107239.04	-947424.52
ALO	-776989.54	134951.44	-1035716.47	-605587.60
TALO	-1019909.74	66819.23	**-1109319.29**	-915827.81

Fig. 4 shows the statistical results regarding the performances of TALO algorithm and other meta-heuristic algorithms for 10 independent runs. As can be seen from Table 2 and Fig. 4, the worst algorithm is ALO algorithm in terms of all metrics. The proposed TALO algorithm has the best cost value in comparison with other algorithms.

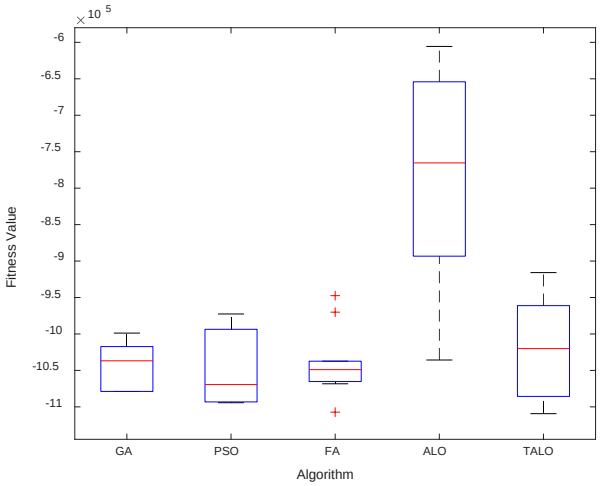


Figure 4: Statistical results of TALO and other meta-heuristic algorithms with 10 runs.

VI. CONCLUSION

In this study, the tournament selection based ALO algorithm (TALO) was presented and the proposed TALO algorithm was adapted to the difficult combinatorial optimization problem known as quadratic assignment problem (QAP). To evaluate the algorithm's performance on QAP, four well-known meta-heuristic algorithms (ALO, GA, PSO, FOA) were used. The comparison outcomes show that the proposed TALO algorithm is able to provide very competitive results. The performance of the proposed TALO algorithm is better than that of the classic ALO algorithm according to all indicators.

REFERENCES

[1] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithms for partitional clustering," *Swarm and Evolutionary Computation*, vol. 16. pp. 1–18, 2014.

[2] Z. Beheshti and S. M. H. Shamsuddin, "A review of population-based meta-heuristic algorithm," *International Journal of Advances in Soft Computing and its Applications*, vol. 5, no. 1. 2013.

[3] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. Addison-We. 1989.

[4] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.

[5] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, 2007.

[6] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput. J.*, vol. 8, no. 1, pp. 687–697, 2008.

[7] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.

[8] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: Artificial bee colony (ABC) algorithm and applications," *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 21–57, 2014.

[9] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[10] K. V Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, vol. 28. 2005.

[11] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.

[12] T. C. Koopmans and M. Beckmann, "Assignment Problems and the Location of Economic Activities," *Econometrica*, vol. 25, no. 1, p. 53, 1957.

[13] T. Peng, W. Huanchen, and Z. Dongme, "Simulated annealing for the quadratic assignment problem: A further study," *Comput. Ind. Eng.*, vol. 31, no. 3–4, pp. 925–928, 1996.

[14] M. R. Wilhelm and T. L. Ward, "Solving Quadratic Assignment Problems by 'Simulated Annealing,'" *IIE Trans.*, vol. 19, no. 1, pp. 107–119, 1987.

[15] D. M. Tate and A. E. Smith, "A genetic approach to the quadratic assignment problem," *Comput. Oper. Res.*, vol. 22, no. 1, pp. 73–83, 1995.

[16] E.-G. Talbi, O. Roux, C. Fonlupt, and D. Robillard, "Parallel Ant Colonies for the quadratic assignment problem," *Futur. Gener. Comput. Syst.*, vol. 17, no. 4, pp. 441–449, 2001.

[17] W. Zhu, J. Curry, and A. Marquez, "SIMD tabu search for the quadratic assignment problem with graphics hardware acceleration," *Int. J. Prod. Res.*, vol. 48, no. 4, pp. 1035–1047, 2010.

[18] F. Hafiz and A. Abdennour, "Particle Swarm Algorithm variants for the Quadratic Assignment Problems - A probabilistic learning approach," *Expert Syst. Appl.*, vol. 44, pp. 413–431, 2016.

[19] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, 2015.

[20] H. Kiliç and U. Yüzgeç, "Improved antlion optimization algorithm," in *2nd International Conference on Computer Science and Engineering, UBMK 2017*, 2017.

[21] V. K. Kamboj, A. Bhadoria, and S. K. Bath, "Solution of non-convex economic load dispatch problem for small-scale power systems using ant lion optimizer," *Neural Comput. Appl.*, vol. 28, no. 8, pp. 2181–2192, 2017.

[22] M. Raju, L. C. Saikia, and N. Sinha, "Automatic generation control of a multi-area system using ant lion optimizer algorithm based PID plus second order derivative controller," *Int. J. Electr. Power Energy Syst.*, vol. 80, pp. 52–63, 2016.

[23] M. Petrović and Z. Miljković, "Biologically inspired optimization algorithms for flexible process planning," *Lect. Notes Mech. Eng.*, pp. 417–428, 2017.

[24] E. Gupta and A. Saxena, "Performance Evaluation of Antlion Optimizer Based Regulator in Automatic Generation Control of Interconnected Power System," *J. Eng. (United States)*, vol. 2016, 2016.

[25] P. Yao and H. Wang, "Dynamic Adaptive Ant Lion Optimizer applied to route planning for unmanned aerial vehicle," *Soft Comput.*, vol. 21, no. 18, pp. 5475–5488, 2017.

[26] N. Chopra and S. Mehta, "Multi-objective optimum generation scheduling using Ant Lion Optimization," in *12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015*, 2016.

[27] S. S. Nair, K. P. S. Rana, V. Kumar, and A. Chawla, "Efficient Modeling of Linear Discrete Filters Using Ant Lion Optimizer," *Circuits, Syst. Signal Process.*, vol. 36, no. 4, 2017.

[28] M. Noraini and J. Geraghty, "Genetic algorithm performance with different selection strategies in solving TSP," *World Congr. Eng.*, vol. II, no. 978-988-19251-4–5, pp. 4–9, 2011.

[29] N. M. Razali and J. Geraghty, "Genetic Algorithm Performance with Different Selection Strategies in Solving TSP," *Int. Conf. Comput. Intell. Intell. Syst.*, 2011.

[30] T. Blickle and L. Thiele, "A Comparison of Selection Schemes Used in Evolutionary Algorithms," *Evol. Comput.*, vol. 4, no. 4, pp. 361–394, 1996.

[31] Yarpiz, Quadratic Assignment Problem (QAP) using GA, PSO and FA. http://yarpiz.com/359/ypap104-quadratic-assignment-problem/ Accessed 8 August 2017.

Parallel Machine Scheduling using Improved Antlion Optimization Algorithm

H.KILIÇ[1], U. YÜZGEÇ[2]

[1] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
haydar.kilic@bilecik.edu.tr

[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
ugur.yuzgec@bilecik.edu.tr

***Abstract* - AntLion Optimization (ALO) algorithm is one recent of the meta-heuristic algorithms that was developed by Mirjalili in 2015. ALO algorithm imitates the antlion's hunting behaviour in its larvae phase. The long run time of ALO algorithm is the biggest disadvantage of this algorithm. To overcome this deficiency, we proposed some improvements on the mechanisms of the original ALO algorithm. In order to improve the ALO algorithm, firstly, the random walking distance was changed as twenty percent of maximum iteration instead of the maximum iteration number in the original ALO algorithm. We defined new movements between boundaries around the antlion on the phase of trapping antlion pits. In addition, the boundary checking process, the catching prey and rebuilding the pit were developed. The parallel machine scheduling problem (PMS) is defined that it is a set of independent jobs to be scheduled on a number of parallel machines. Scheduling process optimizes the production job sequences in terms of the different patterns. When there are the similar type of machines to be existing in multiple numbers, the jobs can be scheduled over these parallel machines at the same time. To show the performance of improved ALO (IALO) algorithm, some of well-known meta-heuristic algorithms were used in comparison works. The obtained PMS results show that the proposed IALO algorithm has very competitive results in terms of the mean, best, worst cost and standard deviation metrics.**

***Keywords* - Parallel Machine Scheduling, Meta-Heuristic Algorithms, Antlion Algorithm.**

I. INTRODUCTION

In real engineering optimization problems, in recent years, the algorithms based on the meta-heuristic approaches have been used as popular solution methods for finding the optimal solutions [1][2]. The meta-heuristic algorithms mostly uses stochastic operators unlike deterministic approaches [3]. Random number generating is the main feature of meta-heuristic algorithms. This means that they use the random operators for finding the global optimal solution in search space.

Scientists often use some behaviors such as animal hunting techniques, nutrition methods, mating habits in the development of such algorithms. There are a lot of studies about the meta-heuristic algorithms and application to real optimization problems [4]. Some of most popular and recent algorithms are Genetic algorithm (GA) [5][6], Differential Evolution (DE) algorithm[7]–[9], Tabu Search algorithm (TS) [10], Simulated Annealing algorithm (SA) [10][11], Particle Swarm Optimization (PSO) algorithm[12]–[14], Artificial Bee Colony (ABC) algorithm[15]–[17], Ant Colony Optimization (ACO) algorithm [18][19], Artificial Immune algorithm (AI) [20]–[22], Harmony Search algorithm (HSA) [23][24], Shuffled Frog Leaping algorithm (SFLA) [25][26], Fish Swarm Algorithm (FSA) [27][28], Bacterial Foraging Optimization algorithm (BFO) [29][30], Biogeography-Based Optimization algorithm (BBO) [31][32], Cuckoo Search algorithm (CSA) [33][34], Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [35][36], Cultural Optimization algorithm (COA) [37], Dragonfly Optimization algorithm (DA) [38][39], Firefly algorithm (FA) [40][41], Fruit Fly Optimization algorithm (FOA) [42], Gray Wolf Optimization algorithm (GWO) [43]–[45], Grasshopper Optimization algorithm (GOA) [46], Gravitational Search algorithm (GSA) [47], [48], Imperialist Competitive algorithm (ICA) [49], [50], Invasive Weed Optimization algorithm (IWO) [51], Moth-Flame Optimization algorithm (MFO) [52][53] and Whale Optimization algorithm (WOA) [54][55].

Antlion optimization algorithm (ALO) basically imitates the hunting technique of antlion. This hunting behavior is investigated by Seyedali Mirjalili and presented as antlion optimization algorithm in 2015 [56]. This algorithm is based on five steps: the random walking mechanism of ants, building a trap, trapping in the antlion's pits, sliding ants towards antlion, catching the prey and rebuilding the pit. In Mirjalili's work [56], there is no analysis regarding algorithm's time consumption, although it provides good results for different optimization problems. The greatest drawback of ALO algorithm is that it solves the optimization problems in the long runtime because of the random walking model.

In this study, we developed ALO algorithm by getting rid of such disadvantages. In the innovations made, the random walking size was used as 20% of maximum iteration number. Besides of this, we added some new movements between boundaries around the antlion in the stage of trapping on antlion pits. By these movements, the ants walk effectively in the region of the antlion in the search space. In the proposed improved ALO algorithm (IALO), the procedure about the catching prey, rebuilding the pit and the boundary checking mechanism have been developed within this study. There are some studies presented in the literature regarding development

or applications of the ALO algorithm. Some of these are: optimal reactive power dispatch solution [57], solving unit commitment problem in smart grid system [58], optimal under voltage load shedding [59], multi-layer perceptrons trainer [60], route planning for unmanned aerial vehicle [61], feature selection [62], segmentation for MRI liver images [63].

The scheduling system is a significant subject that has a big effect in the production processes. It is interested in the allocation of limited resources with time among all jobs. In a scheduling problem, the production job sequences are optimized in accordance with the different patterns. The parallel machine scheduling problem (PMS) is defined as a set of independent tasks to be scheduled on a set of parallel machines. There are a lot of works about solving the PMS problem with meta-heuristic algorithms. Some of them presented in recent years are particle swarm optimization (PSO) algorithm and its different variations [64]–[66], differential evolution (DE) algorithm [67][68], artificial bee colony (ABC) algorithm [69][70].

In this study, we propose the improved antlion optimization algorithm (IALO) to overcome the drawbacks of the ALO algorithm. This proposed IALO algorithm was adapted to PMS problem. To evaluate the performance of IALO algorithm, we used the several well-known meta-heuristic algorithms, such as ALO algorithm, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Firefly Optimization Algorithm (FOA).

II. Antlion Optimization Algorithm (ALO)

The antlion is the common name of the predatory insect species, which take their names from the Myrmeleontidae family about extremely interesting nutritional behaviour in larvae. The larval period of antlion has about 2-years. They are fed as carnivores in this period and usually eat ants. Although they are nurtured as carnivores during the larval period, they are fed as herbivores during the adulthood that they metamorphosed. Fig. 1 illustrates the hunting mechanism of an antlion. The hunting strategy of antlions is a great show which is exhibited in a mathematical order. At the beginning of hunting, antlion prepares a trap for ants by digging a spiral path. It buries itself in the center of the cone-shaped trap and waits for the ants to fall into the trap [56].

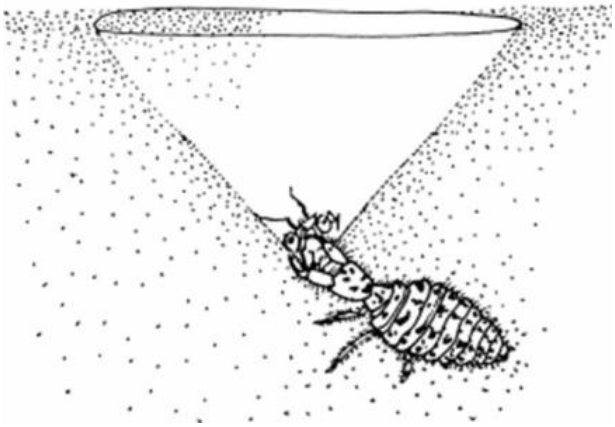


Figure 1: Antlion hunting strategy.

After the ants has fallen into the trap, they try to escape from the trap. But the antlion throws sand from the place that it bury itself to prevent them from escaping from the trap. As a result of the sand throwing, the ants slip down and antlion finally eats them. After this interesting hunting, antlion rebuilds their trap for another ants. The detail information about ALO algorithm can be found from Mirjalili's work [56].

III. Improved Antlion Optimization Algorithm (IALO)

The ALO algorithm finds the optimal point later than other meta-heuristic algorithms and this is a disadvantage for real-world optimization problems. In original ALO algorithm, the random walk route is determined according to the maximum iteration size, and this affects the algorithm to run longer. For this reason, the first improvement in the ALO algorithm is about random walking mechanism. After initializing IALO algorithm, the random walk model is started. The random walking model is obtained for ants as following:

$$X(t) = \begin{bmatrix} 0 \\ cumsum(2r(t_1) - 1) \\ cumsum(2r(t_2) - 1) \\ \vdots \\ cumsum(2r(t_n) - 1) \end{bmatrix} \quad (1)$$

where n denotes 20% of the maximum number of iteration, *cumsum* stands for the cumulative sum, t is the step of random walk, and $r(t)$ is the stochastic function as defined below:

$$r(t) = \begin{cases} 1, & if rand > 0.5 \\ 0, & if\ rand \leq 0.5 \end{cases} \quad (2)$$

In order to keep ant's random walks in the search space, it is normalized by the equation given below:

$$X_i^t = (X_i^t - a_i)(d_i^t - c_i^t)(b_i - a_i)^{-1} + c_i^t \quad (3)$$

where i stands for the variable number's value, t is the iteration number, a denotes the minimum value of the random walk ($a = min(X)$), b denotes the maximum value of the random walk ($b = max(X)$), c is the lower value of the dynamic limit around the antlion, d is the upper value of the dynamic limit around the antlion.

In the period of sliding ants towards the antlion's trap, the ants are shifted toward the antlion by a certain rate of slippage. Some improvements were made on the antlion's pit and on the shifting of the ants by throwing sand. These are given by the mathematical equations below.

$$\left.\begin{matrix} c_i^t = Antlion_i^t + c^t \\ d_i^t = Antlion_i^t + d^t \end{matrix}\right\} if\ 0.75 < opt < 1 \quad (4)$$

$$\left.\begin{matrix} c_i^t = Antlion_i^t - c^t \\ d_i^t = Antlion_i^t - d^t \end{matrix}\right\} if\ 0.5 < opt < 0.75 \quad (5)$$

$$\left.\begin{matrix} c_i^t = -Antlion_i^t + c^t \\ d_i^t = -Antlion_i^t + d^t \end{matrix}\right\} if\ 0.25 < opt < 0.5 \quad (6)$$

$$\left.\begin{matrix} c_i^t = -Antlion_i^t - c^t \\ d_i^t = -Antlion_i^t - d^t \end{matrix}\right\} if\ opt < 0.25 \quad (7)$$

where $Antlion_i^t$ stands for the position of the selected *i-th* antlion at *t-th* iteration, *opt* is a variable chosen randomly. In the stage of sliding ants towards antlion, the radius of ants' random walks is decreased adaptively at each iteration:

$$c^t = c^t . I^{-1} \tag{8}$$

$$d^t = d^t . I^{-1} \tag{9}$$

where I denotes the sliding ratio that is changed in following conditions:

$$I = \begin{cases} 1 + 10^6 \, iter / I_M & if \; 0.95 I_M < iter < I_M \\ 1 + 10^5 \, iter / I_M & if \; 0.9 I_M < iter < 0.95 I_M \\ 1 + 10^4 \, iter / I_M & if \; 0.75 I_M < iter < 0.9 I_M \\ 1 + 10^3 \, iter / I_M & if \; 0.5 I_M < iter < 0.75 I_M \\ 1 + 10^2 \, iter / I_M & if \; 0.1 I_M < iter < 0.5 I_M \\ 1 & otherwise \end{cases} \tag{10}$$

where I_M is the maximum iteration, *iter* denotes the iteration. For each iteration, the positions of ants are determined according to the equation below:

$$Ant_i^t = 0.5 \left(R_A^t (rnd \; index_1) + R_E^t (rnd \; index_2) \right) \tag{11}$$

where R_A^t is antlion selected from the population and R_E^t is elite antlion. Both of them are obtained by Eq.(3) for each iteration. The antlion, selected by the roulette wheel method, does not make any progress for optimization problems with the negative cost values, and after a while, the same antlion for each iteration is selected. To solve this problem, the magnitudes of the cost values are entered into the roulette wheel method as in the following equation:

$$\frac{|f(Antlion_i^{-1})|}{\sum_{j=1}^{n} |f(Antlion_j^{-1})|}, i = 1,2,\cdots,n \tag{12}$$

At the end of the search loop, ants and antlion's cost values are compared for each pair of ant and antlion, and if the ant's cost value is better than that of the antlion, antlion's position is updated as ant's position.

$$if \; f(Ant_i^t) < f(Antlion_i^t), \; Antlion_i^t = Ant_i^t \tag{13}$$

where Ant_i^t sands for the position of the selected *i-th* ant at *t-th* iteration. When the ant goes to the outside of the search space, they are thrown inside the search space again unlike the original ALO algorithm. This idea ensures that the ants take the random positions in the search space.

$$Ant_i^t = low + rnd \times (up - low), \\ if \; Ant_i^t > up \; or \; Ant_i^t < low$$

where *rnd* stands for a random number in interval [0,1], *low* denotes the lower limit and *up* denotes the upper limit of the search space.

IV. PARALLEL MACHINE SCHEDULING (PMS)

The Parallel Machine Scheduling (PMS) problem is one of the combinatorial optimization problems. In PMS problem, the jobs are scheduled over the parallel machines at the same time. McNaughton proposed to schedule on several processors that carry out different tasks to decrease the total cost associated with separate tasks. This proposal is the fundamental of the PMS problem [71]. An example of PMS environment with four machines and nine jobs is shown in Fig. 2.

PMS problem deals with scheduling m parallel machines for n jobs waiting in a queue and the main target of PMS is to minimize the completion time of machine scheduling. The mathematical model of PMS is given below [72]:

$$\begin{aligned} & min \; C_{max} \\ & subject \; to \\ & \sum_{\substack{i=0 \\ i \neq j}}^{n} \sum_{k=1}^{m} x_{ijk} = 1, \forall j = 1,2,\cdots,n \\ & \sum_{\substack{i=0 \\ i \neq h}}^{n} x_{ihk} - \sum_{\substack{j=0 \\ j \neq h}}^{n} x_{hjk} = 0, \; \forall h = 1,2,\cdots,n \; \forall k = 1,2,\cdots,m \\ & C_j \geq C_i + \sum_{k=1}^{m} x_{ijk} \left(S_{ijk} + p_{jk} \right) + M \left(\sum_{k=1}^{m} x_{ijk} - 1 \right), \\ & \forall i = 0,1,\cdots,n \; \forall j = 1,2,\cdots,n \\ & \sum_{j=0}^{n} x_{0jk} = 1, \;\; \forall k = 1,2,\cdots,m \\ & x_{ijk} \in \{0,1\}, \;\; \forall i = 0,1,\cdots,n \; \forall j = 1,2,\cdots,n \; \forall k = 1,2,\cdots,m \\ & C_0 = 0, \;\; C_j \geq 0, \forall j = 1,2,\cdots,n \end{aligned} \tag{15}$$

where,

C_j : Completion time of job *j*

p_{jk} : Processing time of job *j* on machine *k*

S_{ijk} : Sequence dependent setup time to process job *j* after job *i* on machine *k*

S_{0jk} : Setup time to process job *j* first on machine *k*

x_{ijk} : 1 if job *j* is processed directly after job *i* on machine *k* and 0 otherwise

x_{0jk} : 1 if job *j* is the first job to be processed on machine *k* and 0 otherwise

x_{j0k} : 1 if job *j* is the last job to be processed on machine *k* and 0 otherwise

M : A large positive number

V. EXPERIMENTAL RESULTS

In this study, the proposed IALO algorithm has been implemented to an instance of the PMS problem. This instance has been taken from www.yarpiz.com web site [73]. There are four parallel machines and the number of total jobs is 20. The processing time of jobs is a known matrix as $p[20 \times 4]$. For each machine, the values of setup time to process job j after job i on machine have been kept on the $s[20 \times 20 \times 4]$

matrix. This instance has been solved by IALO algorithm and its performance has been compared with the performances of the original ALO algorithm, Genetic algorithm (GA), Particle Swarm Optimization algorithm (PSO), Firefly Optimization Algorithm (FOA)

The codes of IALO and other meta-heuristic algorithms have been run on PC with Intel(R) Core(TM) i5-3230M CPU@2.60GHz RAM/8. Each algorithm has been run 10 times. We used that population size is 20 and maximum number of iterations is 1000. The parameters of meta-heuristic algorithms used for PMS tests are summarized in Table 1. In Fig.2, the PMS problem solution results obtained by the IALO and other well-known meta-heuristic algorithms are shown. As can be seen from the PMS results, the best solution of each algorithm is different each other.

Table 1: Parameters of meta-heuristic algorithms used for PMS.

Algorithm	Parameters
GA	Crossover Coefficient : 0.4 Mutation Coefficient : 0.8 Selection Pressure Coefficient : 5
PSO	Inertia Weight : 1.0 Inertia Weight Damping Ratio : 0.99 Personal Learning Coefficient : 1.5 Global Learning Coefficient : 2.0
FOA	Light Absorption Coefficient : 1.0 Initial Attraction Coefficient: 2.0 Mutation Coefficient : 0.2 Mutation Coefficient Damping R. : 0.98
ALO	Number of Antlions : 20
TALO	Number of Antlions : 20 Tournament Size : 2

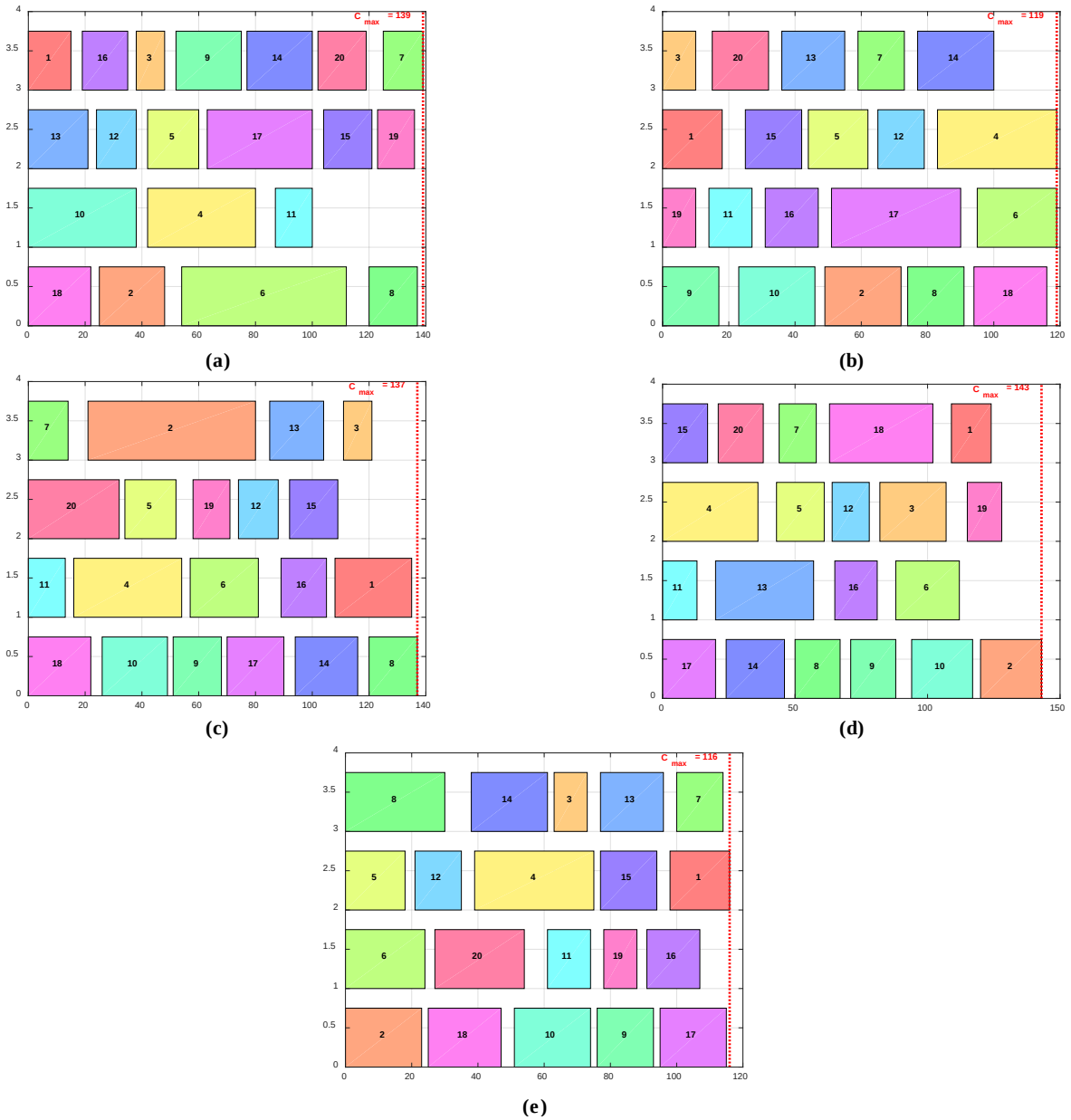


Figure 2: PMS Results obtained by meta-heuristic algorithms, (a) GA, (b) PSO, (c) FA, (d) ALO, (e) IALO.

The results with one time run show that the proposed IALO algorithm provides the best cost value as $C_{max}=116$. Fig. 3 shows the convergence curves obtained by IALO and other meta-heuristic algorithms during solving PMS problem. IALO algorithm's curve is emphasized in bold line. According to this figure, the best convergence curve belongs to the proposed IALO algorithm and original ALO algorithm has the worst performance. In Table 2, PMS results with 10 independent runs summarized for all used algorithms. The best result of each indicator is emphasized in boldface. The performance of IALO algorithm is the best in terms of mean and best cost metrics.

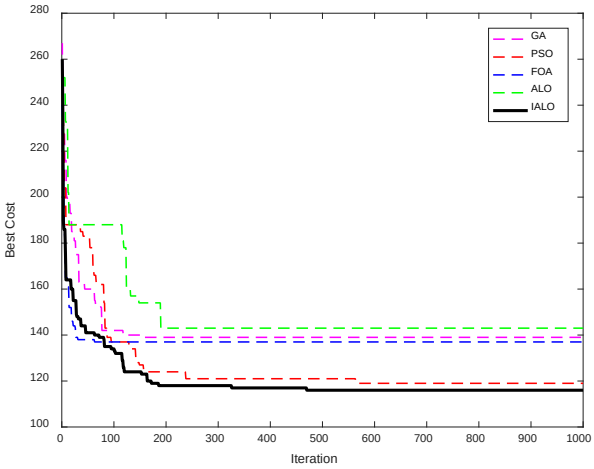


Figure 3: Convergence curves of algorithms.

Table 2: The PMS results with 10 runs of IALO and others.

	Mean Cost	**Best Cost**	**Worst Cost**	**Std. Dev.**
GA	121.4	115	**130**	5.232
PSO	121.3	116	134	**5.143**
FOA	121.1	113	136	7.724
ALO	151.5	142	160	7.075
IALO	**118.3**	**111**	134	7.165

VI. CONCLUSION

In this study, ALO algorithm was developed and the proposed IALO algorithm was applied to the combinatorial optimization problem known as parallel machine scheduling (PMS). To show the algorithm's performance on PMS, four well-known meta-heuristic algorithms (ALO, GA, PSO, FOA) were used. The comparison results show that the proposed IALO algorithm has the best performance.

REFERENCES

[1] I. Boussaïd, J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," in *Information Sciences*, 2013, vol. 237, pp. 82–117.

[2] C. Blum, J. Puchinger, G. R. Raidl, and A. Roli, "A brief survey on hybrid metaheuristics," in *Bioinspired Optimization Methods and their Applications - Proceedings of the 4th International Conference on Bioinspired Optimization Methods and their Applications, BIOMA 2010*, 2010, pp. 3–16.

[3] L. Bianchi, M. Dorigo, L. M. Gambardella, and W. J. Gutjahr, "A survey on metaheuristics for stochastic combinatorial optimization," *Nat. Comput.*, vol. 8, no. 2, pp. 239–287, 2009.

[4] Z. Beheshti and S. M. H. Shamsuddin, "A review of population-based meta-heuristic algorithm," *International Journal of Advances in Soft Computing and its Applications*, vol. 5, no. 1. 2013.

[5] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. Addison-We. 1989.

[6] J. H. Holland, "Adaptation in Natural and Artificial Systems," *Ann Arbor MI Univ. Michigan Press*, vol. Ann Arbor, p. 183, 1975.

[7] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[8] K. V Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, vol. 28. 2005.

[9] U. Yüzgeç, "Performance comparison of differential evolution techniques on optimization of feeding profile for an industrial scale baker's yeast fermentation process," *ISA Trans.*, vol. 49, no. 1, pp. 167–176, 2010.

[10] G. Paul, "Comparative performance of tabu search and simulated annealing heuristics for the quadratic assignment problem," *Oper. Res. Lett.*, vol. 38, no. 6, pp. 577–581, 2010.

[11] T. Peng, W. Huanchen, and Z. Dongme, "Simulated annealing for the quadratic assignment problem: A further study," *Comput. Ind. Eng.*, vol. 31, no. 3–4, pp. 925–928, 1996.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.

[13] R. Poli, J. Kennedy, and T. Blackwell, "Particle swarm optimization," *Swarm Intell.*, vol. 1, no. 1, pp. 33–57, 2007.

[14] Y. Shi and R. Eberhart, "A modified particle swarm optimizer," in *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, 1998, pp. 69–73.

[15] D. Karaboga, "An idea based on Honey Bee Swarm for Numerical Optimization," *Tech. Rep. TR06, Erciyes Univ.*, no. TR06, p. 10, 2005.

[16] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," *Appl. Soft Comput. J.*, vol. 11, no. 1, pp. 652–657, 2011.

[17] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.

[18] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 26, no. 1, pp. 29–41, 1996.

[19] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.

[20] S. A. Hofmeyr and S. Forrest, "Architecture for an Artificial Immune System," *Evol. Comput.*, vol. 8, no. 4, pp. 443–473, 2000.

[21] J. Timmis, A. Hone, T. Stibor, and E. Clark, "Theoretical advances in artificial immune systems," *Theor. Comput. Sci.*, vol. 403, no. 1, pp. 11–32, 2008.

[22] J. Timmis, P. Andrews, and E. Hart, "On artificial immune systems and swarm intelligence," *Swarm Intell.*, vol. 4, no. 4, pp. 247–273, 2010.

[23] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A New Heuristic Optimization Algorithm: Harmony Search," *Simulation*, vol. 76, no. 2, pp. 60–68, 2001.

[24] X. S. Yang, "Harmony search as a metaheuristic algorithm," *Studies in Computational Intelligence*, vol. 191. pp. 1–14, 2009.

[25] M. Eusuff, K. Lansey, and F. Pasha, "Shuffled frog-leaping algorithm: A memetic meta-heuristic for discrete optimization," *Eng. Optim.*, vol. 38, no. 2, pp. 129–154, 2006.

[26] K. K. Bhattacharjee and S. P. Sarmah, "Shuffled frog leaping algorithm and its application to 0/1 knapsack problem," *Appl. Soft Comput. J.*, vol. 19, pp. 252–263, 2014.

[27] X. Li, Z. Shao, and J. Qian, "An optimizing method based on autonomous animats: fish-swarm algorithm," *Syst. Eng. Theory Pract.*, vol. 22, no. 11, pp. 32–38, 2002.

[28] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, "Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 965–997, 2014.

[29] S. Das, A. Biswas, S. Dasgupta, and A. Abraham, "Bacterial Foraging Optimization Algorithm: Theoretical Foundations, Analysis, and Applications," *Found. Comput. Intell. Vol. 3*, vol. 3, pp. 23–55, 2009.

[30] M. S. Li, T. Y. Ji, W. J. Tang, Q. H. Wu, and J. R. Saunders, "Bacterial foraging algorithm with varying population," *BioSystems*, vol. 100, no. 3, pp. 185–197, 2010.

[31] D. Simon, "Biogeography-based optimization," *IEEE Trans. Evol. Comput.*, vol. 12, no. 6, pp. 702–713, 2008.

[32] H. Ma and D. Simon, "Blended biogeography-based optimization for constrained optimization," *Eng. Appl. Artif. Intell.*, vol. 24, no. 3, pp. 517–525, 2011.

[33] X.-S. Yang, "Cuckoo Search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214.

[34] X. S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Computing and Applications*, vol. 24, no. 1. pp. 169–174, 2014.

[35] H. Beyer and B. Sendhoff, "Covariance matrix adaptation revisited–the CMSA evolution strategy–," *Parallel Probl. Solving from Nature–PPSN X*, pp. 123–132, 2008.

[36] M. Willjuice Iruthayarajan and S. Baskar, "Covariance matrix adaptation evolution strategy based design of centralized PID controller," *Expert Syst. Appl.*, vol. 37, no. 8, pp. 5775–5781, 2010.

[37] C. Soza, R. L. Becerra, M. C. Riff, and C. A. Coello Coello, "Solving timetabling problems using a cultural algorithm," *Appl. Soft Comput.*, vol. 11, no. 1, pp. 337–344, 2011.

[38] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2016.

[39] S. R. Sree Ranjini and S. Murugan, "Memory based Hybrid Dragonfly Algorithm for numerical optimization problems," *Expert Syst. Appl.*, vol. 83, pp. 63–78, 2017.

[40] X. S. Yang, "Firefly Algorithm," *Nature-Inspired Metaheuristic Algorithms*, pp. 79–90, 2007.

[41] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5792 LNCS, pp. 169–178.

[42] H. Iscan and M. Gunduz, "Parameter Analysis on Fruit Fly Optimization Algorithm," *J. Comput. Commun.*, vol. 2, no. March, pp. 137–141, 2014.

[43] S. A. Medjahed, T. Ait Saadi, A. Benyettou, and M. Ouali, "Gray Wolf Optimizer for hyperspectral band selection," *Appl. Soft Comput. J.*, vol. 40, pp. 178–186, 2016.

[44] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[45] S. Mirjalili, "How effective is the Grey Wolf optimizer in training multi-layer perceptrons," *Appl. Intell.*, vol. 43, no. 1, pp. 150–161, 2015.

[46] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper Optimisation Algorithm: Theory and application," *Adv. Eng. Softw.*, vol. 105, pp. 30–47, 2017.

[47] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," *Inf. Sci. (Ny).*, vol. 179, no. 13, pp. 2232–2248, 2009.

[48] N. M. Sabri, M. Puteh, and M. R. Mahmood, "A review of gravitational search algorithm," *Int. J. Adv. Soft Comput. its Appl.*, vol. 5, no. 3, 2013.

[49] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive algorithm: An algorithm for optimization inspired by imperialistic competition," in *2007 IEEE Congress on Evolutionary Computation, CEC 2007*, 2007, pp. 4661–4667.

[50] S. Hosseini and A. Al Khaled, "A survey on the Imperialist Competitive Algorithm metaheuristic: Implementation in engineering domain and directions for future research," *Applied Soft Computing Journal*, vol. 24. pp. 1078–1094, 2014.

[51] X. Zhang, Y. Niu, G. Cui, and Y. Wang, "A modified invasive weed optimization with crossover operation," in *Proceedings of the World Congress on Intelligent Control and Automation (WCICA)*, 2010.

[52] S. Mirjalili, "Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm," *Knowledge-Based Syst.*, vol. 89, pp. 228–249, 2015.

[53] G. I. Sayed and A. E. Hassanien, "Moth-flame swarm optimization with neutrosophic sets for automatic mitosis detection in breast cancer histology images," *Appl. Intell.*, vol. 47, no. 2, pp. 397–408, 2017.

[54] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016.

[55] A. Kaveh and M. I. Ghazaan, "Enhanced whale optimization algorithm for sizing optimization of skeletal structures," *Mech. Based Des. Struct. Mach.*, vol. 45, no. 3, pp. 345–362, 2017.

[56] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, 2015.

[57] R. N. S. Mei, M. H. Sulaiman, and Z. Mustaffa, "Ant lion optimizer for optimal reactive power dispatch solution," *J. Electr. Syst.*, vol. 2015, no. Specialissue3, pp. 68–74, 2015.

[58] I. N. Sam'on, Z. M. Yasin, and Z. Zakaria, "Ant Lion Optimizer for solving unit commitment problem in smart grid system," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 8, no. 1, pp. 129–136, 2017.

[59] Z. M. Yasin, I. N. Sam'On, N. A. Salim, H. Mohamad, and N. A. Wahab, "Optimal undervoltage load shedding using ant lion optimizer," *Int. J. Simul. Syst. Sci. Technol.*, vol. 17, no. 41, p. 47.1-47.6, 2017.

[60] W. Yamany, A. Tharwat, M. F. Hassanin, T. Gaber, A. E. Hassanien, and T. H. Kim, "A New Multi-layer Perceptrons Trainer Based on Ant Lion Optimization Algorithm," in *Proceedings - 2015 4th International Conference on Information Science and Industrial Applications, ISI 2015*, 2016, pp. 40–45.

[61] P. Yao and H. Wang, "Dynamic Adaptive Ant Lion Optimizer applied to route planning for unmanned aerial vehicle," *Soft Comput.*, vol. 21, no. 18, pp. 5475–5488, 2017.

[62] H. M. Zawbaa, E. Emary, and B. Parv, "Feature selection based on antlion optimization algorithm," *Proc. 2015 IEEE World Conf. Complex Syst. WCCS 2015*, 2016.

[63] A. Mostafa, M. Houseni, N. Allam, A. E. Hassanien, H. Hefny, and P. W. Tsai, "Antlion optimization based segmentation for MRI liver images," in *Advances in Intelligent Systems and Computing*, vol. 536, 2017, pp. 265–272.

[64] G. Moslehi and M. Mahnam, "A Pareto approach to multi-objective flexible job-shop scheduling problem using particle swarm optimization and local search," *Int. J. Prod. Econ.*, vol. 129, no. 1, pp. 14–22, 2011.

[65] Y.-Y. Chen, C.-Y. Cheng, L.-C. Wang, and T.-L. Chen, "A hybrid approach based on the variable neighborhood search and particle swarm optimization for parallel machine scheduling problems—A case study for solar cell industry," *Int. J. Prod. Econ.*, vol. 141, no. 1, pp. 66–78, 2012.

[66] S. H. Pakzad-Moghaddam, "A Lévy flight embedded particle swarm optimization for multi-objective parallel-machine scheduling with learning and adapting considerations," *Comput. Ind. Eng.*, vol. 91, pp. 109–128, 2016.

[67] R. Zhang, S. Song, and C. Wu, "A simulation-based differential evolution algorithm for stochastic parallel machine scheduling with operational considerations," *Int. Trans. Oper. Res.*, vol. 20, no. 4, pp. 533–557, 2013.

[68] W. L. Wang, H. Y. Wang, Y. W. Zhao, L. P. Zhang, and X. L. Xu, "Parallel machine scheduling with splitting jobs by a hybrid differential evolution algorithm," *Comput. Oper. Res.*, vol. 40, no. 5, pp. 1196–1206, 2013.

[69] D. Kizilay, M. F. Tasgetiren, O. Bulut, and B. Bostan, "A discrete artificial bee colony algorithm for the assignment and parallel machine scheduling problem in DYO paint company," in *Proceedings of the 2014 IEEE Congress on Evolutionary Computation, CEC 2014*, 2014, pp. 653–660.

[70] E. . Caniyilmaz, B. . Benli, and M. S. . Ilkay, "An artificial bee colony algorithm approach for unrelated parallel machine scheduling with processing set restrictions, job sequence-dependent setup times, and due date," *Int. J. Adv. Manuf. Technol.*, vol. 77, no. 9–12, pp. 2105–2115, 2015.

[71] R. McNaughton, "Scheduling with Deadlines and Loss Functions," *Manage. Sci.*, vol. 6, pp. 1–12, 1959.

[72] G. Rabadi, R. Moraga, and A. Al-Salem, "Heuristics for the unrelated parallel machine scheduling problem with setup times," *J. Intell. Manuf.*, vol. 17, pp. 85–97, 2006.

[73] Yarpiz, Parallel Machine Scheduling using Simulated Annealing. http://yarpiz.com/367/ypap107-parallel-machine-scheduling/ Accessed 17 January 2018.

Robot Path Planning using Gray Wolf Optimizer

L.DOĞAN[1], U. YÜZGEÇ[2]

[1] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
lokmandogan34@gmail.com
[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
ugur.yuzgec@bilecik.edu.tr

Abstract **- Path planning problem plays an important role in mobile robot works. The robotic systems use intelligence algorithms to plan the path of the robot from one point to the other point. The main goal of path planning is to find the allowable movements of a robot in an environment with obstacles. These motions involve a path free of collision from the start position to the target position. In this study, Gray Wolf Optimization (GWO) algorithm was adapted to solve robot path planning problem. GWO algorithm imitates the hunting behavior and social leadership of gray wolves in nature. The leadership hierarchy consists of four grey wolf groups: alpha, beta, delta, and omega wolves. This algorithm comprises hunting mechanism with three stages: searching for prey, encircling prey, and attacking prey. In the test simulations of the robot path planning, we used a map with three circular obstacles. GWO algorithm was adapted to this problem. While finding the candidate solutions in path planning, three coordinate points are used between start and target points. For each iteration, these coordinate points are updated by GWO algorithm. If the solution point is in the obstacle zone, then violation is added to the cost function. The performance of GWO algorithm was evaluated with those of meta-heuristic algorithms for solving the robot path planning problem. The results obtained by GWO algorithm show that the optimal path is found for used test map.**

Keywords **- Robot Path Planning, Meta-Heuristic Algorithm, Gray Wolf Optimizer.**

I. INTRODUCTION

In last three decades, meta-heuristic algorithms have become very popular for the optimization problems. Meta-heuristic algorithms are inspired by the evolution concepts or the physics rules or the social behavior of swarms, flocks of animals in nature Meta-heuristic algorithms are classified into, physical based algorithms, evolutionary based algorithms swarm intelligence algorithms, bio-inspired algorithms and other nature-inspired algorithms [1][2]. In the physical based algorithms, solving the optimization problem begins with a single solution and it is updated by physical equations at each iteration. Tabu Search algorithm (TS) [3][4], Simulated Annealing algorithm (SA) [5][6] can be given as examples to physical based algorithms. Genetic algorithm (GA) [3][4] and Differential Evolution (DE) algorithm [9]–[11] are the well known examples of the evolutionary based meta-heuristic algorithms. Some of swarm intelligence algorithms include Particle Swarm Optimization (PSO) algorithm by Kennedy & Eberhart [12][13], Artificial Bee Colony (ABC) algorithm by Karaboga [14][15], Ant Colony Optimization (ACO) algorithm by Dorigo et al. [16][17] and Fish Swarm Algorithm (FSA) by Li et al. [18][19].

The bio-inspired algorithms mimics the activities of biological organisms. The most important examples of such algorithms are Artificial Immune algorithm (AI) [20][21] and Bacterial Foraging Optimization algorithm (BFO) [22]–[24]. Some of the other nature inspired meta-heuristic algorithms are Cuckoo Search Algorithm (CSA) [25][26], Firefly algorithm (FA) [27][28], Fruit Fly Optimization Algorithm (FOA) [29], Gravitational Search Algorithm (GSA) [30][31], Imperialist Competitive Algorithm (ICA) [32][33], Antlion Optimizer (ALO) [34][35], Dragonfly Optimization Algorithm (DOA) [36][37], Whale Optimization Algorithm (WOA) [38].

Gray Wolf Optimization (GWO) algorithm which was proposed by Mirjalili in 2014, imitates the hunting strategy and social leadership of gray wolves [39]. In this algorithm, gray wolves are classified into four levels according to the social hierarchy: alpha, beta, delta, and omega wolves. For example, an alpha wolf is a leader of wolf group, omega wolves are the grey wolves at the lowest level. In addition to the social leadership mechanism, gray wolf hunting strategy is another interesting mechanism of GWO algorithm. Although GWO algorithm is a new meta-heuristic algorithm, the studies about improvement and application on GWO can be found in the literature. Some of the studies are : a modified GWO algorithm based on complex-valued encoding [40], chaotic maps based GWO algorithm [41], a Levy flight-based GWO algorithm [42], optimal control of dc motor using GWO algorithm [43], hybrid maximum power point tracking (MPPT) algorithm with GWO algorithm [44], modified discrete grey wolf optimizer algorithm (MDGWO) for multilevel image thresholding [45].

Robot path planning problem for a mobile robot has been popular especially in the last decades and many approaches have been proposed for a robot in an area with a set of fixed obstacles. In this problem, main objective is to find collision-free trajectories for robots. Mobil robot should reach the target location as fast as possible and as short as possible distance between start and target locations [46]–[48]. The problem of path planning consists of the start point of the robot, the desired target point, the geometric description of the zone including the positions of the obstacles and boundaries of the zone.

In this study, the GWO algorithm is proposed to find the most suitable path from the starting point to the target point without touching any obstacle. To evaluate the performance of GWO algorithm, we used the zone including three circle

obstacles with different radius. GWO algorithm was compared with well-known meta-heuristic algorithms, such as Differential Evolution (DE) algorithm, Particle Swarm Optimization (PSO) algorithm, Artificial Bee Colony (ABC) algorithm and Firefly Optimization Algorithm (FOA).

II. GRAY WOLF OPTIMIZER (GWO)

The Gray Wolf Optimizer (GWO) is based on the behaviors of hunting strategy and social hierarchy of gray wolves. According to the hierarchy of gray wolves, there are four groups, namely alpha, beta, delta, and omega wolves. The leader or dominant wolf is called alpha and alpha wolf follows the other wolves in the group. The alpha is best wolf in terms of managing the group. The second in the social hierarchy of wolf group is beta wolf. Beta helps the leader wolf (alpha) in many activities. Delta wolf has to submit to alpha and beta wolves, but it adjudges the omega wolves. In this group, there are scouts, guards, elders, hunters, and caretakers. Omega wolf is gray wolf at the lowest level [39].

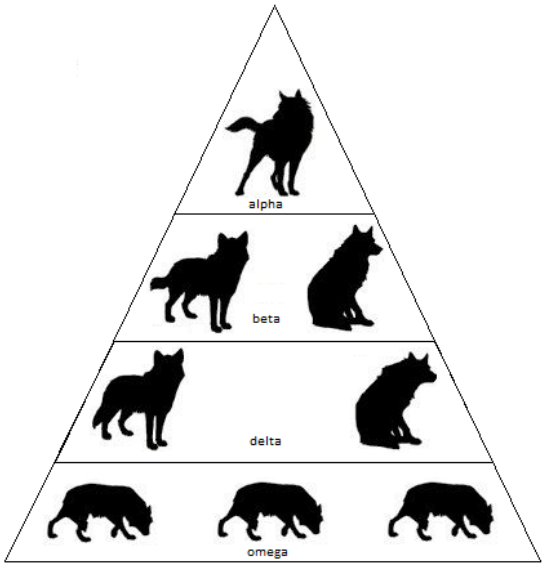


Figure 1: The hierarchy of gray wolves.

The group hunting strategy is another interesting social behavior of gray wolves. In this strategy of the gray wolves, firstly, they recognize the location of prey and encircle it under the leadership of the alpha wolf. In mathematical model of the hunting strategy of gray wolves, it is assumed that the alpha, beta and delta wolves provide better knowledge about the potential location of prey. As a result, the first three best solutions (alpha, beta, delta) are used to update the positions of wolves in GWO algorithm. There is no omega wolves in GWO code [39]. The mathematical model regarding hunting mechanism of gray wolves is given below:

$$\vec{D}_\alpha = \left|\vec{C}_\alpha \cdot \vec{X}_\alpha - \vec{X}_i\right| \quad (1)$$

$$\vec{D}_\beta = \left|\vec{C}_\beta \cdot \vec{X}_\beta - \vec{X}_i\right| \quad (2)$$

$$\vec{D}_\delta = \left|\vec{C}_\delta \cdot \vec{X}_\delta - \vec{X}_i\right| \quad (3)$$

$$\vec{U}_\alpha = \vec{X}_\alpha - \vec{A}_\alpha \vec{D}_\alpha \quad (4)$$

$$\vec{U}_\beta = \vec{X}_\beta - \vec{A}_\beta \vec{D}_\beta \quad (5)$$

$$\vec{U}_\delta = \vec{X}_\delta - \vec{A}_\delta \vec{D}_\delta \quad (6)$$

$$\vec{X}_i = \left(\vec{U}_\alpha + \vec{U}_\beta + \vec{U}_\delta\right)/3 \quad (7)$$

where $\vec{D}_\alpha, \vec{D}_\beta, \vec{D}_\delta$ are distance vector between prey and wolf (alpha, beta, delta), $\vec{X}_\alpha, \vec{X}_\beta, \vec{X}_\delta$ indicate the position vector of the prey for alpha, beta, delta wolves, $\vec{X}_i$ denotes the position vector of gray wolf at i_{th} iteration, $\vec{C}_\alpha, \vec{C}_\beta, \vec{C}_\delta, \vec{A}_\alpha, \vec{A}_\beta, \vec{A}_\delta$ indicate the coefficient vectors of alpha, beta, delta wolves, $\vec{U}_\alpha, \vec{U}_\beta, \vec{U}_\delta$ stand for the trial vector for alpha, beta, delta wolves. The coefficient vectors for alpha, beta and delta wolves are calculated as given below:

$$\vec{A}_\alpha = 2\vec{a}\vec{r}_{\alpha 1} - \vec{a} \quad (8)$$

$$\vec{C}_\alpha = 2\vec{r}_{\alpha 2} \quad (9)$$

$$\vec{A}_\beta = 2\vec{a}\vec{r}_{\beta 1} - \vec{a} \quad (10)$$

$$\vec{C}_\beta = 2\vec{r}_{\beta 2} \quad (11)$$

$$\vec{A}_\delta = 2\vec{a}\vec{r}_{\delta 1} - \vec{a} \quad (12)$$

$$\vec{C}_\delta = 2\vec{r}_{\delta 2} \quad (13)$$

where $\vec{a}$ indicates the vector linearly decreased from 2 to 0 during the optimization, $\vec{r}_{\alpha 1}, \vec{r}_{\beta 1}, \vec{r}_{\delta 1}$ denote the first random vector in [0,1] and $\vec{r}_{\alpha 2}, \vec{r}_{\beta 2}, \vec{r}_{\delta 2}$ denote the second random vector in [0,1].

The hunting mechanism of gray wolf group is illustrated in Fig. 2. The members of gray wolf group update their positions according to the alpha, beta, delta wolves and prey. The gray wolves catch their prey and finish the hunt by attacking the prey. In mathematical model, this situation is defined as decreasing $\vec{a}$ vector given below:

$$\vec{a} = 2 - \frac{2 \cdot Iter}{MaxIt} \quad (14)$$

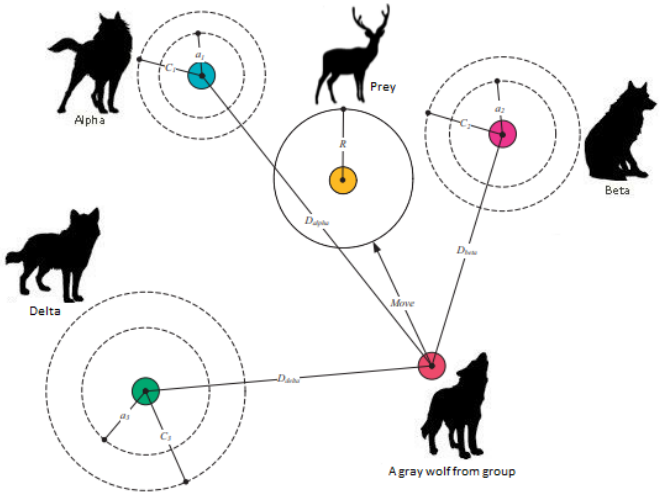


Figure 2: The hunting strategy of gray wolves.

The pseudo code of GWO algorithm is given in Algorithm 1.

Algorithm 1: Pseudo code of GWO algorithm.

```
Initialize the positions of gray wolves
Calculate the cost values of gray wolves
Save the best gray wolf as alpha wolf
Save the second best gray wolf as beta wolf
Save the third best gray wolf as delta wolf
  while (iteration < maximum iteration)
    Decrease a⃗ using Eq. (14)
    for each gray wolf
      Generate the coefficient vectors for alpha, beta, delta
      Calculate the distance vectors using Eqs. (1-3)
      Calculate the trial vectors using Eqs. (4-6)
      Update the position of gray wolf using Eq. (7)
    end for
    Calculate the cost values of updated gray wolves
    for each gray wolf
      if (gray wolf < alpha wolf )
        update alpha wolf
      else if (gray wolf < beta wolf)
        update beta wolf
      else if (gray wolf < delta wolf)
        update delta wolf
      end if
    end for
    Update the elite antlion
    increase iteration one
  end while
return alpha wolf
```

III. Robot Path Planning

The robot path planning problem is a NP-hard optimization problem and this problem is often solved by meta-heuristic algorithms in the literature. The main aim in solving this problem is that the mobile robot should reach from the start point to the target position in the shortest path without touching any obstacles. It consists of the start and target positions, the size of obstacles, the shape of obstacles, the number of obstacles, the zone's boundaries. The objective function of path planning problem is given below:

$$J = \min_{x,y} Q(1 + \beta V) \tag{15}$$

where β is violation coefficient (100), V indicates the violation cost, Q denotes the total distance between start and target points. In calculating the violation for the candidate solution, the following pseudo code was used.

Algorithm 2: Pseudo code of violation's calculation.

```
Violation ← 0
for each obstacle
    Calculate distance vector between the obstacle's center and path
    a ← max (1 − distance / radius_obs , 0)
    Violation ← Violation + mean (a)
end for
```

IV. Experimental Results

To show the performance of GWO algorithm for path planning problem, we have taken an instance scenario from www.yarpiz.com web site [49]. In Fig. 3, this instance scenario is shown. There are three circle-shaped obstacles with different radius in a 6x6 zone. The yellow square indicates the start point of the mobile robot and the green square indicates the target point. We have solved this problem using GWO algorithm and its performance has been compared with the several well-known meta-heuristic algorithms, such as Differential Evolution (DE) algorithm, Particle Swarm Optimization (PSO) algorithm, Artificial Bee Colony (ABC) algorithm and Firefly Optimization Algorithm (FOA).

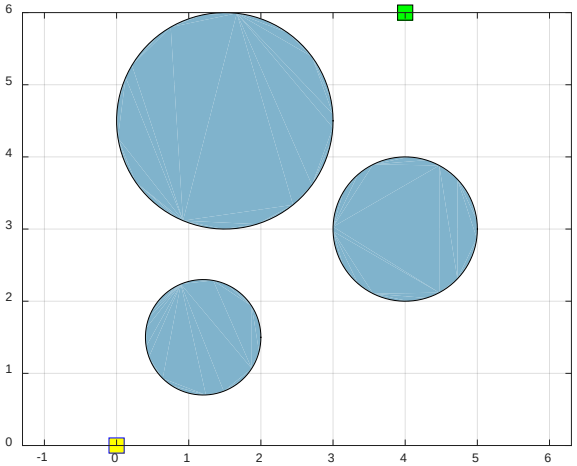


Figure 3: Path planning problem used in this study.

The codes of GWO and other meta-heuristic algorithms have been run on PC with Intel(R) Core(TM) i5-3230M CPU@2.60GHz RAM/8. Population size is 50, maximum number of iterations is 1000. The parameters of meta-heuristic algorithms used for robot path planning problem are summarized in Table 1. Fig.4 shows the best path planning solution obtained at the end of one-time run by the GWO and other meta-heuristic algorithms. As can be seen from this figure, the results of all algorithms are quite close together.

Table 1: Parameters of meta-heuristic algorithms.

Algorithm	Parameters
DE	Lower Bound of Scaling Factor : 0.5 Upper Bound of Scaling Factor : 1.0 Crossover Probability : 0.7 Strategy : rand2bin
PSO	Inertia Weight : 1.0 Inertia Weight Damping Ratio : 0.99 Personal Learning Coefficient : 1.5 Global Learning Coefficient : 2.0
ABC	Number of Onlooker Bees : 50 Abandonment Limit Parameter : round(0.6*NumberOfVar*PopSize)
FOA	Light Absorption Coefficient : 1.0 Initial Attraction Coefficient: 2.0 Mutation Coefficient : 0.2 Mutation Coefficient Damping R. : 0.98
GWO	Number of Antlions : 50

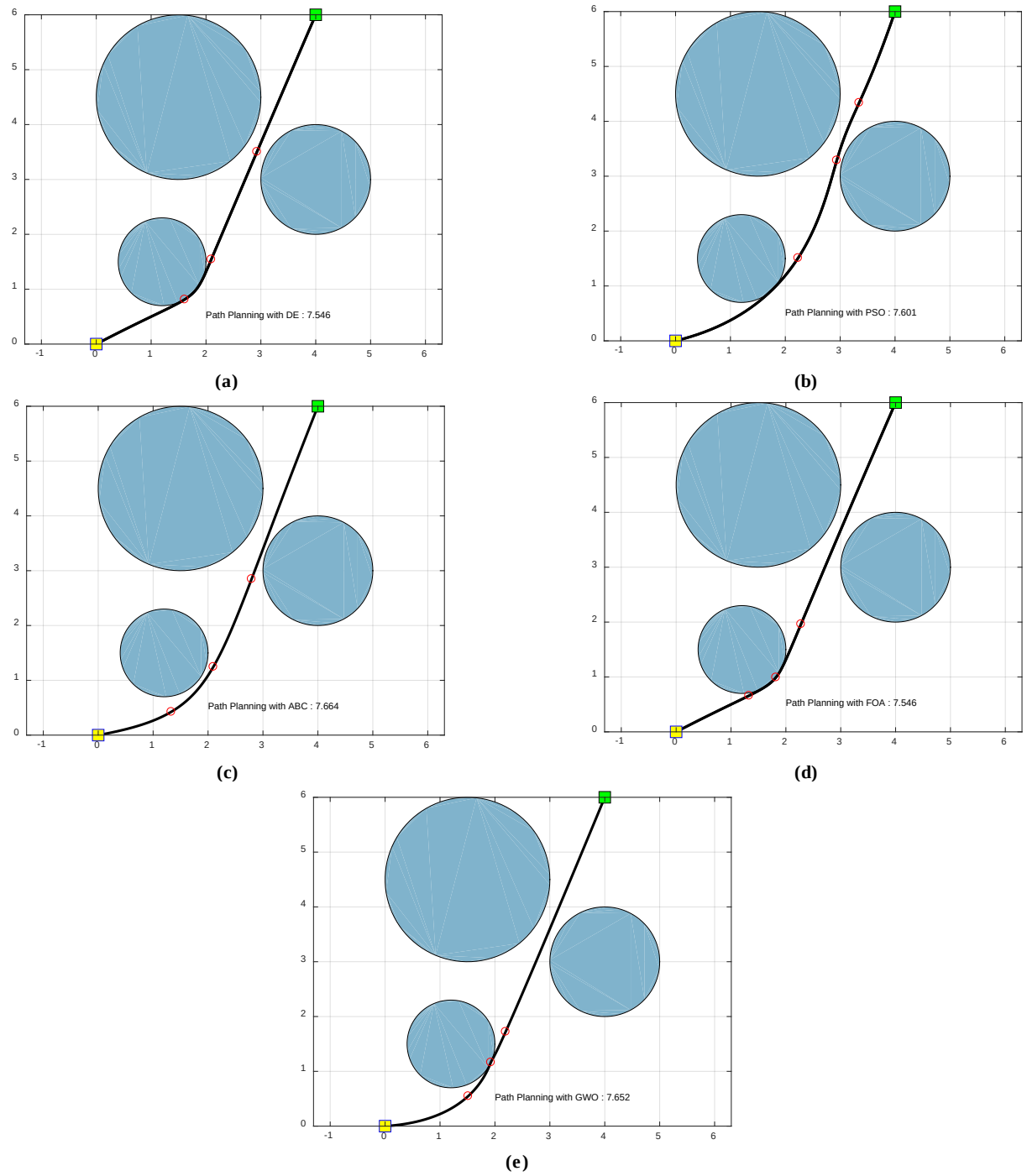


Figure 4: The best solutions obtained by meta-heuristic algorithms, (a) DE, (b) PSO, (c) ABC, (d) FOA, (e) GWO.

The cost value of GWO algorithm is found as 7.652. Fig. 5 shows all solutions of the path planning problem obtained by GWO algorithm. According to this figure, GWO algorithm deals with finding the suitable path with the minimum distance between start and target locations during optimization. Moreover, the paths found by the best current solution at each iteration have very little violation. The convergence curves of GWO and the other algorithms are shown in Fig.6. This figure show that the performance of GWO provides the competitive result and it can be an alternative algorithm for path planning.

V. CONCLUSION

In this study, robot path planning problem was discussed and GWO algorithm was proposed for solving this problem. To evaluate the algorithm's performance on solving path planning problem, four well-known meta-heuristic algorithms (DE, PSO, ABC and FOA) were used. The comparison results show that the proposed GWO algorithm is able to provide very

competitive results. In future, we will add opposition learning to the GWO algorithm to increase its performance.

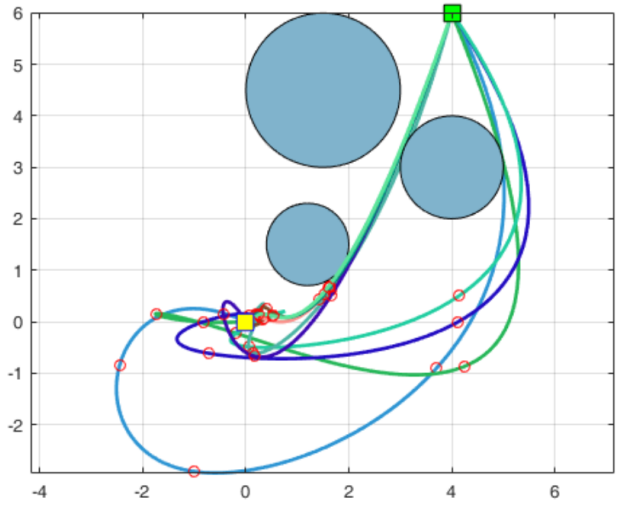


Figure 5: Solutions of path planning problem using GWO algorithm for all iterations.

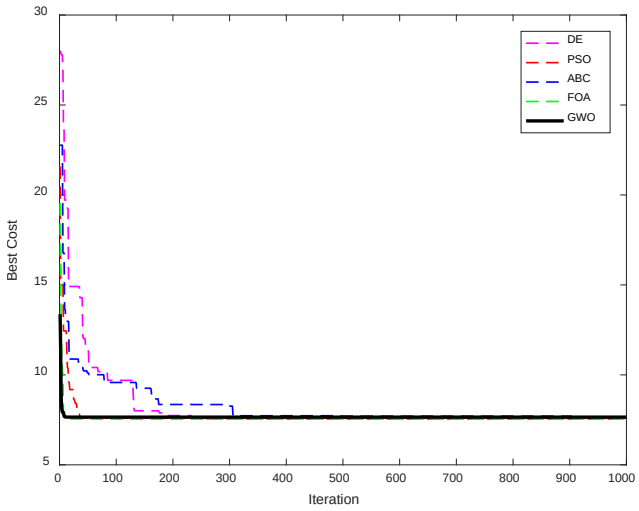


Figure 6: Comparison results of GWO and other meta-heuristic algorithms for robot path planning problem.

REFERENCES

[1] Z. Beheshti and S. M. H. Shamsuddin, "A review of population-based meta-heuristic algorithm," *International Journal of Advances in Soft Computing and its Applications*, vol. 5, no. 1. 2013.

[2] C. Blum, J. Puchinger, G. R. Raidl, and A. Roli, "A brief survey on hybrid metaheuristics," in *Bioinspired Optimization Methods and their Applications - Proceedings of the 4th International Conference on Bioinspired Optimization Methods and their Applications, BIOMA 2010*, 2010, pp. 3–16.

[3] E. Nowicki and C. Smutnicki, "A fast tabu search algorithm for the permutation flow-shop problem," *Eur. J. Oper. Res.*, vol. 91, no. 1, pp. 160–175, 1996.

[4] M. Gendreau, M. Iori, G. Laporte, and S. Martello, "A Tabu Search Algorithm for a Routing and Container Loading Problem," *Transp. Sci.*, vol. 40, no. 3, pp. 342–350, 2006.

[5] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science (80-.).*, vol. 220, no. 4598, pp. 671–680, 1983.

[6] T. D. Mavridou and P. M. Pardalos, "Simulated Annealing and Genetic Algorithms for the Facility Layout Problem: A Survey," *Comput. Optim. Appl.*, vol. 7, pp. 111–126, 1997.

[7] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. Addison-We. 1989.

[8] M. Noraini and J. Geraghty, "Genetic algorithm performance with different selection strategies in solving TSP," *World Congr. Eng.*, vol. II, no. 978-988-19251-4–5, pp. 4–9, 2011.

[9] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[10] K. V Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, vol. 28. 2005.

[11] U. Yüzgeç, "Performance comparison of differential evolution techniques on optimization of feeding profile for an industrial scale baker's yeast fermentation process," *ISA Trans.*, vol. 49, no. 1, pp. 167–176, 2010.

[12] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.

[13] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, 2002.

[14] D. Karaboga, "An idea based on Honey Bee Swarm for Numerical Optimization," *Tech. Rep. TR06, Erciyes Univ.*, no. TR06, p. 10, 2005.

[15] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.

[16] M. Dorigo, V. Maniezzo, and A. Colomi, "Ant system: Optimization by a colony of cooperating agents," *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 26, no. 1, pp. 29–41, 1996.

[17] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Comput. Intell. Mag.*, vol. 1, no. 4, pp. 28–39, 2006.

[18] X. Li, Z. Shao, and J. Qian, "An optimizing method based on autonomous animats: fish-swarm algorithm," *Syst. Eng. Theory Pract.*, vol. 22, no. 11, pp. 32–38, 2002.

[19] M. Neshat, G. Sepidnam, M. Sargolzaei, and A. N. Toosi, "Artificial fish swarm algorithm: a survey of the state-of-the-art, hybridization, combinatorial and indicative applications," *Artif. Intell. Rev.*, vol. 42, no. 4, pp. 965–997, 2014.

[20] S. A. Hofmeyr and S. Forrest, "Architecture for an Artificial Immune System," *Evol. Comput.*, vol. 8, no. 4, pp. 443–473, 2000.

[21] J. Timmis, P. Andrews, and E. Hart, "On artificial immune systems and swarm intelligence," *Swarm Intell.*, vol. 4, no. 4, pp. 247–273, 2010.

[22] S. Das, A. Biswas, S. Dasgupta, and A. Abraham, "Bacterial Foraging Optimization Algorithm : Theoretical Foundations , Analysis , and Applications," *Found. Comput. Intell. Vol. 3*, vol. 3, pp. 23–55, 2009.

[23] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization and control," *Control Syst. IEEE*, vol. 22, no. 3, pp. 52–67, 2002.

[24] K. M. Passino, "Bacterial Foraging Optimization," *Int. J. Swarm Intell. Res.*, vol. 1, no. 1, pp. 1–16, 2010.

[25] X. S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Computing and Applications*, vol. 24, no. 1. pp. 169–174, 2014.

[26] X.-S. Yang, "Cuckoo Search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214.

[27] X. S. Yang, "Firefly Algorithm," *Nature-Inspired Metaheuristic Algorithms*, pp. 79–90, 2007.

[28] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5792 LNCS, pp. 169–178.

[29] H. Iscan and M. Gunduz, "Parameter Analysis on Fruit Fly Optimization Algorithm," *J. Comput. Commun.*, vol. 2, no. March, pp. 137–141, 2014.

[30] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," *Inf. Sci. (Ny).*, vol. 179, no. 13, pp. 2232–2248, 2009.

[31] N. M. Sabri, M. Puteh, and M. R. Mahmood, "A review of gravitational search algorithm," *Int. J. Adv. Soft Comput. its Appl.*, vol. 5, no. 3, 2013.

[32] E. Atashpaz-Gargari and C. Lucas, "Imperialist competitive

algorithm: An algorithm for optimization inspired by imperialistic competition," in *2007 IEEE Congress on Evolutionary Computation, CEC 2007*, 2007, pp. 4661–4667.

[33] S. Hosseini and A. Al Khaled, "A survey on the Imperialist Competitive Algorithm metaheuristic: Implementation in engineering domain and directions for future research," *Applied Soft Computing Journal*, vol. 24. pp. 1078–1094, 2014.

[34] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, 2015.

[35] H. Kiliç and U. Yüzgeç, "Improved antlion optimization algorithm," in *2nd International Conference on Computer Science and Engineering, UBMK 2017*, 2017.

[36] S. Mirjalili, "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2016.

[37] S. R. Sree Ranjini and S. Murugan, "Memory based Hybrid Dragonfly Algorithm for numerical optimization problems," *Expert Syst. Appl.*, vol. 83, pp. 63–78, 2017.

[38] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, 2016.

[39] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.

[40] Q. Luo, S. Zhang, Z. Li, and Y. Zhou, "A Novel Complex-Valued Encoding Grey Wolf Optimization Algorithm," *Algorithms*, vol. 9, no. 1, p. 4, 2015.

[41] M. Kohli and S. Arora, "Chaotic grey wolf optimization algorithm for constrained optimization problems," *J. Comput. Des. Eng.*, 2017.

[42] S. Amirsadri, S. J. Mousavirad, and H. Ebrahimpour-Komleh, "A Levy flight-based grey wolf optimizer combined with back-propagation algorithm for neural network training," *Neural Computing and Applications*, pp. 1–14, 2017.

[43] A. Madadi and M. M. Motlagh, "Optimal Control of DC motor using Grey Wolf Optimizer Algorithm," *Tech. J. Eng. Appl.*, pp. 373–379, 2014.

[44] S. Mohanty, B. Subudhi, and P. K. Ray, "A Grey Wolf-Assisted Perturb & Observe MPPT Algorithm for a PV System," *IEEE Trans. Energy Convers.*, vol. 32, no. 1, pp. 340–347, 2017.

[45] L. Li, L. Sun, J. Guo, J. Qi, B. Xu, and S. Li, "Modified Discrete Grey Wolf Optimizer Algorithm for Multilevel Image Thresholding," *Comput. Intell. Neurosci.*, vol. 2017, 2017.

[46] C. Alexopoulos and P. M. Griffin, "Path planning for a mobile robot," *IEEE Trans. Syst. Man. Cybern.*, vol. 22, no. 2, pp. 1132–1139, 1992.

[47] N. Sariff and N. Buniyamin, "An overview of autonomous mobile robot path planning algorithms," in *Research and Development, 2006. SCOReD 2006. 4th Student Conference on*, 2006, pp. 183–188.

[48] C. Hofner and G. Schmidt, "Path planning and guidance techniques for an autonomous mobile cleaning robot," *Rob. Auton. Syst.*, vol. 14, no. 2–3, pp. 199–212, 1995.

[49] Yarpiz, Optimal Robot Path Planning using PSO. http://yarpiz.com/403/ypap115-path-planning, Accessed 28 March 2018.

Training Multi-Layer Perceptron using Opposition based Learning Spiral Optimization Algorithm

U. YÜZGEÇ[1], C.KARAKUZU[2]

[1] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
ugur.yuzgec@bilecik.edu.tr
[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey
cihan.karakuzu@bilecik.edu.tr

***Abstract* - In this study, the Opposition based learning Spiral Optimization Algorithm (OBLSOA) is presented for training Multi-Layer Perceptron (MLP). The main idea of Spiral Optimization Algorithm (SOA) is based on the dynamic step dimension in its spiral path trajectory. The primary opposition based learning (OBL) concept first was come from the Yin-Yang symbol in the ancient Chinese philosophy. According to OBL concept, if a candidate point is far from the solution, the opposite point of this candidate can be closer to the solution than that point. We applied OBL concept to spiral optimization algorithm for training MLP. OBLSOA comprises two main stages: the first is the opposition-based learning population initialization and the other is opposition-based learning generation jumping. To evaluate the performance of the proposed OBLSOA, we used eight standard datasets including four classification datasets (XOR, balloon, Iris, breast cancer) and three function-approximation datasets (sigmoid, cosine, and sine). The performance proposed OBLSOA was compared with the original SOA for all datasets in terms of the Mean Square Error (MSE) metric. The training and test results show that the proposed OBLSOA is able to be provide very competitive and effective in training MLPs.**

***Keywords* - Multi-Layer Perceptron, Opposition based Learning, Spiral Optimization Algorithm.**

I. INTRODUCTION

In last three decades, the randomize process based meta-heuristic algorithms have become popular algorithms to solve the nonlinear and complex optimization problems [1][2]. These algorithms are inspired by physical event, swarm behaviour, evolutionary mechanism, or hunting strategy of animals. In literature, there are many studies regarding the new methods, improvement the performance of current algorithms, applications on real-world problems. Some of the well-known meta-heuristic algorithms are Genetic Algorithm (GA) [3], Differential Evolution (DE) algorithm [4][5], Particle Swarm Optimization (PSO) algorithm [6], Artificial Bee Colony (ABC) algorithm [7][8], Firefly Optimization Algorithm (FOA) [9][10], Cuckoo Optimization Algorithm (COA) [11][12], Bat Optimization Algorithm (BOA) [13][14], Flower Pollination Algorithm (FPA) [15][16] and Spiral Optimization Algorithm (SOA) [17][18].

The Spiral Optimization Algorithm (SOA) that is proposed by Tamura & Yasuda is a meta-heuristic algorithm inspired by the spiral motion in nature [17][18]. This algorithm is based on searching with the dynamic step size in a spiral trajectory. At initialization stage of SOA, the step size is selected as a large value and this size is reduced in the later phase, so that search agent moves intensely within a smaller region of the search area. SOA consists of two parameters whose names are spiral angle and spiral radius. In general, these parameters are determined values before solving the optimization problem.

Although the meta-heuristic algorithms provide the good performance for different optimization problems, they face with some challenges, such as getting trapped to a local optimal point, having convergence problems, capability of exploration and exploitation in search space, etc. Opposition based learning was proposed by Tizhoosh in 2005 and it is based on the concurrent consideration of an estimate and its corresponding opposite estimate to obtain a better solution in each iteration [19][20]. Opposition-Based Learning based Spiral Optimization Algorithm (OBLSOA) was proposed by Yüzgeç et al. [23] for single objective optimization problems.

In this study, OBLSOA was adapted for training Multi-Layer Perceptron (MLP). To show the performance of the proposed OBLSOA, eight standard datasets including four classification datasets (XOR, balloon, Iris and breast cancer) and three function-approximation datasets (sigmoid, cosine, and sine) were taken from literature. The performance proposed OBLSOA was compared with that of the original SOA for all datasets according to the Mean Square Error (MSE). The training and test results show that the proposed OBLSOA is able to be provide very competitive and effective in training MLPs.

II. SPIRAL OPTIMIZATION ALGORITHM (SOA)

Spiral Optimization Algorithm (SOA) was firstly presented by Tamura & Yasuda [17][18] for only two dimensional optimization problems. This algorithm imitates the analogy of spiral events in nature. For the optimization problems with n dimensional, Tamura and Yasuda improved the SOA. The

SOA has got some advantages: few control variables, ability at local search, fast result, easy adaptation for optimization problem. In SOA concept, there are two important parameters: spiral angle and spiral radius. In general, these parameters are selected as fixed values. There are self adaptive based SOA in the literature. In self adaptive SOA, these two parameters are changed adaptively during optimization process [24].

The main idea of SOA is based on searching with the dynamic step dimension in a spiral trajectory. The step size is determined with the largest value at the initial iteration and then it is reduced by increasing iterations, so that the search agent is moved at the centre of the spiral shape. The length of the step size from iteration to iteration is updated according to the spiral radius. On the other hand, the shape of spiral form is changed using the spiral angle value. Moreover, this parameter affects the distance between two points on the spiral path [23][24]. In Fig.1, the spiral forms are shown for different spiral radius and angles values.

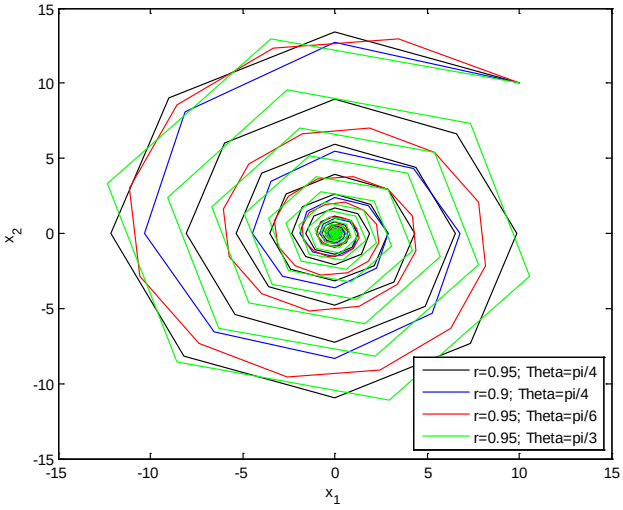


Figure 1: Spiral form with different radius and angle values [23].

The updating procedure of the candidate solution in the population for spiral model is given below:

$$x_{k+1} = r.M(\theta).x_k - (r.M(\theta) - I_n).x_{best} \quad (1)$$

where r indicates the spiral radius, θ denotes the spiral angle, $M(\theta)$ represents the rotation matrix, I_n stands for the unit matrix with n dimensional, x_{best} indicates the best solution. For the optimization problems with n dimensional, the multipoint updating mechanism is presented as given in Eq.(2):

$$x_{(k+1)_i} = r.M(\theta).x_{k_i} - (r.M(\theta) - I_n).x_{best} \quad (2)$$

where $i = 1,2,\dots m$ and m represents the number of problem dimension. The spiral angle around the center at each iteration is between $[0\ 2\pi]$ and a convergence rate of distance between a point and the center becomes $(0\ 1)$. The rotation matrix $M(\theta)$ with two dimensional is given below:

$$M_{2_{(1,2)}}(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad (3)$$

Each two dimensional rotation matrix for n dimensional space is defined as given below:

$$M_{n_{(i,j)}}(\theta_{(i,j)}) = \begin{bmatrix} 1 & & & & & & & \\ & \ddots & & & & & & \\ & & 1 & & & & & \\ & & & \cos(\theta_{i,j}) & & -\sin(\theta_{i,j}) & & \\ & & & & 1 & & & \\ & & & & & \ddots & & \\ & & & & & & 1 & \\ & & & \sin\theta_{(i,j)} & & \cos(\theta_{i,j}) & & \\ & & & & & & 1 & \\ & & & & & & & \ddots \\ & & & & & & & & 1 \end{bmatrix} \quad (4)$$

Whose blank elements mean zero. According to this definition, many rotation matrices are formed by the way selecting two axes consist of each rotation plane during their permutations or combinations [23][24]. The combination of the rotation matrix $M_n(\theta)$ consists of all rotation matrices given in Eq. (4). $M_n(\theta)$ is based on all combination ($n(n-1)/2$ combinations) of two axes. $M_n(\theta)$ is defined as following equation:

$$M_n(\theta_{(i,j)}) = \prod_{i<i} M_{n_{(i,j)}}(\theta_{(i,j)}) \quad (5)$$

The pseudo code of basic SOA is presented below:

Algorithm 1: Pseudo Code of SOA.

Input: Population size (PS), Problem dimension (D), Boundary values (B), Spiral radius, Spiral angle
Output: The best fitness function, The best individuals
Determine the initial population (PS, D, B) randomly
Calculate the cost value of each individual
The best individual ← The individual with the best cost
The best fitness value ← Calculate the cost (the best individual)
While(*iteration* < *Maximum iteration*)
    Calculate $rM_n(\theta)$
    Updating the each individuals using Eq. (2)
    Calculate the new costs (for new individuals)
    Update the best individual
    Update the best cost value
End
Return(The best fitness function, The best individuals)

III. Opposition Based Learning Spiral Optimization Algorithm (OBLSOA)

The idea of opposition-based learning was presented by Tizhoosh [19][21] in 2005. In the concept of opposition-based learning, there is an opposite candidate solution for each candidate solution in the search space. In optimization problems, there are two ways of searching by opposite candidate solutions: defining a function like a map including every solution in the search space or finding opposite candidate solution. As to probability theory, if a candidate solution is far from the global solution, its opposite candidate

solution can be closer to the global solution than that solution. In this study, Opposition based Learning Spiral Optimization Algorithm (OBLSOA) was presented by Yüzgeç et. al [23] for two-dimensional single objective optimization problems. In general OBLSOA consists of two main steps:

- opposition-based population initialization
- opposition-based generation jumping

A. Opposition-based Population Initialization

In case of no information about the solution, the candidates are initialized randomly at the beginning of the optimization process. In opposition based population initialization, the opposition candidates are calculated from candidates in the population. As a result, the individuals assigned by opposition based population initialization are more suitable than those assigned by random population initialization. The procedure is given below:

1. Initial population is determined randomly
2. Calculate the opposite candidate for each candidate in initial population using Eq. (6):
   $$\breve{x}_{i,j} = Ub + Lb - x_{i,j} \quad (6)$$
   where $i = 1,2,\dots PS$, PS is population size, Ub and Lb represent the maximum and minimum boundary values, $j = 1,2,\dots D$, D is dimension, $x_{i,j}$ denotes the i_{th} candidate, $\breve{x}_{i,1j}$ represents the i_{th} opposite candidate at first iteration.
3. Calculate the cost values for the initial population and its opposite-based population and sort both of the populations according to their cost values.
4. The better candidates are selected from the initial population and its opposite.

B. Opposition-based Generation Jumping

At the end of each iteration in solving the optimization problem, the opposite candidate is found according to the pre-determined jumping probability value and the better candidates are selected from combination of the current and the opposite populations for next iteration. Opposition-based generation jumping runs according to comparison result between the jumping rate (J_r) and a random number in range [0 1] as given below:

$$x_{i,g+1} = \begin{cases} \breve{x}_{i,g} = x_{up} + x_{low} - x_{i,g}, & if(rnd_1 < J_r) \\ x_{i,g}, & otherwise \end{cases} \quad (7)$$

The optimal jumping rate should have a problem-oriented value, but the jumping rate is used as a small value for the problems with high-dimensional. The pseudo code of OBLSOA is presented below:

Algorithm 1: Pseudo Code of OBLSOA.

Input: Population size (PS), Problem dimension (D), Boundary values (B), Spiral radius, Spiral angle, Jumping rate
Output: The best fitness function, The best individuals
Determine the initial population (PS, D, B)
Calculate the cost value of each candidate
Search the opposition population
Calculate the cost value of each opposite candidate
Combine both populations and sort according to the cost values
Select the better candidates from merged population
Save the best candidate and its cost value
While(*iteration < Maximum iteration*)
  Calculate $rM_n(\theta)$
  Updating the each individuals using Eq. (2)
  Calculate the new costs (for new individuals)
  Updating the best individual
  Updating the best cost value
  If *(random number < jumping rate)*
    Search the opposition population
    Calculate the cost value of each opposite candidate
    Select the better candidates
  End
End

IV. MULTI-LAYER PERCEPTRON (MLP)

A multi-layer perceptron (MLP) is a class of feed forward neural network (FNN). FNN with one hidden layer is called MLP. MLP comprises at least three layers of nodes. Each node is a neuron with a nonlinear activation function except of the input nodes. In Fig. 2, an example of MLP structure is shown.

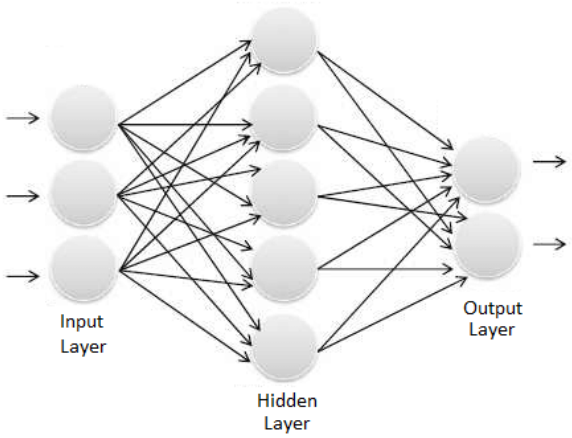


Figure 2: MLP structure with one hidden layer.

In MLP structure, there are the connection weights between the nodes and biases like thresholds on the nodes. In training MLPs, OBLSOA tries to find suitable values for all weights and biases in order to achieve a desirable relation between the inputs and outputs.

V. EXPERIMENTAL RESULTS

In this study, the proposed OBLSOA is used for training the MLPs. The most important variables in training phase of an MLP are the weights and biases in MLP structure. OBLSOA uses the optimization problem's variables as the weights and biases. We have been used Mean Square Error (MSE) as the objective function to evaluate the MLP's training. The average of MSE metric is evaluated for training the MLP structure. The optimization problem about training MLP is given for the OBLSOA as following equation:

$$\min J = \overline{MSE} \quad (8)$$

The process of training MLP using OBLSOA is shown in Fig. 3. To show the performance OBLSOA based MLP trainer, we used four standard classification datasets taken from the University of California at Irvine (UCI) Machine Learning Repository [25]: XOR, balloon, Iris and breast cancer. Besides, three function approximation datasets were taken from [22]: sigmoid, cosine and sine functions. Table 2 and Table 3 summarized some information about the datasets used in this study.

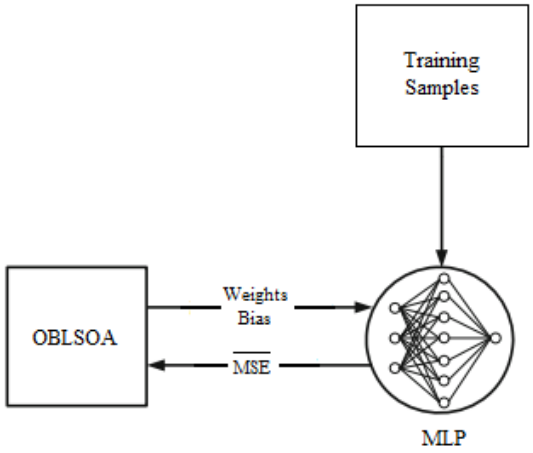


Figure 3: The process of training MLP using OBLSOA.

Table 1: Classification datasets.

Name	Number of attributes	Number of training samples	Number of test samples	Number of classes
XOR	3	8	8	2
Balloon	4	16	16	2
Iris	4	150	150	3
Cancer	9	599	100	2

Table 2: Function-approximation datasets.

Function-approximation datasets	Training samples	Test samples
Sigmoid	61: x in [−3:0.1:3]	121: x in [−3:0.05:3]
Cosine	31: x in [1.25:0.05:2.75]	38: x in [1.25:0.04:2.75]
Sine	126: x in [-2π:0.1:2π]	252: x in [-2π:0.05:2π]

The codes of MLP-OBLSOA have been run on PC with Intel(R) Core(TM) i5-3230M CPU@2.60GHz RAM/8. Population size is 200, maximum number of iterations is 200. In Table 3, the results obtained by SOA and OBLSOA are provided for classification datasets. Fig. 4-6 show the comparison results of SOA and OBLSOA for function approximation datasets.

Table 3: The Results for classification datasets.

Name	SOA		OBLSOA	
	MSE	Classification rate	MSE	Classification rate
XOR	0.00061634	100%	**3.0409e-05**	100%
Balloon	3.5762e-09	100%	6.1233e-07	100%
Iris	0.67253	00%	**0.66707**	00%
Cancer	0.0073029	63%	**0.0030123**	97%

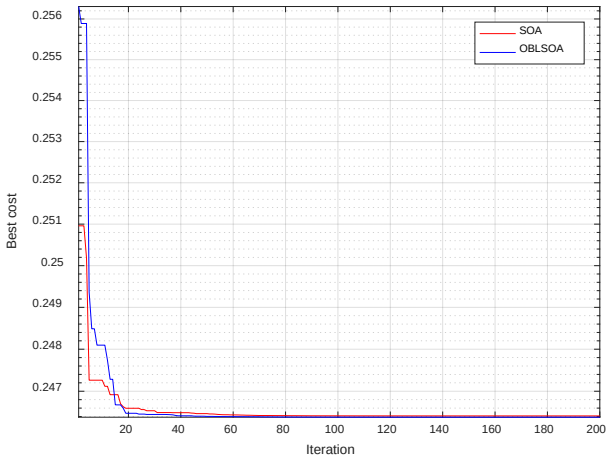


Figure 4: Best cost curves of OBLSOA and SOA for Sigmoid dataset.

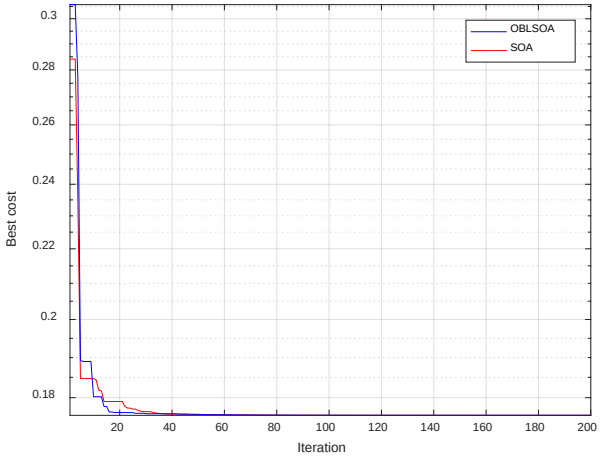


Figure 5: Best cost curves of OBLSOA and SOA for Cosine dataset.

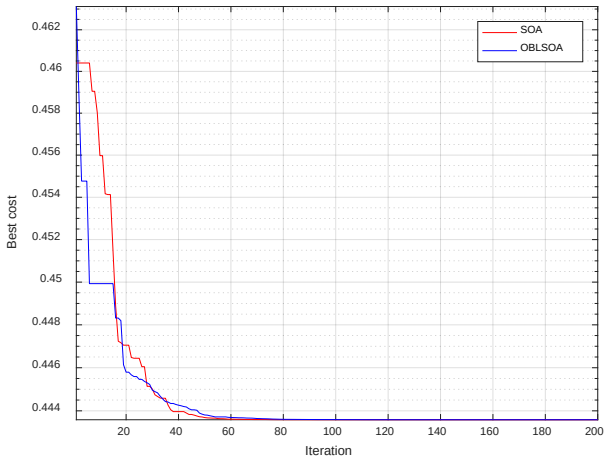


Figure 6: Best cost curves of OBLSOA and SOA for Sine dataset.

VI. CONCLUSION

In this study, the Opposition based learning Spiral Optimization Algorithm (OBLSOA) was proposed for training Multi-Layer Perceptrons (MLPs). The optimization problem of training MLP was presented according to MSE metric for the meta-heuristic algorithms (SOA and OBLSOA). The

proposed OBLSOA was used to optimize the values of weights and biases in MLP structure. To evaluate the performance of MLP-OBLSOA, we used four standard classification datasets (XOR, balloon, Iris and breast cancer) and three function approximation datasets (sigmoid, cosine and sine functions). The OBLSOA's results show that the proposed opposition-based learning concept is able to be effective in training MLPs for different datasets. As can be seen from the results in Table 3, OBLSOA is more effective than original SOA. For the results with function-approximation datasets, OBLSOA and SOA present close convergence curves.

REFERENCES

[1] I. Boussaïd, J. Lepagnot, and P. Siarry, "A survey on optimization metaheuristics," in *Information Sciences*, 2013, vol. 237, pp. 82–117.

[2] Z. Beheshti and S. M. H. Shamsuddin, "A review of population-based meta-heuristic algorithm," *International Journal of Advances in Soft Computing and its Applications*, vol. 5, no. 1. 2013.

[3] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, vol. Addison-We. 1989.

[4] K. V Price, R. M. Storn, and J. A. Lampinen, *Differential Evolution: A Practical Approach to Global Optimization*, vol. 28. 2005.

[5] R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces," *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, 1997.

[6] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.

[7] D. Karaboga, "An idea based on Honey Bee Swarm for Numerical Optimization," *Tech. Rep. TR06, Erciyes Univ.*, no. TR06, p. 10, 2005.

[8] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.

[9] X. S. Yang, "Firefly Algorithm," *Nature-Inspired Metaheuristic Algorithms*, pp. 79–90, 2007.

[10] X. S. Yang, "Firefly algorithms for multimodal optimization," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2009, vol. 5792 LNCS, pp. 169–178.

[11] X. S. Yang and S. Deb, "Cuckoo search: Recent advances and applications," *Neural Computing and Applications*, vol. 24, no. 1. pp. 169–174, 2014.

[12] X.-S. Yang, "Cuckoo Search via Lévy flights," in *2009 World Congress on Nature & Biologically Inspired Computing (NaBIC)*, 2009, pp. 210–214.

[13] X.-S. Yang, "Bat Algorithm: Literature Review and Applications," *J. Bio-Inspired Comput.*, vol. 5, no. 3, pp. 141–149, 2013.

[14] X. S. Yang, "A new metaheuristic Bat-inspired Algorithm," in *Studies in Computational Intelligence*, 2010, vol. 284, pp. 65–74.

[15] X.-S. Yang, "Flower Pollination Algorithms," in *Nature-Inspired Optimization Algorithms*, 2014, pp. 155–173.

[16] X. S. Yang, M. Karamanoglu, and X. He, "Flower pollination algorithm: A novel approach for multiobjective optimization," *Eng. Optim.*, vol. 46, no. 9, pp. 1222–1237, 2014.

[17] K. Tamura and K. Yasuda, "Spiral dynamics inspired optimization," *J. Adv. Comput. Intell. Intell. Informatics*, vol. 15, no. 8, pp. 1116–1122, 2011.

[18] K. Tamura and K. Yasuda, "Spiral Optimization Algorithm Using Periodic Descent Directions," *SICE J. Control. Meas. Syst. Integr.*, vol. 9, no. 3, pp. 134–143, 2016.

[19] H. R. Tizhoosh, "Opposition-Based Learning: A New Scheme for Machine Intelligence," *Comput. Intell. Model. Control Autom. 2005 Int. Conf. Intell. Agents, Web Technol. Internet Commer. Int. Conf.*, vol. 1, pp. 695–701, 2005.

[20] S. Rahnamayan, H. R. Tizhoosh, and M. M. Salama, "Opposition-based differential evolution," *Stud. Comput. Intell.*, vol. 143, pp. 155–171, 2008.

[21] Q. Xu, L. Wang, N. Wang, X. Hei, and L. Zhao, "A review of opposition-based learning from 2005 to 2012," *Engineering Applications of Artificial Intelligence*, vol. 29. pp. 1–12, 2014.

[22] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Let a biogeography-based optimizer train your Multi-Layer Perceptron," *Inf. Sci. (Ny).*, vol. 269, pp. 188–209, 2014.

[23] U.Yüzgeç, T. İnaç, "Opposition based Spiral Optimization Algorithm," *Comp. Sci. and Eng., UBMK2016, Int. Conf.*, pp. 124–128, 2016.

[24] U.Yüzgeç, T. İnaç, "Adaptive Spiral Optimization Algorithm for Benchmark Problems," *Bilecik Seyh Edebali University Science Journal.*, vol. 3, no. 1, pp. 8–15, 2016.

[25] C. Blake, C.J. Merz {UCI} Repository of machine learning databases, 1998.

Kinect Based English Teaching Game for Preschool Children

M.V. ARISOY [1] and E.U. KÜÇÜKSİLLE[2]

[1] Mehmet Akif Ersoy University, Burdur/Turkey, mvarisoy@mehmetakif.edu.tr
[2]Süleyman Demirel University, Isparta/Turkey, ecirkucuksille@sdu.edu.tr

***Abstract*-The use of technological devices in education has become inevitable today as technology is proceeding at a dizzying pace. Technological devices and softwares developed within the scope of these devices have become especially used within the game-based learning system. Especially in pre-school level, to teach some concepts to illiterate children, game based learning system is used.**

In this study, a platform based on Kinect V2 was developed with the aim of teaching English words to pre-school children. The study was tested on a select group of pre-school children. Eventually it has been observed that the English equivalents of the concepts of color and shape have been retained for a longer period of time due to the support of this visual and audio software material. In addition, it has been determined that the motivation of the student is made permanent because it enables to learn by amusing.

Keywords: Learning English, game based education, Kinect, Unity 3D.

I. INTRODUCTION

Nowadays, when the necessity and importance of foreign language learning grows, language teaching is started at a very early age. This age level is reduced to 4 years of age in preschool education. Since pre-school children do not have literacy skills, it is possible to teach them a foreign language other than their native language by using game-based tools. These tools are sometimes methods and materials that the teacher has developed with his / her professional skill, and sometimes they are technology content items. With the work being done, a language teaching game has been developed with the aim of helping children to remember their language skills for a longer time and helping to provide hand-eye coordination. In addition, the developed game was played to the children in the test group and measured as to whether or not it was achieved as desired.

The following sections of the study are in order; Previous studies using Kinect, developed teaching tool, the findings obtained from the test result, results and suggestions.

II. LITERATURE REVIEW

Kinect has found a lot of applications in the field of education. Si et al. (2017) in their work, they talked about a content-based English education system in which Kinect was also involved. Here the person's body movements and speech are captured by Kinect, and then the captured Kinect image is transferred from the background to the virtual media scenes. It is provided that the individual appears to be in that environment. They have created English speaking content related to found environments [1]. Satria et al. (2017) in their work, they tried to teach children English-language equivalents of various animal themes based on the speech recognition feature of Kinect for teaching English to children [2]. Pan, (2017) studied the difference between the use of standard input devices for English vocabulary learning and the interaction with Kinect [3]. In the works of Hsiao et al. (2016) and Kuo et al. (2014) they also developed applications to embody language learning with various body movements made across of Kinect [4, 5]. In Boutsika's (2014) work, he developed an educational game in which Kinect was used to improve the memory and socialization of autistic children [6]. Homer et al. (2014) in their work, they developed a Kinect-based application to increase children's vocabulary and reading comprehension [7]. Kourakli et al. (2017) in their studies, they talked about educational games conducted with Kinect in order to improve the cognitive, motor and academic skills of children with learning disabilities and investigated the positive effects of these games on children [8].

III. KINECT

In recent years, considerable progress has been made in computer-assisted instruction [9-13]. People have accelerated the acquisition of information by increasing interactivity with computers and have seen that game-based instructional paths with computer aided support are more permanent and motivating [14, 15]. In this respect, the Kinect device developed by Microsoft for the Xbox has been used for many educational purposes [15-18].

Kinect sensor detects three-dimensional body movements of the person standing in front of it, face and voice recognition owing to hardware items that contained within such as RGB camera and depth sensor, multi microphone, internal processing core [19]. In this way, human-computer interaction is achieved by voice or body movements without using conventional control devices such as mouse, keyboard.

A. Kinect Working Logic

Apart from its popularity in the entertainment sector (various computer games), the factor that allows Kinect to be used in many disciplines, from education to health, from military to robotics, is, easily financially affordable, and the sensor has the ability to capture data in a variety of ways. The

types of images that Kinect has caught are: Still images, sequence of image frames, video capture [20]. Figure 1 shows the Kinect device, which contains the infrared sensor, color (RGB) camera, depth sensor and microphones.

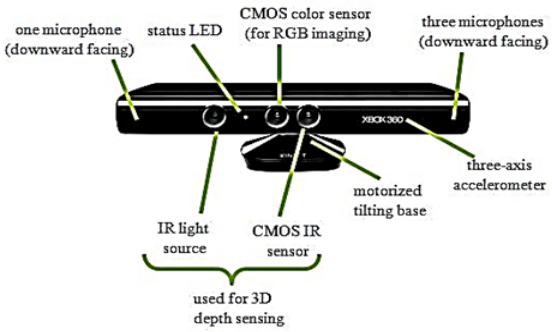


Figure 1: Microsoft Kinect sensor (MS Kinect)[21]

Kinect is simultaneously scanning body movements with the sensors it has and with an open source Software Development Kit (SDK) that allows the device to be used in Windows operating system [22]. To scan and display perceptions in 3D on the screen, Kinect continually sends infrared rays and calculates the reflection speed of the light to reach the depth image [23]. While performing skeletal scanning, Kinect can detect 25 joints on the body for a single person. It has the capacity to detect up to 150 joints at one time (up to 6 persons), due to the ability to perceive more than one person at the same time [22]. Only left / right hand detection is allowed in the work performed.

IV. Developed Framework

The English teaching game was developed using the Unity 3D game engine and Kinect. The aim of the play is to teach the children (4-6 years) the English language equivalents of the concepts of color (12 pieces) and geometric shape (16 pieces) in preschool period. The mentioned software is a game that supports language learning. After the color and shape concepts were given by the teacher, how students learn these concepts has been tested with developed software. The game consists of 4 stages and throughout these stages the child must stay in front of Kinect. The steps related to the functioning of the game are given in Figure-2.

Figure 2: Game stages

At the first stage, one of 12 colors is determined at random and the voice is played by playing the previously recorded audio files. After playing the voice of color, one of the color that is selected randomly between 12 color is assigned to any of the 3 squares that is placed on the screen and to the rest of the squares 2 colors are assigned except for the voiced color. Which square of the voiced color is assigned to is randomly determined. In other words, when a color is spoken, the correct color can be assigned to the first square, but when the next color is passed, the correct color can be assigned to the 3rd square. It is desired to avoid developing an idea of the child that the right color is present in the same square every time although child does not know the color of the played voice.

The child who hears played voice of color, chooses one of the square that thinks the right color by holding the left / right hand as a open position opposite of the Kinect device.

If the right square is selected a warning is displayed in the form of a tick indicating the right answer on the screen and a audible feedback is heard then the child can shift the left hand to the right to move the next color, assign a new color to the squares and execute the new color voice. Also child can shift the right hand to the left to hear previous color again.

In the case of a false answer, a warning is displayed in the form of a cross indicating the wrong answer on the screen, and the child is asked to try again. The screenshot of Step 1 of the game is given in Figure 3.

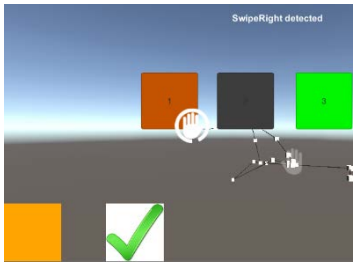


Figure 3: Finding the right color

The vocalizing of the colors and the color assignments to the related squares continued for 12 colors. After the completion of the first phase, the "next scene" statement is commanded by the child and a mini test phase is passed as second phase. In this test scene, there are three selectable objects consisting of a sphere, a cylinder and a cube, a color wheel and a bucket. When the scene is first opened, these 3 objects appear only in white color. When the child gives the command "turn", the color wheel on the screen is rotated. After the end of the rotation process, the stick of the wheel stands on a color and the sound file of that color is played. The color that is said is randomly assigned to one of the three objects standing on the scene, just like in the first stage, and the remaining 2 objects are assigned different colors.

The color wheel is covered and the stick does not seem to stop on which color. However, when the child picks the right color object and leaves the to the basket, the curtain on the wheel is lifted and a tick sign indicating that the correct selection is made is drawn on the top of the screen and an applause feedback is heard. Likewise, when an object in the wrong color is picked up and released, the curtain of the wheel is not opened and a cross is drawn at the top of the screen

indicating that the wrong choice is made and an error sound is heard.

When the child is standing in front of the Kinect, bring the hand in the open position onto the object then by doing the act of closing the hand he can remove that object from its place. There are virtual hand images on the screen that follow the child's real hand shapes. When the child's hand in the release position, the virtual hand is also released (open) or likewise when the child's hand in the grip position, the virtual hand is gripped (close). During the time when the hand is held closed, the object is dragged to the same point as the child's hand moves to the spot on the screen. In other words, it is considered to hold the object virtually. When he brings the object to the place where the bucket is, he opens the hand and allows the object to fall into the bucket.

The process that the actual hand and virtual hand go to the same point at the same time is performed by thanks to Kinect gets the coordinates of the real world and translate it into the Unity 3d scene coordinates. Figure 4 shows a screenshot of the game's second stage.

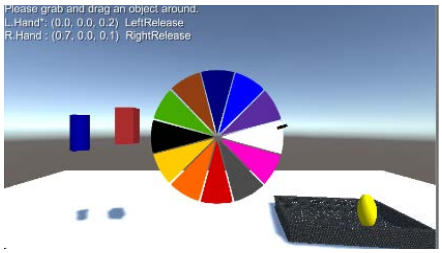


Figure 4: Releasing the objects to the bucket

The third stage of the game is passed by the "next scene" command given by the child. At the third stage, just like in the first stage, the figures at this time are randomly assigned on the squares. In this scene of the game, every shape that is desired to be found is voiced once. For example, when "triangle" is played and the child makes the corresponding selection and then the other shape is played, "triangle" is not played once again, and random selection is made among the remaining shapes. Therefore, a previously voiced shape can be assigned again on the squares at the next steps, but that voice of that shape is performed only once. The same programming logic applied to the assignment of colors at the first stage. That is, every color is voiced only once.

The child who hears played voice of shape, chooses one of the square that thinks the right shape by holding the left / right hand as a open position opposite of the Kinect device.

If the right square is selected a warning is displayed in the form of a tick indicating the right answer on the screen and a audible feedback is heard then the child can shift the left hand to the right to move the next shape, assign a new shape to the squares and execute the new shape voice. Also child can shift the right hand to the left to hear previous shape again.

In the case of a false answer, a warning is displayed in the form of a cross indicating the wrong answer on the screen, and the child is asked to try again. The screenshot of the 3rd stage of the game is given in Figure 5.

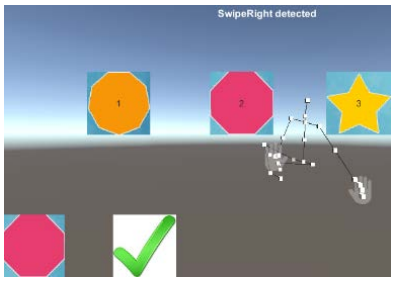


Figure 5: Finding the voiced shape

The vocalization of the shapes and the assignment of the shapes to the respective squares continued throughout the 16 steps. After the end of the 3rd stage, the "next scene" statement is pronounced by the child and the 4th and final stage of the game is switched on.

At this stage, a tangram game was prepared for testing purposes. After selecting the desired shape, by making gripping gesture, of the object from some geometric shapes on the scene, the holded shape can be released to the area on tangram puzzle that having same apperance as the holded shape.

If the figure is tried to be left in a wrong area that does not have the same appearance as itself, even though the hand releasing movement is made holded shape is not transferred to the area and goes to the its first position. After all shapes have been left to their respective fields, a gift box image appears at the place where the tangram puzzle is found and a feedback voice is also returned to the player about winning the game. A screenshot of the last stage of the game is given in Figure 6.

Figure 6: Tangram puzzle

V. RESULTS AND DISCUSSION

The target volume of the developed Kinect-supported English teaching game is pre-school children groups of 4-6 years who are not yet literate. Due to children in this period have not yet learned to read and write, while teaching them a new language in order to further increase their permanence it needs to support this process with a game especially computer aided because of its attractive effect on children of today.

With technology advancing rapidly, children tend to use computer systems even more. Since it is observed that there is a general difficulty in using traditional computer peripherals (gripping the mouse in a right way) such as computer mouse

and keyboard among children, instead of such input devices, the game has been realized with the Kinect device, which is counted as one of today's rising control devices and allows more natural interaction.

Another benefit provided by the developed educational game is the hand-eye coordination in children. Because opening the hand against the Kinect device, choosing an object on the screen, it also supports the development of the child's ability to carry the hand to the point of focus.

Before evaluating the effect of the developed language teaching game, the English equivalents of color and shape concepts were presented verbally and using visual cards for a group of 60 students. Then the students were divided into two groups, the test group and the control group. Only the test group was allowed to play the developed game twice a week for 1 month. At this time, the concept of color and shape is presented only verbally to the control group, again for a month, twice a week.

After completing the necessary training, to assess the effectiveness of the game, because of the target group has not yet learned to read and write, in order to understand how the learners can grasp the information they receive a verbal test method has been applied. First of all, the control group was asked about 5 color questionnaire asking what color the card was given and 5 figure questioning asking what form the wooden block was shown in (asking for English correspondence). For color and shape concepts, the number of correct and incorrect answers given by each child is kept on a table. Later, the same procedures (directing questions and recording on the table) were repeated for the test group. After both groups completed the test, the correct answer rates for both the control group and the test group were calculated as percentages. As shown in Figure 7, the correct response rate of the students in the test group is much higher than that of the control group. In order to measure the correct answer recall times, it was determined by time keeping that how many seconds each student responded. This was done by asking each student questions and recording their response times. Relevant data for correct response times are given in Figure 8. The results in Figure 8 were also in line with Figure 7 and it was seen that the test group could remember the correct answer in a much shorter time. Developed in this regard, the Kinect-supported English teaching game has become an effective learning tool for groups of students aged 4-6 years.

Kontrol Grubu	Öğrenci	Toplam Doğru sayısı	Toplam Yanlış Sayısı	Doğru Cevaplama Oranı (%)
	1	3	7	30
	2	4	6	40
	3	2	8	20
	4	4	6	40
	5	3	7	30

Test Grubu	Öğrenci	Toplam Doğru sayısı	Toplam Yanlış Sayısı	Doğru Cevaplama Oranı (%)
	1	8	2	80
	2	7	3	70
	3	9	1	90
	4	10	0	100
	5	10	0	100

Figure 7: Corect response rate

Kontrol Grubu	Öğrenci	Doğru Cevaba Ulaşma Süresi (sn)	Test Grubu	Öğrenci	Doğru Cevaba Ulaşma Süresi (sn)
	1	7		1	1
	2	8		2	2
	3	6		3	2
	4	7		4	1
	5	8		5	1

Figure 8: Time to access correct response

Apart from students acquisitions, the teacher also referred to the results. It has been reported that the developed training software provides good support for enhancing the academic achievement of the students and at the same time is effective in improving motor skills.

VI. CONCLUSIONS

In this study, a Kinect-based educational software was developed in the Unity 3d game engine to teach English equivalents of color and shape concepts to the children of preschool period. The characteristic of the work is that besides measuring the level of color and shape learning, it also includes speech recognition, which allows the child to pass on to the next scenes, and motion recognition, which allows the transition to other colors and shapes. The targetted goals at the beginning of the work such as increasing the language skill, maintaining consistency, improving hand-eye coordination and taking steps to acquire a new language without even realizing that a language has been learned, have been achieved.

In subsequent versions of the work, the 4th stage tangram puzzle game can be supperted more geometric shape or can be arranged in such a way that children can create animal figures that attract their interest. Also, after the color is pronounced by first computer then by child, resemblance rate can be measured by comparing both pronunciations.

REFERENCES

[1] B. Si, B. Bai and L. Hao, 2017, "Construction of Specialized English Teaching Context under the Content-Based Instruction Concept Based on Virtual Reality", 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI 2017).

[2] F. Satria, H. Aditra, M. D. A. Wibowo, H. Luthfiansyah, M. Suryani, E. Paulus and I. Suryana, "EFL Learning Media for Early Childhood Through Speech Recognition Application", 2017 3rd International Conference on Science in Information Technology (ICSITech), 568-572.

[3] W. F. Pan, "The Effects of Using the Kinect Motion-sensing Interactive System to Enhance English Learning for Elementary Students", Educational Technology & Society, 20 (2), 188–200, 2017.

[4] H. S. Hsiao, J. C. Chen, "Using a gesture interactive game-based learning approach to improve preschool children's learning performance and motor skills", Computers & Education 95 (2016) 151-162.

[5] F. R. Kuo, C. C. Hsu, W. C. Fang and N. S. Chen, "The effects of Embodiment-based TPR approach on student English vocabulary learning achievement, retention and acceptance", Journal of King Saud University – Computer and Information Sciences (2014) 26, 63–70.

[6] E. Boutsika, "Kinect in Education: A Proposal for Children with Autism", 5th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, DSAI 2013, 123-129, 2014.

[7] B. D. Homer, C. K. Kinzer, J. L. Plass, S. M. Letourneau, D. Hoffman, M. Bromley, E. O. Hayward, S. Turkay and Y. Kornak, "Moved to learn: The effects of interactivity in a Kinect-based literacy game for beginning readers", Computers & Education Volume 74, pp 37-49, 2014.

[8] M. Kourakli, I. Altanis, S. Retalis, M. Boloudakis, D. Zbainos and K. Antonopoulou, "Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using Kinect learning games", International Journal of Child-Computer Interaction 11 (2017) 28–39.

[9] D. Schmalstieg, A. Fuhrmann, G. Hesina, Z. Szalavari, M. Encarnação, M. Gervautz, and W. Purgathofer, "The Studierstube augmented reality Project", PRESENCE Teleoperators and virtual environments 11(1), MIT Press, Cambridge, MA, pp. 33-54, 2002.

[10 G. Younes, R. Kahil, M. Jallad, D. Asmar, I. Elhajj, G. Turkiyyah, and H. Al-Harithy, "Virtual and augmented reality for rich interaction with cultural heritage sites: A case study from the Roman Theater at Byblos", Digital Applications in Archaeology and Cultural Heritage, 2017, In Press.

[11] D. Sampaio, P. Almeida, "Pedagogical strategies for the integration of Augmented Reality in ICT teaching and learning processes", Procedia Computer Science 100, pp. 894 -899, 2016.

[12] K. J. Schönborn, G. E. Höst, and K. E. Lundin Palmerius, "Nano education with interactive visualization", Nano Today 11(5), pp. 543-546, 2016.

[13] J. L. Chiu, C. J. DeJaegher, and J. Chao, "The effects of augmented virtual science laboratories on middle school students' understanding of gas properties", Computers & Education 85, pp. 59-73, 2015.

[14] K. K. F. Cheung, M. S. Y. Jong, F. L. Lee, J. H. M. Lee, E. T. H. Luk, J. Shang and M. K. H. Wong, "FARMTASIA: an online game-based learning environment based on the VISOLE pedagogy", Virtual Reality, Vol. 12, No. 1, pp. 17-25, 2008.

[15] Y. H. Chang, J. H. Hwang and R. J. Fang, "A Joyful Kinect-based Learning System", Proceedings of the 2017 IEEE International Conference on Applied System Innovation IEEE-ICASI 2017 - Meen, Prior & Lam (Eds), 1028-1031.

[16] Y. H. Chang, J. H. Hwang, R. J. Fang and Y. T. Lu, "A Kinect- and Game-Based Interactive Learning System", EURASIA Journal of Mathematics Science and Technology Education, 4897-4914, 2017.

[17] M. T. Yang, W. C. Liao, "Computer-Assisted Culture Learning in an Online Augmented Reality Environment Based on Free-Hand Gesture Interaction", Ieee Transactions On Learning Technologies, Vol. 7, No. 2, 107-117, 2014.

[18] R. Fernández, C. von Lücken, "Using the Kinect Sensor with Open Source Tools for the Development of Educational Games for Kids in Pre-school Age", 2015 XLI Latin American Computing Conference (CLEI).

[19] https://en.wikipedia.org/wiki/Kinect (access 26.03.2018)

[20] M. L. Gavrilova, Y. Wang, F. Ahmed and P. P. Paul, "Kinect Sensor Gesture and Activity Recognition", IEEE Consumer Electronics Magazine, 7(1), 2017.

[21] http://fivedots.coe.psu.ac.th/~ad/jg/nui13/kinect.jpg (access 27.03.2018)

[22] J. C. Neto, P. R. Filho, G. P. Silva, N. B. Olegario, J. B. Duarte and V. H. Albuquerque, "Dynamic evaluation and treatment of the movement amplitude using Kinect sensor", IEEE Access, Volume: PP, Issue: 99, 2018.

[23] B. Çubukçu, U. Yüzgeç, "A Physiotherapy Application with MS Kinect for Patients with Shoulder Joint, Muscle and Tendon Damage", 2017 9th International Conference on Computational Intelligence and Communication Networks, 225-228.

WALRUS: A Retro Communication Gadget Based on Internet of Things Technologies

Övünç ÖZTÜRK, Yunus Emre KÜÇÜK, Ahmet YALNIZ

Manisa Celal Bayar University, Manisa/Turkey, ovunc.ozturk@cbu.edu.tr
Manisa Celal Bayar University, Manisa/Turkey, yunusemrekucuk@ogr.cbu.edu.tr
Manisa Celal Bayar University, Manisa/Turkey, ahmetyalniz@ogr.cbu.edu.tr

***Abstract* - In this work, a new system depending on a device that can encode and decode push-button signals, modulated using Morse code conventions, were developed to build a low-cost communication medium based on Internet of Things (IoT). The proposed system consists of two parts: a base station and handheld terminals. The base station is a single board computer with a web application based on Node.js. Handheld terminals are small battery powered devices, developed using MCU's, that can communicate with the base station over the wireless network. They can encode and decode Morse code, and convert to text or speech depending on the configuration of the terminal, which can be extended by using different add-ons, such as an OLED screen or a text to speech module. Communication between terminals is orchestrated by the base station using IoT Technologies like MQTT. The handheld terminals can be used by disabled people as a mean for private conversation, or a gadget for entertainment purposes. The system is an uncomplicated and low-cost communication medium and implemented to find alternative use cases for the IoT technologies.**

***Keywords* – MQTT, Morse Code, Communication, Internet of Things**

I. INTRODUCTION

With the advancement in the era of the internet of things, every item can be connected over wireless technologies. Nowadays, IoT technologies are used in a wide range of application fields including home safety and remote monitoring, agriculture and animal husbandry (measurement of the temperature of the environment in animal farms, detection of unusual conditions), smart city applications (waste bins call for removal, parking spaces report free capacities) and communication (low-cost communication devices).

Internet of Things has three layers: perception layer, transport layer and application layer [1]. The perception layer is responsible for gathering data using wireless sensors. The transport layer integrates the collected data for the usage of upper layer. Finally, application layer offers services to the application or to the end-user for intelligent processing.

This paper applies IoT technologies to implement an efficient and low-cost communication medium. Current remote communication resources (telephone, computer, internet, etc.) are impractical for individuals with visual and hearing disabilities. It is very difficult for a sighted person to communicate with a hearing-impaired person. The purpose of our project is to solve the problem of communication between these individuals easier by using the IoT technologies.

We have developed a device capable of communicating via the Morse code. This is an effective communication method for solving the problem of communication between individuals with visual and hearing disabilities. Morse code [2, 3, 4] is a method of transmitting text information as a series of on-off tones, lights, or clicks that can be directly understood by a skilled listener or observer without special equipment.

The proposed system consists of two parts: a base station and handheld terminals. The base station is a single board computer with a web application based on Node.js. Handheld terminals are small battery powered devices, developed using MCU's, that can communicate with the base station over the wireless network. We used the MQTT protocol for data exchange on this system.

There are similar works in literature. For example, [5] implements a Morse code-based electronic lock using IoT technologies. The novel electronic lock can encode or decode optical signals, modulated using Morse code conventions. This work uses IoT technologies for building a smart home security system. Besides the convenience and cost reduction, the system provides a user-friendly interface for users of all ages.

[6] proposes a mobile solution for linguistic communication with deaf-blind people using Arduino and Android. They use an Arduino microcontroller connected to a vibrating micro-engine, command buttons and a Bluetooth module. The messages are sent/received via Bluetooth. At the end of the translation the user can read the message using the vibration of the Morse code. [6] allows any individual who has a mobile phone to communicate with a deaf-blind people.

On the contrary, we aim to make communication possible between two individuals with visual or hearing disabilities over an efficient, low cost and user-friendly medium. This system does not require users to have mobile phones or any other devices.

Next section describes the Morse code in detail. Section three introduces the WALRUS system and its functionality in detail. Section four presents the components of the system. Finally, section five concludes the paper with a brief talk about possible future work.

II. MORSE CODE

Morse code is a method for transmitting letters, numbers, and characters using standardized sequences of short and long signals called "dots" and "dashes", or "dits" and "dahs". Figure 1 shows the characters encoded by the International Morse Code.

Morse code was created by Samuel Morse in 1835 and started to be used in 1837. Morse code can be transmitted in various forms, such as by sounding, switching on and off the radio signals, the passing electric current through the wires, etc.

We used Morse code as a communication medium in our project Morse alphabet structure allows users to send a message with a single keystroke. All characters can be sent using a single key instead of using a keyboard. This is a more convenient way of communication for disabled users.

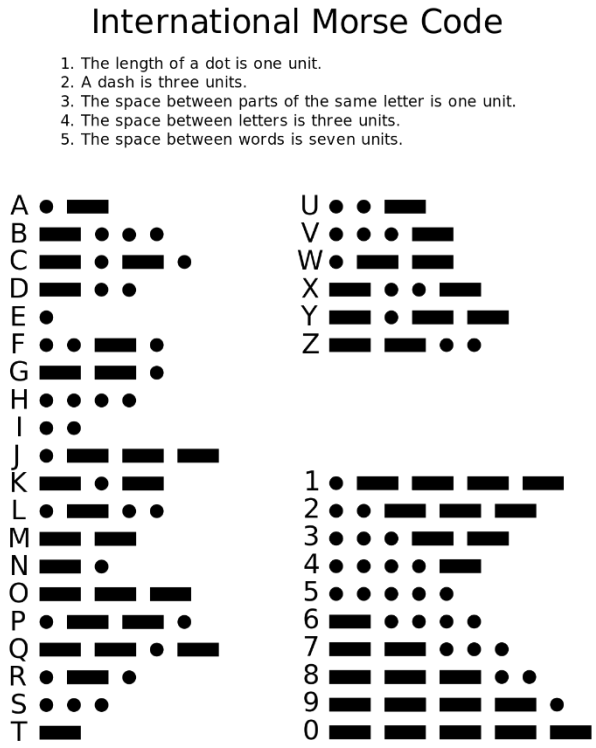


Figure 1: Characters encoded by the International Morse Code.

III. WALRUS SYSTEM

The proposed system, namely Walrus consists of two parts: a base station and handheld terminals (Figure 2).

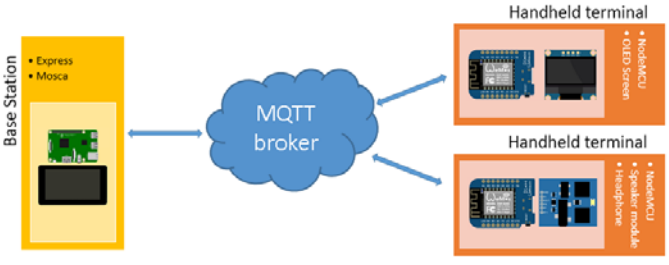


Figure 2: Architecture of the WALRUS system.

We used Raspberry Pi as a gateway in the implementation of the base station. Users connect Raspberry Pi to send their messages to recipients. Raspberry Pi is also used for managing and storing messages. We have two types of handheld terminals:

(a) Handheld terminal having a text to speech component for individuals with visual disability (Figure 3)

(b) Handheld terminal having an OLED screen component for individuals with hearing disability (Figure 4).

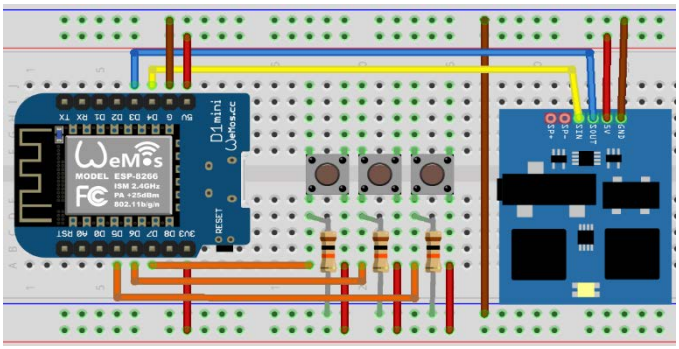


Figure 3: Handheld terminal having a text to speech component.

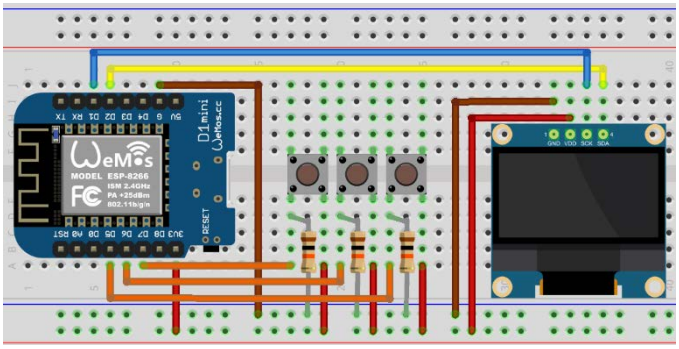


Figure 4: Handheld terminal having an OLED screen component.

The lifecycle of the system is based on four states:
State 1: Scanning and listing wireless networks
State 2: Connecting to a wireless network
State 3: Sending/Receiving messages
State 4: Defining the recipient

We define the encoding in Figure 5 for transitions between states and the encoding in Table 1 for functions in the states.

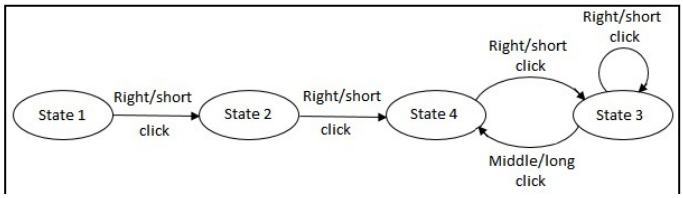


Figure 5: Encoding for the transitions between states.

According the Table 1, the functions of the buttons on the device depends on the state. In State 1 the left button allows scrolling down to switch between wireless network SSIDs. The middle button allows re-scanning of wireless networks. The right button is used to select the wireless network. Then the state is switched to State 2.

Table 1. Encoding for the functions in the states.

Short click				
	State 1	**State 2**	**State 3**	**State 4**
Left	Scroll Down	Morse Dot Input	Morse Dot Input	Scroll Down
Middle	Re-scan	Clear Password	Clear Screen	---
Right	Select	Send Encrypted Password	Send	Select Recipient
Long click				
	State 1	**State 2**	**State 3**	**State 4**
Left	Scroll Down	Morse Dash Input	Morse Dash Input	Scroll Down
Middle	Re-scan	Clear Password	State 4	---
Right	Select	Send Encrypted Password	Send	Select Recipient

In State 2, the left button has two different functions. In this state left button is used to enter the password. If you short click the left button, a short Morse signal is transmitted. If you long click the button, then a long Morse signal is transmitted. The middle button allows you to clear the wireless network password. Finally, the right button allows you to send the encrypted password. Then the state is switched to State 3 and State 4. The functions of these states are executed in parallel.

In State 3, you can send/receive messages. The system supports sending and receiving messages simultaneously. For selecting recipient user long clicks the middle button and switches to State 4. In State 4 the left button allows scrolling down to switch between recipients. The middle button has no function in State 4. Right button allows selecting the recipient and switching to State 3.

In State 3, left button is used to write a message. If you short click the left button, a short Morse signal is transmitted. If you long click the button, then a long Morse signal is transmitted. The right button is used to send the message. When you short click the middle button, then the screen is cleaned.

Figure 6 represents the flowchart of the WALRUS system. When the device is turned on by pressing the power button, The ESP module scans nearby wireless networks (State 1).

If a text-to-speech module is installed on the device, the SSIDs of wireless networks are read with the help of this module. If an OLED screen is installed on the device, the SSIDs of the wireless networks are written on the screen in Latin alphabet. User selects the network and switches to State 2. In State 2, user enters the password of the wireless network in Morse code and sends the password. Then the password is checked. If the password is not correct, the user is redirected to State 1. Otherwise, the state is switched to State 3 and State 4.

In State 4, a recipient is selected. If a text-to-speech module is installed on the device, recipients are read with the help of this module. If an OLED screen is installed on the device, the recipients are written on the screen. User selects the recipient and switches to State 3.

In State 3, the messages are send and received simultaneously. The message is entered using Morse code encoding. If the user wants to change the recipient, then he switches the State 4. The received messages are read or written depending on the module installed on the device.

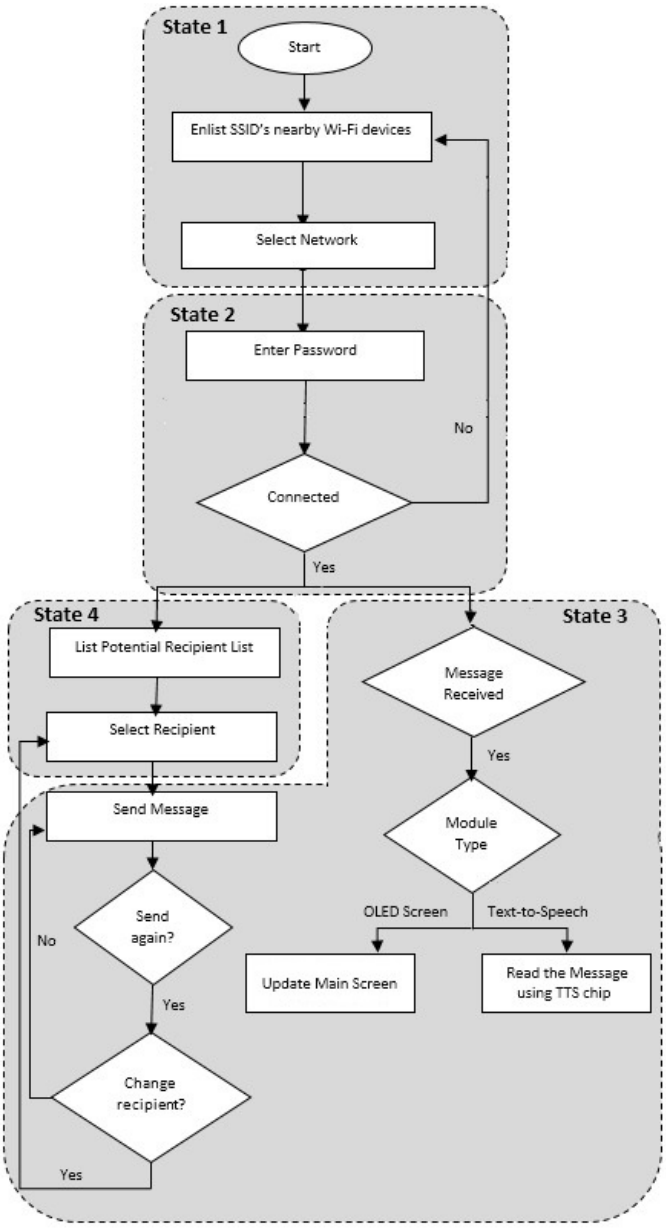


Figure 6: Flowchart of the WALRUS system.

IV. COMPONENTS OF THE SYSTEM

A. WeMos D1 Mini Pro

Wemos D1 Mini Pro (Figure 7) is a development card with an integrated ESP8266 module on it. It is designed for use in IoT projects. It provides eleven 3.3V I/O pins. A 5V output is also available.

The low cost WeMos D1 Mini Pro has a small footprint and supports the I2C protocol [7]. It has an external antenna connector on it, so an external antenna can be connected to increase the power of the device. For these reasons, it is frequently preferred in IoT projects.

The main reason for using the WeMos D1 Mini Pro in this project is that the text-to-speech module needs 5 volts to operate

and Wemos D1 Mini Pro meets this requirement without an external power supply.

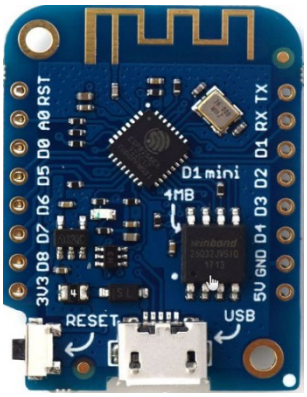


Figure 7. WeMos D1 Mini Pro module.

B. Emic-2 Text-to-speech Module

The Emic-2 text-to-speech module (Figure 8), which performs text-to-speech conversion, is used in the project with the aim of visually impaired users to listen to and control the messages they send.

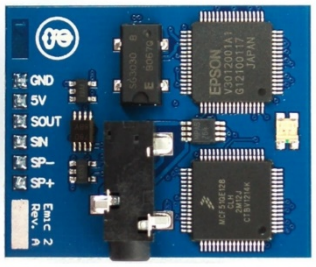


Figure 8. Emic-2 text-to-speech module.

C. I2C OLED Screen

OLED Screen module (Figure 9) is used in the project with the aim of hearing impaired users to write messages and read the incoming messages. The screen module supports I2C protocol therefore it can work with only two connections. In this case, the required I/O pin count is reduced.

Figure 9: OLED Screen.

D. MQTT

MQTT (Message Queuing Telemetry Transport) [8, 9, 10] is an ISO standard publish-subscribe-based messaging protocol. MQTT was invented in 1999 by Arlen Nipper. It is a low resource demanding protocol and it supports asynchronous operation and SSL/TLS security protocols.

MQTT is based on publish/subscribe method. The data is sent to the broker by the publisher, and the broker transmits this data to the subscriber. The MQTT broker uses topics to transmit the message to the relevant subscribers. The topic is a virtual channel that allows a publisher to connect to the relevant subscriber. Both the subject and the message are identified on the publisher side. This message is sent to the topic broker, and the broker redirects the message to the subscriber who subscribed to the topic.

In this project, handheld terminals are both MQTT publishers and subscribers. Each user has a unique topic. Thus, messages are sent only to relevant subscribers.

V. CONCLUSION

Using the internet of things technologies will change the communication between machines and even humans. We developed a new communication gadget by using open source software and hardware, using MQTT protocol for data exchange between the clients. The device is a user-friendly communication gadget, which can have exciting application areas such as lowering the communication barrier between disabled people. It is a modular device, which can be expanded with different output modules. An OLED Screen and a Text to Speech output modules are implemented during the project, but new output modules like Braille Displays will be implemented as a future work.

REFERENCES

[1] L. Atzori, A. Iera, G. Morabito, "The Internet of Things: A Survey", Computer Networks, vol. 54, 2010, pp. 2787-2805.

[2] "Learning Morse Code". Arrl.org. Archived from the original on 20 September 2017. Retrieved 1 December 2017.

[3] K. Mukherjee, D. Chatterjee, "Augmentative and Alternative Communication Device based on Eye-blink Detection and Conversion to Morse-code to Aid Paralyzed Individuals". In Proceedings of ICCICT (2015), Mumbai, India, 15-17 January 2015, pp. 1-5.

[4] L. Peter Carron, "Morse Code: The Essential Language", Radio amateur's library, issue 69, American Radio Relay League, 1986 ISBN 0-87259-035-6.

[5] L. Chin-Tan, S. Tung-Chun, L. Win-Der, "A Novel Optical Morse Code-Based Electronic Lock Using the Ambient Light Sensor and Fuzzy Controller", Applied Sciences 7(2):140, 2017.

[6] A. L. Nogueira Vieira, Fe. Felix Novaes, D. Machado Silva, L. Santos, S. Belozi, T. Castro, "A Mobile Solution for Linguistic Communication with Deaf-Blind People using Arduino and Android", In Proceedings of the Eleventh International Network Conference, Frankfurt, Germany, in July 2016.

[7] D. Paret, C. John, "The I2C Bus: From Theory to Practice", John Wiley & Sons, Inc., 1997.

[8] "ISO/IEC 20922:2016 Information technology -- Message Queuing Telemetry Transport (MQTT) v3.1.1". iso.org. International Organization for Standardization. June 15, 2016.

[9] "MQTT 3.1.1 specification". OASIS. December 10, 2015. Retrieved April 25, 2017.

[10] R. A. Light, "Mosquitto: server and client implementation of the MQTT protocol."

Performance Tuning in Database Systems In High Availability Architecture and Reducing Query Costs Strategies in Oracle Database Management System

N.YUMUŞAK[1] and A.YORULMAZ[2]

[1] Sakarya University, Sakarya/Turkey, nyumusak@sakarya.edu.tr
[2]Sakarya University, Sakarya/Turkey, jayorulmaz@gmail.com

Abstract **- In industry, the priority of services provided by the companies are 7X24 (7 days 24 hours) uninterrupted services. Hence, the main reason why database should provide continuous uninterrupted services. The attributes of database services that provide high accessibility services should be reliability, availability and maintainability; features referred to as RAM (Reliability, Availability, Maintainability) in the literature. In addition to that in this paper, mathematical computation pertaining to the location and accessibility in ITIL processes will also be examined.**

It is also aimed to investigate what should be done at the architectural level to ensure that the commands coming from Structured Query Language (SQL) results in high performance. When database or system administrators wrongly configures the architectures processing the queries, it may result in inadequate power systems that could have operated with high performance. This article then discusses what might be the necessary adjustments in order to use the resources of query systems more efficiently.

Keywords – RAM(reliability availability maintainability), high availability, SQL, performance tuning, ITIL

I. INTRODUCTION

In an ever growing and an ever-changing industry, the protection of lives, quality preservation, energy savings, security and satisfaction of computer users among others are the foremost focus of global competitive industries.

Database systems continues to be and will always be the bedrock of all types of software industry. This goes to show that database system is indeed one of the fundamental components of computer systems.

In this article, how database systems overlap with the concept of reliability, availability and maintainability, also known as the RAM (Reliability, Availability, Maintainability) analysis as in the literature, as well as the philosophy of 7X24 survivability of DBMS that hold data critical computing without tolerance in complex systems will be examined.

The increase in performance of a database results in a noticeable increase in performance of all systems connected to it. However this is not suggestive of the fact that the sole responsibility of performance lies on the database system. In fact the benefits obtained from a software is as a result of the collective contribution of all other working components that make up the software system. What needs to be done to ensure increase in performance should be considered as part of the engineering process (discipline) as a whole.

II. HIGH AVAILABILITY ARCHITECTURE

The word 'Trust' when viewed in the lens of philosophy, we can observe that several research and discussions have been made on it. Prof. Dr. Ahmet Inan summarizes it as follows: " Trust, in relation to conscious beings like us, is reflecting off a "feeling" that is possessed by animals with a complex nervous system. Trust(Reliability) also has a cosmic definition. A creation, placing itself as central point of focus, has its existence based on trust. This we can say is an ontic property of trust(reliability)."[1]

In the engineering discipline, the definition of Trust (reliability) is different from that of social sciences. This is because engineering works relies on calculations. These calculations should be precise, so much so, that we can refer to the calculation as reliable. Reliability is a concept of comparison. In terms of logic, reliability or unreliability carries a very little meaning. The significant question to ask is "to what extent is it reliable?".

Von Braoun and his team while working on missile during World War II said, "A chain cannot be stronger than its weakest ring". He stated that one of all components of the fuse could prevent the success of others.

AGREE (Advisory Group on Reliability of Electronic Equipment) that introduced a new insight into the technology world, claiming that an electronic product that costs $1.0 to produce can be made more reliable if produced at the cost of $2.0. This notion even though expensive results in designing of a less maintenance-requiring product.

There is a subtle distinction between continuity and reliability. This nuance, in terms of continuity is a rate, and a probability in terms of reliability.

Reliability is the likelihood that any component, device or service will not be interrupted in unit time.

The reliability of a system is the probability that it will operate without a fault in a given term and under given

conditions. [2] Reliability, therefore, concerns the likelihood of successful performance for any system.

Regarding database reliability, we can say that in the future it is possible to perform its function correctly in a specified period.

This definition contains four main elements:

• Possibility,
• For what business environment is it valid,
• In the future definite continuous,
• Compliance

The first is that the reliability indices can only be evaluated using probability methods because the future behavior is not apparent. The other three factors in the definition depend on the responsibilities of the system. For example, the operating conditions may be normal or extraordinarily severe conditions. Adequate Performance is an engineering problem. For example, failure of an element may completely disable the system, may have no effect on the behavior of the system or may lead to a decrease in system capacity.

The past 50 years has witnessed an increased momentum in ensuring the credibility of using terminology and thus the result of their reliability analysis is known to be a very necessary factor. The need for safe products is stressed by the constant movement of such technology across systems in various industries such as the military, automobile, aviation as well as the communication and rockets systems [3]. When the multiplicity of such systems is taken into consideration, it becomes obvious that databases play active roles at some point. For highly sensitive structures such as embedded systems, it is unthinkable to leave your credibility to chance.

For continually working (repairable) units, much more than the definition of reliability mentioned above is required. It is desirable that the subject unit is able to work continuously under certain circumstances as well as under specific future circumstances.

A system architecture that provides continuity in reliability will ensure that expected benefits are provided. In this context, the substructure of continuity is reliability with substantial probability.

It provides continuity of results and data for the future.

III. HIGH AVAILABILITY TERMINOLOGY

The terminologies used by many sources in the literature are as follows:

- Meantime to failure - MTTF
- Average time to repair - MTTR
- Meantime between failures - MTBF
- Number of failures at a given time
- The price of the fault

Mean time between failure (MTBF) is the fundamental parameter when system reliability is measured. The measurement bases for this parameter is time.

Ideally the primary target is that MTTF approaches infinity or MTTR approaches zero. This ideal situation means a system that works with infinite zero-repair time.

$$\text{Reliability} = e^{-\left(\frac{Time}{MTBF}\right)}$$

As the MTBF value increases, we can say that the reliability of the system is increasing. These criteria generally are defined parameters. A designer must consider these parameters during the design process. Unpredictable post-design changes, it means higher costs. The way in which these metrics are brought to the appropriate level are given as follows [4]:

- Increase overall component reliability of the product system
- Backup of product components
- Reduction in preventive maintenance period (proactive solutions)
- Increasing and compacting measurement and surveillance possibilities

Table 1:A particular percentage of availability

N.	Perc. Av.	Interrupt Per Year
1	%90.0	36 days 12 hours
2	%99.0	87 hours 36 minutes
3	%99.9	8 hours 46 minutes
4	%99.99	52 minutes 33 seconds
5	%99.999	5 minutes 15 seconds
6	%99.9999	**31.5 seconds**

At the end of this demonstration, about 9 (nine) notations like that of the presentations in Table 1 were obtained. As can be seen, an increase in the 9 (nine) decimal places leads to a significant annual decline. It is obvious that adding nine decimals requires a high HA cost.

The system in Table 1 has a lower level of criticality than that of the system in Level 1, and the level of vulnerability compared with intermittency is low. Test systems are examples of this system. Level 4 on the other hand indicates that a robust system accounts for the less than 1-hour loss. Data Centers are examples of this system. Those with three decimal places and above, are specifically for delicate systems such as banking and medical applications. Level 6 indicates that a system with a high criticality should be designed, which has less than 1minute interruption per year.

We can illustrate this in military applications of defense systems.

The responsibility of providing these values relies on each of the components that make up the entire systems. The level of communication of these components with each other represents the level of responsibility that lies on services. The focus of our article is the responsibility of the continuity of the database server.

Can we measure the high availability of a system? This is a very sophisticated question in practice. This is because,

industrial applications are very complex, dynamic and unstable. [5]

Some standards are needed to measure the high availability of a system. ITIL is one of these standards. ITIL is an abbreviation of Information Technology Infrastructure Library and was named in accordance with infrastructural libraries of information technology. ITIL is a service management methodology developed to manage IT services to the fullest and best quality. [6]

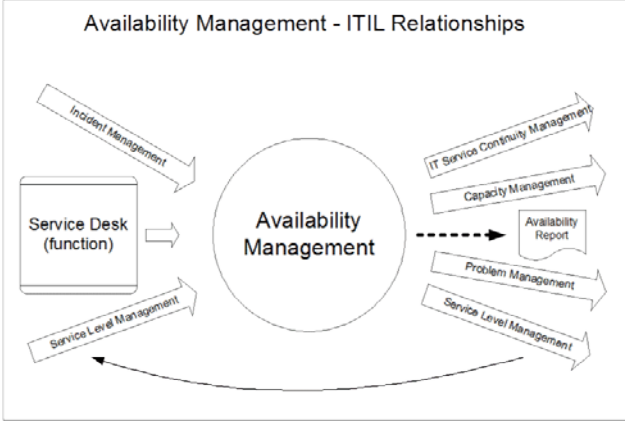


Figure 1:ITIL Availability Management

Part of this management methodology is access management. (Figure-1) This module gives us the continuity of the entire IT system. One of the most critical components in the module is the database. The continuity of the database services is the fundamental component of this module.

Event management (Incident Management) and Help Desk (Service Desk) within the accessibility methodology should proactively provide solutions.[7] The detection time is the time between the moment when a fault occurs in the systemand the time when the failure is detected. The Repair and recovery process results in restoration If these stages are reached(SLA) then we are convinced that system can provide required services.

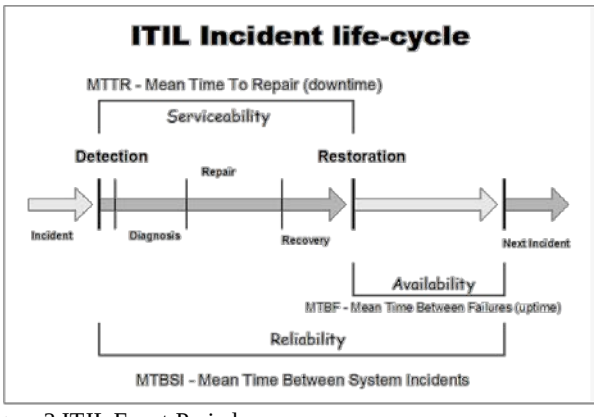


Figure 2:ITIL Event Period

Continuity indicates that the system provided is reliable within a loop. (Figure-2)

IV. CALCULATING HIGH AVAILABILITY IN DATABASE SERVICES

The methods of measuring high availability are concepts that need to be managed by both variety and complexity. For example; we can specify the availability of the database server with an algorithmic method [8]:

```
ALTER FUNCTION ufrDatabaseAvailability
  (
    @MeanTimeBetweenFailures real,  --in hours
    @MeanTimeToRealise real,                --in hours
    @MeanTimeToRepair real                  --in hours
  )
RETURNS REAL
AS

BEGIN
  RETURN

( @MeanTimeBetweenFailures - ( @MeanTimeToRealise
                + @MeanTimeToRepair )

) * 100.000/ @MeanTimeBetweenFailures * 1.0000
END


SELECT  dbo.ufrDatabaseAvailability(2 * 30 * 24, 4, 8)
--99.16666
```

Availability is expressed in the mathematical language [9]

A=1-F (2)

$$F = 1 - f(1-a)^{s+1} \quad (3)$$

In the formula;

A: System Possibility of Standing (Availability)

F: System Failure Probability

a: Possibility of a node of the systemto be standing

s: Number of systemcomponents, separated

f: Number of standby systemcomponents

n: Number of nodules in the system

If our system designs active/active as two nodules, we formulate the formula as follows.

$$A = 1 - F = 1 - (1-a)^2 \quad (4)$$

If the number of nodules increases in an unresolved system, the effect on continuity can be negative. If the accessibility of the nodule based on formula 4 reaches .99, the accessibility of the total system is specified as 4 (four) 9 (nine). This can also be seen from Table 1 that it sums up to 52 minutes, 33 seconds. When adding a nodule without separating the system, this what formula becomes:

$$f = \frac{n(n-1)}{2} \quad (5)$$

$$F = f(1-a)^2 = \frac{n(n-1)}{2}(1-a)^2 \quad (6)$$

If the accessibility of the nodules is .99 then n = 2, then in this case:

$$A = 1 - F \qquad (7)$$

The accessibility of the new system, which has not yet been separated but has increased in the number of nodules, corresponds to a 3 (three) to 9 (nine) notation, and again from Table 1, corresponds to an 8-hour and 46 minutes annual reduction. It is thus, necessary to pay attention to the ratio of systems separated by the number of nodules.

We can increase the continuity by making backup systems in databases. There are three main types of backup mechanisms in two known database management systems (Oracle and MSSQL) in today's industries. Cold, Warm and Hot backup. Cold Backup is a non-standing structure that is backed up asynchronously, separated from the primary product database. Its configuration is identical to the primary product database. Warm Backup is a stand-alone structure that is backed up asynchronously, separated from the product database. The settings are the same as the product database. Hot Backup is synchronous a backup and standing system with the product database.

The shape of these three solutions shown below:(Figure 3)

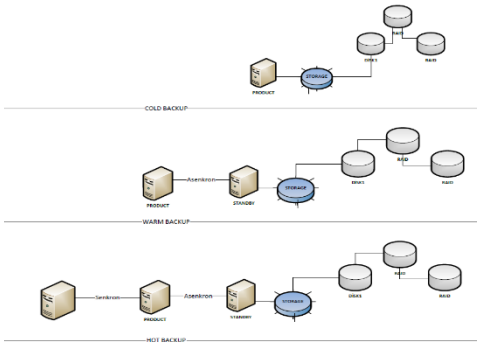


Figure 3:High Availability Database Architecture within Backup

If we take the number of system components (s = 1) separated by the persistence of the database management established with the Cold Backup system and we assume the own accessibility ratios of the components to 0.99:

We get $A=1-(1-0.99)^2$ => 1-0,0001 => 0,9999

From Table-1, it is expected to be approximately 52 minutes and 33 seconds of annual interruption on the line with 9s..

This time, observing that the availability of the database system architecture increases with the addition of separate asynchronous (active-passive) systems.

$$f = \frac{n!}{(s+1)!(n-s-1)} \qquad (8)$$

However, the point we need to be aware of is increasing the number of nodules, but increasing the accessibility with discretization. With the combination approach,

the combination f for 6 modular and two separated systems returns 20 values.

Thus, the system may fail due to 20 routes and any one of these routes. If we calculate the continuity of this:

$$F = f(1-a)^{s+1} \qquad (9)$$

$$F = 20(1-0.99)^3$$

F = .00002 over the likelihood of the systemstops

$$A = 1 - F \qquad (10)$$

As a result, our accessibility ratio yields about 5 (five) 9 results. We have shown that we can predict a continuity of about 5 minutes and 15 seconds per year. In this case, the duration of the downtime of 52 minutes 33 seconds in the cold backup system can be changed to 5 minutes and 15 seconds in 2 separate systems and 6 modular systems in the warm backup system. With the hot backup system architecture, this time we are changing our formula, and we determine the number of nodules as 7 and the number of separated systems as 3. The result of method 7:

F = .0.00000035 over the likelihood of the systemstops

A = 0.99999965, and approximately 6 (six) nine (nine) notches, the system was enforced for an estimated 31.5 seconds of downtime per year.

For an increase in high availability, it is seen that hot backup is the best solutions of all the database services' back up architectures. It seems that database services reduce the accessibility because of the addition of nodules to the servers.

V. REDUCING QUERY COSTS STRATEGIES IN ORACLE DATABASE MANAGEMENT

Structured query language (SQL) is a component in every database system. This function of this language is to performs tasks such as data identification, verbal maneuvering, and motion-session-systemcontrol.[10]

A populated database is stored either on a memory device or on a disk. It's best to keep on a memory device, but that is generally expensive. Storage on the disc is advantageous in that it is a easy, flexible and extendable, but it is much slower. The effective use of these two system components are the first task of a database management system. Oracle DBMS performs this task with the SGA (System Global Area) and other background processes.[11]

One of the parameters that Oracle sets during database setup is the System Global Area (SGA). Subsequently, this memory is fetched from the operating system and used only by Oracle until the database is shut down. In the SGA; database buffer cache, user information, redo log buffers and shared SQL pools can be found.

The sizes of these areas are managed by the database administrator according to the characteristics of the application to be used and it is written to the "init. Ora" parameter file. If the database administrator assigns these values as large or small as necessary, the performance of the database decreases.

.

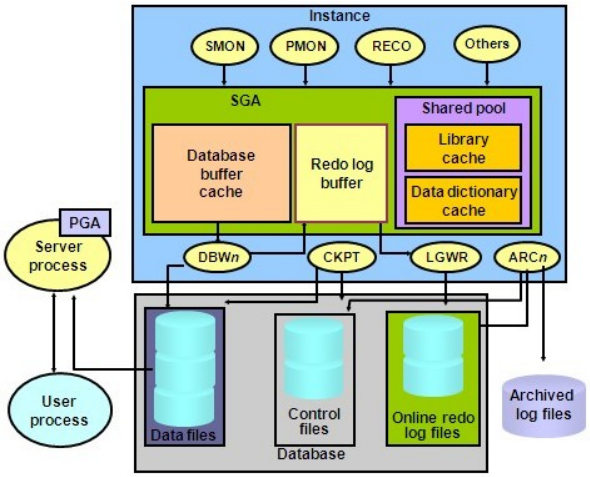


Figure 4:Oracle Database Software Architecture[12]

ASM that resides on the disk and also taken by the operating system or RAW partition together with the infrastructure that ensures access to the disk is called the Oracle "database". Similarly, the memory centered infrastructure is called Oracle "instance". Instance is a term that points to the execution of threads as well as the data from the disk for the active threads. When there is a single database on a system, there can be multiple mounted instances. [Figure 4]

Shared pools: A partition where both SQL and PL / SQL clauses are semantically evaluated (parse) as well as execution plans (text) are done. Additionally, the "data dictionary cache", tables, images and information about the statistical data of the indexes that the structure holds can be found. If the shared memory space is not large enough, then it is necessary to remove again all SQL statements and work plans that are cannot be found in the memory. This situation extends the working time of SQL. The "shared_pool_size" parameter gives the size of the shared memory area in the "init. Ora" parameter file. When we examine the components used in Oracle DBMS, the following memory areas are found

Redo Log Buffer: An intermediate buffer in which changes to the database are kept. This information on the memory that are found in the document "commit" is written to the redo log files. These log files are of great importance as they are used in the database recovery process. If the changes are too small, the I / O count will increase, which causes a reduction in performance. The size of the modified intermediate memory, such as the shared memory area is given by the parameter of the "log_buffer" found in the "init. Ora" file.

Data Buffer Cache: This area is first checked when data is requested. If there is data in this memory area, the number of hits increases, and tasks are performed. If there is no data, the data blocks from the data file (data file) are stored. This is the block of data that is read into the memory area where the most recently used data blocks are located. Before bringing the data into the memory, it determines whether there is space in the memory or not. If there are dirty data in this area, the area will be executed under the process name DBWR. This process opens up space by changing data from disk to data file. How these processes affect, performance is obvious. This is because the latest database contains the most commonly used data on the disk, I / O is significantly reduced to improve database performance. For this reason, the size should be kept large only as required. If it is bigger than necessary, it will decrease performance because extra blank blocks will be read every time. The size of the read data blocks is multiplied by "db_block_size" and "db_block_buffers" in the "init. Ora" parameter file. The "db_block_size" parameter is not arbitrary. The operating system has to be equal or a multiple of the block size. The number of data blocks to be read at the same time is determined by the "db_multiblock_read_count" parameter. At any given time, "db_multibllock_read_count" multiplied by "db_block_size" is read from the information data files and written in the database data files [13].

System global area (SGA) and its related dynamic data can be retrieved with V $ SYSSTAT views (view). If performance-related values are low, the database can be improved by adjusting the size of the SGA.

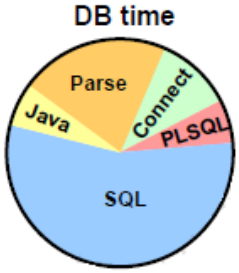


Figure 5: Database Time

In its latest version, Oracle has allowed SGA administration to be put into its core management software. It can do with the parameter named Automatic Shared Memory Management. The fields SGA_RED_POOL_SIZE, LARGE_POOL_SIZE, and JAVA_POOL_SIZE, are called "auto-tuned SGA parameters" in these fields. Oracle Database software can perform performance tuning of these domains by itself.

While the standing database is serving, it tries to use the hardware components such as memory, disk, and processor with full efficiency.

Oracle defines the sum of the elapsed time and the waiting time as the processor unit time. [Figure-5] If the big part of the database time is getting CPU time, then this is indicates that, the structured queries (SQL) on the DBMS to must configure to be more efficient (SQL tuning). If a significant portion of the database time is used spent on waiting, then this indicates that the database should be configured as a system architecture (instance tuning).

Oracle database under normal conditions will spend most of the time on structured query. For this, you need to connect to the system first (link). Pars to the returned query and other activities can then be considered as part of model.

Our primary goal is to reduce database time. The first thing that needs to be done at this point is process monitor and analysis. To be able to monitor, the database must be able to collect statistics. The initial parameter of the database must be set for this. Oracle can collect statistics at three levels: Basic, Typical, and All.

Basic is aggregating the statistics of the reduced parameter group and does not contain suggestions. Typical statistics on the other hand, collection comes as default. It collects more broad parameters than basic statistics collection and adds suggestions. Oracle recommends this level. [14] The statistical level (init.ora) set in the initial parameter file can be confirmed as:

```
SQL> show parameter statistics_level
```

Sample output:

```
NAME              TYPE          VALUE
statistics_level  string        TYPICAL
```

We can analyze some configuration values of DBMS where statistical data can be collected. In this article, aside query improvements, architectural improvements that enhances memory disk usage will also be examined.

It can be a dynamic process to decide what the buffer area of the serving database should be. To be able to make healthy decisions, it is necessary to monitor the usage. Database administrators can decide this with the following command:

```
SQL> SELECT /*+ ordered use_hash(b) */ n.bp_name
BUFFER_POOL,
count(*) CURRENT_BUFFER,
COUNT(*)+COUNT(DECODE(lru_flag, 0, DECODE(tch,
0, null, 1, null, 1))) - COUNT(DECODE(state, 0,
1, DECODE(lru_flag, 8, DECODE(tch, 0, 1, 1,
1))))
IDEAL_MEMORY_SIZE
 FROM
 (
 SELECT /*+ ordered */ p.bp_name, s.addr FROM
X$KCBWBPD p, X$KCBWDS s
 WHERE  s.inst_id = USERENV('Instance') AND
        p.inst_id = USERENV('Instance') AND
        s.set_id >= p.bp_lo_sid AND
        s.set_id <= p.bp_hi_sid AND
        p.bp_size != 0
 )

 n, x$BH b
 WHERE  b.inst_id = USERENV('Instance') AND
        b.set_ds = n.addr
 GROUP BY n.bp_name
 /
```

Sample output:

```
 BUFFER_POO   CURRENT_BUFFE   IDEAL_MEMORY_SIZ
 L            R              E

 DEFAULT      3858261        3989267
1 row selected.
```

As a result of this query, the amount of temporary memory in the database is slightly lower. The odds of being different from the target is 3 percent, which is not critical, but shows that we follow this parameter. What needs to be done is; updating the db_block_buffers parameter. We set ideal memory size. Also, small tables can be kept in memory. The example command we can run for this is as follows:

```
ALTER TABLE <small table> CACHE
```

Database users who run queries in a specified period encounter the same data or query. We can assume that the same query and the result stored in memory will contribute to performance positively since it will allow the same parsed query to be read directly into memory. Oracle VTYS considers this as a way to optimize the library cache space and assign the most appropriate values. By running the following query, we can calculate the number of structured query codes in memory, the probability that formal query codes loaded in memory can be stored in memory:

```
SELECT       sum(PINS-RELOADS)/sum(PINS)*100
"Probability of being in the library cache"
FROM v$librarycache;
```

Sample output:

```
Probability of Library Cache
--------------------------
                99.9701180
1 row selected.
```

The output of this inquiry should be over 99%. The example output shows that SHARED_POOL_SIZE has been assigned the appropriate values. We can also see how many of the queries in the memory are thrown into the library memory without being used again.

```
SELECT sum(PINS) "Running",
sum(RELOADS) "Re-running",
sum(RELOADS)/sum(PINS)
"Percent of Re-running"
FROM v$librarycache;
```

Some query data can be run while the number of bindings is at the top. In terms of performance, it will be efficient if the query results or queries to be sent to the disk, can be kept in the shared pool without access. First of all, we need to find out which queries are being often used, are running and how many database users are using a query. The queries we are going to run for these are:

```
SELECT round(sum(users_opening)/count(*),2)
"Average of number of users one SQL is
executed",
round(sum(executions)/count(*))
"Average of numbers being executed a SQL "
FROM v$sqlarea
```

It is smarter to paste frequently used and large sized objects into the common memory pool. When determining the codes that need to be pasted into the cache, it is necessary to find query objects in the cache. These objects can be found with the following query:

```
SELECT OWNER,NAME,TYPE,SHARABLE_MEM,LOADS,
EXECUTIONS,KEPT
FROM V$DB_OBJECT_CACHE
WHERE OWNER='<sample_own_obj>' and
KEPT='no' AND
TYPE
IN('table','function',procedure',package',pack
age body')
ORDER BY SHARABLE_MEM DESC;
```

As a result of this query, the most called objects from among the database objects are pasted to memory and this contributes significantly to overall performance. How is the code pasted into memory? Below is an explanation:

```
DBMS_SHARED_POOL.KEEP('<own schema.object>')
```

The data dictionary uses Oracle database software for managing the data [15]. Copying the data dictionary, which contains the identification and authorization information of the database users, to the memory will have a positive effect on performance. We can increase the SHARED_POOL_SIZE according to the situation by examining the data dictionary location in our database by calculating the probability of finding this information in memory:

```
SELECT (sum(GETS-GETMISSES))/SUM(GETS) "The
Probability of in Data Dictionary"
FROM v$rowcache

The Probability of in the Data Dictionary
--------------------------
                  .999903176
1 row selected.
```

The sample output shows that the data dictionary is 99% likely to be in memory. It is a good value. Having a figure below 85% signifies a performance problem.

It is quite costly to run the sort operations (query with ORDER BY) which application developers frequently use. This is because additional space and processing are required for sorting. The ranking algorithms that are useful for the business process are the extra performance load. [16] Oracle does sort operations primarily in memory and the parameters SORT_AREA_SIZE and SORT_AREA_RESERVED_SIZE should be sufficient for this. If it is insufficient, it uses temporary tablespace (TEMP TABLESPACE) which translates into a serious performance cost.

During the monitor, you need to be able to use the V $ SQLAREA object to find the most inquiring system. The BUFFER_GETS column of this object gives us an idea of both memory usage and CPU usage. One can learn about which query from the object, V$SQL_TEXT_NEWLINE.

VI. CONCLUSION

The responsibility of ensuring high accessibility in database services continues to rise in recent industrial technologies. The critical step in ensuring high continuity is through intelligent investments in database services, which undoubtedly can have a strong impact on the 7X24 continuity of products.

We emphasized that priorities should be well established in this article as we are overseeing what the database management needs to do at the systemic level of structuring.

Setting the performance is a process that is said to be necessary at all stages and thus System Global Area (SGA) must be carefully designed. It was emphasized that the critical level of the fields in which the queries are managed must be adjusted to the level of the statistics. We discussed what could be done for proactive solutions that give an idea of the performance of the database.

Given that a growth in structural transformation happens alongside a philosophical ideology that, the bigger data the higher access, the performance of a database is now a subject to which much attention and care must be given.

REFERENCES

[1] P. A. İNAM, *«Herşeyin Başı Güven,»* [Online]. Available: http://phil.metu.edu.tr/ahmet-inam/guven.htm. [Accessed: 13 5 2015].

[2] Y. KAPLAN, «*Yüksek Servis Sürekliliği*,» [Online]. Available: http://www.yasinkaplan.com/tr/docs/HA.pdf. [Accessed: 13 5 2015].

[3] A. H. TÜRKAN ve A. H. TÜRKAN, «*Güvenilirlik Analizinde Kullanılan İstatiksel Dağılım Modelleri*,» Adana, 2007.

[4] W. T. a. V. Avelar, «*Mean Time Between Failure : Explanation and Standards*» West Kingston, Rhode Island USA

[5] D. M. Fishman, «*Application Availability: An Approach to Measurement*» 2000.

[6] «What is ITIL® Best Practice?,» [Online]. Available: https://www.axelos.com/best-practice-solutions/itil. [Accessed: 28 12 2015].

[7] «ITIL Availability Management: Beyond the Framework,» CMG, 2006. [Accessed]. Available: http://www.cmg.org/publications/measureit/2006-2/mit33/measureit-issue-4-07-itil-availability-management-beyond-the-framework-by-james-yaple/. [Accessed: 12 28 2015]

[8] W. Brewer, «Database High-Availability: Soup to Nuts,» Simple Talk, 2007. [Online]. Available: https://www.simple-talk.com/sql/database-administration/database-high-availability-soup-to-nuts/. [Accessed: 28 12 2015]

[9] W. H. Highleyman, «Calculating Availability – Redundant Systems,» 2006

[10] Oracle, «Oracle® Database SQL Reference,» Oracle, 2003. [Online]. [Accessed: 8 6 2017]

[11] A. S. R. B. Hitesh KUMAR SHARMA, *SGA Dynamic Parameters: The Core Components*, Database Systems Journal, volume 2, pp. 13-21, 2014.

[12] A. S. Team, «Oracle Performance Tuning : Best Practices,» 26 5 2013. [Online]. Available: http://www.agileload.com/agileload/blog/2013/05/16/oracle-performance-tuning-best-practices:[Accessed: 14 7 2017]

[13] M. Hesson, H. Al-Ameed ve M. Semaka, *Database Performance Issues*, Engineering and Meta-Engineering: ICEME 2012 , Orlando, Florida USA, 2012

[14] O. University, Oracle Database 11g:Performance Tuning, Oracle Corp., 2016.

[15] S.CALLAN, *The Dictionary in the Data Dictionary*,Database Journal,25.9.2007

[16] H.ŞARKIŞLA, H.EVİNGEN, Hadoop MapReduce Algoritmasının Analizi ile Performance Etki Eden Parametrelerin Tespiti ve Optimize Edilmiş Parametreler ile Hadoop Üzerinde Başarım Artımı , *5TH International Symposium On Innovative Technologies and Science*, Architecture and Construction University Baku / Azerbaijan, 147-1257, 29-30 SEPT 2017

Lifestyle Change Treatment with Cloud Based Mobile Tracking Application: VITAE

I. ILTER[1], A. KUT[1], V. MEVSIM[1], S. UTKU[1]

[1] Department of Computer Engineering, Dokuz Eylul University, The Graduate School of Natural and Applied Sciences, Izmir, Turkey, ilay.ilter@ceng.deu.edu.tr
[1] Department of Computer Engineering, Dokuz Eylul University Engineering Faculty, Izmir, Turkey, alp@cs.deu.edu.tr
[1] Department of Family Medicine, Dokuz Eylul University Faculty of Medicine, vildan.mevsim@deu.edu.tr
[1] Department of Computer Engineering, Dokuz Eylul University Engineering Faculty, Izmir, Turkey, semih@cs.deu.edu.tr

***Abstract* - Many people live with various health problems due to smoking, obesity and sedentary life. Especially, many people have severe illnesses such as lung cancer and asthma because of smoking. However, people unfortunately do not quit smoking easily. It is aimed to change the lifestyles of people with the mobile application that will be developed in order to solve these problems. In the proposed system, a cloud based mobile application will be developed under the supervision of a family physician. The treatment plan will be given as a result of the medical history and physical examination who applied to family physicians to quit smoking. Patients will be able to access this treatment plan through mobile application. After that, patients will be able to record and monitor daily cigarettes, nutrition and exercise. The doctor will check the information that the patient has entered and will be able to detect if the patient is not fit in treatment. Accordingly, doctor can update the treatment plan and create the most appropriate treatment model for the patient.**

***Keywords* – mHealth, mobile cloud computing, smoking cessation, healthy life**

I. INTRODUCTION

In recent decades, lifestyle is as an important factor of health. Many problems are caused by unhealthy lifestyle. Problems like metabolic diseases, joint and skeletal problems, cardiovascular diseases, hypertension, overweight, violence and so on. Obesity, smoking and sedentary lifestyle are big parts of unhealthy lifestyle. In addition to that physical inactivity is seen as one of the most important health problems of the 21st century. Being overweight is responsible for more than 1 million deaths each year and 12 million years of illness in the European Region. [1]

Another important element of unhealthy life is smoking addiction. Smoking is an important public health problem in the world and in Turkey, and it has the potential to become addictive because it contains nicotine at a high rate. Every year in the world 6 million people lose their lives due to smoking. This number means that one person dies from cigarettes every 10 seconds. [2] For this reason, the Ministry of Health in Turkey provides important support to quit smoking.

Thanks to the developing technology and the internet, users are oriented to cloud computing technology which is flexible, economical and accessible everywhere. "Cloud Computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (eg., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [3] Cloud computing provides to access resources independently of the location.[4]

In the 21st century intelligent mobile device technology has developed rapidly and has become one of the integral parts of our lives. Daily tasks can be done with mobile technology in more practical, flexible times, in desired environments. Increasing use of mobile technologies in the health field also leads to an increase in the number of mobile health applications. [5] With the use of cloud computing, systematic and transferable patient records that can be accessed from every point. Patient records, date of examination, time of vaccination, time of drug use, feedback and warning systems related to screening, remote and continuous monitoring of health data, treatment decision-making systems, mobile applications can achieve much faster and more accurate results and the interaction between patients, doctors and hospitals is developing rapidly.[6]

II. RELATED WORKS

Cloud computing is one of the latest technologies. Cloud computing is also rapidly increasing in everyday life. Today, the application of cloud computing is so common that it is used even in the healthcare industry. When compared to other industries, the health industry has considerable inadequate technology to improve operational efficiency. [7] One of the main reasons for the use of cloud computing in the health sector is increasing population density and cloud computing to provide the infrastructure for the growth and storage of digital data of patients. Mortality due to cardiovascular diseases is high due to insufficient follow-up of cardiovascular risk factors such as diabetes, physical inactivity and smoking. Healthy lifestyle choices play an important role in the management of these changeable risk factors. [8] According to WHO, Mobile Health (mHealth) is an area of electronic health (eHealth) and it is the provision of health services and information via mobile technologies such as mobile phones and Personal Digital Assistants (PDAs) [9]

Nowadays smartphones can detect many habits of users in their daily lives. In a research conducted, the data gathered from the people's smartphones were analyzed. In this way, it is explained how to determine the most basic lifestyle parameters

related to users' health. Information about how lifestyle parameters that affect health such as physical inactivity, irregular sleeping, stress, etc. can be measured via smartphones. [10] With this research it has been understood which methods to use to determine some data to be used in the project. For example, the number of steps per day.

Cigarette addiction is increasing day by day. This causes serious health problems. For this reason, the number of treatments for smoking cessation is increasing. The treatments are provided via web, telephone, text. Recently, treatments with mobile applications are also increasing. This increases access to smoking cessation initiatives. [11] In a study conducted in the United States, it was requested that 100 selected smoking cessation treatments be used to treat two cigarette addicts. Those with psychotic disorders assessed the first 9 of these practices according to their usability characteristics. As a result, they analyzed the quantitative results and found that three common usability problems. Some of these have been identified as the intensive text content. Others are small icons on the main page for organizing technical symbols and features. [11] In another study, the effect of a text message-based motivation on cigarette smoking on university students was examined. In this research, increasingly used motivational interviewing technique was used by clinical therapists and encouraging researchers to quit smoking, and this technique was adapted to the developing mobile phone technology. In this way, this method will reduce number of hospital visits by using the mobile phone, which has already occupied an important place in daily life. Motivational interviewing technique will be applied in a way that makes the person feel more comfortable. [12] This method has yielded successful results. Smoking and health care cost can be reduced. However, because the text messaging method is used, it can increase the GSM costs. In Australia, overweight and obesity rates have increased over the last 20 years. And mobile phone applications have been seen to play an important role in monitoring and managing the weight of individuals. [13] In the applications examined in this research, the person has only followed nutrition and weight gain. However, since the practices are not under the control of the doctor, the patient may be unable to lose weight in the future.

Healthy life practices provide motivation support for patients to follow their daily activities, to follow their diet, and also to reach their goals. However, most of the applications developed are individual. So it seems that there are practices that do not follow the doctor. Even though the studies are successful, they cannot be used if they are used for therapeutic purposes in changing human profiles.

III. PROPOSED SYSTEM

The proposed solution is to develop a rule-based proposal system for the treatment of patients who refer family physicians to live healthy (quit smoking, to lose weight, healthy diet). This system includes a web application that provides treatment advice based on the patient's information. It is also aimed to develop cloud computing based mobile software where patients can make daily follow-ups and constantly communicate with the doctor. Figure 1 shows the general structure of a software system that allows lifestyle change for patients.

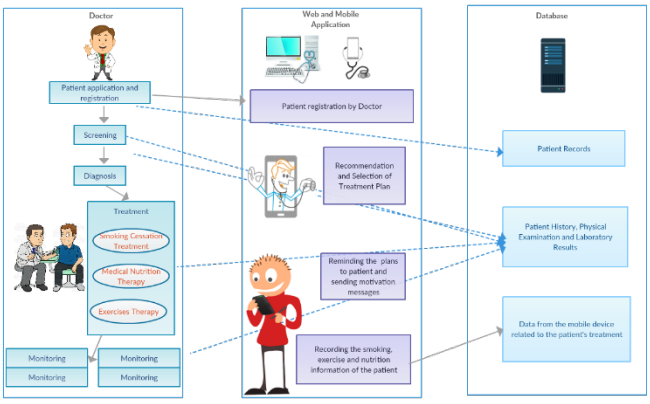


Figure 1. The general structure of lifestyle modification treatment system to quit smoking

A. Motivation

The proposed system mainly consists of two applications, web and mobile. In a web-based application, the patient passes through 3 stages. These;

1) Registration of the patient's general information, recording of the history, physical examination information and laboratory request and completion of the first examination
2) When the patient comes back to the examination together with the laboratory results and the treatment proposal is presented according to the results and the treatment to be applied by the physician is determined
3) Re-examination of the patient for follow-up and continuation, replacement or termination of treatment

The mobile-based application includes three different modules: the Smoking Cessation System, the Exercise System, and the Healthy Nutrition System. These modules include the following items.

- The app will regularly remind the patient of what he needs to do daily (eg, today's smoking day, a 45-minute fast walk, etc.) and record what he / she has done (eg, smoking status, exercise period, weight gain etc.)
- As the patient registers what they do, they will tell the difference between what they need to do and send new recommendations and motivation messages. If the patient reaches the targets, it will send messages that will keep the motivation going.
- The application will also remind the patient of the appointment dates.
- Through the application, the physician will be able to adapt his lifestyle modification treatment and monitor his results.

- Through the application, the patient will be able to ask questions and be in constant contact with the doctor

B. *Mobile Based Software System*

Figure 1 shows the stage in which the mobile software is to be used. Figure 2 shows the architecture of the mobile software system. The mobile software system is classified under three headings. The first title is smoking cessation treatment, the second title is follow-up nutrition, and the third title is follow-up exercise.

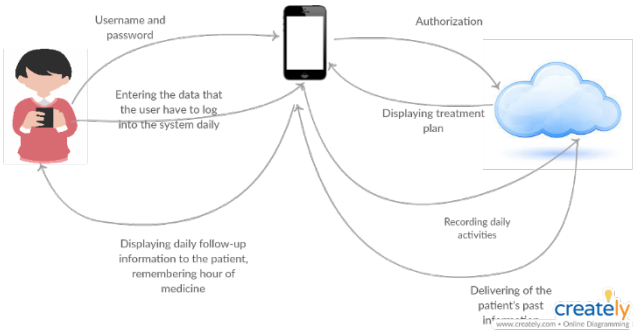


Figure. 2. Architecture of Mobile Software

The patient logs in to the system with his / her username and password if it is registered. If not registered, the user sings in and logs in later. The first time the patient logs in, patient records the sleep and wake time and turns on or off notifications to receive motivation messages. Patient then clicks on the segment he is being treated for. Later, if the doctor has entered a treatment plan in the database, the current treatment plan will be displayed. If not, the plan will be entered. It records daily information in accordance with this plan. The patient can perform daily follow-ups quickly and at any time. In accordance with the information entered by the patient and the follow-up information entered by the doctor, the treatment plan is updated as a result of data analysis methods at specific time intervals. It is aimed to improve patient compliance and success rate of treatment applied in this way. In smoking cessation treatment, the amount of cigarettes the patient smokes during the day, the factors causing smoking and the behaviors he / she makes when he / she wants to smoke are followed. Motivation messages are sent to the patient in the direction of this data. These collected data are stored in the database that the cloud computing platform has provided, and the doctor can follow the data from anywhere and at any time to provide follow-up treatment of the patient, and as a result, update the treatment form of the patient.

IV. RESULTS AND DISCUSSION

The protection and development of health is a great way of gaining a proper eating habit and active life. Besides, it is seen as a cause of many illnesses in smoking dependence. In many studies conducted in the world and in Turkey, treatment methods are being developed to change patients' lifestyles. Along with the developing technology, many web and mobile applications are also developing to support the treatment. Patients have come to the point where they can provide daily follow-ups at home thanks to the developed applications.

According to 2016 data, about 100,000 mobile health applications have been published since the previous year and there are 259,000 mobile health applications available. 72% of these application publishers use analytics, testing, storage and cross-platform tools, but only 23% of these applications have been found to be installed between 5,000 and 50,000. [14] The study [9] investigated 100 smoking cessation practices and found that the main issues in using these applications are the intensive text content, small icons for organizing technical symbols and features on the main page. In the application we developed, the terms were used at a level that the user could understand, the size of the symbols and the density of the contents were designed so that one could use it most efficiently. In study [12], motivational treatment with sms was applied in order to enable university students to quit smoking. However, in the application we have developed, motivational treatment method is provided by summarizing what you do daily at the end of the day and by sending notifications during the day. It is seen that the GSM cost decreases with this method.

Table 1. Comparison of Other Studies and Vitae

	[11]	[12]	Vitae
Daily Tracking	✓	✓	✓
Treatment Follow-up	X	✓	✓
Motivational Treatment	X	✓	✓
Usability without network	✓	✓	✓
Nutrition and exercise tracking	X	X	✓
Smoking cessation treatment	✓	✓	✓
Direct contact with the doctor	X	X	✓
Decision Support System	X	X	✓
Cloud computing	X	X	✓
Motivation messages	✓	✓	✓

Table 1 shows the comparison of other studies with Vitae. The main difference compared to the studies done is whether the patient-doctor is mutual or not. In the majority of the investigated practices, the patient follows the daily amount of smoking, diet and exercise independently of the doctor, and receives motivation messages associated with it. However, with the web and mobile application we have developed, the patient applies and follows the treatment model developed by his / her doctor. The doctor decides whether the treatment is working by evaluating the data that the patient has entered during the day. Subsequent treatment is updated as a result of data mining analysis of the patient's data. This increases the success rate of treatment.

V. CONCLUSION

With the increasing use of smartphones and tablets, the number of mobile applications that are being developed in the healthcare sector is increasing day by day. Mobile applications are available for many things such as monitoring the health of

people, controlling heart rate and blood pressure, calculating calorie intake daily, controlling medication dose and timing, and monitoring the pregnancy process. A web-based system has been developed in which the proposed software system can provide follow-up advice and follow-up of treatment. The patient can communicate directly with the family physician from anywhere and at any time with the possibilities provided by cloud technology and can follow up at the same time in reaching and practicing the treatment given by the family physician. In this study, lifestyle modification therapy was applied by mobile application to smoking cessation and achieves a healthier life. The proposed solution continues to be applied to family physicians and continues to work for worldwide use.

ACKNOWLEDGMENT

This study is a part of the project, which is the project manager of Prof.Dr.Vildan Mevsim, is funded by the Scientific and Technological Research Council of Turkey and VITAE is copyrighted by VINNOVA Health Company.

REFERENCES

[1] Türk Halk Sağlığı Kurumu, http://beslenme.gov.tr
[2] http://www.yesilay.org.tr/tr/bagimlilik/sigara-ve-tutun-bagimliligi
[3] DoD Directive 3020.40, Defense Critical Infrastructure Program, 19 Aug 2005, p. 13, http://www.dtic.mil/whs/directives/corres/pdf/302040p.pdf..
[4] Deepti Sahu, Shipra Sharma, Vandana Dubey, Alpika Tripathi, Cloud Computing in Mobile Applications, International Journal of Scientific and Research Publications, Volume 2, Issue 8, August 2012.
[5] Gülsün Eby, Emel Güler,Akıllı Ekranlarda Mobil Sağlık Uygulamaları, Journal of Research in Education and Teaching, Ağustos 2015
[6] http://www.bthaber.com/saglik-sektorunde-bilisim/saglik-alanindaki-yeni-egilim-bulut-bilisim/1/10631
[7] Vaibhav Kamal Nigam, Shubham Bhatia ,Impact of Cloud Computing on Health Care, International Research Journal of Engineering and Technology (IRJET),2016
[8] Hasan Rehman,Ayeesha Kamal,Saleem Sayani,Pamela B. Morris,Anwar T Merchant,Salim S. Virani,Using Mobile Health (mHealth) Technology in the Management of Diabetes Mellitus, Physical Inactivity, and Smoking , 2017
[9] Burcu Çinaz, Bert Arnrich, Akıllı Telefonlar ile Kullanıcıların Yaşam Tarzı Parametrelerinin Tespiti, Akademik Bilişim'14 - XVI. Akademik Bilişim Konferansı Bildirileri ,Şubat 2014
[10] Ecir Uğur Küçüksille , Fadime Özger , Sevdanur Genç, Mobil Bulut Bilişim ve Geleceği, Akademik Bilişim , 2013
[11] Joelle C Ferron, Mary F Brunette, Pamela Geiger, Lisa A Marsch,Anna M Adachi-Mejia, Stephen J Bartels,Mobile Phone Apps for Smoking Cessation: Quality and Usability Among Smokers With Psychosis, JMR Publications, 2017
[12] Anna Jorayeva, Effects of a text message-based motivational interviewing intervention on cigarette smoking in college students. University of Louisville ThinkIR: The University of Louisville's Institutional Repository, Aralık 2016
[13] Sarah Zaidan, BSc (Hons), MSc (Hons) ,Erin Roehrer, PhD ,Popular Mobile Phone Apps for Diet and Weight Loss: A Content Analysis, JMR Publications , 2016
[14] mHealth app market 2016: Top 10 changes and reasons to be optimistic for the years to come, https://research2guidance.com/mhealth-app-market-2016-top-10-changes-and-reasons-to-be-optimistic-for-the-years-to-come/

Research of the Traffic Flow on the Signal-controlled Intersections using Queuing Theory

F. GUNES[1], S. BAYRAKLI[2] and A. H. ZAIM[1]

[1] Istanbul Commerce University, Istanbul/Turkey, fatgun41@gmail.com
[2] Maltepe University, Istanbul/Turkey, selimbayrakli@maltepe.edu.tr
[1]Istanbul Commerce University, Istanbul/Turkey, azaim@ticaret.edu.tr

***Abstract* – The global problem of the growing city centres has many causes and requiring many different solutions. Existing urban infrastructures are having difficulty to meet the needs within the increasing population and urbanization of the cities. In this context, innovations must be made in many areas of the infrastructure of the cities such as Transport, Energy and Environment cleanliness. The most important increase due to the population growth in urban areas occurs in the traffic intensity. The number of vehicles increasing in traffic brings with it many new problems such as energy, time, and environmental pollution. Several field studies have been done so far to reduce and minimize traffic intensity. In this context, different solutions have been tried, such as speed controls, camera systems equipped with image processing techniques, and measures to reduce the intensity of intersection connections. As a result of interviews and determinations within the scope of this study, it is seen that traffic lights have a serious effect on traffic regulation.**

In this study, the working principle of traffic lights was first investigated. In particular, the optimization and the techniques of signaling times of traffic lights have been examined. Traffic lights, which usually operate for a fixed period of time, can cause the vehicles to remain in traffic for a longer time and increase traffic intensity. Systems that operate according to the vehicle density cause very important improvements if they are operated through the appropriate mathematical model. In our work, queuing theory is mentioned, the application areas of theory are investigated and the data obtained from the field have been tested through this model. According to the queuing theory, the intensity of the intersection arms, the performance of the lights, arrival and the service ratios were obtained. With this study, it is aimed to bring a new approach to the density at the intersection of the city centers.

***Keywords* - Signalized intersections, Smart City, Intelligent Traffic Lights, Queuing Theory**

I. SMART CITIES, TRAFFIC MANAGEMENT SYSTEMS

Traffic intensity causes increased transportation time, excessive fuel consumption, high energy costs, exhaust gas emission and environmental pollution[17]. For this reason, this study investigated the aspects of traffic lights that increase the positive contribution to traffic regulation. As is well known, traffic lights in our country usually work on a timely basis.

This situation poses many problems;

1) Vehicles stay longer in traffic and increases traffic intensity
2) It leads to time loss,
3) Increases fuel consumption,
4) It has a negative contribution to environmental pollution.

A. Intelligent Signalized intersections

Intelligent Signalized intersections are the systems aiming to minimize the average delay times and waiting times of vehicles in a junction. For this purpose, the working times of the traffic lights are automatically adjusted considering the number and density of the vehicles. Traffic lights are optimized according to the parameters such as traffic volume, queuing, etc., and new time periods are applied in real time.

These systems, called dynamic intersection control systems, reduce the waiting time of vehicles by ensuring that light periods are optimized depending on the number of vehicles at the intersections. By integration control devices in the intersection, light periods are automatically determined. Intelligent Vehicle Counting Sensors are used to calculate the densities of vehicles in the intersection of vehicles in real time. Vehicle density data is transmitted to the Dynamic Intersection Control Unit, which is located in the intelligent intersection. This data transmitted to this unit is analyzed in real time by the system, and the light periods at intersections are constantly optimized. Thus, the waiting time of the vehicles in traffic lights is reduced to the minimum. Due to the algorithm that will work in the system, it is considered that both the intersection arms and the complexity of the system are prevented from increasing of the traffic density.

B. Methods developed against the intersection density problem in the literature

The signaling system in an intersection must be very good designed. Vehicle and pedestrian safety must be observed in time distributions, the passage times to the intersection connections must be in accordance with the proportions of the intensities of the directions and the signaled transition times. Hourly, daily, monthly, seasonal changes of current values should be taken into consideration especially in time distributions, and a continuous update should be made because of the change of current values due to time series. The signalized intersections that are unnecessarily,

misplaced or incompatible installed, may increase delays, and as a result, drivers and pedestrians may be forced not to follow the lights [1]. Therefore, the performance of intra-city road transport depends largely on the achievement of road traffic control, and the success of road traffic control on correct signal timing.

A lot of research has been done to improve the signaling systems in the intersections. In studies, artificial neural networks, fuzzy logic methods and decision support systems are usually designed. In 1977, Pappis and Mandani designed a decision-making model using the signal cycle, queue length, approach line graph and extension time as parameters [2]. Tzes and colleagues have made a traffic signal control in 1995 for transportation networks. Tzes, McShane and Kim have developed a simulation model based on fuzzy logic and have compared it with fixed-time signal model [3]. Jongwan Kim designed a simulation model for discrete and 4-phase controlled intersections with the control algorithm, that he developed, and then determined their differences with fixed-time signalization [4]. J.P.Niittymaki has improved the performance of intersections by simulating the control algorithm, that he developed for discrete and two-phase supervised signalized intersections [5]. Hoyer and Jumar have developed a model that deals with variable phase order and compares the traffic volume at intersection links and the adjustment of phase durations depending on the red light signal duration [6]. Huang and colleagues have designed the control of signaling systems with genetic algorithms [7]. Jang and colleagues have contributed to the solution of the traffic congestion in city centers by balancing methodology of queue sizes [8]. Ming Kang and his colleagues have developed a method that can perform real time intersection planning with video processing techniques using RFID sensor, GPS and Camera Sensor in flowing traffic [9]. Babicheva investigated the way to optimize traffic intensity at intersections with queuing theory and made a prediction with a statistical model of green light durations [10]. Anokye and his colleagues measured the performance on the arms of a intersection over real data using the queuing model [11].

In this study we wanted to adapt queuing theory models with real traffic data. We aimed to extract the distributions of the data first by using the intersection data and then to derive the performance criteria such as the arrival and service rates of the arms.

II. QUEUING THEORY

Queuing theory is the mathematical model of the waiting problem in any system such as customers waiting for supermarket payment kiosks, cars waiting in a traffic light, airplane waiting at the airport and Patients waiting in a clinic. Because the customer who comes to the system can not get service right away, the queue comes to the scene. Thus, the queue is a waiting event that occurs because of a limited service and as a result a tail appears at these points [12]. The ideal structure for the enterprises is the incoming customer must leave the service without waiting. Queuing Theory aims to get customers in a system to serve as soon as possible. The purpose of this model is to minimize the cost of servicing and maximize efficiency. The purpose of examining a service is to get some parameters that show the performance of this system. For example, a logical measure of performance is how long a client will wait before serving. Another measure is the percentage of time that the service agent is not used. The first evaluates the system for the customer while the second measure indicates the use rate of the service means. We can understand that when the customer waits more in the queue, so the service system will be full or the opposite of it will be also true. The main actors in a queue situation are "Customer" and "Service".

- Customer arrivals
- Customer service period

The basic elements of a queueing model depend on the following factors [13];

1) Distribution of arrivals and service time
2) Design of service facility. (serial, parallel, network stations)
3) Service discipline/Queue discipline (FIFO, LIFO, service in random order (SIRO))
4) Queue Size, Queue size may be finite or infinite (i.e., a very large queue)
5) Source (customer created) (finite or infinite)

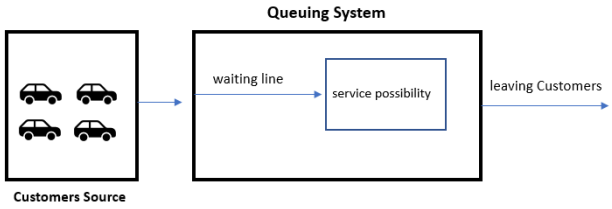


Figure 1 Single Queue, Single Service System

A. *M/M/1/∞ Queue Model*

This work is based on the M/M/1 queuing model. In this model the time between arrivals and service time have exponential distribution. There is only one server in this model. Queue capacity is infinite and the service of the FİFO queuing discipline is used. This model is a good approximation for a large number of queueing systems [11]. Customers in the system are all independent.

B. *Performance parameters of M/M/1/∞ Queue Model*

We can simply explain equations that form an M / M / 1 queuing system as follows;

First we can define the traffic intensity, ρ (sometimes called occupancy). It is defined as the average arrival rate (λ) divided by the average service rate (μ) [11], [19].

Traffic Intensity; should be less than one;

$$p = \frac{\lambda}{\mu} \quad (1)$$

Mean number of customers in the system(N);

$$N = \frac{\lambda}{\mu - \lambda} \qquad (2)$$

Mean number of customers in the queue;

$$Nq = \frac{p^2}{1-p} \qquad (3)$$

Mean time spent waiting in queue;

$$w_q = \frac{p}{(1-p)\mu} \qquad (4)$$

Mean time spent in the System;

$$T = \frac{1}{\mu - \lambda} \qquad (5)$$

III. ANALYZED DATA AND OBSERVED SIGNALIZED INTERSECTION

We tried to make all the performance of the intersection and its arms according to the M / M / 1 queuing model through the real data obtained from the intersection we worked on. As mentioned above, the M / M / 1 model was preferred because the arrival of the vehicles was random and the road has the only one traffic light.

Let's first examine the plan of the signalized intersection in Istanbul-Bostancı where the data were taken;

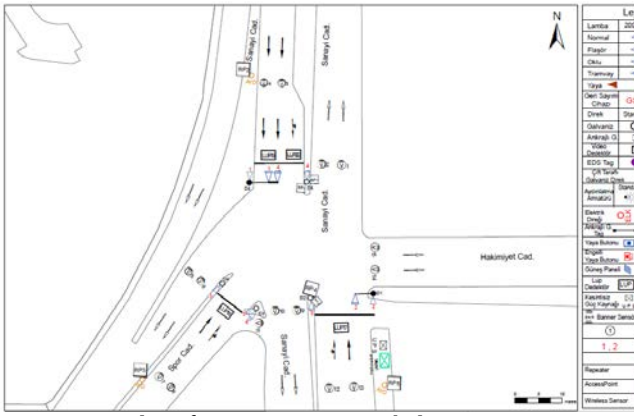


Figure 2 Plan of Intersection in Istanbul-Bostanci

We made our work primarily on a 3-legged intersection. The characteristics of some of the data obtained by the sensors are as follows.

Table 1: Collected Data via sensors from intersection in Bostanci

Faz	Geçiş Tarih ve Saati	Faz Süresi	O1	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
2	01.10.2017 00:00:44	50	0	0	50	50	0	6	0	8	0	0	0	0	0	0	0	0	1	0	1	0	0	0
1	01.10.2017 00:00:49	590	48	47	590	590	46	51	0	25	407	12	0	9	5	0	0	4	7	0	3	1	2	0
3	01.10.2017 00:01:48	141	0	0	141	141	7	22	0	68	0	0	0	0	0	0	0	1	3	0	0	0	0	0
2	01.10.2017 00:02:02	49	0	0	49	49	0	7	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1	01.10.2017 00:02:07	590	67	54	173	83	24	61	171	45	235	4	64	13	8	1	0	7	14	0	1	1	1	1
3	01.10.2017 00:03:06	101	0	0	0	94	24	17	10	49	0	0	0	0	0	0	1	1	1	2	1	0	0	0
2	01.10.2017 00:03:16	49	0	0	14	49	18	30	0	49	0	0	0	0	0	1	0	1	2	0	0	0	0	0
1	01.10.2017 00:03:21	590	52	40	590	590	44	48	3	135	66	30	556	9	6	0	0	6	8	1	3	4	4	1
3	01.10.2017 00:04:20	100	7	0	100	100	16	0	4	31	34	5	100	1	0	0	0	2	0	1	0	1	1	0
2	01.10.2017 00:04:30	50	0	0	50	50	12	18	22	0	4	0	50	0	0	0	0	0	2	1	0	1	0	0
1	01.10.2017 00:04:35	590	47	80	590	590	19	49	29	41	19	33	590	10	16	0	0	5	10	6	6	2	6	0
3	01.10.2017 00:05:34	200	0	0	200	200	48	54	0	24	0	0	200	0	0	0	0	4	5	0	1	0	0	0
2	01.10.2017 00:05:54	50	0	0	50	50	9	4	4	26	20	0	50	0	0	0	0	1	1	1	0	0	0	0
1	01.10.2017 00:05:59	470	49	57	419	419	0	25	0	102	35	18	470	7	7	0	0	0	6	0	4	1	3	0
3	01.10.2017 00:06:46	100	0	0	0	0	13	12	0	6	0	0	100	0	0	0	0	2	2	0	1	0	0	0
2	01.10.2017 00:06:56	50	0	0	0	0	4	4	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0
1	01.10.2017 00:07:01	590	68	62	298	277	2	8	48	117	32	16	590	15	9	1	1	1	2	5	7	1	3	0
3	01.10.2017 00:08:00	480	0	0	480	480	9	43	11	14	22	38	242	0	0	0	0	2	5	2	2	4	2	0
2	01.10.2017 00:08:48	50	0	0	50	50	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	01.10.2017 00:08:53	560	76	86	560	560	70	91	252	17	6	7	355	10	10	0	0	10	16	1	4	2	2	1
3	01.10.2017 00:09:49	480	0	0	480	480	34	81	172	230	22	38	480	0	0	0	0	4	9	1	4	4	7	0
2	01.10.2017 00:10:37	51	0	0	51	51	0	13	3	0	0	0	51	0	0	0	0	0	2	1	0	0	0	0
1	01.10.2017 00:10:42	590	91	94	317	222	66	61	5	49	8	28	590	11	8	2	0	8	7	1	5	2	6	0
3	01.10.2017 00:11:41	479	0	8	0	33	32	53	41	117	64	24	479	0	0	0	1	6	7	2	5	1	5	0
2	01.10.2017 00:12:29	50	0	0	0	10	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0
1	01.10.2017 00:12:34	580	53	80	23	0	51	49	39	58	29	96	458	9	12	2	0	6	10	5	8	1	3	0
3	01.10.2017 00:13:32	480	0	0	8	286	27	66	5	0	0	11	0	0	0	1	1	3	6	1	0	0	2	0
2	01.10.2017 00:14:20	50	0	0	50	50	0	11	3	0	31	0	10	0	0	0	0	0	1	1	0	1	0	1
1	01.10.2017 00:14:25	220	28	44	220	220	29	38	55	209	12	21	220	8	10	0	0	3	7	1	2	0	2	0
3	01.10.2017 00:14:47	231	0	0	231	231	16	28	9	40	31	18	231	0	0	0	0	1	2	1	1	1	2	0
2	01.10.2017 00:15:10	49	0	0	49	49	0	0	6	9	0	0	49	0	0	0	0	0	0	1	1	0	0	0
1	01.10.2017 00:15:15	590	120	109	590	590	20	77	50	13	57	2	590	16	10	0	0	7	15	3	4	2	1	0
3	01.10.2017 00:16:14	231	5	0	231	231	28	64	0	8	4	0	231	1	0	0	0	3	7	0	2	1	0	0
2	01.10.2017 00:16:37	49	0	0	49	49	0	6	7	0	3	0	49	0	0	0	0	0	1	2	1	1	0	0
1	01.10.2017 00:16:42	590	34	39	590	590	64	81	12	156	177	6	590	8	8	0	0	8	12	3	2	2	2	0
3	01.10.2017 00:17:41	480	0	48	480	480	26	67	4	3	0	13	480	0	1	0	0	3	8	1	1	0	3	0
2	01.10.2017 00:18:29	50	0	46	50	50	0	0	9	32	5	0	50	0	0	0	0	0	0	1	1	2	0	0
1	01.10.2017 00:18:34	590	72	85	590	590	39	52	18	110	57	17	220	14	17	0	0	6	8	4	2	2	4	0
3	01.10.2017 00:19:33	481	0	0	481	481	19	39	6	60	3	5	337	0	0	0	0	4	6	1	4	1	1	1
2	01.10.2017 00:20:21	50	0	0	50	50	0	0	0	1	0	0	50	0	0	0	0	0	0	0	1	0	0	0
1	01.10.2017 00:20:26	589	74	70	589	589	55	53	27	111	0	18	589	14	11	0	0	12	9	0	6	0	2	0
3	01.10.2017 00:21:25	200	0	0	200	200	12	46	0	0	0	2	200	0	0	0	0	1	5	0	0	0	1	0
2	01.10.2017 00:21:45	50	0	0	50	50	7	22	0	0	0	0	50	0	0	0	0	0	1	3	0	0	0	0
1	01.10.2017 00:21:50	470	57	66	470	470	41	47	0	158	12	13	359	6	8	0	0	8	10	0	3	1	3	0
3	01.10.2017 00:22:37	100	0	0	100	100	0	13	0	0	0	6	0	0	0	0	0	0	2	0	0	0	1	0
2	01.10.2017 00:22:47	50	0	0	50	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	01.10.2017 00:22:52	470	46	27	470	470	23	29	149	19	4	7	110	11	6	0	0	5	5	2	4	1	2	1
2	01.10.2017 00:23:39	340	2	0	340	340	81	102	3	111	0	0	340	1	0	0	0	3	7	1	2	0	0	0
1	01.10.2017 00:24:13	470	40	30	470	470	78	86	4	29	107	0	470	10	8	0	0	7	10	2	1	1	0	0
3	01.10.2017 00:25:00	120	0	2	120	120	17	29	0	0	103	0	120	0	1	0	0	0	3	0	0	1	0	0
2	01.10.2017 00:25:12	51	0	0	51	51	0	17	2	0	0	4	51	0	0	0	0	0	1	2	0	0	1	0

Explanation of table data;

The table contains phase durations, orders and transit times of the vehicles. V1-V11 sensors show us that from which arms of the intersection the vehicles enter and leave. With the sensors O1-O11, we can obtain the intensities in the arms of the intersection. We can get the order of the phases and their numbers as follows.

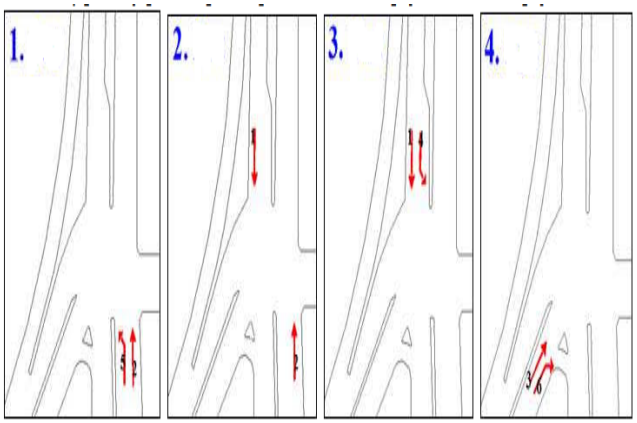


Figure 3 Orders and Structures of Phases

A. Arrival and Service Performance According to M / M / 1 Queue Model

Kadıköy Tatarağa Intersection in Bostancı-Istanbul

		Intersection Arms	Session	ARRIVAL		SERVICE	
				Average Cars	Av. Time (Min)	Average Cars	Av. Time
1.10.2017-	1. Phase	Mehmet Şevki Paşa Cad. - Ali Nihat Tarlan Cad.	00:00 - 07:00	1,76	15,6 (0,26)	3,76	45,81(0,76)
	2. Phase	Ali Nihat Tarlan Cad. side road	00:00 - 07:00	0,2	35,51(0,59)	0,1	5,26(0,87)
	3. Phase	Mehmet Şevki Paşa Cad. one direction	00:00 - 07:00	1,01	26,63(0,44)	1,21	23,38(0,39)
1.10.2017	1. Phase	Mehmet Şevki Paşa Cad. - Ali Nihat Tarlan Cad.	07:00 - 13:00	4,46	14,55(0,24)	7,32	36,39(0,61)
	2. Phase	Ali Nihat Tarlan Cad. side road	07:00 - 13:00	1,14	23,74(0,39)	0,22	11,87(0,19)
	3. Phase	Mehmet Şevki Paşa Cad. one direction	07:00 - 13:00	1,37	30,21(0,50)	1,38	24,52(0,40)
1.10.2017	1. Phase	Mehmet Şevki Paşa Cad. - Ali Nihat Tarlan Cad.	13:00 - 21:00	1,13	18,43(0,31)	8,28	52,61(0,87)
	2. Phase	Ali Nihat Tarlan Cad. side road	13:00 - 21:00	0,55	28,54(0,47)	0,19	12,89(0,21)
	3. Phase	Mehmet Şevki Paşa Cad. one direction	13:00 - 21:00	1,83	31,56(0,52)	2,48	27,74(0,46)

Figure 4 Avg. numbers of cars in the Queue

Arrival and departure service ratios and density at junctions [14]

		Intersection Arms	Session	Arrival Rate	Service Rate	Traffic Intensity	Cars in the System	Cars waiting in the Queue	Mean time spent in the System	Mean time Spent in the Queue
1.10.2017-	1. Phase	Mehmet Şevki Paşa Cad. - Ali Nihat Tarlan Cad.	00:00 - 07:00	6,77	4,94	1,37	3,7	5,07	0,55	0,75
	2. Phase	Ali Nihat Tarlan Cad. side road	00:00 - 07:00	0,33	0,41	0,80	2,5	2,25	0,6	0,82
	3. Phase	Mehmet Şevki Paşa Cad. one direction	00:00 - 07:00	2,29	3,2	0,72	2,8	2,1	1,25	0,92
1.10.2017	1. Phase	Mehmet Şevki Paşa Cad. - Ali Nihat Tarlan Cad.	07:00 - 13:00	18,5	12	1,54	3	4,5	0,16	0,25
	2. Phase	Ali Nihat Tarlan Cad. side road	07:00 - 13:00	2,92	3,56	0,82	4	5,2	0,18	0,36
	3. Phase	Mehmet Şevki Paşa Cad. one direction	07:00 - 13:00	2,74	2,56	1,07	3,76	2,97	1,38	1,1
1.10.2017	1. Phase	Mehmet Şevki Paşa Cad. - Ali Nihat Tarlan Cad.	13:00 - 21:00	8,69	10,35	0,84	11,62	10,35	1,31	1,19
	2. Phase	Ali Nihat Tarlan Cad. side road	13:00 - 21:00	1,17	0,91	1,29	4,6	5,85	3,92	5,02
	3. Phase	Mehmet Şevki Paşa Cad. one direction	13:00 - 21:00	3,52	3,68	0,96	1,86	1,21	0,53	0,35

Figure 5 The performance of the Signalized System at Intersection Bostanci

IV. CONCLUSION

Regardless of the traffic conditions at the crossroads, careless time assignments can cause vehicles in both the intersection approach arms and the central island to lose a lot of time at the intersection. This increases the average vehicle delay at the junction and decreases the junction capacity leading to blockage of the traffic in the junction after a while. One of the most important issues related to the design and operation of signaled intersections is the assignment of the signaling times of the signaling system at the intersections [15], [16], [20].

When we pay attention to the results according to the queuing model that we referenced it is seen that the service time is longer in some cores than the arrival times. This shows the result of traffic jamming at intersection links and the traffic lights are not sufficiently optimized. A poorly planned signalized system increases the number of vehicles waiting in the queue. In order to avoid this we need to adjust the phase times of the denser arms more properly and take into account the other arms as well[18].

We are working on the mathematical model that will optimally determine the signal periods of the intersections. In particular, we plan to dynamically adjust both phase durations and phase sequences, and we intend to benefit from heuristic algorithm methods. In summary, we plan to further optimize the system by taking into consideration the data and results obtained up to now in the next stage, and plan to finalize the improvement at the intersection with analytical model and simulation.

REFERENCES

[1] Polat, Y., Kuloglu, S.D. 2002. Mikrodenetleyici Kontrollü Dısarıdan Süre Ayarlı Sabit Zamanlı iki Fazlı Trafik Sinyalizasyonu. Lisans Tezi, SDÜ. Teknik Egitim Fakültesi, Isparta, 22s.

[2] Pappis, C.P.,Mamdani, E.H., "A Fuzzy Logic Controller for a Traffic Junction", IEEE Transactions on systems, Man and Cybernetics, , 707-717, (1977)

[3] Tzes A., McShane and Kim, S., "Expert Fuzzy Logic Traffic Signal Control for Transportation Networks", Instute of Transportation Engineers 65th Annual Meeting, Denver USA, 154-158, (1995).

[4] Kim, Jongwan, "A Fuzzy Logic Control Simulator for Adaptive Traffic Management", Proc IEEE International Conference on Fuzzy Systems, 1519-1524, (1997).

[5] Niittymaki, Jarkko, P., "Isolated Traffic Signals- Vehicle Dynamics and Fuzzy Control", Ph.D. Thesis, Helsinki University of Technology, Civil and Environmental Engineering, (1997).

[6] Hoyer, R., Jumar, U., " Fuzzy Control Traffic Control of Traffic Lights", Proc. IEEE International Conference on Fuzzy Systems, 1526-1531, (1994).

[7] Zhen-Jin Huang, Chun-Gui Li, Zeng-Fang Zhang, "Traffic signal control based on genetic neural network algorithm", 2009 March

[8] Kitae Jang, Hyungjoo Kim ve Gwun Jang, "Traffic Signal Optimization for Oversaturated Urban Networks: Queue Growth Equalization", August 2015

[9] Ming Kang Shi, Hong Jiang ve Song-Huan-Li, "An Intelligent Traffic-Flow-Based Real-time Vehicles Scheduling Algorithm at Intersection", 2016

[10] Babicheva, "The use of queuing theory at research and optimization of traffic on the signal-controlled road intersections", 2015

[11] Martin Anokye, A.R. Abdul-Aziz, Kwame Annin, "Application of Queuing Theory to Vehicular Traffic at Signalized Intersection in Kumasi-Ashanti Region, 2013 Ghana

[12] M. TİMOR, "Yöneylem Araştırması ve İşletmecilik Uygulamaları", İstanbul Üniversitesi Basımevi, İstanbul-2001, S: 434

[13] Elements of Queuing Systems, http://staff.um.edu.mt/jskl1/simweb/intro.htm

[14] Y. Liu, K. Lee, Modeling Signalized Intersection Using Queueing Theory, Department of Electronic and Computer Eng, 2009

[15] Yetis Zayi Murat, "A new approach for fuzzy traffic signal control", July 2002

[16] Y. Zayi Murat, M. Başkan, "İzole Sinyalize Kavşaklardaki Ortalama Taşıt Gecikmelerinin Yapay Sinir Ağları ile Modellenmesi", January 2006

[17] V. Kostakos, "Traffic in the smart city: Exploring the potential of city-wide sensing to augment a traffic control center", 2013

[18] Ming Kang Shi, Hong Jiang ve Song-Huan-Li, "An Intelligent Traffic-Flow-Based Real-time Vehicles Scheduling Algorithm at Intersection", 2016

[19] Ms. Avani.Joshi, Dr. Dhirendra Mishra, "Review of Traffic Denstiy Analysis Techniques", July 2015

[20] Gurnoor Walia, Kuljit Kaur,"A Review on Intelligent Public Transport System for Smart City",2015

[21] Akıllı Ulaşım Sistemleri İSBAK Raporu, 2016

Real-Time Monitoring and Control of The SDLC Process on a Single Automation in Core Banking Applications

Ahmet Sefa OZTAS[1], Eren YEMEN[2], Ercan TUZUN[3]

[1] Vakıf Participation Bank Inc. R&D Center, İstanbul/Turkey, ahmetsefa.oztas@vakifkatilim.com.tr
[2] Vakıf Participation Bank Inc. R&D Center, İstanbul/Turkey, eren.yemen@vakifkatilim.com.tr
[3] Vakıf Participation Bank Inc. R&D Center, İstanbul/Turkey, ercan.tuzun@vakifkatilim.com.tr

Abstract **- In core banking applications, delivering innovative ideas in a fast and reliable manner is extremely important for any corporation or organization. In the existing scenario, the result of increased competition between the companies and products, the core banking industry needs to better respond to dynamic market requirements, by offering new initiatives and services to better the user experience. This process needs to be minimal and with customer support interaction with technology in mobile and internet banking. In the past few years of financial industry, engineering education and knowledge are easy, intuitive for the user and provide the perfect customer experience. In the digital banking platform, the cutting-edge technology innovations and financial user experience designs create DevOps culture.**

DevOps, which is a conceptual framework for the re-integration of development and operations in Information Systems, extends the agile methodology to create applications and deliver them across the environment in a fast and automated manner to improve the performance and the quality assurance. The approaches of Continuous Integration (CI) and Continuous Delivery (CD) are crucial for the practices of application development and release management. The objective of this paper is to create a proof of concept for management of systems that have a critical importance; quick and continuous copying of application files and database objects in a single package between Development, Test, Pre-Production and Production environments, integration with Load Balancer Systems, integration with request and call management applications, integration with source code management platforms, implementation of static code analysis and code review process, integration with automatic antivirus scan before testing, with only one automation for testing composed with built result application files for the creation automatic test process and definitions and only one platform structure that can make a new version deployment a release to many application architectures.

Keywords **- Continuous Integration, Continuous Testing, Continuous Delivery, DevOps**

I. INTRODUCTION

The management of complex systems must be continuous and uninterrupted in large-scale companies in the main banking industry. Although the primary target is to deliver innovative ideas quickly and safely to bank's customers, the product should respond better to market needs. All companies that develop their software in-house or use platform, they want to follow the daily practice of users during the operation and provide them with more effective software. In software applications, when the idea is created; Dev, Test, PreProd and Prod environments follows planning, production and using processes, which is defined as Software Development Life Cycle (SDLC). This process allows the software team return to any progress in the cycle and resume [1, 2]. A professional development of the SDLC process, puts this process on a corporate model. If this cycle is following by an automation software, to improve performance and quality assurance the company must have an architecture which quickly develop applications and automatically deliver them [3, 4].

To date, different software development process models have been created to produce quality and mature software efficiently. They focus on organizing the sequence of progresses to provide time and cost criteria on the generated models. However, when looking at the failure rates in software projects, it is seen that the desired result cannot be obtained. In A Replicated Survey of IT Software Projects Failuries named study published by El Amam and Koru, the real success of rate of software project is %37,3 [5]. The ratio is very low. The main reason for this situation is that multiple roles and responsibilities are needed to operate the SDLC process. There are many positions such as project managers, system analysts, business analysts, customers and software engineers. Efficient and effective communication of these people is the source of a successful project.

Technologies, requirements, ideas, tools and timelines are constantly changing. To be successful, the information must be accurate, easily found and delivered to all members in real time. Modern software development processes consists the basic concepts of adaptability to change data based project management [6]. In order to ensure real-time monitoring and control of the progress in the software development process by users, an automation fully integrated into SDLC processes and environments must be created.

The process basically works between Development, Test, Pre-Production and Production environments. The development environment is where the software engineer performs the development and testing processes and also the system analyst tests are performed. In the test environment, test and quality control engineers determine and fix system bugs. The end user tests the system with near-truth data in PreProd

environment and gives approval. After this approval, parts of the software are available in Prod environment. The main function of the architect is to ensure that continuous and uninterrupted dissemination of a developing application or a new product is safely and without a problem. Development system automation should ensure that software components are passed without errors and risk. In addition, if there is an error, with backup mechanisms to restore the system to previous versions should ensure uninterrupted operation of the system. With automation, the system can be managed and monitored.

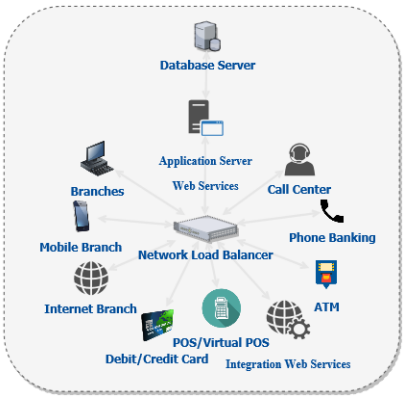


Figure 1: Application Deployment Architecture

This article describes a single platform structure that can be fast, uninterrupted, backed up and instantly intervened with a single automation to service channels (Web service applications, web applications, mobile applications, call center, internet branch, ATM, card systems, branch applications) for layered architectures in the main baking applications of financial institutions (Figure 1).

II. Modern Software Development Architecture

Modern software development architecture includes Development, Test, Pre-Production and Production environments. Application and database objects can be transferred between these environments in a single package quickly and uninterrupted through automation.

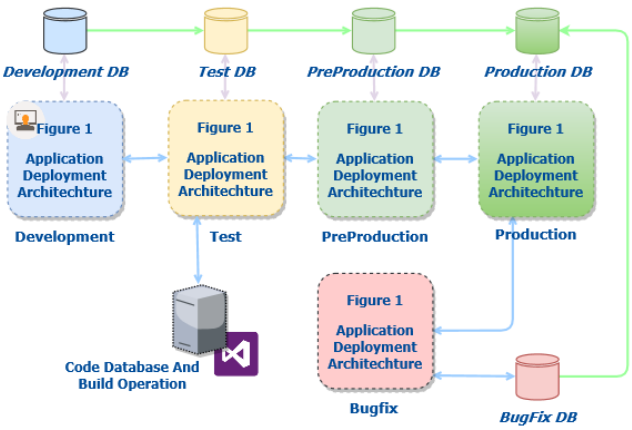


Figure 2: Software Development Life Cycle Environments

The development environment is where the software developer begins to develop software. New applications to be added to the banking system and changes to be made to the system are made in development environment. At the same time, software developers test their developed software in this environment. During testing and analysis, the bugs are resolved apart from development. Fix the bugs should be done in the development environment, after the test, it must be transferred to the test environment. Developed software components are transferred from the development environment to the test environment when they are ready to be tested. Test engineers and analysts perform their tests in this environment. The bugs and deficiencies detected in the system are reported to the software developer. No changes should be made by directly interfering with test environment codes. Software components transferred to the test environment pass through the approval mechanism before being transferred to the PreProd environment. The end user makes final controls with near-truth data before the Production environment.

Software components that pass the end user test successfully are transferred to Production environment. Errors in this environment can cause system interruptions. Testing in the prep environment is important for the speed of process. After that, it is desired that dissemination should be uninterrupted and continuous. BackUp & Recovery possibility have been added to automation for safe recovering and continuous system. The system performs two types of back up operations: full back up and default. In case of system error, the system recover and continues to work uninterrupted.

Dissemination is the process of transfering developed software components to environments. Alternative applications can be used for deployment as well as these processes can be done manually. However, Companies with large system architecture should perform this complex process with an application and follow the process. The methods of dissemination are different for each environments. The operation of dissemination to an environment without control is performed by the tasks. (one per hour or every two hours) The main structure of software development environments used in automation is shown in Figure 2. There are differences in the banking sector in the dissemination to other sectorrs.

The application code that is tested in the source code pool development environment must be moved into the source code of the test environment. Code moved to the test environment must be transferred to client and application servers after compilation and deployment. Database objects with application codes must also be transferred from the development database environment to the test database environment in the same way. This method should also be used in transfer to PreProd environment. When transferring the Prod environment the method for database objects same but different for applications objects. In Prod environment transitions, Application objects running before Prod must be transferred to the Prod environment with changes and version controls.

III. Change Management Automation System

Software development process stages must be regular. The creation of a request for the application to be developed is the firs phase of the process. Project management tool helps the

developer create demand by business departments. İssue is opened via project end demand management system when error is encountered. All system works on issue concepts. When the issues review and fixed, it is closed over the system. It is reported that the error was detected and fixed. Enables professional follow-up of errors transferred by the test engineer to the software engineer during software development. The test user can add the importance and urgency order of the situation in the error issue. Who is responsible for the issue can be added. Gives advantages in solution processes in complex projects. Deployment automation works integrated with project management tool. Actions from development to production can be triggered through project management tool.

After the development phase is completed, The developed software components must be transfer between environments. It is great need to use an application to follow the system because it is complex. So we should use automation system, system provides perfect progress of the process. This system basically performs the task of transferring software components between SDLC environments.

Developments are made with a request. Therefore, software objects that have completed their development can be converted into a package associated with a change request. Transfer of the generated package to the environments must be with a confirmation mechanism which includes managers. Application transfer to Prod environment is subject to authorizations. Therefore, packages that complete the flow must be transferred to Prod environment under a version management department control.

A. Deployment Packages

Resource code pool needed to work together in the software development lifecycle. This application can be open source or professional. All software engineers, system analysts nd other required persons within the organization can be able to share code. History of code changes should be saved. There must create environment code to hold environment codes and can be code merge to switch between environments. This tool helps keep source codes in one place and allows software developers to develop on common files. The outputs generated by compiling these source code is called the application object. These objects creates part of version package. The last state of the application object of the requested code component or of a special version can transfer between environments with this automation system. Application objects can be list in automation interface and requested application objects can be added to the dissemination package.

Database objects can be add to the package with an easy to use interface. Stored procedure, table, function and database queries do not belong to a group that van be add to a package as an update object. These objects are not affect by changes made to the database after they are added to the package. (If there is no update for objects)

A lock mechanism is used in packets in order to prevent the change moment coinciding with the approval moment of the packets sent to the approval. The object in a package can be locked for non-use in another package, because the same object in two packages can be transfer between the environments to prevent problems in the event of trying to move. It takes time for a package to transfer Production environment. A package can be managed by more than one user because of when package owner can not be reached, other user manages the package. But only one user can work on package at the same time.

Automation in which change request are kept in the systems should be considered different from these systems. This system contains only the contents of the software exchange. Emergency switch feature has been added to automation system to be used at the required times. This option is used for transfer package between environments quickly in emergency.

Database objects must be created regularly. Automation system controls compliance with this order. For example, creating a function about a table without creating table and index structure will result in error. Therefore, the work order feature is used in the structure of dissemination packages and the working order of the database objects can be determined when preparing the package.

B. Users Role Description and Approval Mechanism

Required roles for managing package migrations between environments are Software Engineer, Configuration & Version Specialist, Service Manager and Database Admin. Software engineer creates software change package, starts migration between environments, follows migration to Production environment, restarts the process in case of change. Approvals and controls are made by authorised users during the migration of package.

Users must assign a role according to their authority. The authorizations should be determined according to the project's needs and the progress of the system. For example, the applications and objects belonging to database layer in a deploymant package must go through the control of an administrator. In this direction, there are two basic authorizations: Application object authorizaion and Database objecy authorization.

IV. Release And Deployment Management System

Customer, account, transaction and system configuration information of the Bank is kept in banking systems. Difference from other systems; process must take place very quickly, system and service continuity is close to % 100, data consistency, system and data security is high level. Another difference is that banking services can be made through distribution channels such as ATM, internet and mobile banking. Continuity and security needs are important with the widespread use of these distribution channels. In general, the average continuity rate target of banks with critical financial systems is 99,99% [7]. This rate means 52.56 minutes' interruption per year. A lower level is %99,9 and this rate means 8.76-hour interruption per year. This time is very high for the banking sector [8]. For this reason, banks want to establish hardware and software systems that will keep the interruption amounts to a minimum.

Automation to be used in the banking system should be an application that complies with the procedures under many security measures and allows continuous migration without interruption of the Production environment. By creating a

between environment flow chart with user role definitions, project request is created in project management application. Software engineer develops the request in development environment. Testing after completion of development. İf tests are successful, the project request is sent to the PreProd environment and the request status is updated. Then, the request Service Manager's approval should be expected and after approval, the new status should be determined according to the structure of the objects. At the same time, the package is prepared to be migrated to Prod environment. If the package contains a database object, it must be first be approved by the database administrator. The new status of the package must be updated.

The automation system can migrate one or more version packages to production environments. The system performs channel-based parsing of version packages through its decision mechanisms. At the same time, it prioritizes among the packages by determining the criticality levels of the application and database objects contained in the version packages with specific collections. Therefore, the interruption times are too small to be tried. The production environment contains a real and continuous system. Because they can not be intervened in a working system, these transitions must be made at certain time intervals.

If the package is urgent state, with the approval of the service manager and database administrator, it can migrate directly to the Production environment. The process is completed as a result of the database specialist approving the package and physically run the package objects in a Production environment.

The newest version of migrations to automation system varies according to application types. General application types: Desktop applications, Windows services, new version migrations of version files for ATM machines, new version migrations for WCF services and web services. Migrations occur from simple copy logic. Version migrations are made in accordance with certain criteria, such as file extension, file size end code comparison, as a result of environment synchronization of environment DLL.

Version migrations for web based application type are added to the automation system. The output of the code contained in the application type differs from other application types. File extensions, such ass .resx, .dll, .aspx, .cab, .js, .cshtml and .asax. These files made by the software developer by entering the package number with the package change request. When the migrate to the Production environment, the server is removed from Load Balance system and is not affected system. After the Application Pool is up, include the system. Prevents interruption of the systems. This process is the same for all migration types.

ATM systems are an application for customers banking operations. The system must be running continuously and should be updated with new technologically ideas. Such as credit card transactions, money transfers and cash deposit performs from ATM's and there must no interruption in this application where cash transactions take place. Therefore, the version file is created. After the current version file is transferred to the ATM machine, checks every 15 minutes for a new version file. When the system process are finished, it starts updating and remains Temporarily out of service. This type of updates are made when there is no intensity.

The background has been positioned as a windows service in the core banking system, some of the batch transactions that have been working periodically for a certain period of time, working periodically. Depending on the service intervals and criticality levels, upgrading should be done carefully and without interruption. During the transition windows services are stopped and wait for a while, then the process ends. Once the automation system has passed, it checks to see if the service is running. If the service receives an error, it will send a mail to the user and trigger the service to restart.

Desktop applications are used for core banking operations at branch and general directorate locations. Client application objects for implementation are downloaded to user computers and distributed. When distributing each user separately, distribution is carried out with the help of agents employees working on the user machines via the branches and the headquarters servers located together. Transition operations for the servers located in the bank branches do not interrupt the system at any time of the day.

The application layer is grouped on a channel basis and integrated with load balancer systems. Since the automation system is also integrated with load balancer systems, during the transition, the load on the servers is reset to load balancer and the transition process is performed. The server is included in the load balancer-site by performing the necessary tests by automation.

The average size of the packets transferred to the branch servers is determined as 350 application object and 200 mb packet size. As the bank's transaction volume and the number of products increase, the application object numbers and package sizes will increase proportionally. As the number of branches and users increases, delivery times to server and user machines of client application objects will be extended due to limited network lines. In order to be able to bypass this situation, version packages are passed through compression algorithms and divided into 10 equal parts and copied to the branch servers at the same time in parallel. When the copy operation is completed, the packages are reassembled in the branch servers. This process is done automatically by algorithms used in automation. A 95% gain in copying times was achieved. In the following years, the increase of branch and user numbers prevented the problem in the automation system. At the same time, as the distribution periods have shortened, the operational loads have decreased considerably. Due to the fast transitions, the changes have been transferred and used in very short time to live environments.

V. DISCUSSION

The applications are working within the necessary security mechanism. The certification process, which allows applications to run, is built on the server where the application is located. Certification authorities should be able to easily view the usage times and certificate location information through the automation system.

Production environment servers have a minimal impact on version transitions since the banking system has working

environments, and even more important is the fact that there is no security vulnerability at all. Packages should be scanned through an antivirus before system architecture passes. Secure package distribution should be included in the automation system.

VI. CONCLUSION

Designing efficient and effective software development processes is a challenging problem, and solutions to the variety of risks to software project success are a widely debated topic. A recent movement in the software development industry attempts to bring operations and development teams into closer collaboration by relying on automated systems as the interface for project artifacts and knowledge between these teams. This process aims to ensure a smooth transition of developed software to quality assurance and maintenance teams. Though this process improvement has proven beneficial to teams in achieving continuous deployment strategies the authors also propose that the fundamental principles of task automation knowledge transfer can be applied to internal communications and incremental deliverables, enabling near real-time communication at all phases of the SDLC. A package in a condition of pre-production transitions 7 different application types continuous and uninterrupted by application administrator. The package history report of the deployment should be stored. Back up and recovery is performed via automation. In this way, during an incorrect migration, the application can be restored to the old version. Load balancer control can be performed via the system so that there is no interruption during the deployment to the servers. CPU and RAM information of application servers, hard disk capacity, application pool status, load balance status and service status are displayed. In addition, the necessary reports are displayed on the system in case of a possible error or supervision for examining the software changes being made. Versions are made for different application types. At the same time, integration with IIS is provided for each server in the automation. Integration of any application type into the system can be easily accomplished since the automation system is based on definition. In the core banking applications, real-time monitoring and control of the SDLC process on a single automation has been achieved due to this automation.

VII. REFERENCES

[1] W. W. Royce, "Managing the Development of Large Software Systems: Concepts and Techniques", into *Proceedings of the 9th International Conference on Software Engineering*, Los Alamitos, CA, USA, 1987, pages. 328–338.

[2] T. E. Bell ve T. A. Thayer, "Software Requirements: Are They Really a Problem?", into *Proceedings of the 2Nd International Conference on Software Engineering*, Los Alamitos, CA, USA, 1976, pages. 61–68.

[3] "DevOps: Born in the Cloud and Coming to the Enterprise". [Online]. Available at: https://www.gartner.com/doc/1458129/devops-born-cloud-coming-enterprise. [Access: 30-March-2018].

[4] A. Schaefer, M. Reichenbach, ve D. Fey, "Continuous Integration and Automation for Devops", into *IAENG Transactions on Engineering Technologies*, Springer, Dordrecht, 2013, pages. 345–358.

[5] K. E. Emam ve A. G. Koru, "A Replicated Survey of IT Software Project Failures", *IEEE Softw.*, c. 25, number 5, ss. 84–90, Sept. 2008.

[6] C. A. Cois, J. Yankel, ve A. Connell, "Modern DevOps: Optimizing software development through effective system interactions", into *2014 IEEE International Professional Communication Conference (IPCC)*, 2014, pages. 1–7.

[7] N. Bajgoric, "Server operating environment for business continuance: framework for selection", *Int. J. Bus. Contin. Risk Manag.*, c. 1, number 4, pages. 317–338, Jan. 2010.

[8] S. Pandey ve S. Nepal, "Modeling Availability in Clouds for Mobile Computing", into *2012 IEEE First International Conference on Mobile Services*, 2012, pages. 80–87.

Forensic Analysis of Persistent Data Storages Analyzing NTFS Formatted Drives

H. G. OZGUR[1] and S. SAHIN[1]

[1]Izmir Institute of Technology, Izmir/Turkey, huseyinozgur@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, serapsahin@iyte.edu.tr

***Abstract* - Computational forensic is a far-reaching field for criminal and civil laws in this age. Aim of this project is to develop an educational tool to analyze persistent storage devices to find possible evidences. Evidences are found as metadata information on these drives. These metadata information are collected in a comma separated value-CSV format document. This CSV file can be imported to a database and can be easily analyzed. By the classical forensic investigator tools mostly can analyze only one media at a time, but by this way more than one digital evidence can be comparatively analyzed. Also, deleted files may hold evidences, so recovering of deleted files is an additional topic for this study. The structured codes and related documents have shared on github as an open project; to give a chance to extension of this study by new projects and we hope that the product of this study can be useful tool for computational forensic courses.**

***Keywords* - Computational forensic, Hard-drive Analysis, Metadata, Recovering Files Deleted**

I. INTRODUCTION

The digital age defines a new virtual world which includes many type of activities, operations and relations among many type of individuals. The individuals quite likely use digital devices for some reason. These devices can be selected as target or source for the criminal activities and they must be defined and verified by digital evidences according to laws and regulations. Digital forensic is a very specific area which focuses on required specific analyzing tool and their usage principles and methodologies to discover evidences. The digital evidence or electronic evidence; according to Eochan [6] can be any information stored or transmitted in digital form which can be used as a proof material to a court case. The very important point is that the authenticity of the digital evidence must be protected and should be provable.

If an investigator wants to find evidences which are in digital form, she starts with analyzing of hard-drives, because digital evidences are most commonly held in static storage units. Analyzing of static storage units such as hard drives, memory sticks etc. is basis of digital forensic. In this project, a tool is developed to analyze the persistent data storages, and first focus is to find all patterns which may constitute evidence for a case. Even evidences are deleted, or corrupted, or modified, they usually leave a trace on these units and its file system structured records. Due that reason the second focus of this project is to recover of deleted files in storage unit.

Analyzing hard-disks to search evidences is required to have technical and conceptual understanding. The layers of a storage device from application to its hardware must be known. A hard-drive abstraction involves four parts which are physical layer, partitions on device, used file system, and its semantic layer. Physical layer reflects device itself and it is split into one or more partition. Each partition needs a bridge (abstraction layers to have a connection) between files on system hardware and user's application. File system provides this bridge. Semantic layer has different design structures according to different file systems.

The following Figure 1 shows overall steps of finding evidences on hard-drives:

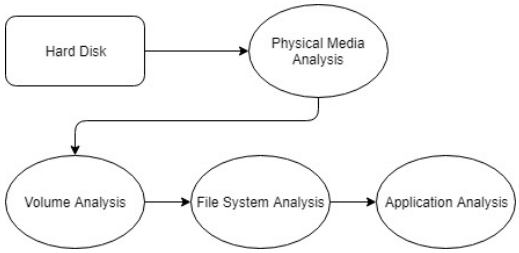


Figure 1 Steps of finding evidence on hard-drives [1]

- In the physical media analysis, the locations of the disk partitions are determined, and file formats of the partitions are detected.
- Each file system has its own structure. For NTFS formatted drives, beginning of the master file table (MFT) records is in the volume analysis part, and some information about the volume are obtained, then sector by sector analysis is started.
- In the file system analysis part, metadata in MFT record are extracted. All records of MFT as metadata are written on comma separated value - CSV document. In this project, also deleted files are recovered. Then, each file is analyzed in analysis phase.

The second chapter of this article includes the background information about the physical media analysis methodology. Third chapter explains the volume analysis of NTFS structure. Fourth chapter presents the developed solution tool for file system analysis over previous steps. At last the conclusion chapter explains the success results of this study and future works to enlarge the efficiency domain of this tool.

II. PHYSICAL MEDIA ANALYSIS

At the beginning of the hard-drive analysis, master boot record (MBR) is analyzed. MBR is located always in the first sector of booting device. Hard-drive partition information starts at the offset 0x1BE in MBR table and its size is 16 bytes. There are four partition table entries in MBR. Last two bytes of MBR sector are 0x55 and 0xAA.

Structure of the partition table entry explained in [2], [3], here:

* Important bytes of MBR for forensic analysis

- Byte 1*: 0x80 → active(bootable), 0x00 → inactive (not bootable)
- Byte 2, 3, 4: Location of the beginning of the partition
- Byte 5*: Type of the partition (e.g.0x07 is NTFS)
- Byte 6, 7, 8: Location of the end of the partition
- Bytes 9, 10, 11, 12*: Logical block addressing of start sector in little endian format, and it is measured in blocks
- Bytes 13, 14, 15, 16: Number of sectors in the partition

III. VOLUME ANALYSIS

A. *NTFS Structure*

When a hard-drive is formatted as NTFS, the space which involves the first 16 sectors is known as partition boot sector ($Boot) which includes boot metadata file. It includes bootstrap code at the first 15 sectors, and then continues with initial program loader (IPL) of the boot sector. Also, it contains a copy of boot sector at the last sector to make system more reliable.

The Master File Table (MFT) holds records about each file and directory on the NTFS formatted hard drive. There are 2 mirrors of MFT, if one of them is damaged, then the other one is used to recover. The location of Master File Table in a NTFS volume is not in the predefined clusters. In the boot sector, there are two fields which are BIOS parameter block - BPB and Extended BPB [7]. These fields describe the location information of MFT. In this way, if there is a problem occurred in MFT's own sector, operation can be transferred to the mirror of MFT.

NTFS Boot Sector	Master File Table	File System Data	Master File Table Copy

Figure 3 Structure of NTFS drives

The above Figure 3 indicates the structure of NTFS formatted drive. And following table 1 shows the layout of MBR. Boot sector begins with jump instruction which is EB 52 90. Then it is followed by Original Equipment Manufacturer - OEM ID which is used to identify the name and version number of the operating system that formatted the volume.

Table 1 Structure of Master Boot Record (MBR) [4]

Offset	**Length of Field**	**Meaning**
0x00	3 Bytes	Jump Instruction
0x03	8 Bytes	OEM ID
0x0B	25 Bytes	BPB
0x24	48 Bytes	Extended BPB
0x54	426 Bytes	Bootstrap Code
0x1FE	2 Bytes	End of the Sector Marker

IV. FILE SYSTEM ANALYSIS

A. *NTFS Master File Table Analysis*

MFT table stores a record for each file or folder which is on an NTFS volume [8]. Also, MFT stores records about itself. Each record is 1KB.

These records consist of a header and a few attributes as you can see in Figure 5.

MFT Entry Header	Attribute	Attribute	Attribute	Unused space

Figure 5 Structure of an MFT record

The following table 2 represents the layout of an entry of MFT. Some important metadata given in this table are:

- Magic number indicates the beginning of the MFT record. Its value is always "46 49 4C 46" which is represented in hexadecimal form.
- LSN changes when the record is modified.
- Sequence number indicates that how many times this MFT record has been reused.
- Hard link count indicates the number of entries that is referencing this record.

Table 2 Layout of an MFT record

Offset	**Size**	**Description**
0x00 – 0x03	4	Magic Number: 'FILE'
0x04 – 0x05	2	Offset to the update sequence
0x06 – 0x07	2	Number of entries in fixup array
0x08 – 0x0f	8	$LogFile Sequence Number(LSN)
0x10 – 0x11	2	Sequence Number
0x12 – 0x13	2	Hard Link Count
0x14 – 0x15	2	Offset to first attribute
0x16 – 0x17	2	Flags: 0x00: Unallocated space, free to be over written, 0x01: record in use, 0x02: directory
0x18 – 0x1b	4	Used size of MFT entry
0x1c – 0x1f	4	Allocated size of MFT entry
0x20 – 0x27	8	File reference to the base FILE record
0x28 – 0x29	2	Next attribute ID
0x30 – 0x1000		Attributes and fixup value

1) *MFT Attributes*

Table 3 shows possible attributes which can be included in an MFT record. Each attribute has different structure and consist of different metadata about the file.

Table 3 MFT record attributes

Type	OS	Name
0x10		$STANDARD_INFORMATION
0x20		$ATTRIBUTE_LIST
0x30		$FILE_NAME
0x40	NT	$VOLUME_VERSION
0x40	2K	$OBJECT_ID
0x50		$SECURITY_DESCRIPTOR
0x60		$VOLUME_NAME
0x70		$VOLUME_INFORMATION
0x80		$DATA
0x90		$INDEX_ROOT
0xA0		$INDEX_ALLOCATION
0xB0		$BITMAP
0xC0	NT	$SYMBOLIC_LINK
0xC0	2K	$REPARSE_POINT
0xD0		$EA_INFORMATION
0xE0		$EA
0xF0	NT	$PROPERTY_SET
0x100	2K	$LOGGED_UTILITY_STREAM

For instance, the structure of $STANDARD_INFORMATION attribute is presented in table 4:

Table 4 The structure of $STANDARD_INFORMATION attribute

Offset	Size	Description
~	~	Standard Attribute Header
0x00	8	C Time – File Creation
0x08	8	A Time – File Altered
0x10	8	M Time – MFT Changed
0x18	8	R Time – File Read
0x20	4	**DOS File Permissions**
0x24	4	Maximum Number of Versions
0x28	4	Version Number
0x2C	4	Class ID

All the MFT attribute structures can be found in [9].

The other important point is about the size of the file content which is addressed by MFT record. The files are either resident or nonresident according to its size. When a file content can fit within the same cluster of MFT file record, it is called resident attribute. Otherwise, it is called non-resident attribute and it can spread on different clusters.

For example, $DATA attribute contains the either location of the content of file or content of the file itself. If it is non-resident, then content of the file is stored in different location in the disk. There are different addressing methods for fragmented, un-fragmented, and compressed files. To obtain file content, data-run is used to address file content. The reader can reach more detailed information on data-run from [5].

Almost all attributes start with Standard Attribute Header. This header includes important information about attributes. You can see the layout of the header in the table 5 given below:

Table 5 Standard Attribute Header

Offset	Size	Description
0x00	4	Attribute Type
0x04	4	Length (including header)
0x08	1	Non-resident flag
0x09	1	Name length
0x0A	2	Offset to the Name
0x0C	2	Flags
0x0E	2	Attribute ID
0x10	4	Length of the Attribute
0x14	2	Offset to the Attribute
0x16	1	Indexed flag
0x17	1	Padding
0x018	2N	The Attribute's Name
2N+0x18	L	The Attribute

Some important metadata which part of the header are:

- Attribute Type: For example, if its value is 0x10, then it refers to $Standard_Information attribute.
- Length: It gives us total length of attribute.
- Non-resident flag: If its value is 0x00, then given attribute is resident. If it is 0x01, then the attribute is non-resident. MFT record size is fit for resident attributes. If size is not enough, then the attribute is located somewhere else in the volume.
- Offset to the Attribute: This indicates the location of the attribute.

2) Forensics Tool Developed in Scope of the Project

Forensic tool development for persistent data storages starts with reading the first sector of the physical media. In that step, program checks whether MBR (Master Boot Record) is included or not. If MBR is included, then program looks for NTFS formatted partition information in MBR, in that way, location of NTFS boot sector is found, and analysis continues in that location. If MBR is not included, then program checks whether the first sector is NTFS boot sector or not. If it is NTFS boot sector, analysis continues in the first sector.

NTFS boot sector includes some metadata about the volume. These metadata are extracted to a CSV file. Also, NTFS boot sector is used to locate beginning address of the MFT (Master File Table). Program continues its execution in at that location.

Most of the work is done in MFT records. Each MFT record is analyzed to obtain metadata of each file or folder which is stored in the storage. Some Metadata of each record are extracted to a CSV file. Each line of the CSV file includes metadata of a folder or a file. One of the metadata indicates whether the file is deleted or not. If the file deleted, and there is not an overwrite on its location, then program tries to recover the deleted file. But the recovery process is implemented for the file contents which have been stored in continuous clusters.

3) Similar Products and Novelty of the Project

There are several similar products. But most of them are not open source. To make an expressive comparison, some open source products has been selected.

This project can be compared with "Analyze MFT[1]" solution. The software projects can be compared according to the principles of software engineering such as documentation, learnability, changeability, interoperability etc. Table 7 shows the differences when these two projects compare with these principles.

Table 7 Differences of Analyze MFT, and the Project

	Comparison of Analyze MFT and Our Project
Documentation	Detailed documents are provided by our project but Analyze MFT has not been well documented. It just includes how a user can implement it on her system.
Learnability	Our project comes with both detailed documentation, and well commented codes. Learning it is easy for both programmers, and Investigator.
Changeability	Any programmer can change our code after reading its documents and reading its code easily. Making any change on Analyze MFT is hard, because learning it takes time.
Evolvability	Evolvability takes time for both applications. Both analyze NTFS formatted drives. Expanding their scope requires analyzing structure of different file formats. Pros of our program that it provides sector reader class, and csv documentation class. These two classes can be used to analyze any other file formats. Also, after reading documentation, it gives an idea about the processes to analyze a disk.
Interoperability	Our application comes with some features which are sector reading, recovering deleted files, csv documenting. These features can be used separately inside any other application, and the other applications can use our program to get metadata of MFT records. Analyze MFT can be used CSV documentation, anomaly detection etc. inside an application.

Table 7 has been created according to our usage experiences of Analyze MFT.

The Analyze MFT which has been developed very similar in the scope of our project. But, our application also provides recovering of deleted files. Codes of MFT project has not been commented well, so it is hard-to understand the code. But, almost each line of our code includes an explanation part. The Analyze MFT has not enough documentations and our project is published with its documents as well.

There are also some other applications like Autopsy, but they have been developed as open product for end users, and they include different functionalities. For example, Autopsy can be used for hash filtering, web artifacts extraction etc. It is provided with usage training, and commercial support for forensics analysts. Our project can be used by developers and forensics end users, but main purpose is that users can easily understand the processes of analyzing a digital source. After understanding, anyone can develop more comprehensive forensic analysis environment with our project.

4) Improvable Aspects and Future Work

Improvable aspects of the project:

- Recovering fragmented files
- Different kind of MFT attributes can be analyzed in detail, for example creation times or read times of files can be exported and relation between them can be observed
- CSV documentation content can be extended
- An GUI can be designed
- An option to recover deleted data files can be added
- An option to view desired sectors as hexadecimal can be added

Future work:

- Different file systems like FAT, ext2, ext3, ext4 etc. can be analyzed
- Non-persistent storages can be analyzed, different requirements and procedures are required to collect evidence from a volatile memory

V. CONCLUSION

Scope of the analyzing persistent data storages is a beginning objective on NTFS file systems, but it could involve a wide range of file formats. Each file format has its own structure. Finding documentation about any kind of file format can be hard. Most of the documentations include only limited part of the structure of the desired file format.

Values of the project:

- Analyzing storages to find evidences is a sensitive process. First, the evidence properties of the collected data should be provable. Second, the secrecy or privacy issue of analyzed data should be guaranteed. Data can be belonging to government or persons. Due this reason, forensic analysis job is mostly done in national boundaries by governmental agencies. This project can be starting point to develop national forensic analysis tool.
- The current forensics tools are imported, and this causes high costs to buy product and its training for users. Developing national product cuts off these costs.

[1] https://pypi.python.org/pypi/analyzeMFT/2.0.19

- As a nation we would like to have more forensic professionals. Some of the universities and computer engineering departments include related courses. This national tool can be used training purposes in lectures and used as support laboratory material.

- This open tool can be enhanced by anyone to develop different kind of forensic facilities.

The most valuable point in this project is to provide environment for investigators to create associations inside different kind of collected data. Because the developed program gives a CSV file as an output. This CSV file can be imported to a relational database, and investigators can create necessary queries to see relations among different metadata records on single or many of different storage units.

All in all, forensic science is the topic which certainly must be worked on it. The necessary lectures are given by different educational organizations. The purpose of this project is to create a lecture material and as a future objective to start an open project to develop a national free of charge forensic storage analyzing tool.

VI. REFERENCES

[1] File System Forensic Analysis - By Brian Carrier

[2] http://www.bydavy.com/2012/01/lets-decrypt-a-master-boot-record/

[3] https://www.youtube.com/watch?v=BG1gQ4Ta79M

[4] http://www.cse.scu.edu/~tschwarz/coen252_07Fall/Lectures/NTFS.html

[5] http://homepage.cs.uri.edu/~thenry/csc487/video/66_NTFS_Data_Runs.pdf

[6] Casey, Eoghan (2004). Digital Evidence and Computer Crime, Second Edition. Elsevier. ISBN0-12-163104-4

[7] Hardware White Paper, FAT32 File System Specification, General Overwiev of On-Disk Format, Version 1.03, Dec. 6, 2000, Microsoft Corporation, https://staff.washington.edu/dittrich/misc/fatgen103.pdf

[8] The NTFS File System, https://technet.microsoft.com/en-us/library/cc976808.aspx

[9] Richard Russon Yuval Fledel, NTFS Documentation, http://inform.pucp.edu.pe/~inf232/Ntfs/ntfsdoc.pdf

EDU-VOTING: An Educational Homomorphic e-Voting System

L. TEKIN[1], H. G. OZGUR[1], B. SAYIN[1], A. KARATAS[1], P. SENKULA[1], E. IRTEM[1], and S. SAHIN[1]

[1]Izmir Institute of Technology, Izmir/Turkey, leylatekin@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, huseyinozgur@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, burcusayin@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, arzumkaratas@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, pelinsenkula@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, emreirtem@iyte.edu.tr
[1]Izmir Institute of Technology, Izmir/Turkey, serapsahin@iyte.edu.tr

Abstract - As an instrument of democracy, voting is a critical issue. Although paper-based voting systems are still used commonly, e-voting systems have started to substitute under favor of improvements in the technology. This situation gives rise to need for secure, reliable, and transparent e-voting systems to make people trust. To do this, there are some security requirements that should be concerned and satisfied such as privacy, fairness, verifiability etc.

This study has an educational intuition that analyzes those requirements, theoretical background information related to cryptographic schemes behind them and creates a place-based e-voting design which was implemented for kiosk voting. As a contribution, Paillier homomorphic cryptosystem is used in our system. Moreover, our study includes a detailed criticism for the implemented system in terms of chosen cryptosystems and design modules with security and e-voting requirements.

Keywords: e-voting, Paillier homomorphic cryptosystem, privacy, secrecy, multi-party computation, elliptic curve cryptography, PKI, digital signature, open source software.

I. INTRODUCTION

Because of the advanced technology, e-voting becomes a hot topic that many countries and organizations try to adapt their voting systems to electronic environment. Especially, small organizations prefer to use e-voting systems because of its practicability. The most critical issue here is that how well the system is designed. Naturally, the need for a reliable, secure, and adaptable system remains.

Key requirements for an e-voting system implemented on any scale includes (i) *inalterability:* once a vote is casted, it cannot be modified. (ii) *non-reusability:* a voter must have only one valid vote. (iii) *eligibility:* only eligible voters should cast a vote. (iv) *fairness:* unless the voting process ends, counting process cannot be started. (v) *individual verifiability:* every voter should keep track of whether her own casted vote still in the system or not. (vi) *universal verifiability:* everyone can verify the correctness of the whole voting process and results according to announced system keys and data. (vii) *privacy:* votes cannot be correlated with voters with the help of vote anonymization. (viii) *authentication:* checking the credentials of anyone to see whether the proffered identity consistent or not. (ix) *integrity:* while applying some operations on data, protecting the accuracy and consistency of it. (x) *coercion-resistance:* providing a fake credential for every voter to use in any possible coercion case. (xi) *receipt-freeness:* attackers cannot find any receipt of a voter's casted vote. (xii) *secrecy*: ensuring that no one can read the message except the intended receiver.

In order to build a well-designed e-voting system, key requirements mentioned above should be satisfied. A variety of cryptographic algorithms propose solutions to solve security problems and meet requirements. According to the necessities of the domain, in which an e-voting system will be adapted, the most suitable algorithms should be chosen and applied.

The aim of the following study is to analyze the requirements for a local e-voting system (e.g. Student Association Elections for a university) using homomorphic encryption scheme, cryptographic background and possible vulnerabilities which may cause system to fail. Furthermore, this study was performed with an educational approach and has a modular architecture. Related open source code and completed documentation placed on GitHub[10]. In accordance with mentioned purposes, Helios system [1] was examined and used as a related study.

The remainder of the paper is organized as follows. Section 2 presents related e-voting studies. In Section 3, we give an enhanced explanation of our system design. A comprehensive summary of our work which also includes the selected cryptosystems and satisfied requirements for each system module is given in Section 4.

II. RELATED WORK

Helios [1] is the first web-based and open-audit e-voting system. It is wholly transparent that anyone can create an election on the system and get a result from there. Additionally, it is completely trackable by anyone (open-audit) so the voting process can be maintained securely and fairly. Due to these reasons, we selected this system as a related work for our study. We examined the whole system with its cryptographic background and tried to create a novel design according to our selected security requirements. Implementation steps of the Helios system is as follows:

- Admin creates an election and determines the election dates and name.
- ElGamal cryptographic scheme is used to construct the system private and public key.
- Admin arranges the ballot.
- Admin adds voters to the system together with their names

and email addresses.

- System creates a password for each voter. This password is sent to voters via e-mail.
- Admin suspends the system and takes the election details as a JSON format.
- When the voting process starts in the specified time, a notification together with the election link is sent to voters.
- When voters click to the link, a java applet is run.
- The casted votes are encrypted, and hash value of the encrypted vote is send to voters and kept in a bulletin board.
- When the voting process ends, casted votes are anonymized by shuffling with a proof.
- Admin decrypts the shuffled ciphertext by verifying with a proof.
- After counting process, auditors can reach to all election data and verify the whole process with the generated proofs.

Zeus [8] is another verifiable e-voting system, and it is based on Helios. Neither a new workflow, nor any new e-voting protocols are described in the scope of this project. Proven systems and ideas are used with some contributions as follows; (i) an extension is added to handle all kind of voting system, (ii) all cryptographic content is wholly documented, and (iii) audit system is improved to check both integrity of the vote, and authenticated communication channel.

Sailau [9] is another e-voting system. It is like a paper-based voting. Its difference from paper-based voting is that the voter does not mark the ballot on paper, instead she takes a smart card from the registration table and casts her vote on the voting terminal which is in a different place. Selection of the voter is saved into the smart card. Afterward, the voter inserts the smart card to the electronic box in order to make her vote recorded to the system. A four-digit number is returned to the voter, this number is used for verification after election is over. Then she signs the electronic poll book with her id card.

Unlike the mentioned related works, our proposed e-voting system is based on a Kiosk voting. The aim of it is to create a research environment for developing a viable e-voting system. It differs from the referenced Helios system by means of both selected cryptographic scheme i.e. Paillier Homomorphic cryptosystems and application domain as kiosks. Further information related to our system is explained in Section 3.

III.SYSTEM DESIGN

A. System Overview

The proposed e-voting system consists of seven essential modules which are public key infrastructure (PKI) implementation, secret sharing, registration, election preparation, vote casting, vote management and tallying.

Flowchart of our proposed system is shown in *Figure 1*. Functionality of system components are explained in detail below.

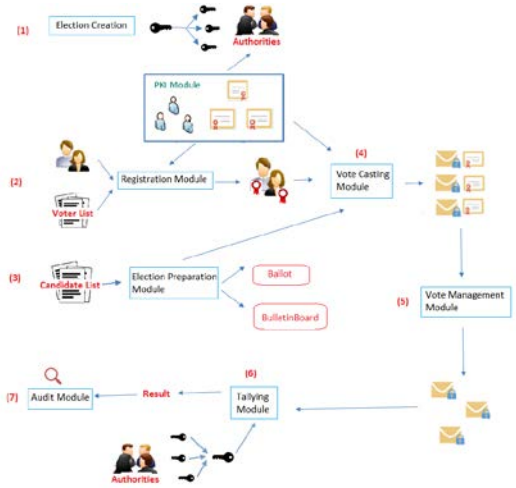


Figure 1. System Overview

Stage 1 - Election Creation: Admin determines election dates and name, then creates the election.
Eligible voters, candidates and authorities are predefined in the system with their names and email addresses and certificates with the help of PKI module. Paillier homomorphic scheme is used to construct the system private and public key. Then, private key is partitioned to shares and distributed to authorities.

Stage 2 – Registration module: This process starts at predefined registration time. Eligible voters who want to cast a vote must register to system. System creates and gives a registration code for eligible vote and a fake registration code to detect coercion case for each registered voter.

Stage 3 – Election Preparation module: After voter registration process ends, election preparation process starts. Predefined candidates are obtained from database as candidate list and an identification number is given for each. Next, ballot that contains the list of candidates is created. Later, the number of eligible voters is counted. Following this, a bulletin board is created.

Stage 4 – Vote Casting module: Voting process starts at defined voting time. When a voter casts a vote, it is encrypted by system public key of Paillier Homomorphic Cryptosystem. Then, login panel is opened, and username and registration code are asked. If the login is successful, voter signs the encrypted vote by her private key (Elliptic Curve Digital Signature – ECDSA) which is related her PKI certificate. Hash of the registration code, encrypted vote, and vote signature is concatenated. Then hash value of this concatenation is created and given to voter for tracking her vote. Voters can query their casted votes from the bulletin board via using given hash value.

Stage 5 – Vote Management module: When the voting process ends, casted votes are verified by using registration code entered at casting time. If real registration code is used, the vote is verified as valid vote, else if coercion (fake) code is used, then the vote is verified as invalid vote. Thereafter, valid votes are anonymized by shuffling.

Stage 6 – Tallying module: Anonymized votes are counted without decryption by the facility of homomorphic Paillier cryptosystem. After counting process of the encrypted votes, each

authority sends its private key share to others by Shamir's secret sharing scheme [7]. Authorities who build the system private key starts to decrypt the votes (by multiparty computation).

Stage 7 – Audit module: Universal verifiability is satisfied by bulletin board. Bulletin board shows election steps and related counts and percentages for registered voters, before and after shuffling number of valid and invalid votes, and the result of election. For the auditing purposes; after decryption process ends, auditors can reach to all election data with last results from database, hence verify the whole process step by step with the system details and keys.

Each module as mentioned here above will be explained detailed in following subsection 3.4.

B. System Requirements

In our system, we addressed all fundamental requirements described in introduction part to satisfy the needs of an e-voting system: (i) inalterability, (ii) non-reusability, (iii) eligibility, (iv) fairness, (v) individual verifiability, (vi) universal verifiability, (vii) privacy, (viii) authentication, (ix) integrity, (x) coercion-resistance, (xi) receipt-freeness, (xii) secrecy. *Table 1* in conclusion part shows an overview to specify which requirements were satisfied and which cryptosystems were used in which module. Furthermore, we have an additional constraint for this design that the one election can be carried out at same time interval.

C. Assumptions

There are some assumptions to make our system viable, and simulate the core functionalities more smoothly:

1. Admin is a trusted actor.
2. Authorities, candidates and voters are known by system election poll before starting time of specific election.
3. Each voter and authority have a public-private key pair and certificate in system creation time. Those certificates are verified with using public key of Certification Authority (CA) for each interaction with voters, which is default module in our system to satisfy authentication requirement.
4. Authorities are honest in the whole election process. Authorities use principles of Shamir's scheme [7] and they are responsible for saving the decryption key shares as secret. However, in this simulation the shares are stored in text files.
5. Whole system is embedded in only one Kiosk machine. However, modules can be distributed by different architectures in future. In that case there would be a secure communication problem. For this reason, encrypted communication is used as a default to create secure communication channel between modules in this design and implementation.
6. Security requirements and risks of network and communication protocols are out of the scope.

D. System Modules

i. Stage 1 - PKI Module

PKI [11] is essential for an e-voting system because it ensures authentication, non-repudiation and integrity with Elliptic Curve Digital Signature Algorithm (ECDSA). Each certificate is signed by CA and includes related authority's, or related voter's public key. The PKI module includes two methods as generation and verification of signature. A key pair for CA and a signed certificate for each user are created in the system before election time. In the registration process, signed certificate of related user is verified by CA verification. There is no registration process for authorities, they are already defined in the system as specified in assumptions (section 3.3).

According to implementation of PKI scenario, voter should have signed certificate from CA. As a novelty, here we used ECDSA to generate signature with using CA's private key in the signing process. It is used for verification of generated signature with public key of CA by voter registration module.

Elliptic curve cryptography uses smaller key size for the same level of security when compared to other cryptosystems. Therefore, it is implemented in the project. One of the NIST recommended elliptic curves can be chosen for ECDSA in this implementation. Koblitz [2] shows detailed structure of elliptic curves, and FIPS PUB 186-4 [3] shows both detailed structure of elliptic curve and recommended elliptic curves that contain choices for private key length and underlying fields.

ii. Stage 2 - Secret Sharing

In this study, Shamir's (k, n)-threshold secret sharing scheme [7] has been applied to implement sharing and reconstructing the private key of the system. If there is one authority in the system to keep the private key, this can lead to key escrow problem since the result of all counting operation will depend on only this authority. To handle key escrow problem, the key should be distributed to multiple authorities. Thus, the private key of the system is split into n shares and distributed to n authorities, and any threshold k or more authorities together can recover the private key. To simulate the scheme, each authority has a public-private key pair and certificate for authentication and perform the following steps: Firstly, each authority saves its share given by a dealer to a text file in its local device in sharing step. Then, in reconstruction step, it sends its own share to other authorities by encrypting with their public keys via Elliptic Curve Diffie-Hellman message transfer protocol [4] to satisfy the secrecy requirement. So, authorities check certificates of other authorities before sending the shares for authentication. The reconstruction step is used by Tallying module at stage 6 (section 3.4.7 includes details).

The system security and reliability are increased by using an efficient threshold cryptosystem because at least k authorities are required to reconstruct the secret which means that an individual authority cannot count the votes with only its share and cannot obtain any early results which could affect the remaining voters. So, fairness increases. As our future work, digital signature scheme can be added to ensure the integrity and authenticity of the shares, and verifiable secret sharing techniques can also be incorporated to provide the consistency of the shares.

iii. Stage 3 - Registration Module

Registration is the first step for eligible voters. In real system, signed certificate of voter is brought by the voters via portable

devices such as smart cards or flash memories. However, voter information and signed certificates are stored in our system for simulation purpose. Eligible voters have to register by filling registration request form with their genuine information and certificate name. Then they have to sign the form to be a registered voter. This module performs the following tasks in registration time specified at the election creation (stage 2 as described in Figure 1):

- checks certificate validity by using PKI module
- checks whether the information sent via form compromises the certificate information and stored voter information.
- generates two unique 6-digit alphanumeric codes for every registered voter: (i) registration code (satisfies *authentication* and *eligibility* requirements) and (ii) coercion code (ensures *coercion-resistance requirement*). It is explained in vote casting module at section 3.4.5 that how to satisfy those requirements. The uniqueness of both registration and coercion codes helps to increase voter *privacy* by adding randomness factor for a casted vote.

iv.Stage 4 - Election Preparation Module

After registration stage, a preparation should be done for voting process. In election preparation module at stage 3 as shown in Figure 1; six fundamental tasks must be performed, which are (i) collecting information of predefined candidates (ii) creating identification numbers to represent the candidates in database, (iii) generating ballot which contains candidate list, (iv) creating bulletin board, (v) counting number of eligible voters and (vi) showing the number of eligible voters on the board.

v.Stage 5 - Vote Casting Module

The steps of a vote-casting can be explained in six steps as below:

Step 1: During voting process, a voter first selects a candidate on ballot without logging in to system. This is important to protect voter privacy.
.

Step 2: Then identifier number of the selected candidate is encrypted with system public key. Paillier homomorphic cryptosystem [5] is used for encryption of the casted vote.

Step 3: After encrypting the vote, the voter should login to system. Username and registration codes are needed for login process. If the voter is under coercion, she should use the fake registration code (coercion-resistance requirement). If the voter first casts the vote with fake registration code, system accepts it and does not allow the voter to cast a fake vote again. If there is no coercion case, the voter should enter the actual registration code.

Step 4: After specifying the login credentials, system checks whether login is successful or not. This check operation provides authentication and eligibility requirements. If it is successful, then the validity of the casted vote is checked. System controls if there is already a valid casted vote by this voter. If so, system does not accept the casted vote and shows a message that says, "you have already casted a vote". If not, the casted vote should be validated before saving to system. So, non-reusability requirement is satisfied. Also, system does not give any receipt to the voter to prove how she voted to meet the receipt-freeness property.

Step 5: Signing is needed for validating casted vote. The voter must sign encrypted vote to make it valid. Signature generation steps here uses ECDSA [3] to assure secrecy and integrity requirements. The most important point for signing a vote is that each voter has her own private, and public key pair as explained in PKI module. Voter's private keys are long enough, so they cannot memorize. There is some hardware such as smart cards to store private keys. But, for simulation, we did not use any hardware. To simulate signing of the encrypted vote, voters' private keys are stored in the database. In the real system, this module can be exchanged with a storage material to keep private keys.

Step 6: After signing operation, system concatenates the hash of the entered registration code, encrypted vote, and signed encrypted vote. Then, system takes the hash value of this triple concatenation. The resulting hash value is saved to the database and returned to the voter.

These six steps are necessary to cast and save a vote. Additional functionalities are included in this module according to requirements. For example, we added a "Query My Vote" panel to our system. By using this module, voters can query their casted vote to see whether it is saved to system or not by checking the hash value given to them at Step 6. This option satisfies individual verifiability requirement. If the given and queried hash values are equal, then it is proved that no one changes her vote which means that inalterability requirement is assured.

vi.Stage 6 - Vote Management

After vote casting process ends, casted votes should be managed with satisfying privacy, secrecy, and integrity requirements, and using coercion-resistance requirement. Vote management is done in two steps: vote verification and shuffling.

Vote verification process has two objectives which are (i) check whether the vote was casted by a registered voter and (ii) check whether the vote was casted under coercion. These objectives are checked by the following steps. First, it collects the casted votes. Then it determines the validity of the votes by checking the signatures on each vote through certificate (authentication requirement) and the related registration codes entered to system at casting time of that vote (eligibility requirement). If the authentication is done via coercion code, it is inferred that the vote is casted under coercion (coercion-resistance requirement). Votes have to be anonymized to satisfy privacy requirement. Therefore, each vote is made disjoint from its signature as in seen Figure 1. Then, it is put into invalid-vote pool if there is coercion. Otherwise, it is put into valid-pool, so real registration code was used at casting time. At the end, counts of valid and invalid votes are shown on the bulletin board for universal verifiability.

Anonymized valid votes are shuffled before tallying process to destroy the order of casted votes. This operation makes sure that there is no link between anonymized votes and the corresponding voters anymore. Shuffling is the last step to ensure voter privacy requirement.

In shuffling process, all anonymized votes are given to a shuffler. The shuffler uses a blind function to perform shuffling. It

is performed in four steps [6]: (i) keep the given votes as ciphertext, (ii) keep the permuted order of them as a vector, (iii) re-encrypt the ciphertext and the permutation vector with Paillier encryption scheme and (iv) decrypt it to get shuffled votes. We performed only one level shuffling, but more than one level can be used for extra security. Naturally, shuffling process is used with a commitment scheme to add a verification step such as Groth's protocol [6]. However, we did not apply any commitment scheme so there is no prover in our shuffling process, yet. Shuffling process in our system assures that secrecy and integrity of the votes are preserved without revealing permutation order.

vii.Stage 7 - Tallying Module

All the computation on computer-based systems is related with data. Data processing can produce wide range of results. Not only efficiency of processing, but also secrecy of the data is essential. Also, personal information of users must be used just for intended computation for privacy. Therefore, confidentiality of data should be regarded at the environment in which data is processed. In e-voting systems, the vote confidentiality can be served via processing encrypted form of votes without decryption. In this context, vote counting refers calculating sum of all encrypted votes, so encrypted votes must be processed. Homomorphic cryptosystems allow to handle encrypted data.

Both Paillier, and ElGamal are considered partially homomorphic cryptosystem. Because, for given any two ciphertexts, they support the generation of a ciphertext associated with either the addition or the multiplication of the two underlying plaintexts, but not both. Paillier supports addition of the plaintexts, ElGamal supports multiplication of the plaintexts [12].

According to Paillier additive property, decryption of the products of two or more ciphertexts result in the sum of their corresponding plaintexts. Also, votes are encrypted using system public key, and encrypted votes are multiplied in our system by homomorphism facility. It means that sum of encrypted votes is calculated by this way. Before decryption of the result, private key must be reconstructed among authorities. Reconstruction procedure is explained in secret sharing part (3.4.2). Decryption of this sum is only done by one who builds the system private key.

As a novelty our contribution is using of Paillier Homomorphic Cryptosystem to create an e-voting solution. Our system design is built according to the facilities of Paillier. But, we have to know its efficiency from this implementation. As presents in Figure 2, more number of voters causes increasingly growing elapsed time. These results are computed on a computer with 2.4 GHZ CPU speed, 16GB memory capacity, Arch Linux operating system.

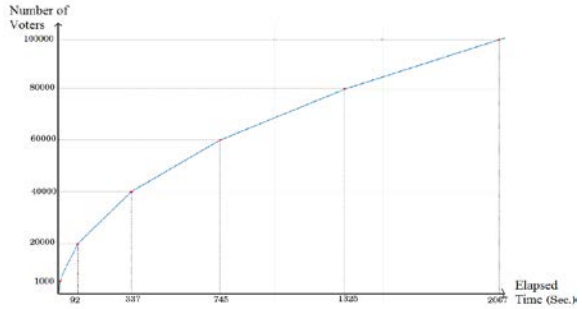


Figure 2: Relation between Elapsed Time and Number of Voters in Paillier Counting Process

The reason of decreasing efficiency is that Paillier counting method multiplies each vote to calculate encrypted result, it means that more number of vote is more elapsed time.

viii. Stage 7 - Audit Module

Anyone can monitor each step of election from Bulletin Board as universal verifiability requirement. Bulletin Board shows "how many registered voters / how many valid and invalid votes exist and how many votes are counted and results" etc. Additionally, for auditing purpose; after election result is announced, auditors can reach to all up-to-date election data from database, therefore verify the whole process step-by-step with system details and keys.

E. Construction of Design

We would like to give very brief information about our construction. Because all detailed documentation is accessible from GitHub [10] link.

We implemented software engineering principles to have modular design. System has three layered structure as database, object and user interface layers. Java programming language was used as implementation environment with oracle database connection.

IV. CONCLUSION AND FUTURE WORK

In this paper, we have provided the design and implementation of a secure and efficient local e-voting system based on the homomorphic encryption which guarantees the desirable security requirements like privacy, eligibility, verifiability, receipt-freeness etc. *Table 1* shows the detailed list of satisfied requirements and used cryptosystems for each module of our system design.

Table 1: An overview of modules in terms of requirement satisfaction and cryptosystems

	System Requirements												Cryptosystems						Stage
Modules	Inalterability	Non-reusability	Eligibility	Fairness	Individual verifiability	Universal verifiability	Privacy	Authentication	Integrity	Coercion-resistance	Receipt-freeness	Secrecy	Shamir' s secret sharing scheme	SHA3	Paillier homomorphic cryptosystem	ECDSA	ECDH message transfer protocol	Groth' s Shuffling protocol without commitment	
PKI							✓	✓						✓		✓			
Secret Sharing				✓								✓	✓			✓	✓		(1)
Registration			✓				✓	✓		✓				✓		✓			(2)
Election Preparation																			(3)
Vote Casting	✓	✓	✓		✓		✓	✓	✓	✓	✓	✓		✓	✓	✓			(4)
Vote Management							✓		✓	✓		✓		✓	✓			✓	(5)
Tallying				✓		✓						✓			✓				(6)
Audit						✓													(7)

Our system has been built with educational purposes by implementing cryptographic algorithms, taking into consideration requirements of e-voting system and basic security functions.

First as presents in Figure 2, e-voting implementation and design should be evaluated according to cost of Paillier. Different homomorphic solutions can be also tested with this modular design and measure their efficiencies.

The Kiosk that voters use to vote is not connected to a network. That's why, it is protected against a vast number of cyberattacks.

For the future scenarios, system modules can be spread to different locations or untrusted servers. System have to provide solutions to prevent network attacks. The secure channel has to be guaranteed among the communication of modules. Naturally the cost of this additional cryptographic implementations will be analyzed, and whole protocol verification and validation will be realized.

However, the system is already subjected to social engineering attacks to steal the registration code. For the future work, codes can be sent via e-mail, or SMS against codes the written down by the registered voter.

REFERENCES

[1]. Ben Adida. 2008. Helios: web-based open-audit voting. In Proceedings of the 17th conference on Security symposium (SS'08). USENIX Association, Berkeley, CA, USA, 335-348.

[2]. Neal Koblitz, *Elliptic Curve Cryptosystems*, Mathematics of Computation Volume 48, Number 177, January 1987, Pages 203-209

[3]. Federal Information Processing Standards Publication 186-4

[4]. Alfred Menezes, Elliptic Curve Public Key Cryptosystems, Kluwer Academic Publishers, 1993.

[5]. Craig Gentry. 2010. Computing arbitrary functions of encrypted data. Commun. ACM 53, 3 (March 2010), 97-105. DOI=http://dx.doi.org/10.1145/1666420.1666444

[6]. Stephanie Bayer and Jens Groth. 2012. Efficient zero-knowledge argument for correctness of a shuffle. In *Proceedings of the 31st Annual international conference on Theory and Applications of Cryptographic Techniques* (EUROCRYPT'12), David Pointcheval and Thomas Johansson (Eds.). Springer-Verlag, Berlin, Heidelberg, 263-280. DOI=http://dx.doi.org/10.1007/978-3-642-29011-4_17

[7] Adi Shamir. How to Share a Secret. Communications of the ACM, 22(11):612–613, 1979.

[8] Georgios Tsoukalas, Kostas Papadimitriou, Panos Louridas and Panayiotis Tsanakas. From Helios to Zeus. USENIX Journal of Election Technology and Systems(JETS), Volume 1, Number 1, 2013.

[9] Douglas W. Jones. Kazakhstan: The Sailau E-Voting System

[10] https://github.com/Edu-Voting

[11] Niels Ferguson, Bruce Schneier, Tadayoshi Kohno, Cryptography Engineering: Design Principles and Practical Applications, Wiley, 2010.

[12] Fang-Yu Ra, On the Security of a Variant of ElGamal Encryption Scheme

Design and Implementation of Greenhouse Automation System with Matlab GUI

F. KATIRCIOGLU[1]

[1] Duzce University, Duzce/Turkey, ferzankatircioglu@duzce.edu.tr

Abstract - Easy availability and cheapness of microcontrollers, intelligent sensors and wireless technology equipments have paved the way for different engineering and industrial applications of control systems. Conventional closed greenhouses, which are one of these spaces, have been converted into high-tech plant production units. A low-cost greenhouse automation system which can be monitored at computer environment by virtue of an interface created in Matlab GUI programming language is proposed in this study. To this end, a model greenhouse system taking into consideration temperature, air humidity, light intensity and soil's physical quantities as input parameters was made. The output parameters used in the model greenhouse were heating, ventilation, irrigation and lighting control. It was determined that the system created by comparing with conventional control systems, was more advantageous in terms of usage, monitoring and programming. Ease of use was provided to the user due to control and monitoring of physical quantities in the greenhouse by virtue of the Matlab GUI interface. The most prominent characteristic of the offered system is that it can easily be used for other agriculture and poultry stockbreeding areas by making small changes in the software infrastructure.

Keywords - Greenhouse Automation, Matlab GUI, Control and Monitoring Systems.

I. INTRODUTION

Easy availability and cheapness of microcontrollers, intelligent sensors and wireless technology equipments have paved the way for different engineering and industrial applications of control systems. Conventional closed greenhouses, which are one of these spaces, have been converted into high-tech plant production units.

A low-cost greenhouse automation system which can be monitored at computer environment by virtue of an interface created in Matlab GUI programming language is proposed in this study. To this end, a model greenhouse system taking into consideration temperature, air humidity, light intensity and soil's physical quantities as input parameters was made. The output parameters used in the model greenhouse were heating, ventilation, irrigation and lighting control.

The greenhouse realized as a model was made in dimensions of 150x75x135cm from 20x20mm profiles. The Medium Density Fiberboard (MDF) is paved the lower part of its and then covered with soil. An electro pneumatic ventilation cover was placed on the roof of the greenhouse. The following equipment has been made so that requests can be met according to the plant species planned to be produced.

Signal sensing elements: In order to effectively manage the greenhouse, the temperature and relative humidity must be measured continuously. DHT 11 is used as temperature and air humidity sensor [1]. It is a basic, ultra low-cost digital temperature and humidity sensor. It uses a capacitive humidity sensor and a thermistor to measure the surrounding air, and spits out a digital signal on the data pin. The soil moisture sensor is used to measure the amount of moisture in the soil or the level of a liquid on a small scale. It is used to immerse the measuring points in the measurement. Technical specifications operating voltage: 3.3V-5V, output voltage and current 0- 4.2V and 35 mA [1].

Heating System: The heating elements applied electric heating in greenhouse and the resistance element is placed diagonally across the two side walls of its, technical specifications are given in detail in Table 1.

Pneumatic cover and fan ventilation system: Greenhouse ventilation is carried out with the aim of preventing greenhouse temperature elevations, controlling the greenhouse relative humidity and providing the necessary CO2 to plants for food products to be produced by photosynthesis. With greenhouse ventilation, the movement of fresh air on the soil surface increases the gas exchange between soil and air and the circulation in the soil. For this, a fan ventilation system is used when the temperature difference decreases and when the wind movement stops. In addition, a mechanical ventilation system has been realized by putting pneumatic covers in the roof of the greenhouse which supports this system [2].

Humectation: In summer the temperature inside the greenhouse is not kept at the values that plants will not be adversely affected by natural or forced ventilation. Humidification systems are needed for this [3]. In the realized greenhouse, 24 V ultrasonic mist maker fogging humidity device is used. In the greenhouse, a fan is added behind the humidifier for homogenous spread of the produced noodle, technical specifications are given in detail in Table 1.

Lighting: Plants need plenty of light during development periods, but less light during rest periods [4]. To adjust the amount of light in the greenhouse appropriately, two pieces of 64-LED fleurasan lamps were used in daylight form.

Drip Irrigation: The plants grown in the greenhouse can't take the natural water they need because they are in the closed environment [5]. In order to meet the need of water, greenhouse is used as a method of irrigation by means of dripping with perforated pipes. Whether or not the plant needs water is determined in accordance with the information from the sensors. The watering process is started or ended automatically by a small pump motor.

Matlab (MATrix LABORATORY) was developed by C.B Moler in 1985 for use in mathematics [6]. It is an interactive packaged programming language with general purpose that can be used in numerical computation, data analysis and graphics processing. Matlab has been used in the analysis of engineering systems especially in the fields of artificial neural networks, optimization, data acquisition, database, filter design, fuzzy logic, and it offers an excellent environment for users with these features.

Applications were performed on Matlab R2015a with Intel (R) Core (TM) i5-4200U CPU 2.30 GHz processor and 6 GB Ram. Problems can be solved by writing programs that run on the command line, as well as visual software consisting of objects such as forms and buttons, using the Matlab GUI (Graphical User Interfaces) development tool [7].

Information about the structure of the graphical user interface designed in this section. The desired temperature, soil moisture and air humidity values are entered in the user interface in the greenhouse environment, and then the "START" button is pressed, as can be seen in Figure 1.

Table 1. This caption is centered.

	Device Name	Voltage	Current	Power	Technical Specifications
INPUTS	Soil Moisture Sensor	5V DC	35 mA	175 mW	0-4.2V output analog soil probe size: 6 cm * 2 cm.
	Air Humidity Sensor	3.5-5V DC	5 mA	2.5 mW	20-90 %RH measurement range, Body size 15.5mm x 12mm x 5.5mm, humidity readings with 5% accuracy
	Temperature Sensor	3.5-5V DC	5 mA	2.5 mW	Rated for full 0° to +50°C range, Linear + 10.0 mV/°C scale factor, Body size 15.5mm x 12mm x 5.5mm, temperature readings ±2°C accuracy
OUTPUTS	Heater	220V AC	2A	350W	
	Lighting	220V AC	36 mA	8W	64 LED Fluorescent lamp
	Humectation	24V DC	800 mA	20W	Suitable water depth 3-6 cm
	Pneumatic Cover	24V DC			
	Fan	12V DC	0.20 A	2.40W	80x80x25MM 2500 RPM
	Drip Irrigation	220V AC		4W	4.5x3x3 inch 300L/H capacity

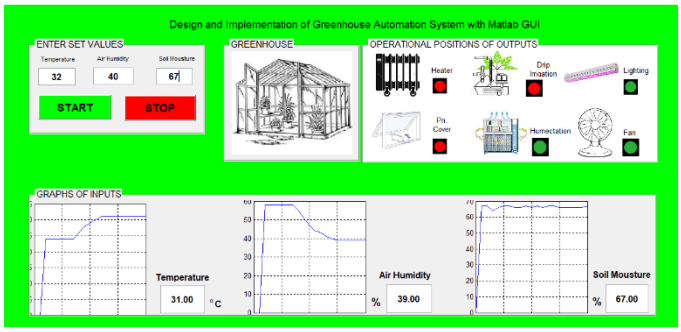


Figure 1. Matlab GUI interface

Instantaneous values of temperature, soil and air humidity data for the greenhouse are numerically displayed in the "Measurement results" window. At the same time, the change of the data depending on the time is recorded on the "GRAPHS OF INPUTS". Transmission of data is provided between Matlab GUI and Arduino by Serial Communication protocol. With this program, the data are analyzed, displayed and recorded instantaneously in the Matlab environment.

There are signal lamps on the upper right of the interface that indicate the current operating status of the greenhouse outputs. If the red lamp which is next to the corresponding output element is activated, it is not working and if the green lamp is on it means that it is working.

It was determined that the system created by comparing with conventional control systems, was more advantageous in terms of usage, monitoring and programming. Ease of use was provided to the user due to control and monitoring of physical quantities in the greenhouse by virtue of the Matlab GUI interface. The most prominent characteristic of the offered system is that it can easily be used for other agriculture and poultry stockbreeding areas by making small changes in the software infrastructure.

REFERENCES

[1] Ardunio Official Site [Internet Resource], http://www.arduino.cc/ [Access Date: 30.03.2018].

[2] D. L. Dupont, U.S. Patent No. 4,815,365. Washington, DC: U.S. Patent and Trademark Office, 1989.

[3] M. M. Hocagil, H. H. Öztürk, "Seralarda sıcaklık ve bağıl nem kontrolü üzerine bir araştırma", *Tarım Makinaları Bilimi Dergisi*, 1(3), 2005.

[4] M. Ciğer, "Bilgisayar kontrollü, internet destekli sera otomasyonu", Yüksek Lisans Tezi, Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Tarım Makinaları Anabilim Dalı, Adana, 2010.

[5] S. A. Gnan, "Meyve Fidanı Çoğaltılmasında Kullanılan Köklendirme Seralarının Otomasyonu", Yüksek Lisans Tezi, SD Ü. Fen Bilimleri Enst., Bahçe Bitkileri AD, Isparta, 2002.

[6] T. Haigh, "Cleve Moler: Mathematical software pioneer and creator of Matlab, *IEEE Annals of the History of Computing*, 30(1), 87-91, 2008.

[7] The MathWorks, Inc., Creating graphical user inter-faces, The MathWorks, Nantick, MA, 2002.

Gray Image Enhancement with Regional Similarity Transformation Function (RSTF)

F. KATIRCIOGLU[1] and Z. CINGIZ[1]

[1] Duzce University, Duzce/Turkey, ferzankatircioglu@duzce.edu.tr
[1]Duzce University, Duzce/Turkey, zafercingiz@duzce.edu.tr

Abstract* - Image Enhancement is a required and indispensable technique in order to improve the image quality of digital images. Much as the digital cameras and mobile phones are available everywhere, lack of clearness on the side textures, emerging of dark or bright areas and creation of noise occurs due to reasons such as failure of camera foci, lack of lighting and atmospheric disturbances. As such, it is necessary and beneficial to develop an effective improvement algorithm in digital images addressing such negative issues and noises. The fundamental function of image improvement is to generate a new density value for each pixel value in the image through utilization of the transformation function after the density value of each pixel of introduction image is received. The proposed conversion function is named Regional Similarity Transfer Function (RSTF) in the study and the conversion is applied by taking into account the density distribution similarity between neighbor pixels. The intuitional optimization technique preferred mostly in engineering applications recently, named Gravitational Search Algorithm (GSA) has been utilized with an eye to optimize the parameter values of the proposed RSTF conversion function (1). An objective evaluation criterion was employed with a view to measure the quality of the images by finding the parameters of the conversion function with GSA. The objective function three performance measures-namely entropy value of images, sum of edge densities and number of edges- of which were combined, was preferred (2). Our experimental results reveal the fact that the proposed method efficiently eliminates the noises received from the images while increasing the image quality.

Keywords – Image Enhancement, Gray Image, Regional Similarity Transformation Function.

I. INTRODUCTION

Image Enhancement is a required and indispensable technique in order to improve the image quality of digital images. Much as the digital cameras and mobile phones are available everywhere, lack of clearness on the side textures, emerging of dark or bright areas and creation of noise occurs due to reasons such as failure of camera foci, lack of lighting and atmospheric disturbances. As such, it is necessary and beneficial to develop an effective improvement algorithm in digital images addressing such negative issues and noises.

The fundamental function of image improvement is to generate a new density value for each pixel value in the image through utilization of the conversion function after the density value of each pixel of introduction image is received. The proposed conversion function is named Regional Similarity Transfer Function (RSTF) in the study and the conversion is applied by taking into account the density distribution similarity between neighbor pixels.

$$g(i,j) = T[f(i,i)] \tag{1}$$

In Eq. (1) while *f(i,j)* is input image, *g(i,j)* is output image *T* represents the transformation function applied to the pixel at the point *(i,j)* in the image. The proposed transformation function is RSTF which carries out a transformation on one pixel taking into account the similarity feature between neighboring pixels and the regional density distribution.

$$g(i,j) = \frac{[a * f(i,j) - (Max + b * S(i,j))]}{\varphi(i,j) + d} + \frac{S(i,j)^c + f(i,j)}{2} \tag{2}$$

The expression *S(i,j)* in Eq. (2) represents the similarity with the pixels around the pixel at *i, j* coordinates and the process sequence which is following Eq. (3)-Eq. (6) is calculated by performing operation.

In order to find the similarity value of the pixels in the 3x3 mask, it is required first to find distances between pixels within mask in order to achieve the relation matrix. When it is assumed that there are two pixels as *k* and *l* within the mask, the distance between two pixels as P_k and P_l is determined by using the following[1].

$$d_{kl} = |P_k - P_l| \tag{3}$$

The exponential function in the Eq. (4) has been used to calculate the similarity value of the two pixels [2].

$$S_{(k,l)} = exp^{(-\frac{d_{kl}}{D})} \tag{4}$$

In the Eq. (4), *D* value is the normalization coefficient and the results achieved can be graded by assigning 32, 64, 128 and 255 values. 9 pixels within the mask creates a 9x9 similarity relation matrix among themselves and with their adjacent, and the relevant similarity relation matrix is provided in the Eq. (5).

$$\begin{bmatrix} S(1.1) & \cdots & S(1.9) \\ \cdots & \cdots & \cdots \\ S(9.1) & \cdots & S(9.9) \end{bmatrix} \tag{5}$$

The arithmetic mean of the relation matrix provided above in the Eq. (5) is calculated and the similarity value to be included within the similarity image for central pixel is calculated according to the Eq. (6) [2].

$$S(i,j) = \frac{1}{81} * \sum_{k=1}^{9} \sum_{l=1}^{9} S_{kl} \tag{6}$$

This value obtained is assigned to the newly created similarity image by considering the address of the central pixel on the real image and the mask panning is realized for the next operation.

This operation is continued until the similarity value of the last pixel in the lower-right corner of the real image is found.

The *Max* value in Eq. (2) gives the most used gray level value in the image. The pixel which is worked on the transform function has a large distance to this *Max* value provides propagation and value-enhancing behavior.

$\varphi(i,j)$ in the RSTF function is the standard deviation in the local area of *nxn* and is shown in mathematical expression Eq. (7). Here, the *nxn* value refers to the local area area and the application uses fields such as 3x3, 5x5. $m(i,j)$ is the average value of the pixels in these area and is given in Eq. (8) [3].

$$\varphi(i,j) = \sqrt{\frac{1}{nxn}\sum_{x=0}^{n-1}\sum_{y=0}^{n-1}(f(x,y) - m(i,j)^2} \quad (7)$$

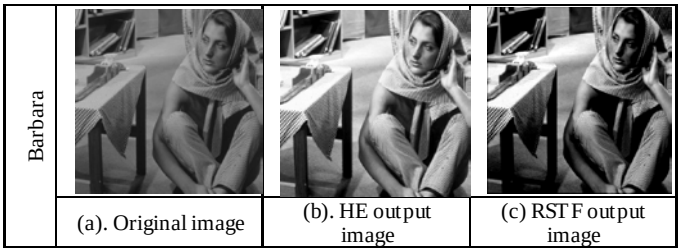
$$m(i,j) = \frac{1}{nxn}\sum_{x=0}^{n-1}\sum_{y=0}^{n-1} f(x,y) \quad (8)$$

In Eq. (2) a, b, c and d are the parameter values and are applied in order to optimize these four parameter values of RSTF by Gravitational Search Algorithm (GSA) [4].

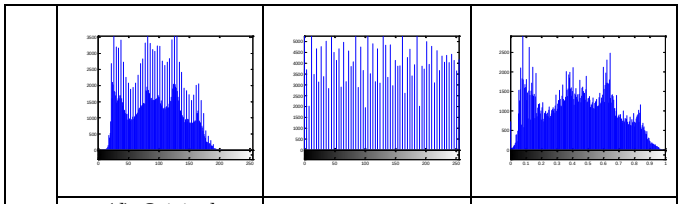
The optimization of *a*, *b*, *c* and *d* parameter values of the transformation function used was carried out by GSA. The steps of the recommended method are as follows:

1. Randomly generate the values of *a*, *b*, *c* and *k* of the transformation function used,
2. Calculate the similarity input image, *Max*, the standard deviation value $\varphi(i,j)$ within the 3×3 regional areas, and the mean value $m(i,j)$ and apply the transformation function to the input image,
3. Optimize *a, b, c and k* parameters using GSA [5].
4. Repeat the steps 2-3 until the improvement criterion is achieved.

The value ranges of *a*, *b*, *c*, and *d* parameters in the RSTF were determined as $a \in [0,1.5]$, $b \in [0,0.5]$, $c \in [0,1]$ and $d \in [0.5,1.5]$ by taking into consideration the previous studies. In the proposed method, the following values were determined for GSA; G_0=100, α=20, *maximum iteration*=20 and *number of agent*=20. Barbara image in 512x512 jpeg format, Peppers in 256x256 jpeg format and Tire in 288x253 jpeg format were used to evaluate the effect of RSTF's performance.



(a). Original image | (b). HE output image | (c) RSTF output image



(d). Original histogram | (e). HE histogram | (f). RSTF histogram

Figure 1: Visual and histogram curves.

Table 1. Objective evaluation results

Image	Metod	Parameters	PSNR	Time
Barbara	HE		14.88	3.26
	RSTF	a=0.901 b=0.106 c=0.979 d=1.375	15.805	3.36sn

Well known Histogram Equation (HE) were preferred to compare the performance of proposed RSTF [6]. Peak Signal-to-Noise Ratio (PSNR), Mean, Best, Standard Deviation (SD) and Time results of the evaluation criterion obtained as a result of running the algorithm 30 times for different limit parameters, are given in Table 1.

Applications were performed on MATLAB R2015a with Intel (R) Core (TM) i5-4200U CPU 2.30 GHz processor and 6 GB Ram. For Barbara image, input and output images of the proposed RSTF method, histograms were included. When Figure 1 is examined, the method created brightness on these images, and an effective improvement was achieved. The success of this method was increased by optimizing the parameters of the transformation function used with GSA.

In the classic He method, the contrast did not come out smoothly for dark areas with increased brightness. Illumination of dark areas reduces the clarity of the view. When carefully examined in Figure 1, it is seen that the proposed RSTF method darkens dark areas and increases clarity. In addition, the curvilinearity of the original histogram is multiplied.

The intuitional optimization technique preferred mostly in engineering applications recently, GSA has been utilized with an eye to optimize the parameter values of the proposed RSTF conversion function. An objective evaluation criterion was employed with a view to measure the quality of the images by finding the parameters of the conversion function with GSA. The objective function three performance measures-namely entropy value of images, sum of edge densities and number of edges- of which were combined, was preferred. Our experimental results reveal the fact that the proposed method efficiently eliminates the noises received from the images while increasing the image quality.

***Keywords* – Image Enhancement, Gray Image, Regional Similarity Transformation Function.**

REFERENCES

[1] BIBLIOGRAPHY F. Katircioglu, "Renkli görüntülerin bağıntı matrisine dayalı ayrıştırılması ve kenar algılama", A.I.B.Ü. Fen Bilimleri Enstirtüsü, 2007.

[2] R. Demirci, F. Katircioglu, "Segmentation of color images based on relation matrix", *In 2007 IEEE 15th Signal Processing and Communications Applications*, Eskişehir, (2007).

[3] C. Munteanu, A. Rosa, "Towards automatic image enhancement using genetic algorithms", *In Evolutionary Computation, 2000. Proceedings of the 2000 Congress*, Vol. 2, pp. 1535-1542 IEEE.

[4] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, S. "GSA: a gravitational search algorithm", *Information sciences*, 179(13), 2232-2248, 2009.

[5] Z. Ye, M. Wang, Z. Hu, W. Liu, "An adaptive image enhancement technique by combining cuckoo search and particle swarm optimization algorithm", *Computational intelligence and neuroscience*, 13, 2015.

[6] S. M. Pizer, S. E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, K. Zuiderveld, "Adaptive histogram equalization and its variations", *Computer vision, graphics, and image processing*, 39(3), 355-368, 1987.

Trends in Cloud-Based Learning Management Systems

H. ÖZCAN[1] and B. G. EMİROĞLU[2]

[1] Amasya University, Amasya/Turkey, hozcan@gmail.com
[2]Kırıkkale University, Kırıkkale/Turkey, bulentgursel@gmail.com

***Abstract* - Every new development in the field of software engineering opens new possibilities for education practices and engenders new opportunities for learning environments. In the world of rapidly developing information technology, cloud computing has given new directions to the structures of traditional learning management systems. Today online learning, as a form of distance learning or e-learning, has developed the capacity of organizations to reach more students than in a traditional classroom setting via scalable online services mainly thanks to cloud-based software solutions. Moreover, organizations have become more equipped with the help of sharing resources and gained enrichment by avoiding large expenditure on hardware and software for the required learning management systems. This work-in-progress study tries to analyze and summarize some trending features directly related with cloud-based learning management systems.**

***Keywords* – Software features, infrastructure, trends, cloud-based, learning management systems**

I. INTRODUCTION

While a typical traditional Learning Management System (LMS) provides a platform through a local server within the organization and needs installing and upgrading by an experienced user or the vendor, the architecture of a cloud-based system allows storing information on the cloud, supports remote access through the Internet, and preferably employs multiple servers to improve the accessibility of content without any installation or maintenance within the organization to operate. This study attempts to introduce some elements of cloud-based LMSs and review the trend features in terms of software characteristics.

II. CLOUD-BASED SERVICES

In recent years, the number of consumers of cloud-based online services has increased in the world. Statistics report that the number of internet users accessing cloud computing services is projected to hit 3.6 billion in 2018, up from 2.4 billion in 2013 [1]. Although some hybrid cloud computing services exist, main models can be classified in three categories in terms of logical abstraction level: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

The SaaS model allows users to connect to cloud-based applications over the Internet. In fact, it is a software distribution and licensing model that enables the use and access of applications on remote servers. In this model the system is almost completely under the vendor's control [2]. Google Apps, Dropbox and Office 365 are some examples of these applications.

As for the PaaS model, it allows consumers to use a platform as a service. In this model, the entire system, including networking, operating system and server applications, is installed by the service provider, and maintained to ensure the performance as well as the overall service uptime. Therefore, consumers do not need to separately license each component that makes up the system, instead, rent all system components as a platform on which software can be developed, tested and deployed to production efficiently. Windows Azure, Heroku, and Apache Stratos are some examples of PaaS.

When it comes to the IaaS model, the system basically has the most flexible structure that sits on the bottom layer of the cloud computing base. In this model system resources, including the central processing unit, memory, storage, network devices, can usually be split and resized on the consumer side, with maximum flexibility, via an API or a control panel. Simply, the IaaS constitutes the basis on which SaaS and PaaS operate. Cisco Metapod, DigitalOcean and Google Compute Engine are examples of IaaS.

III. CLOUD-BASED LMSS

As cloud-based services are gaining a lot of popularity mainly because of the scalability and accessibility, they are widely used by LMS providers as well. A report [3] indicates that global LMS market was measured at over $3 billion in 2016 and expected to grow at a compound annual growth rate of 24% during the years 2016 – 2020. According to the report, cloud-based LMS usage, accordingly, becomes increasingly widespread and the market is growing with the rise of the cloud-based services. Another study reports that in the future it will be inevitable for more LMSs to run on cloud [4].

Although there are some drawbacks using a cloud-based LMS, such as relying on vendors to apply latest data security standards and needing a reliable internet connection, there are numerous advantages associated with cloud-based LMSs, including flexibility to instantly make updates, built-in security settings, greater accessibility, high flexibility to scale on demand, disaster recovery, availability, rapid production without compromising on reliability, low maintenance costs without the need for internal IT support and greater connectivity channels between learners and teachers.

IV. TREND FEATURES OF CLOUD-BASED LMSS

Some of the trend features in cloud-based LMSs include mobile access support, integrated cloud storage, application integrations, enhanced security and gamification.

A. Mobile Access Support

The future of mobile internet usage is expected to increase nearly sevenfold between 2016 and 2021 [5]. As it is accepted that the future learning environment is also mobile-based [6], it is highly expected that LMS users will access the next generation of LMSs from the mobile devices. A study confirms that future LMSs will be available through mobile devices [7]. Today, many of the popular cloud-based LMSs such as CourseMill, Docebo, Edmodo, eFront, ExpertusONE LMS, G-Cube LMS, iSpring Learn and Litmos LMS offer mobile LMS applications accessible via mobile devices for convenient access of course materials.

B. Integrated Cloud Storage

Another salient trend feature in cloud-based LMSs is the integrated cloud storage. A study reports that this technology will have wider acceptance, especially by the increase in demand for cloud-based solutions [8]. Considering that organizations, mainly use LMSs for document organizations and file sharing [9], it is not surprising, therefore, many popular cloud-based LMSs have cloud-based file repositories at various capacities for students and teachers. These file repositories usually come with tools for document management, sharing and collaboration facilities. In fact, integrated cloud storage provides organizations a single unified place to store, access, upload, create, edit, and share course materials in a secure and flexible way. LMSs such as Canvas, Docebo, Edmodo and ScholarLMS provide consumers with cloud-based file repositories.

C. Application Integrations

Modern LMSs offer a variety of opportunities to integrate the system with other platforms and third-party applications. One of the main opportunities is to adopt user accounts of another system for the current authentication needs to increase end user authentication flexibility using the standards such as LDAP and SAML. For example, some popular cloud-based LMSs such as Litmos, CourseMill, Docebo, eFront, G-Cube and iSpring offer SAML and LDAP support for end user authentication integration. Another integration is related to LMS contents. This brings opportunities to import contents from outside and make them more accessible by means of popular technical standards such as SCORM, AICC and PENS. Adobe Captivate Prime, CourseMill and Docebo are of the ones that support AICC and SCORM standards. Moreover, Schoology accepts QTI content packages from other systems. Many of the cloud-based LMSs also support the next generation SCORM - Tin Can API, which is also known as Experience API, for an enhanced integration of a modern communication approach with other systems into LMS platforms.

D. Enhanced Security

Since most of the cloud-based LMS solutions are based on the SaaS model, they are more in need of technical and structural security measures as well as updates and tests before being made available to consumers. In today's cloud-based LMSs, among popular measures there are integrated anti-spam services, strong password policies, IP blocking capabilities and restriction of registration to specific domains or regions. Some examples of cloud-based LMSs that suggest such security measures in accordance with industry standards: LatitudeLearning, Litmos LMS, TalentLMS, CourseMill, Docebo, eFront and Adobe Captivate Prime.

E. Gamification

In addition to providing accessible and secure LMSs, it is also important to create incentive mechanisms to enhance learner engagement. It has become compulsory for cloud-based LMSs to formulate structures that are consistent with evolving teaching strategies expected by organizations. One of the popular mechanisms to improve learners' engagement has become "gamification". Gamification is a trend involving activities that try to apply game mechanics in non-game contexts to increase learner interaction, motivation and participation [10]. It is observed that some LMSs offer various applications for this necessity. For example, Canvas, Schoology, TalentLMS, The Academy LMS, UpsideLMS, iSpring Learn provide gamification features creating competition between learners using some elements such as badges, points, completion boards, levels and rewards.

REFERENCES

[1] "Consumer cloud computing users worldwide 2018 | Statistic," *Statista*. [Online]. Available: https://www.statista.com/statistics/321215/global-consumer-cloud-computing-users.

[2] M. Brandel, "Cloud security: The basics," *Network World*, 15-Jun-2010. [Online]. Available: https://www.networkworld.com/article/2211914/saas/cloud-security--the-basics.html.

[3] "eLearning Market Trends and Forecast 2017-2021," *Docebo*. [Online]. Available: https://www.docebo.com/resource/elearning-market-trends-and-forecast-2017-2021.

[4] N. Radwan, N. Senousy, and El Din, M. Alaa, 2014. Current Trends and Challenges of Developing and Evaluating Learning Management Systems. *International Journal of e-Education*, e-Business, e-Management and e-Learning, vol. 4, issue 5.

[5] "Mobile internet usage worldwide" *Statista*. [Online]. Available: https://www.statista.com/topics/779/mobile-internet.

[6] R. Magdalene, & D. Sridharan, 2018. POWERING E-LEARNING THROUGH TECHNOLOGY: AN OVERVIEW OF RECENT TRENDS IN EDUCATIONAL TECHNOLOGIES. *The Online Journal of Distance Education and e-Learning*, vol. 6, issue 1, pp.60.

[7] N. Cavus, "Investigating mobile devices and LMS integration in higher education: Student perspectives," *Procedia Computer Science*, vol. 3, pp. 1469–1474, 2011.

[8] W. C. Su, and S. E. Chang, 2014. Integrated cloud storage architecture for enhancing service reliability, availability and scalability. In Information Science, *Electronics and Electrical Engineering (ISEEE)*, 2014 International Conference on vol. 2, pp. 764-768. IEEE.

[9] L.Hershey and P. Wood, 2011. Using the blackboard CMS to develop team work skills in undergraduate marketing principles class. *Academy of Educational Leadership Journal*, vol.15 issue 1., p.57.

[10] P. Herzig, M. Ameling and A. Schill, 2012. A generic platform for enterprise gamification. In Software *Architecture (WICSA) and European Conference on Software Architecture (ECSA)*, 2012 Joint Working IEEE/IFIP Conference on pp. 219-223. IEEE.

A Usability Analysis of Edmodo Learning Management System: The Case of a University

H. ÖZCAN[1] and B. G. EMİROĞLU[2]

[1] Amasya University, Amasya/Turkey, hozcan@gmail.com
[2]Kırıkkale University, Kırıkkale/Turkey, bulentgursel@gmail.com

***Abstract* - With the rapid development of computer science, today many universities in their academic departments are using various educational software systems to provide class-related resources, manage the learning activities, and organize the contents of courses. Although these software systems vary widely, the most needed one has tended to be the learning management system. Learning management systems have gained popularity in recent years in conducting online academic programs more commonly to develop and support off-campus education. Learning management systems make it easier for students and teachers to share resources and communicate with each other. As technology influx has expanded alternatives for educational software solutions in distance education organizations, the need for usability testing for learning management systems has increased for better learning experience. This study determined to measure how students perceive the usability of a cloud-based learning management system that is widely used by students and faculty members in many universities.**

***Keywords* – Software, system usability, perceived usability analysis, students, cloud-based learning management system**

I. INTRODUCTION

Learning management system (LMS) platforms differ according to their characteristics and capabilities such as streaming sessions live; integration of third party applications; providing communication tools, customizable modules, personalized dashboards, content management tools; assessing student performance, creating flexible learning paths, reporting progresses, presenting tools to conduct polls and quizzes, etc [1,2]. The number of LMS platforms that use cloud computing technologies to deliver software infrastructure to more organizations is increasing day by day. One of the well-known cloud-based LMS platforms in the world can be referred as Edmodo that has been around since 2008. Edmodo is an online networking application for students, teachers, parents and administrators. It offers completely cloud-based learning management infrastructure for continuing classroom activities for millions of users around the world in a virtual environment, with a set of software modules and various features for the interaction between students and teachers. Today, more than 90 million users worldwide in over 350,000 organizations across 190 countries benefit from Edmodo LMS [3,4]. Recent statistics and insights [5] show that the system has gained considerable popularity among students and teachers at Amasya University (AU) as well. The built-in insights board of the university account reports that the number of students enrolled in the Edmodo courses of AU has reached about 5,600 since 2014. The board also shows that currently about 800 students have been enrolled in the Edmodo LMS in the spring semester of 2018. This number amounts to approximately 5.3% of the total students in the university though the university administration does not have an official policy or suggestion about its use. Despite its widespread adoption in the university, there is, currently, a lack of studies investigating the students' perceived usability for enhanced learning experience. This study determined to examine students' perception of usability of the web application side of Edmodo in the university.

II. METHOD

In this study, the Edmodo LMS was tested through the collection of both quantitative and qualitative data. For the quantitative data collection, the System Usability Scale (SUS) was used to measure how students perceive the usability of the system [6]. The system usability scale is a Likert-type questionnaire of 10 items with five response options ranging from "Strongly agree" to "Strongly disagree" and widely used to evaluate services including software and web-based applications [6,7], even with a small group of participants, at least 12-14 [8]. A group of 20 fourth grade university students taking Project Management class from the Department of Computer Education & Instructional Technology in AU participated in the study during the spring semester of 2017-2018. Participants were briefed on the purpose and objectives of the study and given information about the procedure. The participants were invited to perform 10 tasks using the web application of Edmodo in the classroom. The tasks are directly related to the fundamental features and included: logging into an account, joining a course, using student backpack, completing and turning in an assignment, commenting on a wall post, answering quiz questions, responding to a questionnaire, sending a direct message, checking if the notifications were enabled or not and logging out of the system.

Participants were then asked to respond to the system usability scale questionnaire which was proved to be a useful tool for assessing web application usability in a reliable way [8]. Finally, the participants were requested to provide demographic information including: year of study, age, gender, department and years of experience with the platform. The SUS score for each participant was computed and presented descriptively. Independent samples t-test was conducted to attempt to unravel whether female and male students differed in terms of SUS scores. In addition, the Pearson correlation was

conducted to reveal the relationship between the SUS scores and the Edmodo experience.

As for the qualitative data collection, participants were prompted to comment on each of the tasks and write down the difficulties, if any, they had experienced with the use of Edmodo. All the activities were undertaken during lecture sessions to ensure participation in the study. The observational part of the study was based on screen video recording analysis of participants' interactions with the software during the completion of the given tasks.

III. Results

The system usability scale score (M=82.3, SD=4.1) showed that the perceived usability of the Edmodo web application was high as shown in Figure 1.

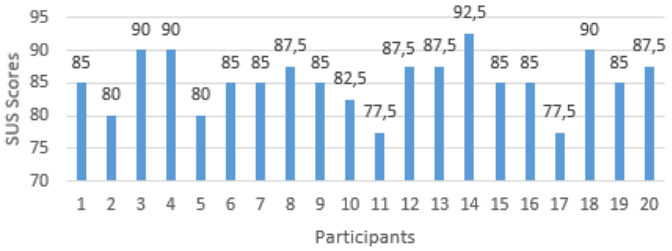


Figure 1: Participants' SUS scores.

In terms of gender, the female participants have higher SUS scores than the male participants as summarized in Table 1. However, the independent samples t-test showed that there is no statistically significant difference between the female (M=86.3, SD=3.8) and male (M=84.3, SD=4.4) participants; $t(18)=1.1$, $p>.05$.

Table 1: Descriptive statistics of the SUS scores in terms of gender and experience.

	Group	n	*M*	*SD*
Gender	Female	10	86.3	3.8
	Male	10	84.3	4.4
Experience	2 years	2	85.0	.0
	3 years	3	85.0	5.0
	4 years	15	85.3	4.4
	Total	20	85.3	4.1

As for the experience, while the participants with 2-year and 3-year experience had equal SUS scores, the ones with 4-year experience had the higher SUS scores than the both others. However, the one-way ANOVA results indicated that there is no statistically significant difference between the participants from 2-year experience group (M=85.0, SD=.00), 3-year experience group (M=85.3, SD=5.0), and 4-year experience group (M=85.3, SD=4.4); $F(2,17)=.01$, $p>.05$.

Participants' comments also reflected that they were highly satisfied with the system. Despite the overall satisfaction the study additionally found that a few participants were unhappy with the direct messaging system and the group wall. A participant stated, "I had difficulties to find the assignment that you presented to the class". Three of the participants stated that they had a bit of difficulty finding the last quiz on the wall. Two participants reported that while they were trying to send a message to the teacher, they mistakenly sent a message to the group and it was published on the wall. According to the participants (n=2) themselves, the direct message sending interface had to be more distinctive.

IV. Discussion

The study indicates that web application side of Edmodo provides a positive student experience and a high level of perceived usability. Even though, according to the literature [9-11], perceived gender differences in computer science indicates male participants are generally more inclined towards computers than females, this study shows that there is no significant difference between genders in relation to perceived usability of the system. According to the study, experience does not seem to be significantly associated with the perceived usability of the software. The results also imply that at least two-year experience is adequate for gaining satisfied user experience on Edmodo. These findings reveal the wider usability among both experienced and less experienced students and suggest user acceptance of the software to some extent. The major limitation of this study is the number of participants, which may not be an accurate representation of the entire population. Moreover, the low number of participants made it difficult to make significant differences, if any, according to gender and experience. Larger future studies are needed to replicate the study and then better elucidate the perceived usability.

References

[1] P. Oliveira, C. Cunha and M. Nakayama, "Learning Management Systems (LMS) and e-learning management: an integrative review and research agenda", *Journal of Information Systems and Technology Management*, vol. 13, no. 2, 2015.

[2] R. K. Ellis, "A field guide to learning management systems", *American Society for Training and Development Tech. Rep.*, 2009.

[3] "Interesting Edmodo Facts and Statistics", *DMR*, 2018. [Online]. Available: https://expandedramblings.com/index.php/edmodo-facts-statistics. [Accessed: 28-Mar-2018].

[4] "How we got started," *About Edmodo*. [Online]. Available: https://www.edmodo.com/about. [Accessed: 28-Mar-2018].

[5] "School and District Insights", *Edmodo*, 2018. [Online]. Available: https://amasya.edmodo.com. [Accessed: 28-Mar-2018].

[6] J. Brooke, "SUS-A Quick and Dirty Usability Scale," In P. W. Jordan, B. Thomas, B. A. Weerdmeester, A. L. McClelland, Eds., *Usability Evaluation in Industry*, Taylor and Francis, London, 1996.

[7] J. R. Lewis. *Usability Testing*. IBM Software Group, p. 71, 2006.

[8] T. Tullis, and J. Stetson, "A comparison of questionnaires for assessing website usability," *Proc. of the Usability Professionals Association (UPA) 2004 Conference*, 2004, pp. 1–12.

[9] J. Cooper, "The digital divide: the special case of gender", *Journal of Computer Assisted Learning*, vol. 22, no. 5, pp. 320-334, 2006.

[10] He J, Freeman L. "Are men more technology-oriented than women? The role of gender on the development of general computer self-efficacy of college students", *Journal of Information Systems Education*. 2009:21(2):202–212.

[11] S. Beyer, K. Rynes, J. Perrault, K. Hay and S. Haller, "Gender differences in computer science students", *ACM SIGCSE Bulletin*, vol. 35, no. 1, p. 49, 2003.

MultiMedia Application and IPv6 Addressing in Soft Switches

A. HUSSIAN[1] A. ABDURAHMAN[2] and A. A. URBANEK[3]

[1]Prof. Dr. in Istanbul University, Istanbul/Turkey, adbul.hussian@istanbul.edu.tr
[2]Msc. Eng. Osmanli University, Istanbul Turkey, amar1.rahman@gmail.com
[3]Dr. in Osmanli University, Istanbul Turkey, adam.a.urbanek@gmail.com

Abstract*-In this research, we studied the new *generation of telephone exchanges, and discussed how to convert the public number of the recipient into the new IPv6 system, within the telephone exchange and the possibility of serving recipient with the multimedia system.

The rapid development of telecommunication systems over the past decades has led to the latest development in the technology and the volume of services provided by the telephone exchange to recipients through the development of the software architecture.

The presence of modern microprocessors, their speed of work and their ability to implement a greater number of operations have created the possibility of providing many The idea of the work is to find the possibility of transferring the recipient number from the normal numbering system to the IPv6 network system by rearranging the distribution of the address again, eliminating the DHCP feature and redistributing the fields to suit the new numbering. Considering that any country is a major server and has a fixed address of the new address and small sub-division takes its title from the main server, and these sub-subdivisions can give new label to the recipients belonging to them.

***Keywords* - multimedia, IPv6, network, receiving, address.**

I. INTRODUCTION

THE new technology for exchange is dependent on the guiding of the direction routers architecture of distribution of multimedia services (MMS) for all customers, the assumption architecture supposes that the exchange will consist of main server that will be responsible for installation and setup of connections between the terminals (customers) and require additional functionality beyond edge routing.

Demanded by customers across industries, fig (1) describes the assumption diagram of the next generation.

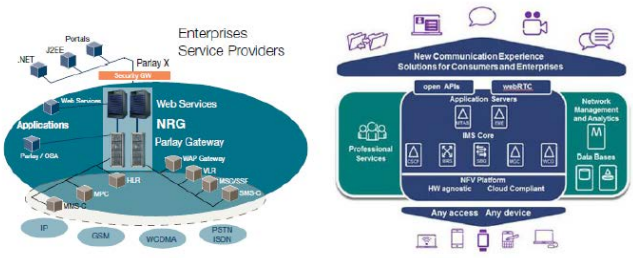


Figure 1: the assumption diagram of the next generation.

These functions are: IP/MPLS edge routing, Broadband Remote Access Server, Ethernet aggregation, Mobile IP, Session Border Gateway, Security (IPSec) and Intelligent Traffic Management, which help to support the terminals with application services fig. (2).

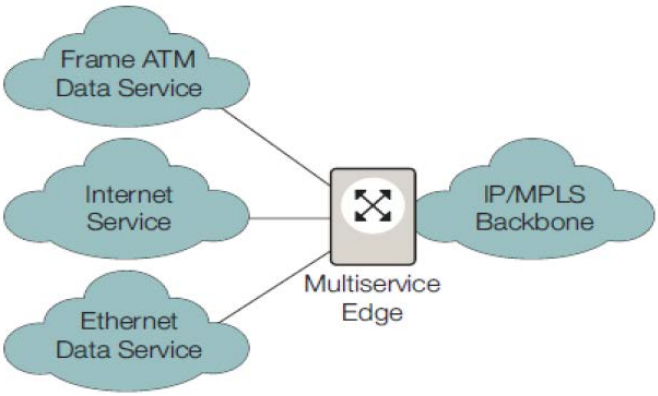


Figure 2: the terminals with application services

To These functions can be deployed in any combination for a variety of today's bandwidth intensive and revenue-generating applications.

The server versatility enables deployment in any type of network architecture, e.g., the traditional overlay architecture with an Ethernet aggregation between access subscribers nodes and main server or fully consolidated architecture where the Ethernet aggregation.

The server of soft switch function offers very wide range of protocols (such as PPPoA, PPPoE, DHCP,…etc) with dynamic service selection, multiple contexts, static and dynamic bindings, over-provisioning, bandwidth management and subscriber flow control.

In addition, the server supports Clientless IP (CLIP) subscribers without requiring PPPoE usernames for DHCP environments, and Intelligent Traffic Management(ITM) via Deep Packet Inspection and Heuristic Analysis IPSec for establishing secure tunneling communication including the use in Mobil based station connections fig. (3).

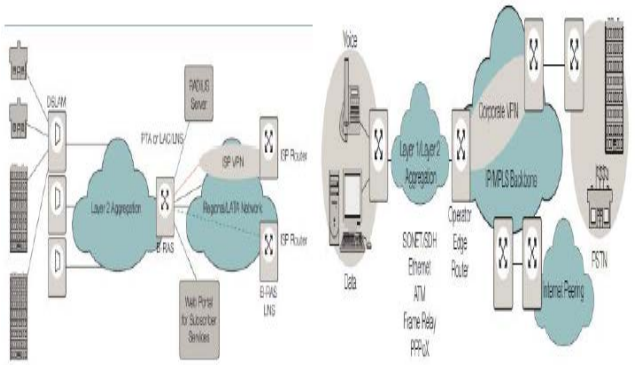


Figure 3: Mobil based station connections

Leaders of industry companies in the development and delivery of networking, mobility, cloud - formed global business and technology partnership create the networks of the future. This offers development functions such as: routing, data centers, networking, cloud, mobility, management and control, and global services capabilities. The next-generation strategy of companies will lead growth, accelerate innovation, and speed digital transformation for packets of data between terminals. This terminal has special IP address. In general, IPv4 were there but in new generation it will be IPv6 which consist of 128 bits. This packet includes three parts (16 bits for main servers, 48 bits for DHCP for terminals, 64 bits for addressing terminals), but for new assumption the new addressing will be in new arrangements.

II. THE RESEARCH OBJECTIVE

The research aims are:

- How next generations of exchanges will be numbered with Ipv6.
- How the exchanges arrange the customer for multimedia with new numbering.
- How the customer can take the benefit of MMS (voice, video, data, remote control ..etc.).

III. THEORETICAL STUDY

An IPv6 address consists of 128 bits divided into eight16-bit packets. Each packet is converted to 4-digit Hexadecimal numbers separated by colon symbols ':'.

The principle of simulation is to make new numbering from IPv6 in the soft switch and this will be the general number of the customer. We will suppose that any country as the main soft switch (main server) and the exchanges as sub soft switches (sub servers). Of course the soft switch has two connection directions one with another soft switch (countries) and the other with sub soft switch servers in first stage, in next stages it will be with other sub soft switches as in fig (4).

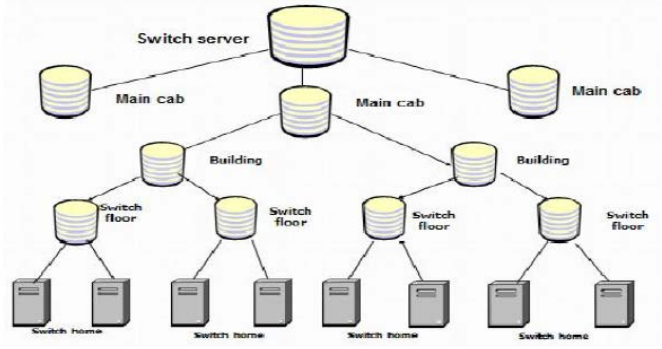


Figure 4: Sub soft switch servers

Any main soft switch has own IPv6 address, this address will equivalents in general exchange code number. This equivalent can be described as in fig (5)

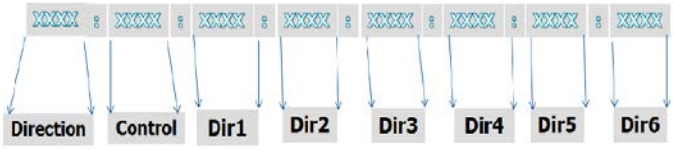


Figure 5: Equivalent describe

1- Direction: reserved for international domains services and the internet application, connected with all main soft switches.
2- Control: reserved for control stage (call server – internet server –video server– other controls server–…. etc.).
3- Dir1: describes the countries address code for international and states used.
4- Dir2: describes the code cities.
5- Dir3: describes the area codes in cities.
6- Dir4 and Dir5: describe the owner number code for customers.
7- Dir6: describes the accessory code for terminal equipment services.

IV. PRACTICAL STUDY

The simulation will describe main servers (main soft switch) and sub servers which describe many exchanges ,the main server has a special own IPv6 number (defined in dir2 part) ,also the sub servers will have Ipv6 number (defined in dir3 part),and the terminal's will get addresses inside of sub servers rang and consist of two ranges (dir4 and dir5), the last stage (dir6) will be to define the equipment at the customers add to this equipment (internet – TV- Radio - …etc) will be serviced depending on how many services the customer needs and the kind of customers (persons – companies – factories …etc).

This simulation will consist of:

1- Main server: equivalent's the main exchange and we will be Define the required applications to serve the nodes fig (6).
2- Node: equivalent's the sub server(exchange), and it will be a management for the packets which will contain applications and will send them to terminal's (customers) fig (7).

3- Terminal: equivalent the customer, it will be a management router (wire – wireless), and how many applications the customer needs.

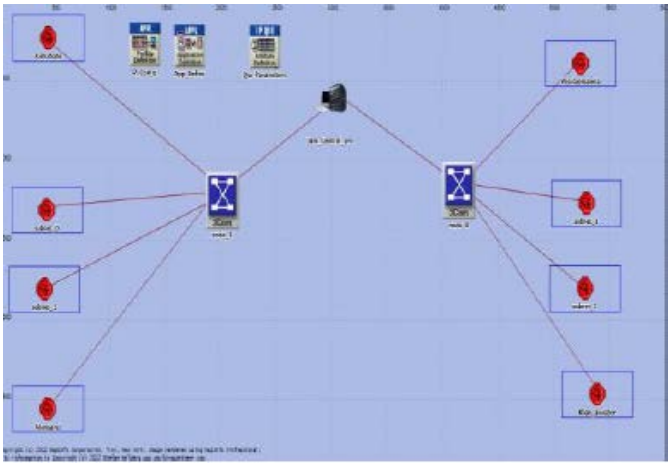


Figure 6: The required applications to serve the nodes

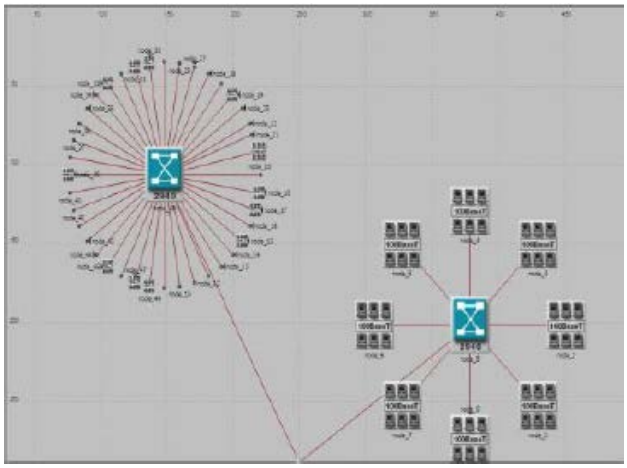
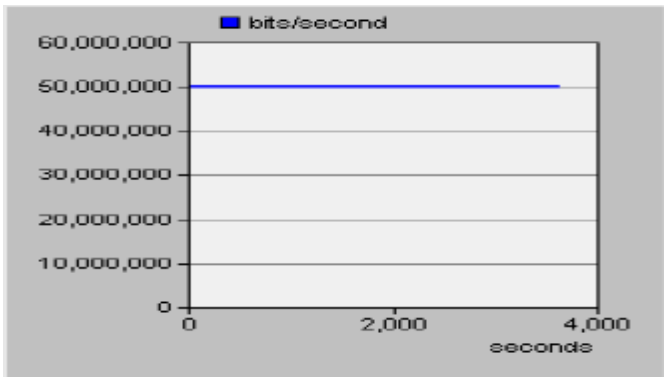


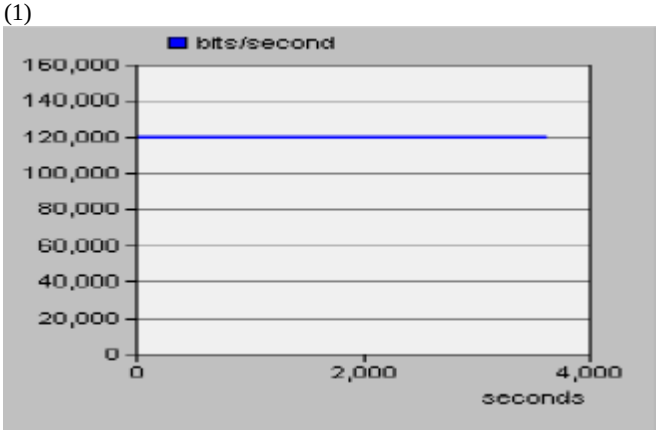
Figure 7: Management for the packets

We will use Opnet14.5 program for this simulation, for main server in fig and for sub servers will adjust them with the features of network and will be as follow:

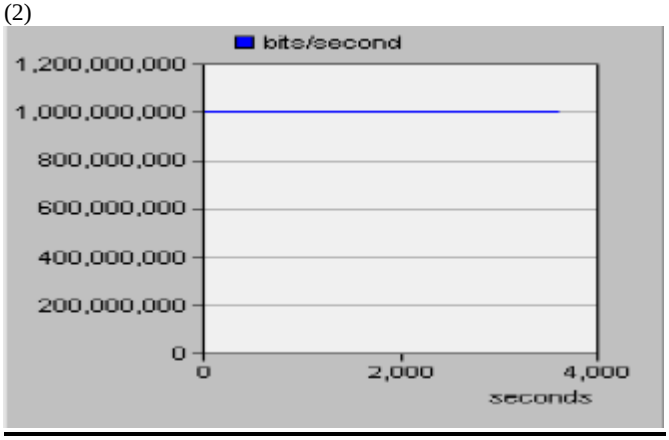
- The main server will be the source of MMS (Phone – VoIP – TV – TV child – Internet – remote control –… etc.), and will have fixed IPv6 address (example 10:0:0963:0:0:0:0:0).
- The main two routers will be a distribution station for sub servers (exchanges) to guide the packets for the correct addresses.
- The main switches will be the directed servers for customers (sub exchanges), have own address, give addresses for customers and provide them with MMS
- The terminals (customers) will be routers (4, 8, etc) ports as the customer needs of MMS, each one's has a special address and the last packet of address (0-255) bits will be for the ports of switch, each port gives one service of MMS as defined in it. The terminal equipment for all customers will be like a main router which has multi-port and its own address. Each port will have its own sub address from low side part of IPv6 address (LSP - 8 bits) this port will give the services for terminals (video – voip – internet – data -…etc).



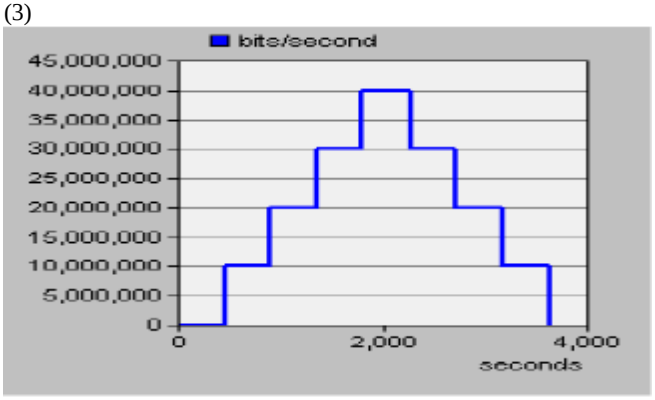
(1)



(2)



(3)



(4)

Figure 8: The variable and the carrier of packets

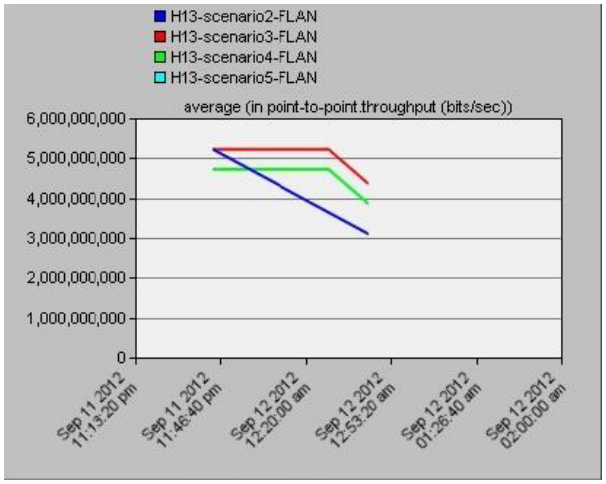
The wide range of packet for any terminal need more size of bits and time that means we need more speed of data analyzing and direction with its own addresses.

The simulation will run in Opnet for one hour during four periods of time we will change main server, switches, and terminals. During those four intervals we use the protocols (ATM, IP, RIP, Frame Relay, X.25, OSPF, IGRP, BGP4). The variable and the carrier of packets (signal holding) will be defined as in fig (8).

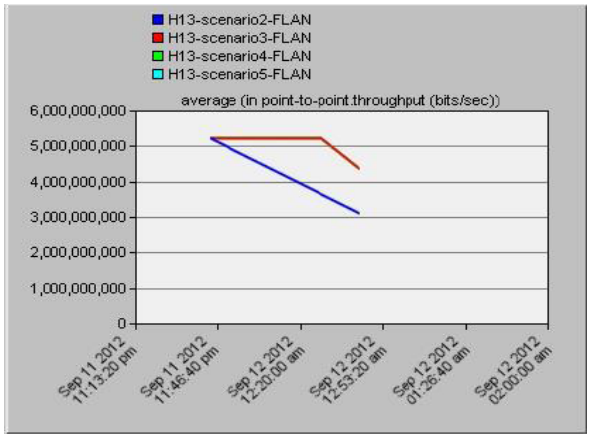
we will get the changed results for this theoretical assumption:

- The source exchange (main server) is difference depending on the main processors kind the application contains it, the support protocols and ability to gives addresses for next stages of sub switches management (DHCP).
- The interfaces units (main switches) are full management also. The unit could get new address from source exchange (bridge management).
- The units get new addresses from the source exchange (we could put fixed Ipv6 address) and could give the terminals addresses with the applications.
- The clients' terminals get addresses and use the applications between them.

From simulation program the results of data flow for all cases of scenarios indicate the differences between the curves depending on data analyzing in the source exchange for both ways fig. (9).



CP to SubNet



SubNet to CP

Figure 9: the source exchange

Fig. (10) indicates the data terminal rate packet used in the router for services (calls, video conf., internet, …).

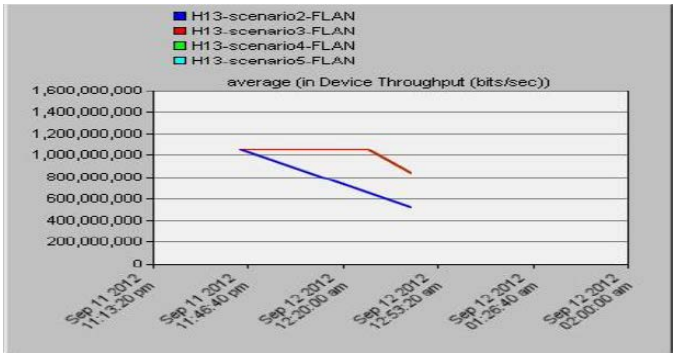


Figure 10: the data terminal rate packet

Fig (11) indicates the data transfer rate between two terminals used in all services (application services):

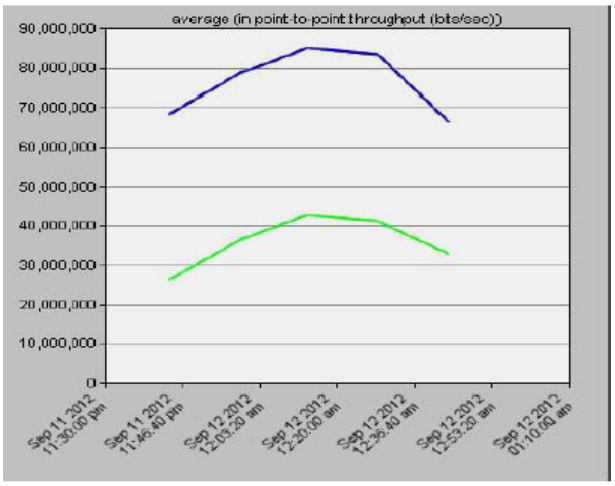


Figure 11: the data transfer rate between two terminals

Fig. (12-13) indicate the data packet success rate transferred between all terminals for services.

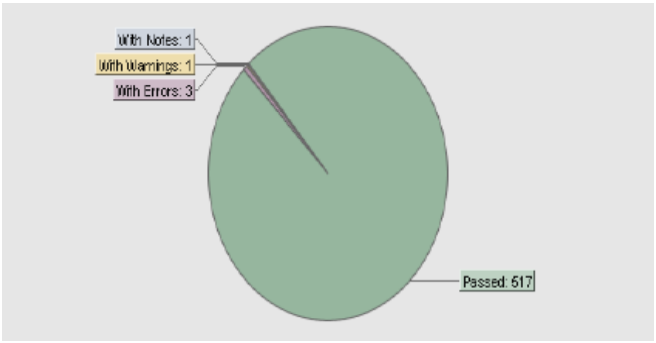


Figure 12: the data packet success rate transferred

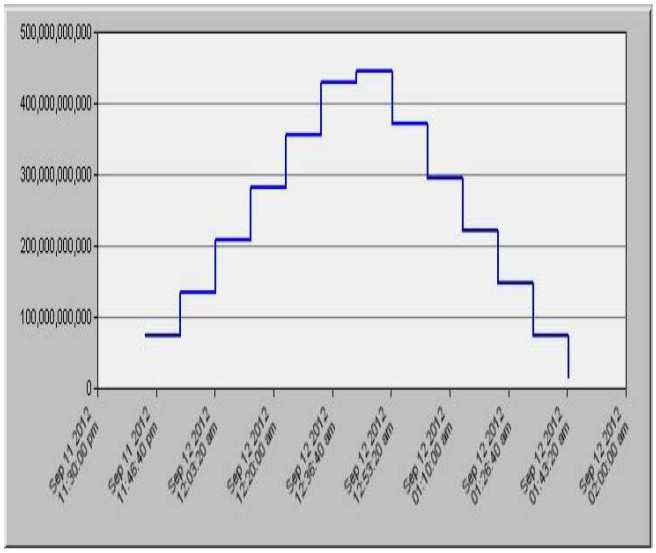


Figure 13: the data packet success rate transferred

For main soft switch gives the interface unit (main router) in subnet Ipv6 addresses for next stages (10:0:963: 22:0:0: 0:0, 10:0:963:23:0:0:0:0, 10:0:963:24:0:0:0:0, …. etc.) and for terminals we can get new Ipv6 addresses (10:0:963: 22:0:0: 0:1, 10:0:963:23:0:0:0:2, 10:0:963:24:0:0:0:1, 10:0:963:26:0:0:0:2, 10:0:963:29:0:0:0: 2, … etc.).

These units(routers) will send a packet for each next stage (router) which contains (video – VoIP – internet – data -…etc.), and inside it any service will redirect for the ports as defined inside in.

The fig (14) describes the packets delay when the simulation uses new IPv6 numbering, the time delays will run respectively, because of the interface unit management for IPv6 numbering.

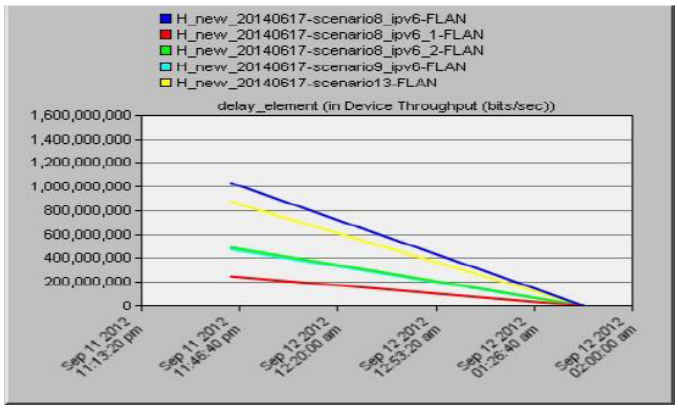


Figure 14: the packets delay

The fig (15) describe. the amount of data transferring before and after terminals numbering. We can see the differences; the interface units have the directions for terminal by new numbering.

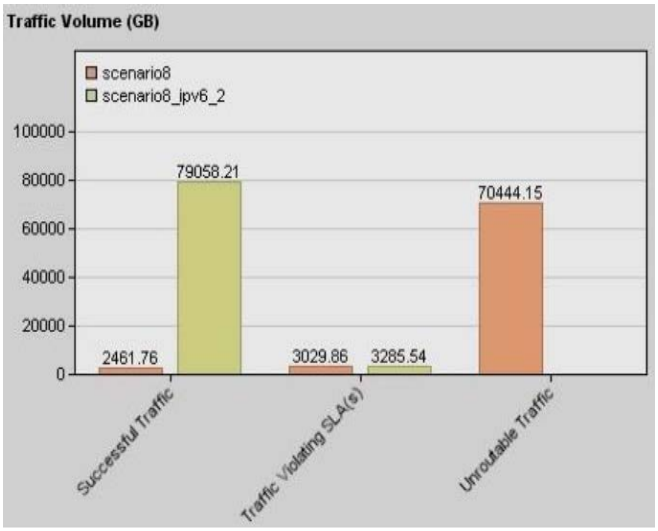


Figure 15: the amount of data transferring

This SIMULATION SHOWS THE ABILITIES TO USE THE FIXED OF IPv6 NUMBERING WITHOUT DHCP PART. THIS GIVES US TO REARRANGED IP ADDRESSING FOR TERMINALS (CUSTOMERS) WHEN USING THE IPv6, AND COULD BE DIVIDED INTO PARTS OF ADDRESSING FOR SPECIAL FORM IN SOFT SWITCHES (EXCHANGES) AS DESCRIBED ABOVE.

In recent technology the soft switches could provides the customers the application services by using internet with telephone (example: 4G for mobile and soft technology for land phone (Ericsson-EAR amp)), and the CPU for dealing with this technology changed from the standard case to server case (example: Ericsson-HP combined (APZ 212/50-APZ 212/55)).

The new technology search for how could increase the packet for customers to serve them with MMS. That means we need more speed for CPU (5G: Ericsson-HP (APZ 212/60)), and more size for packet to customer (Ericsson – Cisco), this will give mobile more separated of application without net and for land phone which will give the applications with separated ports of router.

The future technology for exchanges (mobile – land phone) is to serve the customers MMS services on telecommunications networks with many signal rates by increasing the data flow rate and using IPv6 in telecommunication networks to guide the addresses for all customer.

V. THE RESULTS

The telecommunication networks development to server technology will be different in soft switch size, data transfer packet size, ability of carrying many applications at the same time, data losing rate, cable size, maintenance services and cost ,and the result for applications in simulation gives us more facilities for using, much more quantities of applications, more easier for maintenance and low costing of money, the new addressing will help the telecom networks to be easier in hardware structure to serve all applications and redirection more easy to customer.

VI. THE SUMMARY

The development of telecommunication networks gives these benefits:

1. Data transferring packets and application services has more size with new addressing in IPv6.
2. The number of customers will be much more with the new addressing and control system.
3. Connection time will be short.
4. MMS applications serve all customers.
5. The equipment needed by the PBX will be small. Speed dealing for data will be more to the telecom networks.
6. The cost of communication cables is lower.
7. Data processing speed will be greater for telecommunication networks
8. Maintenance and control of remote equipment becomes easier wherever the recipient location (subscriber) is.
9. We can give any specific IPv6 address to recipients (subscribers)
10. We can provide better applications, better maintenance and high quality to all recipients.

VII. REFERENCES

[1] H. Poor, *An Introduction to Signal Detection and Estimation.* New York: Springer-Verlag, 1985.

[2] C. Bradley, *securing media streams in an Asterisk-based environment and evaluating the resulting performance cost*; Confidentiality, Integrity and Availability, 8th January 2007.

[3] F. Tanjila, G. Rajvir, *Comparison of WiMAX and ADSL by Streaming Audio and Video Content, fixed and mobile Internet access*, Spring 2011.

[4] K. Xiaomin, *Modelling and Performance Studies of Integrated Network Scenarios in a Hospital Environment; integrated approach*, past 2008.

[5] S. Khaled., H. Jamil., A. B. Zubair., H. Fahd., *Video-on-Demand (VoD) deployment over hospitality networks, VoD services over IP-based networks*,14 July 2011.

[6] Spyros Denazis, *Enabling live video streaming services realization in telecommunication networks using P2P technology*, University of Patras, Greece, Ph.D., 2011.

[7] K. Salah, *An analytical approach for deploying desktop videoconferencing*, IEE Proceedings Communications 2006.

[8] Ineo Quest, *Quality of experience for media over IP*, White Paper, Available: http://ftp.ineoquest.com/pub/, June 2005.

[9] K. Salah, *On the deployment of VoIP in Ethernet networks: methodology and case study*. International Journal., 2007

[10] Ineo Quest, *Quality of experience for media over IP,* White Paper, Available: http://ftp.ineoquest.com/pub/, June 2005.

[11] K. Salah, *On the deployment of VoIP in Ethernet networks: methodology and case study*. International Journal. 2007.

[12] S. L. Tompros, S. Denazis, *Interworking of heterogeneous access networks and QoS provisioning via IP multimedia core networks. Computer Networks,* Vol. 52, Issue 1, pp. 215-227, 2008.

[13] Meng ZHANG, Qian ZHANG, Lifeng SUN, Shiqiang YANG, *Understanding the Power of Pull-Based Streaming Protocol: Can We Do Better?* IEEE JSAC, 2007.

[14] *Distributed Scheduling Scheme for Video Streaming over Multi-Channel Multi-Radio Multi-Hop Wireless Networks,* IEEE Journal on Selected Areas in Communications, vol. 28, no. 3, pp. 409-419, Apr. 2010.

[15] 5G Available: https://www.ericsson.com/en/news-and-events/press-center/media-kits/5g

[16] Ericsson, Available:https://www.ericsson.com/en/cases/2015/telkomsel-indonesia.

[17] Ericsson, Available https://www.ericsson.com/en/ericsson-technology-review/archive/2017/devops-fueling-the-evolution-toward-5g-networks.

[18] Telecomes, Available; https://uat.telecoms.com/45223/global-tech-brands-go-live-with-ipv6.

[19] Ericsson, Available: https://www.ericsson.com/en/ericsson-technology-review/archive/2017/devops-fueling-the-evolution-toward-5g-networks.

[20] Ericsson, Available: https://archive.ericsson.net/service/internet/picov/get?DocNo=2/28701-FGC1010723&Lang=EN&HighestFree=Y.

[21] Falesia Available: http://www.falesia.pl/pdf/se1200.pdf.

Trellis Combining Approach for Demodulate and Forward Based Multiple Access Relay Channels

Ö.ÖZDEMİR

Selcuk University, Konya/Turkey, ozgur@selcuk.edu.tr

***Abstract* – Network coding has emerged as a paradigm for communication systems in which each intermediate node in the network generates and transmits output data by encoding previously received packets in a manner to increase the bandwidth efficiency of the system. The multiple access relay channel (MARC) is one of the real-world reference network coded communication scenarios where multiple users transmit data to a common destination through the use of one or multiple relays. In this paper, performance of demodulate and forward protocol which is among the prominent digital relaying techniques is considered in MARC systems. A convolutional code is used as the channel code. Maximum likelihood (ML) detection and user selective relaying are utilized in order to decrease the performance degradation due to decision errors at the relay. The simulation results obtained for Rayleigh fading channels have shown that the examined joint channel-network coding approach is superior with respect to non-cooperative system by providing full diversity gain.**

***Keywords* - Network coding, multiple access relay channel, demodulate and forward protocol, selective relaying.**

I. INTRODUCTION

COOPERATIVE diversity which is one of the most popular research areas in communication systems has been proposed for next generation wireless networks to meet the requirements of increased spectrum and power efficiency, improved system capacity and diversity, extended network coverage and reduced outage probability [1-2]. Enabling single-antenna nodes in a multi-user environment to collaborate with each other in a manner that forms virtual macro antenna arrays is the main motivation behind cooperative communication systems [3]. This approach provides an efficient way to receive and combine multiple version of the message from the source and intermediate terminals called relays, which undergo independent fading, at the destination and acquire a more reliable estimation of the source signal.

Analog and digital relaying protocols are among the various cooperative signal processing methods exploited at the relay terminals to provide spatial diversity gain via a virtual antenna array [2-3]. In analog relaying protocols, an amplified copy of the signal from the source is just scaled and forwarded without decoding while in digital relaying protocols the relay nodes detect, re-modulate and re-transmit the received signal. The main drawbacks of analog and digital relaying based wireless cooperative communication strategies are the noise amplification at intermediate nodes and error propagation due to the detection errors at the relay terminals respectively. On the other hand, elimination of the requirement of expensive RF chains in practice and their proneness for using with coding techniques and network protocols are the main superior aspects of digital relaying techniques over analog relaying approaches. In this paper, the discussion is restricted to the digital relaying.

Decode and forward (DCF), compress and forward (CF) and demodulate and forward (DMF) protocols are among the prominent digital relaying techniques. CF and DCF protocols require highly complex encoding, decoding and compression processes at the relay terminals while in DMF protocol the relay terminals carry out only demodulation process on the received signal. The DMF approach, which significantly reduces system complexity, power and time consumption, has been used in the literature with several channel codes of different types. In [4] and [5], DMF protocol is used with Low Density Generator Matrix [6] and Repeat Accumulate codes [7] respectively. The performance of DMF protocol is considered in [8] for incremental relaying where relays help via the destination feedback only when necessary. The impact of uncertain channel state information (CSI) on the performance of DMF relaying protocols with higher order modulation formats is studied in [9]. In [10], a maximum likelihood decoder of the distributed Alamouti code for the DMF based cooperative system is derived.

After the publication of the seminal article [11], network coding has become one of the important frameworks in wireless cooperative systems which depends on the principal idea of transmitting smart combinations of the received data from multiple source nodes over the network [12-13]. A practical reference scenario utilizing ideas from network coding is the so-called Multiple Access Relay Channel (MARC) where multiple user terminals transmit independent information to a common destination with the help of one or multiple relay nodes. The MARC model has attracted a lot of interest due to its potential applications in ad hoc networks and uplink transmission for both 4G and 5G cellular systems. In [14] the error rate performance and optimal power allocation of a class of network coding schemes known as analog network coding is studied for MARC model. A code design procedure is considered in [15] for a kind of asymmetric MARC in which the users and relay have different channel conditions. In [16],

authors have investigated linear network coding construction at the relay in the MARC system so as to minimize the average packet loss rate at the destination.

In this study, performance of DMF protocol is investigated in MARC model where convolutional coding is used as the baseline code. The rest of this paper is organized as follows. In Section-II, the system model is given. Decoding Algorithm employed at the receiver which is based on trellis combining is defined in Section-III. The mode selection algorithm is given in Section-IV. Performance analysis is provided in Section V. Section VI contains numerical results. Finally, the concluding remarks are provided in Section VII.

II. System Model

The two-user MARC network shown in Figure 1 in which the users share a relay in transmitting data to the common destination applying DMF strategy is considered. All of the terminals are equipped with a single antenna and operate in half-duplex mode i.e., they cannot transmit and receive at the same time using the same frequency bandwidth.

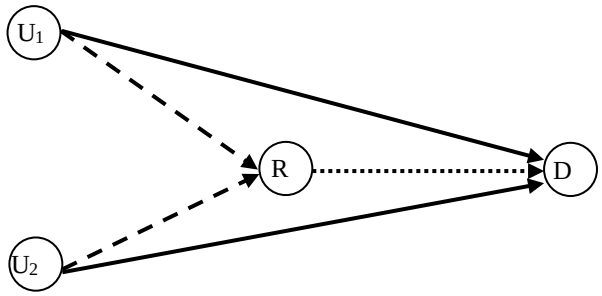


Figure 1: Multiple Access Relay Channel Model.

A (p, k) convolutional code with rate R = p/k is employed at the users to encode the user data stream $d_i = \left[d_i^1 \quad d_i^2 \quad \dots \quad d_i^L \right]$ (i = 1, 2) with length N which is splitted into data blocks $d_i^j = \left[d_{i,1}^j \quad d_{i,2}^j \quad \dots \quad d_{i,k}^j \right]$ (j = 1, 2, ... , L) with length L = N/k. The data blocks are applied to the convolutional encoder where k and p are the number of input and output bits of the encoder respectively. The codeword at the output of the encoder for input data block d_i^j is represented by $c_i^j = \left[c_{i,1}^j \quad c_{i,2}^j \quad \dots \quad c_{i,p}^j \right]$ while the codeword stream for input the data stream d_i can be denoted by $c_i = \left[c_i^1 \quad c_i^2 \quad \dots \quad c_i^L \right]$. $x_i = \left[x_i^1 \quad x_i^2 \quad \dots \quad x_i^L \right]$ is the symbol stream which is obtained from symbols blocks $x_i^j = \left[x_{i,1}^j \quad x_{i,2}^j \quad \dots \quad x_{i,p'}^j \right]$ (j = 1, 2, ... , L) that corresponds to the codewords c_i^j. Here $A_M = \exp\left(j2\pi m / M \right)$ (m = 0, 1, …, M-1) is the M-PSK constellation used in transmission and $p' = p / \log_2 M$. In this study, BPSK is used for modulation.

We consider a quasi-static flat Rayleigh fading scenario where the channel gains h_{ij} (i ∈ {U_1, U_2, R}, j ∈ {R, D}) are zero-mean complex Gaussian random variables with variance $\sigma_{ij}^2 / 2$ per dimension which remain constant during transmission of a block and vary independently in the next transmission. It is assumed that the channel gains are perfectly known by the corresponding receiver terminals and the destination terminal knows all of the channel gains $(h_{U_1R}, h_{U_2R}, h_{U_1D}, h_{U_2D}, h_{RD})$ in the system. The communication protocol is divided into three phases with equal time duration where the users broadcast their data blocks in sequence during the first two phases and the third phase is allotted to the relay terminal. The additive white Gaussian noise at terminal j in phase p (j ∈ {R, D}, p ∈ {1, 2, 3}) is denoted by z_j^p and modeled as an independent zero-mean complex Gaussian random variable with variance $N_0 / 2$ per dimension. The instantaneous and average SNR between terminal i and j are defined as $\gamma_{ij} = E_s \left| h_{ij} \right|^2 / N_0$ and $\overline{\gamma}_{ij} = E_s \sigma_{ij}^2 / N_0$ respectively, where E_s denotes the transmitted energy per symbol.

In the first phase, the user U_1 transmits its data and the received signals at the relay and destination can be written as

$$y_D^1(n) = h_{U_1D} \sqrt{E_s} x_1(n) + z_D^1(n), \tag{1}$$

$$y_R^1(n) = h_{U_1R} \sqrt{E_s} x_1(n) + z_R^1(n) \tag{2}$$

respectively. Similarly, in the second phase U_2 broadcasts its own encoded data sequence and we model the received signals by the relay and destination as

$$y_R^2(n) = h_{U_2R} \sqrt{E_s} x_2(n) + z_R^2(n) \tag{3}$$

$$y_D^2(n) = h_{U_2D} \sqrt{E_s} x_2(n) + z_D^2(n), \tag{4}$$

respectively. The third phase of the communication protocol depends on the detection method. In maximum likelihood (ML) detection [17] based MARC, after demodulating the received signals in the first and second phase with $\hat{x}_{U_i}(n) = \mathrm{sgn}\left\{ \mathrm{Re}\left\{ y_R^i(n) h_{U_iR}^* \right\} \right\}$, (i ∈ {1, 2}), the network encoded BPSK symbol stream obtained by $x_R(n) = \hat{x}_{U_1}(n) \hat{x}_{U_2}(n)$ is transmitted to the destination node. Here, the operators sgn(.), Re(.) and (.)* indicate sign of a real number, real part of complex expression and complex conjugate respectively. Then, the received signal at the destination in the third phase can be given as

$$y_D^3(n) = h_{RD} \sqrt{E_s} x_R(n) + z_D^3(n) \tag{5}$$

In user selective relaying based approach [18], the information of the some of the user terminals are transmitted to the destination through both direct and relay channel utilizing ideas from network coding while the other terminals employ only the direct link. The optimal subset of all user terminals whose transmissions will be listened and retransmitted by the relay are determined by the destination terminal in a manner to minimize the overall end to end bit error rate (BER) of the system and sent to the relay through the feedback link [18].

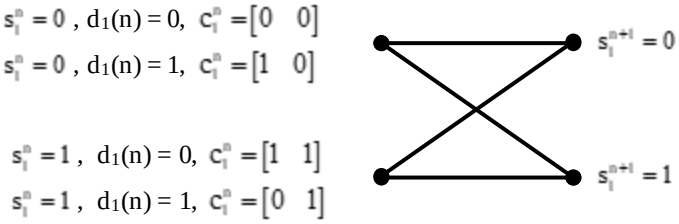
When user selective relaying is utilized in a MARC model with two user terminals, four different modes m_0, m_1, m_2 and m_3 can be described for the relay. In mode m_0, the relay makes no transmission at the last phase and the communication is accomplished through the direct links in the first and second phase. In mode m_1, the relay listens only the transmission of U_1 and re-transmits only the data of U_1 while in mode m_2, the relay receives and broadcasts only the data sequence of U_2. In mode m_3, the relay listens the transmissions of both U_1 and U_2, therefore broadcasts the combined packet in the last phase. In this case, the network encoded symbol $x_R(n)$ can be given for user selective relaying as

$$x_R(n) = \begin{cases} 0 & \text{for } m_0 \\ \hat{x}_{U_1}(n) & \text{for } m_1 \\ \hat{x}_{U_2}(n) & \text{for } m_2 \\ \hat{x}_{U_1}(n)\hat{x}_{U_2}(n) & \text{for } m_3. \end{cases} \quad (6)$$

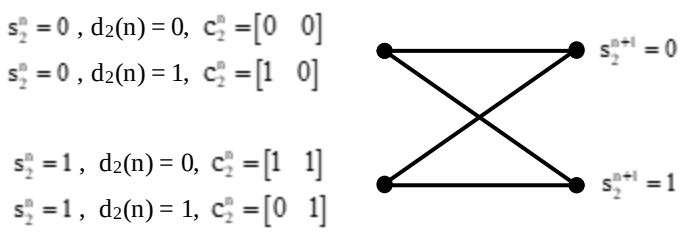
III. Trellis Combining Based Joint Channel-Network Decoding for DMF Protocol

In order to be able to explain the trellis combining concept for convolutional encoding more clearly, a transmission scenario where the users encode their data utilizing the simple R = 1/2 convolutional code with generator matrix $(3, 1)_{octal}$ shown in Figure 2 is considered. The destination obtains users' data by applying Viterbi or MAP algorithm to the new trellis formed by employing a technique called trellis combination in which both the direct and relay signals are utilized in calculation of transition probabilities.

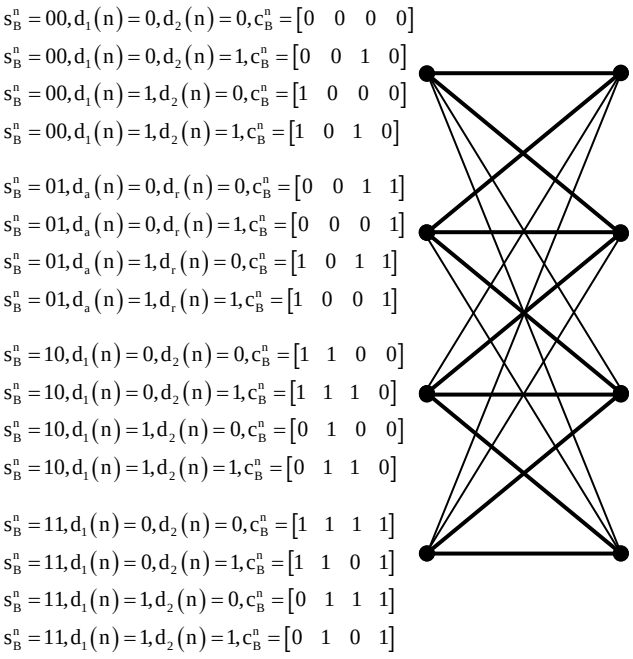
The trellis combination approach provides a method to represent the encoder transitions of both users in a single diagram. The combined trellis shown in Figure 2 has four states represented by $s_B^n = s_1^n s_2^n \in \{00, 01, 10, 11\}$. In this state enumeration, the first and the second numbers s_1^n and s_2^n correspond to trellis states $s_1^n, s_2^n \in \{0,1\}$ at the user U_1 and U_2 respectively. For instance, the transition $s_B^n = 00 \rightarrow s_B^{n+1} = 10$ on the combined trellis corresponds to the transition from state 0 to the next state 1 in the trellis of user U_1 and the transition from state 0 to the next state 0 in the trellis that corresponds to U_2 for a given time instant. The probability of this transition can be calculated by



(i) The trellis diagram of convolutional code for user U_1



(ii) The trellis diagram of convolutional code for user U_2



(iii) Trellis Combinations

Figure 2: The procedure for trellis combining

$$p\left(s_B^n = 00 \rightarrow s_B^{n+1} = 10\right) = p\left(s_1^n = 0 \rightarrow s_1^{n+1} = 1\right) p\left(s_2^n = 0 \rightarrow s_2^{n+1} = 0\right) \times p\left[x_R(2n-1) = -1 \mid x_1(2n-1) = 1, x_2(2n-1) = -1\right] \times p\left[x_R(2n) = 1 \mid x_1(2n) = 1, x_2(2n) = 1\right] \quad (7)$$

where $p\left(s_1^n = 0 \rightarrow s_1^{n+1} = 1\right)$ is the probability of transition from state 0 to state 1 in user terminal U_1, $p\left(s_2^n = 0 \rightarrow s_2^{n+1} = 0\right)$ is the probability of transition from state 0 to state 0 in user terminal U_2 at time instant n. These probabilities can be given by

$$p\left(s_1^n = 0 \rightarrow s_1^{n+1} = 1\right) = p\left[d_1(n) = 1\right] \times p\left[y_D^1(2n-1) \mid x_1(2n-1) = -1\right] \times p\left[y_D^1(2n) \mid x_1(2n) = 1\right] \quad (8)$$

$$p\left(s_2^n = 0 \rightarrow s_2^{n+1} = 0\right) = p\left[d_2(n) = 0\right] \times p\left[y_D^2(2n-1) \mid x_2(2n-1) = 1\right] \times p\left[y_D^2(2n) \mid x_2(2n) = 1\right] \quad (9)$$

$$p\left[x_R(2n-1)=-1 \mid x_1(2n-1)=1, x_2(2n-1)=-1\right]= \\ (1-\varepsilon_R)\,p\left[y_D^3(2n-1) \mid x_R(2n-1)=-1\right]+ \\ \varepsilon_R\,p\left[y_D^3(2n-1) \mid x_R(2n-1)=1\right] \quad (10)$$

$$p\left[x_R(2n)=1 \mid x_1(2n)=1, x_2(2n)=1\right]= \\ (1-\varepsilon_R)\,p\left[y_D^3(2n) \mid x_R(2n)=1\right]+ \\ \varepsilon_R\,p\left[y_D^3(2n) \mid x_R(2n)=-1\right] \quad (11)$$

The term ε_R in expressions (10) and (11) is the probability of erroneous transmission from the relay which can be expressed by

$$\varepsilon_R = \varepsilon_1(1-\varepsilon_2)+\varepsilon_2(1-\varepsilon_1) \quad (12)$$

Here, ε_1 and ε_2 are the probabilities of error in the first and second phase transmissions respectively which are calculated with $\varepsilon_1 = Q\left(\sqrt{2\gamma_{U_1R}}\right)$ and $\varepsilon_2 = Q\left(\sqrt{2\gamma_{U_2R}}\right)$ for BPSK modulations.

IV. The Mode Selection Algorithm in User Selective Relaying Based Marc Model

In this section, the mode selection algorithm that determines the transmission strategy of the relay in the last phase is defined.

Let $\bar{\mathbf{d}}_{U_i} = \left[\bar{d}_{U_i}(1) \;\; \bar{d}_{U_i}(2) \;\; \ldots \;\; \bar{d}_{U_i}(L)\right]$ ($i \in \{1, 2\}$) be the data block of user U_i obtained at destination. The decision rule for the transmission strategy of the relay in the last phase can be given by

$$m_k = \arg\min_{k\in\{0,1,2,3\}}\left(P_{U_1}^k + P_{U_2}^k\right) \quad (13)$$

where $P_{U_1}^k$ and $P_{U_2}^k$ ($k \in \{0, 1, 2, 3\}$) are the BERs of the users U_1 and U_2 at the destination terminal for uncoded transmission respectively for m_k strategy. Note that taking BER of uncoded transmission instead of BER of coded scheme does not effect the mode selection procedure since the amount of the decrease in BER is the same for each of the mode probability. In (13), $P_{U_1}^k$ and $P_{U_2}^k$ can be given by

$$P_{U_1}^0 = Q\left(\sqrt{2\gamma_{U_1D}}\right),\; P_{U_2}^0 = Q\left(\sqrt{2\gamma_{U_2D}}\right) \quad (14)$$

$$P_{U_1}^1 = \varepsilon_1 P_{prop}^{U_1} + (1-\varepsilon_1)P_{coop}^{U_1},\; P_{U_2}^1 = Q\left(\sqrt{2\gamma_{U_{S_2}D}}\right) \quad (15)$$

$$P_{U_1}^2 = Q\left(\sqrt{2\gamma_{U_1D}}\right),\; P_{U_2}^2 = \left[\varepsilon_2 P_{prop}^{U_2} + (1-\varepsilon_2)P_{coop}^{U_2}\right] \quad (16)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \exp(-t^2/2)dt$, $P_{coop}^{U_i}$ and $P_{prop}^{U_i}$ are the bit error rates of user terminals U_i ($i \in \{1, 2\}$) at the destination terminal when the relay makes an correct or erroneous decision respectively. These probabilities can be calculated by [19]

$$P_{U_i}^{coop} = Q\left(\sqrt{2\left(\gamma_{U_iD}+\gamma_{U_iR}\right)}\right) \quad (17)$$

$$P_{U_i}^{prop} = Q\left(\frac{\gamma_{U_iD}-\gamma_{U_iR}}{\sqrt{\left(\gamma_{U_iD}+\gamma_{U_iR}\right)/2}}\right). \quad (18)$$

When a network encoded packet is transmitted from the relay (m_3 strategy), the bit error rate at the destination can be given as [18]

$$P_{U_i}^3 = \sum_{\substack{\bar{d}_{U_{S_i}}(n),d_{U_{S_i}}(n),d_R(n)\in\{0,1\} \\ d_{U_{S_i}}(n)\neq\bar{d}_{U_{S_i}}(n)}} \Pr\left\{d_R(n) \mid d_{U_i}(n), d_{U_2}(n)\right\}\times \\ \Pr\left\{\bar{d}_{U_1}(n), \bar{d}_{U_2}(n) \mid d_{U_i}(n), d_{U_2}(n), d_R(n)\right\} \quad (19)$$

where

$$\Pr\left\{\bar{d}_{U_{S_1}}(n), \bar{d}_{U_{S_2}}(n) \mid d_{U_{S_1}}(n), d_{U_{S_2}}(n), d_R(n)\right\} = \\ \Pr\left\{\gamma_S\left[\bar{d}_{U_{S_1}}(n), \bar{d}_{U_{S_2}}(n), \bar{d}_{U_{S_1}}(n)\oplus\bar{d}_{U_{S_2}}(n)\right] \mid \gamma_S\left[d_{U_{S_1}}(n), d_{U_{S_2}}(n), d_R(n)\right]\right\} \quad (20)$$

$$\gamma_S(a,b,c) = (-1)^a\gamma_{RD} + (-1)^b\gamma_{U_{S_1}D} + (-1)^c\gamma_{U_{S_2}D} \quad (21)$$

$$\Pr\left\{d_R(n) \mid d_{U_{S_1}}(n), d_{U_{S_2}}(n)\right\} = \begin{cases} 1-\varepsilon_R & \text{for } d_R(n) = \tilde{d}_R(n) \\ \varepsilon_R & \text{for } d_R(n) \neq \tilde{d}_R(n) \end{cases} \quad (22)$$

$$\tilde{d}_R(n) = d_{U_{S_1}}(n)\oplus d_{U_{S_2}}(n) \quad (23)$$

V. Performance Analysis

In this section, BER performance analysis of the MARC scheme is investigated for DMF protocol in case of convolutional encoding with trellis combining approach.

The BER of the proposed model P_E can be written as

$$P_E = \sum_{k=0}^{3} P_{M_k}\left(\hat{P}_{U_1}^k + \hat{P}_{U_2}^k\right) \quad (24)$$

where $k \in \{0, 1, 2, 3\}$, P_{M_k} is the probability of selection of mode m_k, $\hat{P}_{U_1}^k$ and $\hat{P}_{U_2}^k$ are the BERs of the users U_1 and U_2 at the destination terminal for coded transmission respectively. The instantaneous values of $\hat{P}_{U_1}^0$ and $\hat{P}_{U_2}^0$ be given as

$$\hat{P}_{U_1}^0 < \sum_{d=d_{free}}^{\infty} B_d \left[2\sqrt{p_1(1-p_1)}\right]^d, \hat{P}_{U_2}^0 < \sum_{d=d_{free}}^{\infty} B_d \left[2\sqrt{p_2(1-p_2)}\right]^d \quad (25)$$

where $p_1 = Q\left(\sqrt{2\gamma_{U_1D}}\right)$ and $p_2 = Q\left(\sqrt{2\gamma_{U_2D}}\right)$ for BPSK constellation, d_{free} is the free distance of the convolutional code and B_d are the coefficients of the bit weight enumerating function (BWEF) of the baseline convolutional code [20].

$\hat{P}_{U_1}^2 = \hat{P}_{U_1}^0$ and $\hat{P}_{U_2}^1 = \hat{P}_{U_2}^0$ since the data of U_1 and U_2 are sent to destination directly in mode m_2 and mode m_1, respectively. The other probability terms can be calculated by

$$\hat{P}_{U_1}^1 < \sum_{d=d_{free}}^{\infty} B'_d \left[2\sqrt{p_3(1-p_3)}\right]^d, \hat{P}_{U_2}^2 < \sum_{d=d_{free}}^{\infty} B'_d \left[2\sqrt{p_4(1-p_4)}\right]^d \quad (26)$$

where B'_d are the coefficients of the bit weight enumerating function (BWEF) of the combined code, [17]

$$p_3 = p_3^1 + p_3^2 + p_3^3 \quad (27)$$

$$p_3^1 = \Pr\left\{t_0(n) < T_1, t_1(n) < -T_1\right\} \quad (28)$$

$$p_3^2 = \Pr\left\{t_0(n) < -T_1, t_1(n) > T_1\right\} \quad (29)$$

$$p_3^3 = \Pr\left\{t_0(n) + t_1(n) < 0, -T_1 \le t_1(n) \le T_1\right\} \quad (30)$$

$$t_0(n) = \mathrm{Re}\left\{\frac{2y_D^1(n)\left[y_D^1(n-1)\right]^*}{N_0}\right\} \quad (31)$$

$$t_1(n) = \mathrm{Re}\left\{\frac{2y_D^3(n)\left[y_D^3(n-1)\right]^*}{N_0}\right\} \quad (32)$$

$$T_1 = \ln\left[(1-\varepsilon_1)/\varepsilon_1\right] \quad (33)$$

Although the probability term p_4 is calculated by using a same way given by the expressions in (27-33), the terms $t_0(n)$ and T_1 are calculated by

$$t_0(n) = \mathrm{Re}\left\{\frac{2y_D^2(n)\left[y_D^2(n-1)\right]^*}{N_0}\right\} \quad (34)$$

$$T_1 = \ln\left[(1-\varepsilon_2)/\varepsilon_2\right] \quad (35)$$

in determination of the term $\hat{P}_{U_2}^2$.

VI. Simulation Results

In this section, computer simulation results for performances of DMF protocol based MARC models are presented in quasi-static flat Rayleigh fading channels. In these computer simulations, the data block length is chosen as L = 100. Two types of channel models, symmetric and asymmetric, have been investigated in the simulations performed. In the symmetric channel model, variances of all links are taken as one. In the first asymmetric channel model, we choose the variances of the channels as $\sigma_{U_1D}^2 = \sigma_{U_2D}^2 = \sigma_{RD}^2 = 1$, $\sigma_{U_1R}^2 = \sigma_{U_2R}^2 = 10^3$ for investigation of the case in which the distance between users and relay is too small with respect to the distance between relay and destination. On the other hand, in the second asymmetric channel model, we investigate the case where the distance between relay and destination is too small with respect to the distance between relay and user terminals by selecting the channel variances as $\sigma_{U_1D}^2 = \sigma_{U_2D}^2 = \sigma_{U_1R}^2 = \sigma_{U_2R}^2 = 1$ and $\sigma_{RD}^2 = 10^3$. The proposed relaying schemes are compared with the conventional noncooperative system that involves direct transmission between the source and the destination. For fair comparison we set E_S = 2/3 and E_S = 1 for MARC schemes and direct transmissions respectively. Performance comparisons have been carried out at a bit error rate of 10^{-3}.

It is seen from the performance curves given in Figure 3 for ML based detection that an apparent increase in slope exists with respect to direct transmission and full diversity order is achieved in case of both symmetric and asymmetric channel models. The best performance is obtained in case of asymmetric scenario where the relay is close to the users. The SNR gains provided by this asymmetric channel model with respect to the asymmetric channel model where the relay is close to the destination and symmetric scenario are approximately 1 and 1.5 dB respectively.

In Figure 4, performance curves for user selective relaying based MARC model are proposed. It is seen from these curves that full diversity gain is obtained for user selective relaying.

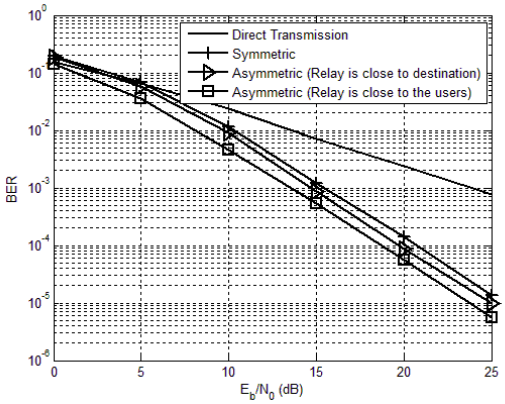


Figure 3: BER performance results for ML detection

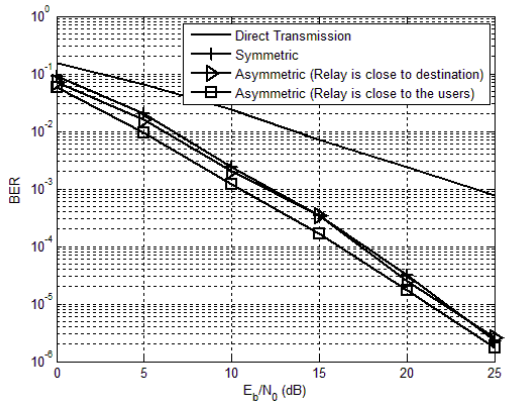


Figure 4: BER performance results for user selective based detection

The SNR gain in case of user selective relaying with respect to the ML based detection where the relay is always transmits network encoded data is about 2 dB. The best performance is obtained for asymmetric case where relay is close to destination.

VII. CONCLUSION

DMF approach which significantly reduces not only the system complexity but also the power and time consumption is applied to MARC communication scenario. A convolutional code is employed as the channel code. A decoding algorithm based on trellis combining is presented. The simulation results have demonstrated that the proposed decoding algorithm based on trellis combining provides full diversity gain for both ML detection and user selective based MARC scenario.

REFERENCES

[1] T. M. Cover and A. A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Info. Theory*, vol. 25, no. 5, pp. 572–584, Sept. 1979.

[2] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity Part I and Part II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[3] J. N. Laneman, G. W. Wornell, and D. N. C. Tse, "An Efficient Protocol for Realizing Cooperative Diversity in Wireless Networks," Proc. *IEEE ISIT*, Washington, DC, June 2001, pp. 294.

[4] A. W. Eckford, J. Chu and R. Adve, "Low complexity cooperative coding for sensor networks using rateless and LDGM Codes," *IEEE International Conference on Communications*, 2006, pp. 1537–1547.

[5] A. W. Eckford and R. Adve, "A practical scheme for relaying in sensor networks using repeat-accumulate codes," *Conference on Information Sciences and Systems-CISS*, 2006, pp. 386–381.

[6] T. R. Oenning and J. Moon, "A low density generator matrix interpretation of parallel concatenated single bit parity codes," *IEEE Transactions on Magnetics*, vol. 37, no. 2, pp. 737–741, Mar. 2001.

[7] D. Divsalar, H. Jin, and R. J. McEliece, "Coding theorems for 'Turbo-Like' codes," *36th Annual Allerton Conf. on Comm., Control, and Computing*, 1998, Monticello-U.S.A, pp. 201–210.

[8] K.-H. Liu, "On the Performance of Time-Orthogonal Incremental Relaying Based on Demodulate-and-Forward With Distributed Channel Access," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 2, pp. 737–747, February 2012.

[9] R. Annavajjala, A. Maaref and J. Zhang, "Demodulate-and-Forward Relaying with Higher Order Modulations: Impact of Channel State Uncertainty," *IEEE International Conference on Communications*, 2010, pp. 1–5.

[10] A. Bansal, M. R. Bhatnagar and A. Hjorungnes, "Decoding and Performance Bound of Demodulate-and-Forward Based Distributed Alamouti STBC," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 702–713, February 2013.

[11] R. Ahlswede, N. Cai, S. Y. R. Li and R. W. Yeung, "Network Information Flow," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[12] R. Bassoli, H. Marques, J. Rodriguez, K. W. Shum and R. Tafazolli, "Network Coding Theory: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 4, pp. 1950–1978, Fourth Quarter 2013.

[13] S. T. Başaran, G. K. Kurt, M. Uysal, and İ. Altunbas, "A Tutorial on Network Coded Cooperation," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2970–2990, Fourth Quarter 2016.

[14] K. Ntontin, M. D. Renzo, A. I. Pérez-Neira, and C. Verikoukis, "Analog Network Coding in the Multiple Access Relay Channel: Error Rate Analysis and Optimal Power Allocation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3015–3032, June 2015.

[15] L. Liu, Y. Li, C. Yuen, Y. L. Guan and Y. Sun, "Joint Source-Channel Coding for Asymmetric Slepian-Wolf Multiple Access Relay Channel," *IEEE Wireless Communications Letters*, vol. 6, no. 5, 2017, pp. 642–645.

[16] M. Dai, H. Y. Kwan, and C. W. Sung, "Linear Network Coding Strategies for the Multiple Access Relay Channel with Packet Erasures," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 218–227, January 2013.

[17] P. Liu and I. Kim, "Average BER analysis for binary signallings in decode-and-forward dissimilar cooperative diversity networks," IEEE Trans. Wireless Commun., vol. 8, no. 8, pp. 3961-3968, August 2009.

[18] Ö. Özdemir, "User selective relaying in multiple access relay channels," IEEE 20th International Conference on Software, Telecommunications and Computer Networks, pp. 1–5, Sept. 2012.

[19] F. A. Onat, A. Adinoyi, Y. Fan, H. Yanikomeroglu, J. S. Thompson and I. D. Marsland, "Threshold Selection for SNR-based Selective Digital Relaying in Cooperative Wireless Networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 11:4226-4237, 2008.

[20] S. Lin and D. J. Costello, "Error Control Coding, Fundementals and Applications", Pearson Educational International, 2nd ed, 2004.

A Review: Mobile Communication Past, Present Future

M.ÇAKMAK[1] and Z. ALBAYRAK[2]

[1] Karabuk University, Karabuk/Turkey, muhammetcakmak@karabuk.edu.tr
[2] Karabuk University, Karabuk/Turkey, zalbayrak@karabuk.edu.tr

Abstract **- Nowadays, the content used on the internet is increasingly provided by mobile technologies. Mobile communication, which started with only voice transmission and then continued to undergo a major transformation with messages, mails, images and videos. As the amount of data accessed from mobile media increases day by day, researchers continue to be interested in the network access area, network access speed, data download/upload speed, and data security areas in mobile technologies. In this study, we explained First Generation (1G), Second Generation (2G), Third Generation (3G), Forth Generation (4G) and Fifth Generation (5G). Finally compered them briefly.**

Keywords **– 1G, 2G, 3G, 4G, 5G, Mobile Communication**

I. INTRODUCTION

MOBILE communications technology has undergone a very rapid transformation in recent years. The basic contribution in this growth of mobile communication is increased data requirement. The changes in mobile communication technology, which undergo a rapid change, are expressed as generations. These generations are 1G, 2G, 3G, 4G and 5G. Every generation has some new values, techniques, structure, capacity values and innovations that distinguish it from other generations.

First Generation (1G) mobile communications technology begins with the communication of analog signals. With Second Generation (2G), mobile communications have turned into digital signals. With 2G, innovations such as high sound quality, encryption and transmission of message data have emerged. It started data transmission in mobile communication with General Packet Radio Service (GPRS) 2.5 G and Enhanced Data Rates for GSM Evolution (EDGE) 2.75G. With the Third Generation (3G), known as the Universal Mobile Telephone System (UMTS), speed data transmission, video calling and mobile internet access have gained momentum. With High Speed Packet Access (HSPA), also known as 3.5G and Evolved High Speed Packet Access (HPSA +), also known as 3.75G, low latency and high speed data have been achieved for faster internet. WiMAX, Long Term Evolution (LTE), Long Term Evolution Advanced (LTE-A) which are fourth generation (4G) technologies, can be used for high speed applications such as High Definition TV, video conferencing, 3D TV. With the Fifth Generation (5G), which is expected to be officially launched in 2020, the internet of objects, smart home, driverless car, high resolution data transfer drone, very fast data transfer with cloud technology, virtual reality and enhanced reality, Industry 4.0 and remote surgical applications area will continue to be affected.

In this research study, we have prepared a detailed review of the different generations of mobile communication systems. In this study, all mobile communication technologies from 1G to 5G are examined in detail and the obtained data are shown comparatively on the tables.

II. FIRST GENERATION (1G)

The first generation mobile communication technology, 1G, was first introduced in the US by Advance Mobile Phone System (AMPS). [1] 1G has used analog data communications for voice calls up speeds of 2.5 Kbps. [2] 1G allows users to make voice calls in only one country. 1G had got some problems about communication infrastructure. It had small capacity, unreliable handoff and feeble voice connection. [3] 1G uses main network as Public Switched Telephone Network (PSTN) and circuit switching technology. 1G has got 800-900 MHz frequency rate. Table 1 shows the properties of 1G.

Table 1: Properties of 1G [4]

Emerge	1970-1984
Data Volume	2.5 Kbps
Data Sending Technique	Analog
Standarts	AMPS
Multiplexing Technlogy	FDMA System
Switching Technology	Circuit Technology
Applications	Only voice data
Forward Erroe	PSTN Network
Frequency	800-900 MHz

III. SECOND GENERATION (2G) - GSM (GLOBAL SYSTEM FOR MOBILE)

The 2G systems used for more audio applications, designed in the 1980s, were established on digital technology, including digital signal cultivationing techniques. [5] 2G technology is also known as Global System for Mobile (GSM). When it was first used, GSM systems offered the 25 MHz frequency spectrum at 9000 MHz. In 1991, 2G cellular telecom networks in Finland were commercially marketed under the GSM

standard. [6] With 2G technology, phone calls were digitally encrypted. This digital encryption allowed the data to be retrieved and read only by the intended recipient. 2G technologies use Time Division Multiple Access (TDMA) or Code Division Multiple Access (CDMA) techniques. [7] Today's GSM systems work in the 900 MHz and 1800 MHz bands worldwide, except for the Americas where they operate on the 1900 MHz band. In Europe, GSM technology has provided uninterrupted roaming in all countries. [6]

2G technologies handle some data capabilities such as sending fax and short message up to 9.6 Kbps, but 2G is not adequate for web browsing and the other multimedia applications. The omnipresence of the GSM standard reach international roaming common between mobile phone operators, permitting subscribers to use their mobile phones in many parts of the world. [8] 2G technology allows batteries to last longer because digital signals consume less battery power. The use of digital signals provides privacy and security in sending data and voice calls. [9] Comparison of 1G to 5G systems is shown in Table 2 below.

Table 2: Comparison of 2G & 5G [10]

2G	5G
Used after 1990	Emerge will come till 2020
It is successor to 1G	It will used after 4G
Use digital data technique	Establish on internet protocol IPv6
Especially used for mobile connections	Mobile communication and wireless internet
Don't use quality of service	Use high quality of service
Handsets are provided with messaging facility (SMS, MMS and picture message)	Handsets are provided with ultrahigh HD video calling
Don't take multiple application simultaneously	Take unlimited number of parties simultaneously
Band width per frequency up to 200Khz	Band frequency channel 5Ghz -28Ghz
Network type is PSTN	Network type is internet
Speed limit up to 64Kbps	Speed limit up to 1Gbps

IV. Second Generation (2.5G) - GPRS (General Radio Packet Service)

2.5G is a cellular wireless technology build up in between its precursor, 2G and its successor, 3G. "2.5G" is an unofficial label, manufactured solely for marketing purposes, different from "2G" or "3G" which are formally describe standards based on those delimitate by the International Telecommunication (ITU). [11] General Radio Packet Service (GPRS) as known a 2.5 G technology that can render data rates from 56 Kbps up to 115 Kbps. [12] Also 2.5G technology supports WAP, MMS, SMS services. [13]

V. Second Generation (2.75G) - EDGE (Enhanced Data rates for GSM Evolution)

EDGE is an shortening for Enhanced Data rates for GSM Evolution. It is the enriched generation of GPRS technology. Also EDGE works in GSM networks. EDGE provides faster transfer of data and information. EDGE technology was invented by AT&T. EDGE technology is give wayed over GSM owing to its resilience to provide packet switch data and circuit switch data. EDGE transfers data in slight seconds compare to GPRS Technology. For instance a normal text file of 40KB is transferred in merely 2 seconds qua compared to the transfer from GPRS technology, which is 6 seconds. The most important benefit of using EDGE technology is one does not need to install any additional hardware and software on account of make use of EDGE Technology. Users were not charged any additional fees due to the use of EDGE technology.

VI. Third Generation (3G)

3G was established in late 2000. [14] It serves transmission rate up to 2Mbps. 3G is the mobile communication technology that emerges after 2G technology. 3G has emerged due to International Telecommunication Union (ITU) standards. 3G offers a broader range of services to network operators and enables greater network capacity with higher spectral efficiency. [15] With 3G services, high-quality voice calls, video calls and broadband wireless data transmission are available on mobile area networks. Spectral efficiency refers to the amount of data that can be transferred over a digital system. HPSA is a mobile communications technology that extend and enhances the performance of Universal Mobile Telecommunications System (UMTS) technology. High Speed Packet Access (HSPA) allows access to downlink of 14 Mbps and uplink to 5.8 Mbps. [16]

3G technology is based on TDMA and CDMA technology. 3G enables the use of GPS (Global Positioning System) and video conferencing services. [14] 3G can directly and indirectly support CDMA, TDMA, FDMA, IMTS-DS (direct spread) and IMT-MC (multi-carrier) technologies. 3G technology is also compatible with 2G technology. The goal of 3G is to provide more coverage and widespread penetration with the optimal investment cost. [17]

VII. Third Generation (3.5G-3.75G) HSDPA (High-Speed Downlink Packet Access) and HSUPA (High-Speed Uplink Packet Access)

High-Speed Downlink Packet Access (HSDPA) is a UMTS-based 3G technology that allows for high data rates. The WCDMA infrastructure used by HSDPA allows for more bandwidth than 5MHz. [18]

3.75G technology is used to denote the future of 3G technology. High Speed Uplink Packet Access (HSUPA) is a UMTS / WCDMA technology. The HSUPA technology is directly related to HSDPA. HSUPA can provide services for applications requiring high data rates, such as advanced applications, mobile email services, and real-time games. HSUPA enables speeds from 1.4 Mbps to 5.8 Mbps with UMTS

/ WCDMA technology. [16] Comparison of 3G to 5G systems is shown in Table 3 below.

Table 3: Comparison of 3G & 5G. [19]

3G	5G
High capacity broad band data connection	Base on internet protocol 6
Data transfer speed is up to 30 Mbps	Data transfer speed is up to 10/50 Gbps
Use digital navigation	Use virtual private network
Use WCDMA, UMTS, CDMA 2000 technology	Use Beam Division Multiple Access (BDMA) or Filter Bank Multi Carrier (FBMC) technology
Frequency band: 800/850/900/1800/1900/2100 Mhz	Frequency band: 1.8/2.6 Gz Expected: 30/300Ghz
Seamless roaming	IP based mobility
Voice, Data, Video Calling	Ultra High Definition Video, Virtual Realty

VIII. FORTH GENERATION (4G)

The first trial for 4G was conducted in Tokyo, Japan, 2005. 4G technology is an IP based network system. 4G offers a download speed of 100 Mbps. [6] The main objective of 4G technology is to provide services such as voice over IP, large data and internet with high quality, capacity and cost reduction. With 4G, current and existing network communication techniques are brought together to provide free roaming among cellular networks. The connection speed is 100 Mbps on mobile phones and 1 Gbps on Wi-Fi networks. [20] It is also the same size as the WiMAX band width. 4G ensure not only telecommunications services, but also data and multimedia services.

Pre-4G technologies like mobile WiMAX and first-release 3G Long Term Evolution (LTE) have been permitting on the market since 2006 and 2009 respectively. [21] When 4G technology is compared with HSUPA, HSDPA or WLAN technology, the following approach can be taken.

3G and HSDPA and HSUPA < 4G

3G and WLAN <4G

3G and HSDPA and WLAN <4G

The aim of 4G replace to the entire core of cellular networks with a single worldwide cellular network completely standardized based on the IP for video, Voice over IP (VoIP) and multimedia services.

Some applications with 4G; [22]

1. With mobile TV applications, a direct TV channel can be accessed via the subscriber's phone.
2. The subscriber can easily send a video over the phone to another user
3. Subscribers can make video conferences with each other.
4. With location based services, the subscriber can be informed about the user and traffic situation.
5. Mobile ultra broadband access and multi-carrier transmission
6. Mobile WiMAX technology

Comparison of 3G to 5G systems is shown in Table 4 below.

Table 4: Comparison of 4G & 5G [23]

4G	5G
4G has used since 2000	5G will active in 2020
4G provides dynamic information access for wearable devices	5G provides dynamic information access, wearable devices with AI capabilities
Data transfer speed is up to via LTE-A 3Gbps	Data transfer speed is up to 10/50 Gbps
Use OFDMA/SC-FDMA,SOFDMA technology	5G will provide very high speed as well as efficient use of available band width to the consumers
Online gaming, High Definition TV	Ultra High Definition Video, Virtual Realty
Forward Error Correction: Turbo Codes	Forward Error Correction: Low density parity check codes

IX. FIFTH GENERATION (5G)

Rapid evolution in cellular networks will continue at 5G, officially expected to be in service in 2020. [24] 5G get defieances that are not effectively addressed by 4G i.e. higher capacity, higher data rate, lower End to End latency, massive device connectivity, reduced cost and consistent Quality of Experience provisioning. [15] These challenges are shown in Table 2 and the solution of 5G is shown against these challenges.

5G users promise: [5]

1. Lower latency, wider coverage, and higher data rates
2. Lower battery consumption
3. Simultaneous and multiple data transfer
4. 1Gbps data rate in mobility
5. More data security
6. World Wide Web (wwww) service
7. Applications supported by structured intelligence applications.

In mobile and wireless networks, it is likely that bubble volume will continue to increase rapidly over the next 20 years. [15] By 2020, the number of devices to be included in the cloud system is estimated to reach 50 billion. With the rapid increase in the number of devices, increasing capacity, energy efficiency, cost and improving the use of spectrum are causing difficulties. 5G technology provides a vision for solutions to these problems: [25]

1000 times more data volume per domain

10 to 100 times more connected devices

Typical user data rate increased by 10 to 100 times

10 times longer battery life for Massive Machine Communication (MMC) devices

End to End (E2E) delay time reduced by 5 times

5G offer challenges, facilitators to address challenges and design of fundamentals in Figure 1.

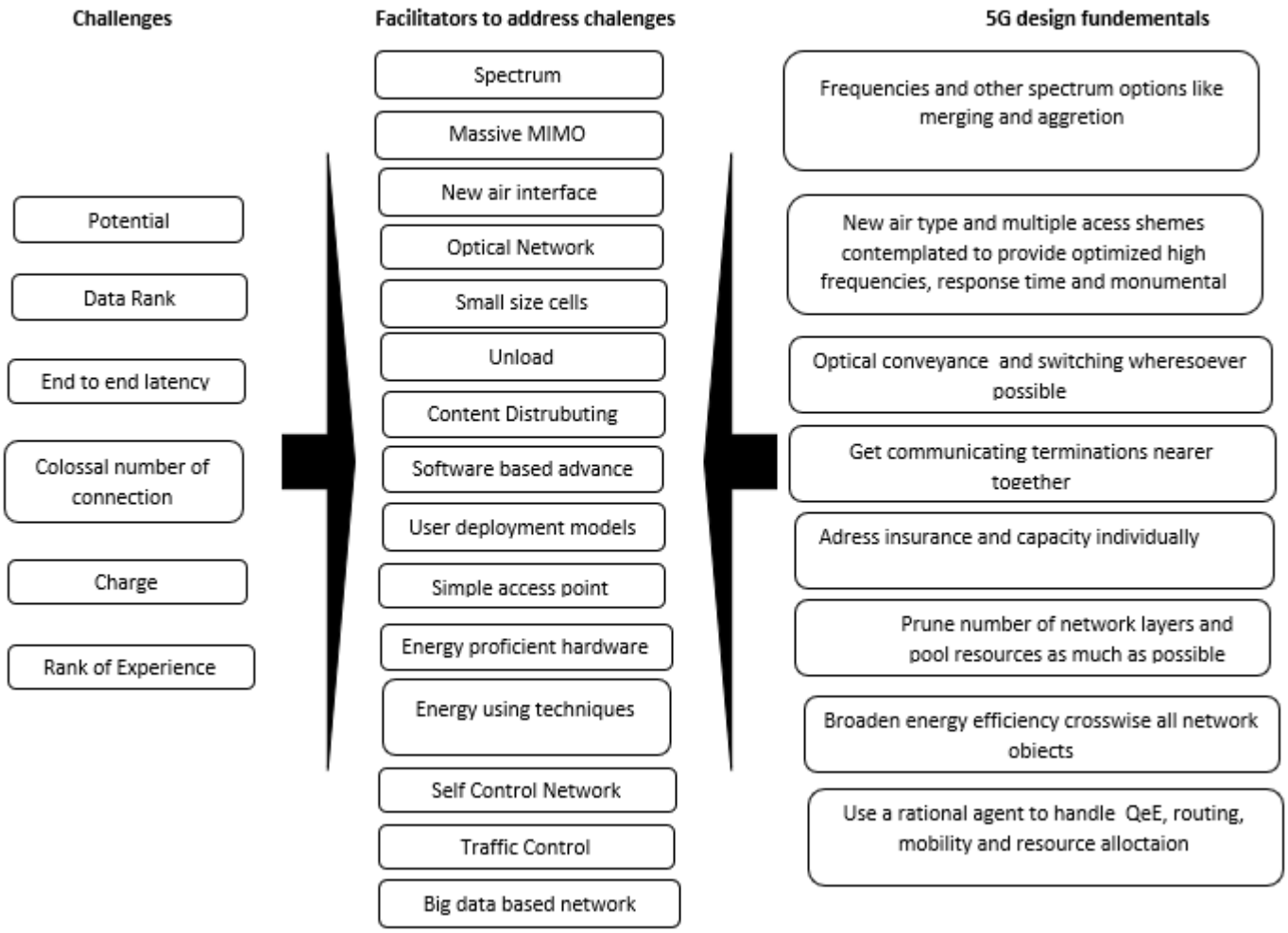


Figure 1: Challenges, facilities, fundamentals [24]

We can show the developments from 1G to 5G up to today with the development diagram in Figure 2.

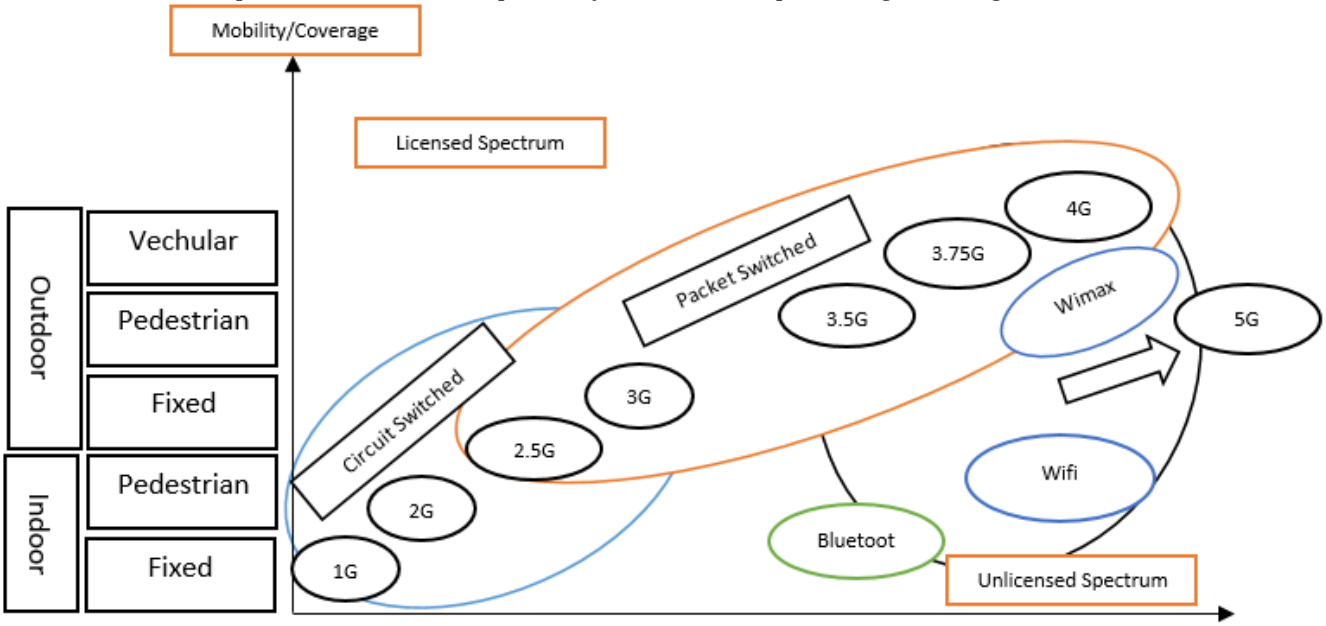


Figure 2: Evolution of Wireless Technology [5]

X. CONCLUSION

In this paper, mobile communication technologies from 1G to the next generation 5G have been examined. 5G based on 4G technology. 5G technology will offer many services very quickly and reliably. 5G technology will deliver significant contributions to many mobile service problems in daily life with performance, efficiency and speed.

REFERENCES

[1] V. Pereira and T. Sousa, "For more detailed reading and to reference , please use : Vasco Pereira , Tiago Sousa , Paulo Mendes , Edmundo Monteiro , ' Evaluation of Mobile Communications : From Voice Calls to Ubiquitous Multimedia Group Communications ', in Proc . of the 2nd Intern," no. June, pp. 1–19, 2004.

[2] N. Dumbre Monali Patwa Kajal Patwa, "5G WIRELESS TECHNOLOGIES-Still 4G auction not over, but time to start talking 5G," *Int. J. Sci. Eng. Technol. Res.*, vol. 2, no. 2, pp. 2278–7798, 2013.

[3] J. K. Sao, V. Thaiwait, K. Mahilane, and S. Kumar, "A BRIEF CASE STUDY ON 5G CELLULAR TECHNOLOGY Figure 1 : Mobile Cellular Network Evolution Timeline [3]," pp. 1–8.

[4] R. Prasad, M. Alam, and J. R. Farserotu, "Wireless Next Generation : 4G ?," vol. 5, pp. 305–306, 2000.

[5] S. Kumar, G. Gupta, and K. R. Singh, "5G: Revolution of future communication technology," *Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015*, pp. 143–147, 2016.

[6] S. Dekleva, J. P. Shim, U. Varshney, and G. Knoerzer, "Evolution and Emerging Issues in mobile wireless networks," *Commun. ACM*, vol. 50, no. 6, pp. 38–43, 2007.

[7] S. Y. Hui and K. H. Yeung, "Challenges in the Migration to 4G Mobile Systems," *IEEE Commun. Mag.*, vol. 41, no. 12, pp. 54–59, 2003.

[8] M. R. Bhalla and A. V. Bhalla, "Generations of mobile wireless technology: a survey," *Int. J. Comput. Appl.*, vol. 5, no. 4, p. 7, 2010.

[9] T. Zahariadis and D. Kazakos, "(R)evolution toward 4G mobile communication systems," *IEEE Wirel. Commun.*, vol. 10, no. 4, pp. 6–7, 2003.

[10] A. Kumar and M. Gupta, "A review on activities of fifth generation mobile communication system," *Alexandria Eng. J.*, 2016.

[11] J. Kim, D. Kim, and S. Choi, "3GPP SA2 architecture and functions for 5G mobile communication system," *ICT Express*, vol. 3, no. 1, pp. 1–8, 2017.

[12] B. Clerckx, A. Lozano, S. Sesia, C. van Rensburg, and C. Papadias, "3GPP LTE and LTE-Advanced," *EURASIP J. Wirel. Commun. Netw.*, vol. 2009, no. 1, p. 472124, 2009.

[13] L. Bos and S. Leroy, "Toward an all-IP-based UMTS system architecture," *IEEE Netw.*, vol. 15, no. 1, pp. 36–45, 2001.

[14] Y. Park and T. Park, "A survey of security threats on 4G networks," *GLOBECOM - IEEE Glob. Telecommun. Conf.*, 2007.

[15] A. Gupta, S. Member, R. K. Jha, and S. Member, "A survey of 5G network: Architecture and emerging technologies," *IEEE access*, vol. 3, 2015.

[16] P. Datta and S. Kaushal, "Exploration and comparison of different 4G technologies implementations: A survey," *2014 Recent Adv. Eng. Comput. Sci. RAECS 2014*, pp. 6–8, 2014.

[17] M. J. Chang, Z. Abichar, and C. Y. Hsu, "WiMAX or LTE: Who will lead the broadband mobile internet?," *IT Prof.*, vol. 12, no. 3, pp. 26–32, 2010.

[18] X. Li, A. Gani, R. Salleh, and O. Zakaria, "The Future of Mobile Wireless Communication Networks," *2009 Int. Conf. Commun. Softw. Networks*, pp. 554–557, 2009.

[19] V. Kumar and N. Mishra, "5G : Future of Wireless Network," vol. 1, no. 5, pp. 632–634, 2014.

[20] J. Govil and J. Govil, "4G mobile communication systems: Turns, trends and transition," *2007 Int. Conf. Converg. Inf. Technol. ICCIT 2007*, pp. 13–18, 2007.

[21] I. F. Akyildiz, D. M. Gutierrez-Estevez, R. Balakrishnan, and E. Chavarria-Reyes, "LTE-Advanced and the evolution to Beyond 4G (B4G) systems," *Phys. Commun.*, vol. 10, pp. 31–60, 2014.

[22] I. F. Akyildiz, D. M. Gutierrez-Estevez, and E. C. Reyes, "The evolution to 4G cellular systems: LTE-Advanced," *Phys. Commun.*, vol. 3, no. 4, pp. 217–244, Dec. 2010.

[23] A. Gohil, H. Modi, and S. K. Patel, "5G technology of mobile communication: A survey," *2013 Int. Conf. Intell. Syst. Signal Process. ISSP 2013*, pp. 288–292, 2013.

[24] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5G network architecture," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 65–75, 2014.

[25] I. F. Akyildiz, S. Nie, S. C. Lin, and M. Chandrasekaran, "5G roadmap: 10 key enabling technologies," *Comput. Networks*, vol. 106, pp. 17–48, 2016.

Effects of Sandstorms on Vehicular-to-Road Visible-Light Communication

K. J. EBRAHIM, A. AL-OMARY

Member IEEE

Abstract: Visible light communication (VLC), popularly known as light fidelity (Li-Fi), is a promising alternative to overcome the limitations of radio-wave communication. VLC is a green technology which uses light-emitting diode (LED) illumination to transmit data without needing fibre cables. VLC is applicable for both indoor and outdoor communication. In this study, we investigate the effect of sandstorms on VLC via simulating a vehicular-to-road VLC (V2LC) outdoor application. Sandstorms are a weather phenomenon which frequently occurs in the Arab peninsula and other parts of the world; in this context, researchers have not thus far addressed the effect of sand particles, which absorb and scatter light, on VLC. Our simulation is conducted using MATLAB software, and the results show that the effect of sandstorms on VLC is similar to that of fog and rain as investigated by other researchers. However, sandstorms are also different in terms of the nature of sandstorm particles, with different sizes and refraction indices when compared with rain and fog particles. We also find that high-density-clay sandstorms, among other types of storms, most severely affect VLC communication and limit the transmission range. Other low- and medium-density storms less severely affect VLC while exhibiting a relatively larger communication range.

Keyword: Visible light communication (VLC), Li-Fi, Vehicular-to-road VLC (V2LC) system, Optical wireless communication (OWC)

I. INTRODUCTION

The need for wireless connection is rapidly increasing due to rapid growth in social media, mobile applications, and entertainment applications. The available radio band is currently saturated to the extent that it cannot cover our communication needs in the near future, and further, radio waves contribute significantly to electromagnetic pollution. Meanwhile, Wi-Fi technology (which is a form of radio-wave communication) has certain limitations, and thus, it cannot be used in many areas; for example, Wi-Fi is not suitable for use in hospitals, airplanes, petrochemical plants, hazardous environments, and underwater submarines [1] [2]. The primary challenge to Wi-Fi usage is interference and signal dropping. Moreover, providing Wi-Fi access points everywhere is a costly affair, and this has led to researchers seeking better and cost-effective solutions to radio communication. Further, such solutions should be cost-effective and environment-friendly. Thus, researchers have begun to investigate to the possibility of using visible light for optical wireless communication (OWC). Visible light is a suitable alternative solution for radio waves because it is easily accessible in homes and public spaces, and further, it has a large bandwidth which can be exploited for communication. In addition, it is cost-effective since there is no need for 'licensing' or new infrastructure. Researchers have thus begun to propose visible light communication (VLC), also known as light fidelity (Li-Fi), as an alternative to radio communication. VLC was first proposed as a new communication technology in 2003 by researchers at Nakagawa Laboratory, Keio University, Japan. VLC solves the saturation problem in the radio-frequency band by considering every light source as a free access point. While VLC can be used for both indoor and outdoor communications, the challenges facing its implementation include the influence of weather on VLC and the consideration that VLC involves line-of-sight transmission; the signal can drop due to obstacles in the communication channel.

Most studies have focused on the use of VLC for indoor purposes. Meanwhile, some researchers have studied the effect of fog and traditional (non-LED) light sources on 'outdoor' VLC models [3]. In this backdrop, in this study, we examine the effect of different weather phenomena (such as sandstorms, which occur in many Arab countries) on an outdoor VLC application.

II. LITERATURE REVIEW

Several researchers have previously studied VLC for outdoor applications, and some studies have focused on the effects of weather phenomena such as fog, haze, and interference with sunlight on VLC. Hu et al. [4] followed the experimental methodology framework to examine the use of visible red-light lasers for VLC. Their study found that when a visible red-light laser used for free-space optical (FSO) communication, there was some attenuation of the received power, which was caused by absorption and scattering of laser light due to the presence of gas molecules and aerosol particles. Mahalati and Kahn [5] experimentally studied the effect of haze, fog, and rain, which were classified under weather influences, on FSO links using imaging receivers. They focused on link attenuation and image blooming, and they noticed that when the fog cover was 'thick' or 'medium', there was an increase in attenuation along with image blooming. Moreover, they found that when the fog was light, the attenuation was less, but image blooming still affected the link performance. Beshr et al. [6] conducted experiments in several locations around the world to study the effect of the sunlight on VLC. They examined the performance, bit error rate (BER), and signal-to-noise ratio (SNR) of VLC in Cairo and Glasgow during the year, and they noticed that sunlight differently affects VLC; sunlight in Cairo more strongly affected VLC than in Glasgow during summer. In addition, they noticed that at maximum sunlight intensity, the VLC performance and SNR decreased while the BER increased.

III. RESEARCH METHODOLOGY

In this study, we investigate the influence of weather on outdoor VLC, which to our knowledge has not previously been addressed in other studies. We conduct simulations of a vehicular-to-road VLC (V2LC) system application under different weather conditions using MATLAB R2013a. The simulation approach utilizes the V2LC model proposed by Kumar et al. [7], and we use this model to investigate the effects of sand and dust storms on VLC. In this V2LC model, a red-LED traffic light is used as the transmitter, and the car is considered as the receiver, as shown in Fig. 1. The height h of the traffic light is assumed to be 3 m and the distance between the receiver and the traffic light (x) is set to a minimum of 2.05 m, below which there is no line of sight between the transmitter and receiver. Further, Umax and Umin denote the range of possible communication, with Umin being closer to traffic signal. Moreover, a road width of 10 m and length of 100 m is set as the simulation design range. In addition, the temperature in our simulation is set to 32 ° C. Several different weather scenarios are simulated to examine their effects on VLC:

1. Simulating VLC in clear weather.
2. Simulating VLC in low-, medium-, and high-density sand and dust storms.

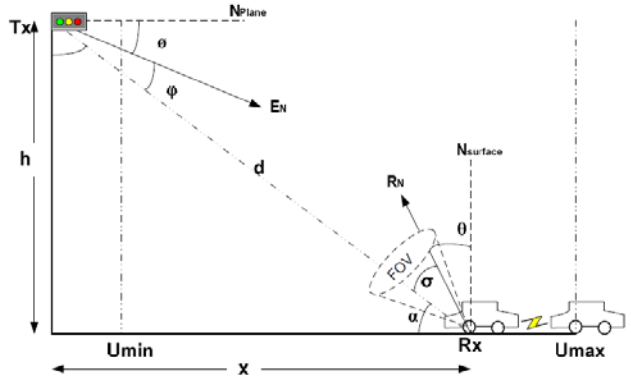


Fig. 1. Vehicular-to-road visible light communication (V2LC) model proposed by Kumar et al. [7]

3.1. Sandstorm Types

Sand and dust storms are weather phenomena caused by turbulent and strong winds, and in dry climates or deserts, such winds move loose soil into the air. Such storms affect horizontal visibility and reduce this visibility to less than 1 km. Figure 2 represents sandstorm statistics issued by the Bahrain Meteorological Office. The figure depicts the number of days in a year in Bahrain in which massive dust or sandstorms occurred, which reduced visibility to below 1 km, for the period between 1990 and 2013. The worst year was 2008, with sandstorms raging for 18 days in this year.

Sandstorms contain different types of dust including mineral dust. The Arab peninsula is affected by sandstorms with the following particle types:

1- Mineral dust which mainly contains clay particles with sizes less than 2 μm.
2- Silt particles with sizes between 2 and 74 μm.
3- Sand and gravel particles with sizes larger than 74 μm [8].

We used these sandstorm data in our simulations, along with other parameters.

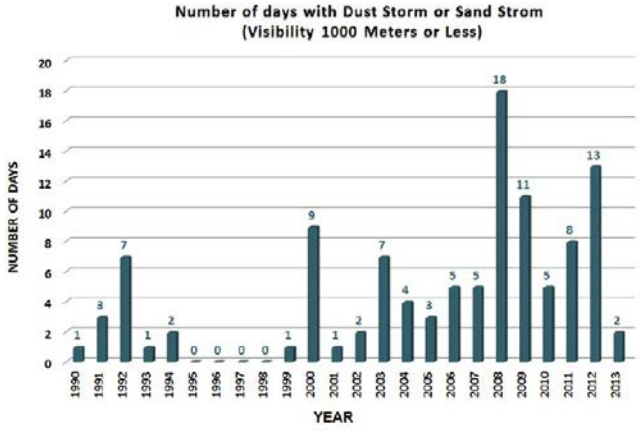


Fig. 2. Number of days in a year with dust storms or sandstorms leading to visibility of 1 km or less over a period of 24 years [9]

3.2. Simulation Parameters

VLC uses an LED as the transmitter and a photodiode as the receiver. The photodiode converts the received illumination into current; therefore, it is necessary to measure the power received to calculate the signal power. We used the following equation to obtain the power gain in the receiver:

$$P_r = P_t \cdot \frac{(m_l+1)A}{2\pi d^2} cos^{m_l}(\theta) . T_s(\Psi) . g(\Psi) . cos(\Psi) , 0 \le \Psi \le \Psi_{con} \quad (1)$$

Here,

P_t: Transmitted power from LED
m_l: Order of Lambertian emission
A: Photodetector area
d: Distance between transmitter and receiver
θ: Angle of irradiance
Ψ: Angle of incidence
$T_s(\Psi)$: Signal transmission coefficient of optical filter
$g(\Psi)$: Gain of optical concentrator
Ψ_{con}: Receiver field of view (FOV)

VLC connectivity can strongly affected by weather. The presence of molecular species and aerosols in the atmosphere affects VLC; these molecules and particles cause absorption, scattering, and scintillation of light. Moreover, they attenuate the transmitted light signal, which reduces the received power in the photodiode. Therefore, we investigate light attenuation to calculate the actual received power in the photodiode. In this regard, Chen [10] studied the attenuation of electromagnetic radiation by haze, fog, clouds, and rain. In addition, Buys and Janssen [11] used the Beer–Lambert–Bouguer law to calculate the attenuations in the light signal during rain or snow. Moreover, Tamošiūnaitė et al. [12] used the Beer–Lambert–Bouguer law to calculate the atmospheric attenuation due to humidity. Therefore, we also decided to use the Beer–Lambert–Bouguer law in our calculation following these studies in every proposed scenario.

$$P_r = P_t . e^{(-l\varepsilon)} \qquad (2)$$

Here,

P_t: Transmitted power
P_r: Received power
l: Distance between transmitter and receiver
ε: Extinction coefficient

In addition, the performance of the VLC can also be affected by sunlight when the sunbeams are in the same direction as the visible-light link.

To apply the Beer–Lambert–Bouguer law, we first decided to calculate the extinction coefficient ε. We determined that Rayleigh scattering (molecular scattering), aerosol absorption, and aerosol scattering (Mie scattering) form the main contributions to the extinction coefficient. Therefore, we studied each scattering type and absorption separately for use in our calculation.

The attenuation from Rayleigh scattering is due to the interaction between the illuminating light wave and atmospheric particles which are smaller than the light wavelength [13] [14]. Rayleigh scattering yields an extinction coefficient which can be calculated as follows:

$$\beta_m(\lambda) = \frac{24\pi^3}{\rho\lambda^4} 10^3 \left(\frac{[n(\lambda)]^2 - 1}{[n(\lambda)]^2 + 2}\right)\left(\frac{6 + 3\delta}{6 - 7\delta}\right) \qquad (3)$$

Here,

$\beta_m(\lambda)$: Molecular scattering coefficient (km^{-1})
λ: Wavelength (μm)
ρ: Molecular density (m^{-3})
δ: Depolarization factor of the air (≅0.03)
$n(\lambda)$: Refractive index of air.

Parameter $\beta_m(\lambda)$ can be approximated as

$$\beta_m(\lambda) = A\lambda^{-4} \qquad (4)$$

Here,

$$A = 1.09 * 10^{-3} \frac{P}{P_0}\frac{T_0}{T} \quad km^{-1}m^4 \qquad (5)$$

and

P: Atmospheric pressure (mbar)
P_0: 1 013 mbar
T: Atmospheric temperature (K)
T_0: 273.15 K

In addition, aerosol absorption involves the interaction of the illuminating light with solid or liquid atmospheric particles such as ice, dust, smoke, and fog [13] [14]. These particles float in the atmospheric air or fall ground wards with very low speeds. Their sizes usually range between 10^{-2} and 100 μm [13]. The following equation was used to calculate the extinction coefficient for aerosol absorption:

$$\alpha_n(\lambda) = 10^5 \int_0^\infty Q_a\left(\frac{2\pi r}{\lambda}, n''\right)\pi r^2 \frac{dN(r)}{dr} dr \quad km^{-1} \qquad (6)$$

Here,

λ: Wavelength (μm)
$dN(r)/dr$: Particle size distribution per unit of volume (cm^{-4})
n'': Imaginary component of the refractive index, n, of the considered aerosol
r: Radius of the particles (cm)
$Q_a(2\pi r/\lambda, n'')$: Absorption cross-section for a given type of aerosol

Moreover, when the particle size is less than or equal to the wavelength of the light beam, the beam undergoes Mie scattering. In order to calculate the extinction coefficient for Mie scattering, we used the following equation:

$$\beta_n(\lambda) = 10^5 \int_0^\infty Q_d\left(\frac{2\pi r}{\lambda}, n'\right)\pi r^2 \frac{dN(r)}{dr} dr \quad km^{-1} \qquad (7)$$

Here,

λ: Wavelength (μm)
$dN(r)/dr$: Particle size distribution per unit volume (cm^{-4})
n': Real component of the refractive index n of the aerosol
r: Radius of the particles (cm)
$Q_d(2\pi r/\lambda, n')$: Scattering cross-section for a given type of aerosol

Here, we mention that the attenuation due to aerosol scattering and absorption is difficult to estimate due to variation in the concentration, composition, and sizes of the particles. Therefore, we used the abovementioned mathematical equations to calculate the effect of sand and dust storm particles on VLC.

To solve (6) and (7), we need to determine the sand and dust storm density. Consequently, we used the visibility factor V to determine the density as follows [15] [16] [17] [18] [19]:

$$V = \frac{5.5 \times 10^{-4}}{N a_e^2} \quad km \qquad (8)$$

Here,

a_e: Equivalent particle radius (m)
N: Particle density (number of particles/m^3)

From the visibility, we obtained the density as

$$N = \frac{5.5 \times 10^{-4}}{V a_e^2} m^{-3} \qquad (9)$$

Finally, since aerosol absorption and scattering depend on the particles sizes, refractive index, and illuminating wavelength, we investigated the particles sizes for each scenario along the corresponding refractive indices of these particles. Here, we remark that the refractive index consists of two components: the real component n' which is responsible for scattering, and the imaginary component n'' which is responsible for absorption [13].

TABLE 1 Input parameters used in the simulation

LED Power	30 mW
Number of LEDs	620
Receiver area, A	9×10^{-5} m^2
Distance between source and receiver	2.05 m
Simulation range dimensions	lx =100 m, ly = 10 m, lz = 3 m
Field of view (FOV) of receiver	60°
Semi-angle at half power	60°
Photodiode responsibility	0.43 A/W
Gain of optical filter, Ts	1
Refractive index of lens at photodiode	1.5
Boltzmann's constant, k	$1.3806488 \times 10^{-23}$ J/K
Fixed temperature	273.15 K
Fixed pressure	1013 mbar
Refractive index of medium (Air)	1.003 + 0i
Electronic charge, q	$1.602176487 \times 10^{-19}$ C
Background power	0.009444 W
Bandwidth	10×10^6 Hz
Shunt Resistance, R	20×10^6 Ω
Red LED wavelength	0.644 μm
Visibility	Variable depending on type of storm
Atmospheric pressure	Variable depending on type of storm
Atmospheric temperature	32 °C
Particle refractive index	Variable depending on type of storm
Particle radius	Variable depending on type of storm

IV. SIMULATION SCENARIOS

Five different simulation experiments with different simulation scenarios were conducted to investigate the effects of sandstorms and dust storms on VLC. The investigated VLC simulation scenarios are as follows:

1. Clear weather
2. Storm with low-, medium-, and high-density clay particles
3. Storm with low-, medium-, and high-density silt particles
4. Storm with low-, medium-, and high-density sand particles
5. Storm with low-, medium-, and high-density gravel particles

We chose these scenarios based on a study by the National Centre of Meteorology and Seismology (NCMS) in the United Arab Emirates, which lists the types of the particles found in storms in the Arab Gulf countries. For a low-density storm, we assumed a visibility of 5 km, for a medium-density storm, the visibility was assumed as 1 km, and for a high-density storm, we assumed a visibility of 0.1 km.

Table 2 lists the parameters used for the different types of storms.

TABLE 2 Parameters used for different types of sandstorms

Type of storm	Visibility	Atmospheric pressure	Particle refractive index	Particle radius
No storm (Clear atmosphere)	10 km	1017 mbar	1.000271 + 0i	4.8×10^{-9} cm
Low-density-clay storm (Montmorillonite) particles	5 km	1000 mbar	$1.523 + i3.8 \times 10^{-5}$	1×10^{-4} cm
Medium-density-clay storm (Montmorillonite) particles	1 km	700 mbar	$1.523 + i3.8 \times 10^{-5}$	1×10^{-4} cm
High-density-clay storm (Montmorillonite) particles	0.1 km	500 mbar	$1.523 + i3.8 \times 10^{-5}$	1×10^{-4} cm
Low-density-silt-particle storm	5 km	1000 mbar	1.4718 + 0i	36×10^{-4} cm
Medium-density-silt-particle storm	1 km	700 mbar	1.4718 + 0i	36×10^{-4} cm
High-density-silt-particle storm	0.1 km	500 mbar	1.4718 + 0i	36×10^{-4} cm
Low-density-sand-particle storm	5 km	1000 mbar	2.288 - 0.0568i	74×10^{-4} cm
Medium-density-sand-particle storm	1 km	700 mbar	2.288 - 0.0568i	74×10^{-4} cm
High-density-sand-particle storm	0.1 km	500 mbar	2.288 - 0.0568i	74×10^{-4} cm
Low-density-gravel-particle storm	5 km	1000 mbar	$1.550 + i1.0 \times 10^{-4}$	74×10^{-4} cm
Medium-density-gravel-particle storm	1 km	700 mbar	$1.550 + i1.0 \times 10^{-4}$	74×10^{-4} cm
High-density-gravel-particle storm	0.1 km	500 mbar	$1.550 + i1.0 \times 10^{-4}$	74×10^{-4} cm

V. RESULTS AND DISCUSSION

The received power, SNR, and BER were calculated and compared between the clear-weather conditions and storms with different particles and particle densities. Table 3 lists the received power, SNR, and BER for clear weather with a visibility of 10 km. Here, we recall that two factors contribute to signal degradation: distance and line of sight. First, when the distance between the transmitter and the receiver is 2.864 m, a maximum power of -10.221 dBm is received with an SNR of 100 dB and BER of 4.58×10^{-46}. For a distance of 100 m, the power reduces to - 38.014 dBm, the SNR drops to 57.872 dB, and the BER to 2.76×10^{-27}. Second, the received signal deteriorates due to signal weakness over a short distance when there is no line of sight access. When the distance between the transmitter and receiver is 2.05 m, there is no longer a line of sight between them, and therefore, no power is received, with SNR and BER becoming zero. These degradations in power due to the experimental setup are illustrated in Fig. 1.

TABLE 3

Received power, signal-to-noise ratio (SNR), and bit error rate (BER) for clear weather (no storm)

	Clear Weather		
	Min (at 100 m)	Max (at 2.864 m)	Distance ≤ 2.05 m
Received Power	-38.014 dBm	-10.221 dBm	-∞ dBm
SNR	57.872 dB	100.820 dB	- dB
BER	2.76×10^{-27}	4.58×10^{-46}	-

Tables 4, 5, and 6 list the received power, SNR, and BER for the four different types of storms examined in the study. We note that for high-density storms, the clay storm most severely affects VLC, wherein the received power is least when compared with those for the other storm types; this because clay particles most strongly absorb and scatter light. The other storm types relatively less severely affect VLC, as can be inferred from the table entries.

TABLE 4

Received power for sandstorms comprising different particles

Type of Sandstorm	Low-density		Medium-density		High-density	
	Min	Max	Min	Max	Min	Max
Clay Storm	-38.459 dBm	-10.233 dBm	-40.237 dBm	-10.28 dBm	-60.245 dBm	-10.854 dBm
Silt Storm	-38.324 dBm	-10.230 dBm	-39.564 dBm	-10.26 dBm	-53.517 dBm	-10.663 dBm
Sand Storm	-38.320 dBm	-10.230 dBm	-39.542 dBm	-10.26 dBm	-53.297 dBm	-10.656 dBm
Gravel Storm	-38.321 dBm	-10.230 dBm	-39.546 dBm	-10.26 dBm	-53.338 dBm	-10.658 dBm

TABLE 5

Comparison of signal-to-noise ratio (SNR) values for different sandstorm particles

Type of Sandstorm	Low-density		Medium-density		High-density	
	Min	Max	Min	Max	Min	Max
Clay Storm	56.995 dB	100.807 dB	53.478 dB	100.754 dB	13.537 dB	100.151 dB
Silt Storm	57.261 dB	100.811 dB	54.811 dB	100.774 dB	26.990 dB	100.354 dB
Sand Storm	57.270 dB	100.811 dB	54.855 dB	100.774 dB	27.430 dB	100.361 dB
Gravel Storm	57.268 dB	100.811 dB	54.846 dB	100.774 dB	27.348 dB	100.359 dB

TABLE 6

Comparison of bit error rates (BERs) for different sandstorm types with different particle densities

Type of Sandstorm	Low-density		Medium-density		High-density	
	Min	Max	Min	Max	Min	Max
Clay Storm	6.5×10^{-27}	4.642×10^{-46}	2.3×10^{-25}	4.897×10^{-46}	9.8×10^{-08}	8.972×10^{-46}
Silt Storm	5.0×10^{-27}	4.623×10^{-46}	5.9×10^{-26}	4.799×10^{-46}	1.0×10^{-13}	7.316×10^{-46}
Sand Storm	5.0×10^{-27}	4.622×10^{-46}	5.7×10^{-26}	4.796×10^{-46}	6.5×10^{-14}	7.268×10^{-46}
Gravel Storm	5.0×10^{-27}	4.622×10^{-46}	5.7×10^{-26}	4.796×10^{-46}	7.0×10^{-14}	7.277×10^{-46}

Figure 3 depicts the received power for clay and sandtorms as a function of distance, while Fig. 4 depicts this for gravel and silt. From these figures, we note that the clay storm most severely affects communication, followed by sandstorms.

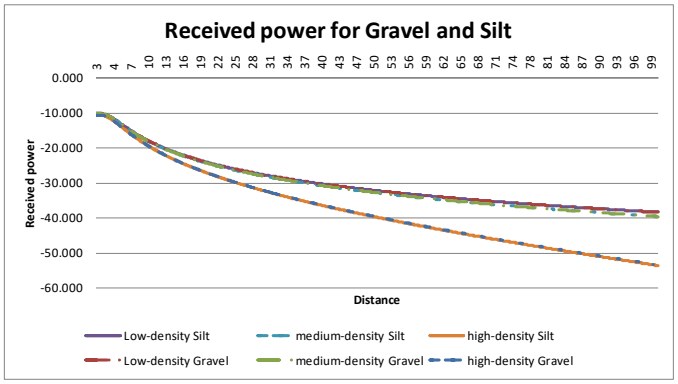


Fig. 3. Comparison of received power for varying clay- and sand-particle-density sandstorms as a function of distance

As regards the required SNR for VLC, Ab-Rahman et al. [20] have previously determined that the minimum acceptable SNR is 15 dB. The SNR of 15 dB corresponds to a received power of around -60.084 dBm at a distance of 90 m for a high-density-clay storm and the same power for a distance of around 100 m for heavy sandstorms, as shown in Fig. 3. From Figs. 5 and 6, we note that the SNRs for high-density silt, sand, and gravel particles are more than the minimum acceptable SNR at 100 m. Thus, the received power can carry to more than 100 m for these particles even when the storm density is high. Moreover, for medium- and low-density storms, the SNR is easily more than 50 dB, from which we can conclude that the effects of these storms on VLC are minor, and the communication distance can easily be more than 100 m.

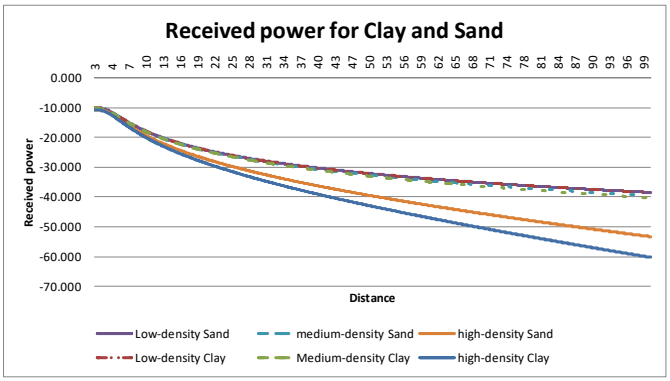


Fig. 4. Comparison of received power for varying gravel- and silt-particle-density sandstorms as a function of distance

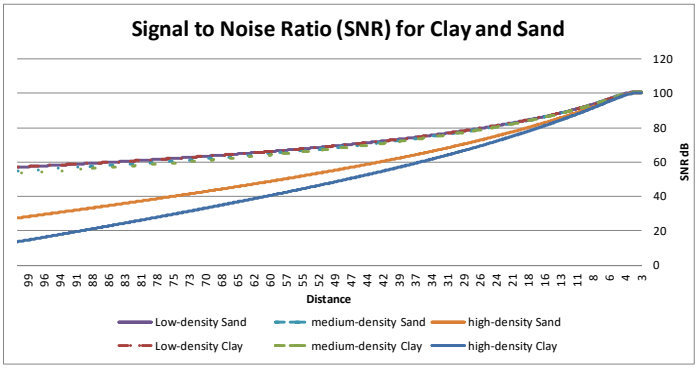


Fig. 5. Comparison of signal-to-noise ratios (SNRs) for varying clay- and sand-particle-density sandstorms as a function of distance

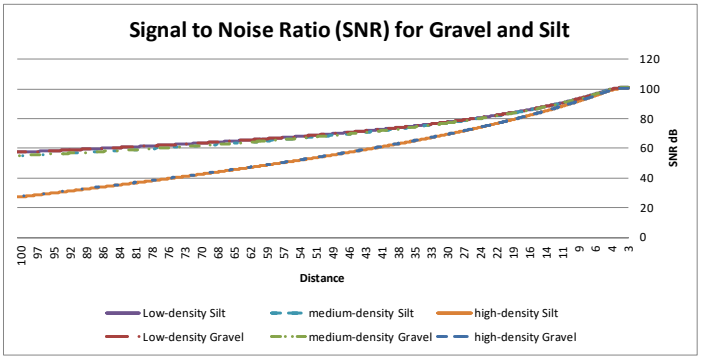


Fig. 6. Comparison of signal-to-noise ratios (SNRs) for varying gravel- and silt-particle-density sandstorms as a function of distance

Figure 7 depicts the BER values as a function of distance for clay and sandstorms, while Fig. 8 depicts the BER for gravel and silt storms as a function of distance. The BER is not significantly affected by low- and medium-density storms, and the signal coverage range is fairly large. On the other hand, for high-density-clay storms, the coverage range does not exceed 100 m since the SNR drops to less than the minimum acceptable value beyond this distance, with a corresponding increase in BER.

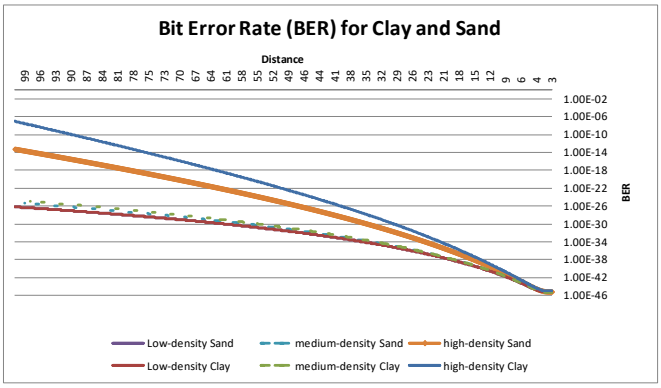


Fig. 7. *Comparison of bit error rate (BER) for varying clay- and sand-particle-density sandstorms as a function of distance*

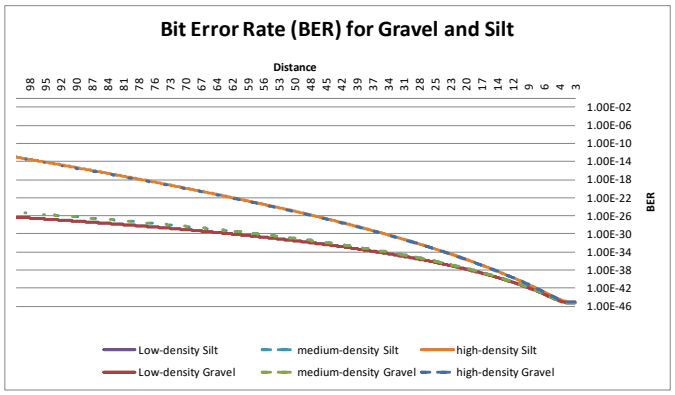


Fig. 8. *Comparison of bit error rate (BER) for varying gravel- and silt-particle-density sandstorms as a function of distance*

VI. CONCLUSION

Visible light communication (VLC) was first proposed as a new communication technology in 2003 by researchers at Nakagawa Laboratory, Keio University, Japan. VLC has since been considered an alternative solution for radio-wave communication because of its ease of use in homes and public spaces (light sources are easily available), and further, VLC offers a very high bandwidth. In this context, many studies have investigated the factor of visibility for VLC; however, there has been no study on the effects of sand and dust storms on VLC. Thus, in this work, we applied a mathematical model to investigate the effect of dust and sand particles on VLC; the effect of such storms was studied via simulation of a vehicular-to-road VLC (V2LC) outdoor application. We found that sandstorms and dust storms affect VLC connectivity, and their influence varies according to the density of the storm. The four types of storms used in the simulation, clay storm, sandstorm, gravel storm, and silt storm, are the most frequently occurring storms affecting the Arab peninsula region. These storms comprise different particles with different sizes and different refractive indices, which affect light communication via light absorption or scattering. We further determined that high-density-clay storms severely affected VLC. The maximum distance corresponding to the minimum acceptable SNR in this case was around 90 m. On the other hand, the maximum distance for high-density silt, sand, and gravel storms was around 100 m. With medium- and low-density storms, the coverage distance is considerably more than 100 m, which means that these storms do not seriously affect VLC performance. Our results show general similarities with those of past studies on the effects of fog and rain on VLC, e.g., Mahalati and Kahn [5]; on the other hand, we also needed to consider the nature of sandstorms with different refraction indices and particle sizes. VLC technology can be implemented for outdoor application, except for 5% of the year at dawn, as measured by the Bahrain Meteorological Directorate. In further studies, it is required to investigate the effect of replacing the photodiode with an image sensor for outdoor VLC. Further, different outdoor applications of VLC also need to be studied with the use of actual prototype VLC systems.

Acknowledgments

We would like to thank Editage (www.editage.com) for English language editing.

REFERENCES

[1] George, J. J., Mustafa, M. H., Osman, N. M., *et al.*: 'A survey on visible light communication', Int. J. Eng. Comp. Sci., 2014, **3**, (2), pp. 3805-3808.

[2] Jaiswal, N. S., and Chopade, P. S.: 'Review of Li-Fi technology: New future technology-light bulb to access the Internet', Int. J. Sci. Eng. Res., 2013, **4**, (12), pp. 36-40. ISSN: 2229-5518

[3] Kumar, N., Terra, D., Lourenço, N., *et al.*: 'Visible light communication for intelligent transportation in road safety applications', *7th IEEE International Wireless Communications and Mobile Computing Conference*, Istanbul, July 2011, pp. 1513-1518

[4] Hu, G. Y., Chen, C. Y., and Chen, Z. Q.: 'Free-space optical communication using visible light', J. Zhejiang Univ. Sci. A, 2007, **8**, (2), pp. 186-191

[5] Mahalati, R. N., and Kahn, J. M.: 'Effect of fog on free-space optical links employing imaging receivers', Opt. Express, 2012, **20,** (2), pp. 1649-1661

[6] Beshr, M., Andonovic, I., and Hussien, M.: 'The impact of sunlight on the performance of visible light communication systems over the year', Proc. SPIE 8540, Unmanned/Unattended Sensors and Sensor Networks IX, 2012, 85400F. doi:10.1117/12.978935

[7] Kumar, N., Nero, L. A., and Aguiar, R. L.: 'Visible light communication for advanced driver assistant systems, *FCT project VIDAS'*, 2006, Retrieved from: http://www.av.it.pt/conftele2009/papers/38.pdf, accessed 17 December 2014

[8] Dust, National Centre of Meteorology & Seismology Ministry of Presidential Affairs, 2011, Retrieved from: http://www.ncms.ae/ncms-book-details.html?id=05#page/1, UAE, accessed 8 January 2015

[9] Dust (n. d.), Retrieved from: http://www.bahrainweather.gov.bh/en/web/guest/dust, accessed 8 January 2015

[10] Chen, C. C.: 'Attenuation of electromagnetic radiation by haze, fog, clouds, and rain, Report Prepared for United States Air Force Project Rand', 1975, Retrieved from: https://www.rand.org/content/dam/rand/pubs/reports/2006/R1694.pdf. RAND CORP SANT A MONICA CA, accessed 17 January 2015

[11] Buys, J. H., and Janssen, L. H.: 'Comparison of simultaneous atmospheric attenuation measurements at visible light, mid-infra-red (3–5 μm) and

millimetre waves (94 GHz)', IEEE Proceedings H Microwaves, Optics and Antennas IEEE Proc. H. Microw. Opt. Antennas UK, 1981, **128**, (3), pp. 131-136. IET Digital Library

[12] Tamošiūnaitė, M., Žilinskas, M., Tamošiūnas, S., *et al.*: 'Atmospheric attenuation due to humidity', Electromagnetic Waves, Prof. Vitaliy Zhurbenko (Ed.), *ISBN: 978-953-307-304-0, InTech*, doi: 10.5772/21430

[13] P.1817: Propagation data required for the design of terrestrial free-space optical links (2012), *ITU (International Telecommunication Union) is the United Nations specialized agency for information and communication technologies–ICTs*, retrieved from: http://www.itu.int/rec/R-REC-P.1817-1-201202-I/en, accessed 5 February 2015

[14] Taylor, A. E.: 'Illumination fundamentals' Lighting Research Center, Rensselaer Polytechnic Institute, 2000

[15] Alhaider, M. A.: 'Radio wave propagation into sandstorms-system design based on ten-years visibility data in Riyadh, Saudi Arabia', Int. J. Infrared Milli., 1986, **7**, (9), 1339-1359

[16] Chu, T. S.: 'BSTJ brief: Effects of sandstorms on microwave propagation', Bell Syst. Tech. J., 1979, **58**, (2), pp. 549-555. doi: 10.1002/j.1538-7305.1979.tb02234.x

[17] Elabdin, Z., Islam, M. R., Khalifa, O. O., *et al.*: 'Development of mathematical model for the prediction of microwave signal attenuation due to duststorm,' IEEE International Conference on Computer and Communication Engineering, Kuala Lumpur, May 2008, pp. 1156-1161

[18] Elshaikh, Z. E. O., Islam, M. R., Khalifa, O. O., *et al.*: 'Mathematical model for the prediction of microwave signal attenuation due to duststorm', Prog. Electromagn. Res. M, 2009, **6**, pp. 139-153. doi:10.2528/PIERM09021906

[19] Goldhirsh, J.: 'Attenuation and backscatter from a derived two-dimensional duststorm model', IEEE Trans. Antennas Propag., 2001, **49**, (12), pp. 1703-1711

[20] Ab-Rahman, M. S., Shuhaimi, N. I., Azizan, L. A., *et al.*: 'Analytical study of signal-to-noise ratio for visible light communication by using single source', J. Comp. Sci., 2012, **8,** (1), pp. 141-144. doi:10.3844/jcssp.2012.141.144

Solving University Course Timetabling Problem Using Ant Colony Optimization: An Example of Mersin University Engineering Faculty

S. ASLAN[1] and C. ACI[1]

[1] Mersin University, Department of Computer Engineering, 333434, Mersin, Turkey, aslansemir26@gmail.com
[1]Mersin University, Department of Computer Engineering, 333434, Mersin, Turkey, caci@mersin.edu.tr

Abstract **- Building effective schedules in academical institutions considering the wishes and needs of administrative staff, professors and students at the same time is a rather difficult and time-consuming activity for staff involved in this work. Despite improvements in software and hardware technology in recent years, charts are still manually created in many educational institutions and the desired efficiency has not achieved. In this study, the course chart of Mersin University Engineering Faculty was built using Ant Colony Optimization (ACO) technique. While the course schedule was being formed, 9 departments, 24 common classrooms, 105 faculty members, 239 courses, 14.374 students who have attendance obligations and 8286 students who have not attendance obligations were taken into consideration. In the placement of the courses, adaptation to ACO algorithm has been achieved by targeting the maximum lecture minimum classroom usage. The appropriate hours of the lecturers were accepted as strict constraints and other cases were added to soft constraints. All courses of Mersin University Engineering Faculty have placed the course schedule to appropriate classrooms at the rate of 99% using ACO technique, and 17 classrooms of common 24 classrooms were determined to be sufficient for educational activities.**

Keywords **- Course scheduling; ant colony optimization; timetabling.**

I. INTRODUCTION

The problem of course schedule varies from country to country due to the diversity in academic systems, even from universities within the academic system of the same country. Timetabling problems in the class of Np-Hard problems such as other scheduling problems have caught the intense attention of researchers since about 1950's. Many algorithms have been developed for solving these problems [1]. Due to these reasons, the time scheduling problem has been worked intensively for many years in the operational research literature and still maintains its popularity [2]. Generally, the important parameters of schedules in universities are lecturers of the university, number of students who are required to attend classes and classroom capacity. The course schedule is prepared by the lecturer or secretary of the department. Consequently, this implementation causes to interruption on lecturer's work time and reduce their productivity.

In addition to the demands of the lecturers on the problem of scheduling the course, the number of students who have attendance obligations is added to the number of students who have not attendance obligations. The necessity of this process ensures the adequacy of the classroom capacity in case of students who have not attendance obligations willingness to join the class. Furthermore, it was also aimed to eliminate disruptions that physical conditions could cause. In this study, Ant Colony Optimization (ACO) algorithm has been modeled with MATLAB programming language on Mersin University Engineering Faculty lessons and the course schedule was established.

II. ANT COLONY

Real ants are capable of finding the shortest path from a food source to the nest without using visual cues. Also, they are capable of adapting to changes in the environment, e.g. finding a new shortest path once the old one is no longer feasible due to a new obstacle. Consider Fig. 1A: ants are moving in a straight line that connects a food source to their nest. It is well known that the primary means for ants to form and maintain the line is a pheromone trail. Ants deposit a certain amount of pheromone while walking, and each ant probabilistically prefers to follow a direction rich in pheromone. This elementary behavior of real ants can be used to explain how they can find the shortest path that reconnects a broken line after the sudden appearance of an unexpected obstacle has interrupted the initial path (Fig. 1B). In fact, once the obstacle has appeared, those ants which are just in front of the obstacle cannot continue to follow the pheromone trail and therefore they have to choose between turning right or left. In this situation, we can expect half the ants to choose to turn right and the other half to tum left. A very similar situation can be found on the other side of the obstacle (Fig. 1C). It is interesting to note that those ants which choose, by chance, the shorter path around the obstacle will more rapidly reconstitute the interrupted pheromone trail compared to those who choose the longer path. Thus, the shorter path will receive a greater amount of pheromone per time unit and in turn, a larger number of ants will choose the shorter path. Due to this positive feedback (autocatalytic) process, all the ants will rapidly choose the shorter path (Fig. 1D). The most interesting aspect of this autocatalytic process is that finding the shortest path around the obstacle seems to be an emergent property of the interaction between the obstacle shape and ants distributed behavior: although all ants move at approximately the same speed and deposit a pheromone trail at approximately the same rate, it is a fact that it takes longer to contour obstacles on their longer side than on their shorter side which makes the pheromone trail accumulate quicker on the shorter side. It is the ant's preference for higher pheromone trail

levels which makes this accumulation still quicker on the shorter path[3].

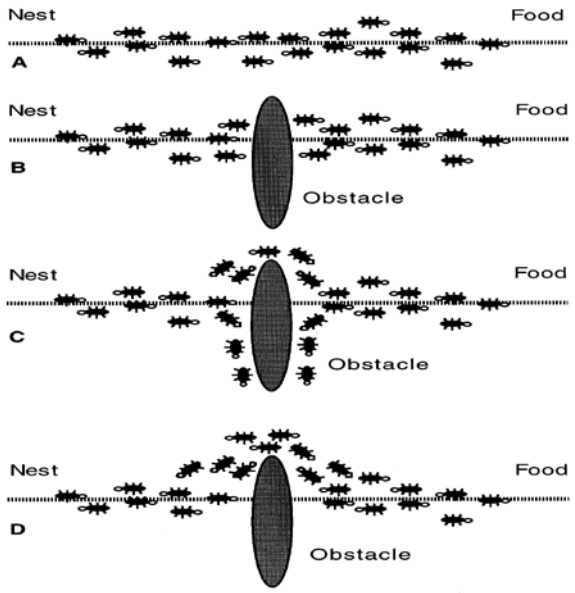


Figure 1: Ant colony food search behavior

III. ANT COLONY OPTIMIZATION ALGORITHM

ACO takes inspiration from the foraging behavior of some ant species. These ants deposit pheromone on the ground in order to mark some favorable path that should be followed by other members of the colony. ACO exploits a similar mechanism for solving optimization problems [4].

Ant System (AS): It is the first ACO algorithm proposed in the literature. Its main characteristic is that, at each iteration, the pheromone values are updated by *all* the *m* ants that have built a solution in the iteration itself. The pheromone, associated with the edge joining cities i and j, is updated as follows:

$$k\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \sum_{c=1}^{m} \Delta\tau_{ij}^{k} \quad (1)$$

Where ρ is the evaporation rate, m is the number of ants, and $\Delta\tau_{ij}^{k}$ is the quantity of pheromone laid on edge (i, j) by ant k.

$$\Delta\tau_{ij}^{k} = \begin{cases} \frac{Q}{L_k} & if\ used\ edge\ (i,j) in\ this\ tour, \\ 0 & otherwise, \end{cases} \quad (2)$$

Where Q is a constant, and L_k is the length of the tour constructed by ant .

In the construction of a solution, ants select the following city to be visited through a stochastic mechanism. When ant k is in city i and has so far constructed the partial solutions s^{ρ}, the probability of going to city j is given by:

$$P_{ij}^{k} = \begin{cases} \frac{\tau_{ij}^{\alpha} \times \eta_{ij}^{\beta}}{\sum_{c_{il} \in N(s^{\rho})} \tau_{ij}^{\alpha} \times \eta_{ij}^{\beta}} & if\ c_{il} \in N(s^{\rho}) \\ 0 & otherwise \end{cases} \quad (3)$$

Where $N(s^{\rho})$ is the set of feasible components; that is, edges (i, l) where l is a city not yet visited by the ant k . The parameters α and β control the relative importance of the pheromone versus the heuristic information, which is given by:

$$\eta_{ij} = \frac{1}{d_{i,j}}, \quad (4)$$

Where $d_{i,j}$ is the distance between cities i and j [4].

IV. ADAPTATION OF ACO SYSTEM TO SCHEDULING PROBLEM

When ACO algorithm is applied, appropriate hours of the lecturers, course-hour eligibility and lecturer-course hour parameters are added to the equation. The roulette wheel method has been applied for selection and it is added in ACO algorithm. The resulting equation and the meanings of the parameters according to this equation are shown below.

1. Choosing Classroom

When modeling is done, the course-classroom distance matrix is normalized to 0-1 (Figure 2). A random course is selected from unselected courses. According to the selected course, (3) is applied for classroom selection. In Equation 3, t is the classroom which is not yet visited by the ant k. Appropriate classrooms are selected.

Classroom

239x24 double

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	0.0395	0.0353	0.0318	0.0185	0.0367	0.0339	0.0339	0.0216	0.0185	0.0185	0.0339	0.0367	0.0255	0.0339	0.0185	0.0185	0.0216	0.0353	0.0367	0.0216	0.0185	0.0185	0.0311	0.0367
2	-0.00...	0.0382	0.0347	0.0214	0.0396	0.0368	0.0368	0.0246	0.0214	0.0214	0.0368	0.0396	0.0284	0.0368	0.0214	0.0214	0.0246	0.0382	0.0396	0.0246	0.0214	0.0214	0.0340	0.0396
3	0.0374	0.0332	0.0297	0.0164	0.0346	0.0318	0.0318	0.0195	0.0164	0.0164	0.0318	0.0346	0.0234	0.0318	0.0164	0.0164	0.0195	0.0332	0.0346	0.0195	0.0164	0.0164	0.0290	0.0346
4	0.0374	0.0332	0.0297	0.0164	0.0346	0.0318	0.0318	0.0195	0.0164	0.0164	0.0318	0.0346	0.0234	0.0318	0.0164	0.0164	0.0195	0.0332	0.0346	0.0195	0.0164	0.0164	0.0290	0.0346
5	-0.00...	1	0.0365	0.0232	-0.0014	0.0386	0.0386	0.0263	0.0232	0.0232	0.0386	-0.0014	0.0302	0.0386	0.0232	0.0232	0.0263	1	-0.00...	0.0263	0.0232	0.0232	0.0358	-0.00...
6	0.0379	0.0337	0.0302	0.0169	0.0351	0.0323	0.0323	0.0200	0.0169	0.0169	0.0323	0.0351	0.0239	0.0323	0.0169	0.0169	0.0200	0.0337	0.0351	0.0200	0.0169	0.0169	0.0295	0.0351
7	-0.00...	-0.00...	0.0388	0.0255	-0.0037	-8.75...	-8.75...	0.0286	0.0255	0.0255	-8.75...	-0.0037	0.0325	-8.75...	0.0255	0.0255	0.0286	-0.0023	-0.00...	0.0286	0.0255	0.0255	0.0381	-0.00...
8	-0.00...	0.0398	0.0363	0.0230	-0.0012	0.0384	0.0384	0.0262	0.0230	0.0230	0.0384	-0.0012	0.0300	0.0384	0.0230	0.0230	0.0262	0.0398	-0.00...	0.0262	0.0230	0.0230	0.0356	-0.00...
9	0.0367	0.0325	0.0290	0.0157	0.0339	0.0311	0.0311	0.0188	0.0157	0.0157	0.0311	0.0339	0.0227	0.0311	0.0157	0.0157	0.0188	0.0325	0.0339	0.0188	0.0157	0.0157	0.0283	0.0339
10	0.0396	0.0354	0.0319	0.0186	0.0368	0.0340	0.0340	0.0218	0.0186	0.0186	0.0340	0.0368	0.0256	0.0340	0.0186	0.0186	0.0218	0.0354	0.0368	0.0218	0.0186	0.0186	0.0312	0.0368
11	-0.02...	-0.01...	-0.0133	1	-0.0182	-0.01...	-0.0154	-0.0032	1	1	-0.01...	-0.0182	-0.0070	-0.0154	1	1	-0.00...	-0.0168	-0.01...	-0.0032	1	1	-0.0...	-0.01...
12	-0.01...	-0.01...	-0.0072	0.0339	-0.0121	-0.00...	-0.0093	0.0370	0.0339	0.0339	-0.00...	-0.0121	-8.751...	-0.0093	0.0339	0.0339	0.0370	-0.0107	-0.01...	0.0370	0.0339	0.0339	-0.0...	-0.01...
13	-0.01...	-0.01...	-0.0089	0.0356	-0.0138	-0.01...	-0.0110	0.0388	0.0356	0.0356	-0.01...	-0.0138	-0.0026	-0.0110	0.0356	0.0356	0.0388	-0.0124	-0.01...	0.0388	0.0356	0.0356	-0.0...	-0.01...
14	-0.02...	-0.01...	-0.0133	1	-0.0182	-0.01...	-0.0154	-0.0032	1	1	-0.01...	-0.0182	-0.0070	-0.0154	1	1	-0.00...	-0.0168	-0.01...	-0.0032	1	1	-0.0...	-0.01...
15	-0.01...	-0.01...	-0.0119	0.0386	-0.0168	-0.01...	-0.0140	-0.0018	0.0386	0.0386	-0.01...	-0.0168	-0.0056	-0.0140	0.0386	0.0386	-0.00...	-0.0154	-0.01...	-0.0018	0.0386	0.0386	-0.0...	-0.01...

Course

Figure 2: Course-Classroom Matrix

2. Choosing Course

Equation 5 is applied to assign the remaining courses in the selected classroom. Where l is a classroom not yet visited by the ant k. Parameters are added to the equation and calculated. Selected courses are assigned to the classroom.

This is the symbolic representation and the added parameters.

$$P_{ij}^{k} = \begin{cases} \frac{\tau_{ij}^{\alpha} \times \eta_{ij}^{\beta} \times \mu_{:}^{\theta} \times \lambda_{:}^{\delta} \times \nu_{:h}^{\gamma}}{\sum_{c_{il} \in N(s^{\rho})} \tau_{ij}^{\alpha} \times \eta_{ij}^{\beta} \times \mu_{:}^{\theta} \times \lambda_{:}^{\delta} \times \nu_{:h}^{\gamma}} & if\ c_{il} \in N(s^{\rho}) \\ 0 & otherwise, \end{cases} \quad (5)$$

τ_{ij}: The pheromone matrix between classroom i and lecture j,

η_{ij}: The visibility matrix between classroom i and course j,

$\mu_{:}$: Matrix of theoretical lecture hours,

$\lambda_{:}$: Weekly total free time of lecturers,

$\nu_{:h}$: The matrix of daily leisure hours of the faculty members,

α: Relative importance of τ_{ij}

β: Relative importance of η_{ij}

θ: Relative importance of the $\mu_{:}$

δ: Relative importance of the $\lambda_{:}$

γ: Relative importance of the $\nu_{:h}$

V. DATA USED IN IMPLEMENTING THE PROGRAM

The details of data used are shown in Figure 3. Data were taken from Information Technologies Center of Mersin University. The numbers belong to the spring semester of Engineering Faculty in 2015-2016 Academic year.

University	Mersin University
Department	Engineering Faculty
Department Number	9
Common Classrooms	24
Faculty Members	105
Courses	239
Compulsory Course	159
Elective Course	50
Students Who Have Attendance Obligations	14.374
Students Who Have Not Attendance Obligations	8286

Figure 3: Data properties

VI. RUNNING THE PROGRAM

Data that were taken from Information Technologies Center of Mersin University are imported to the program by being set it to be convenient to MATLAB. MATLAB is an interactive computer program that serves as a convenient "laboratory" for computations involving matrices [5]. The first course selection has been done randomly and a rand classroom has been chosen from the course-classroom matrix. The reason for this choice is to make sure that the algorithm can select different classrooms without getting stuck on local optima. After the classroom is selected, course hours are started according to (5) of suitable courses and it is sent to the function of applied roulette wheel selection. A course is chosen in order to place it in the classroom. After the classroom is chosen, the same process is carried out for another course selection and until all of the courses are placed or all the classrooms are visited, the process continues. All the ants do this process. After all processes have finished, the best course schedule is recorded and the global pheromone is updated. The next iteration is passed. After all of the iterations are completed, the course schedule is saved and the program ends.

The pseudo-code of the algorithm that is given information above is presented in Figure 4.

Figure 4: Pseudo code for course schedule ACO algorithm.

```
-----------------------------------------------
for maxIt
  for maxAnt
            rand(course)
            for maxClassroomNumber
            apply equation (3)
            roulette wheel selection for Classroom
                for weeklyPeriod
                apply equation (5)
                roulette wheel selection for the
                course
                end
            end
  end
global pheromone update
print course schedule
end
-----------------------------------------
```

- One week is divided into 50 time slot.
- It is accepted as there will be 5 courses before noon and 5 courses afternoon every day.
- One lesson period is accepted as 40 minutes.
- 1-10 defines Monday, 11-20 defines Tuesday, 21-30 defines Wednesday, 31-40 defines Thursday and 41-50 defines Friday.
- Courses that have been placed in a classroom can be seen in Figure 5.

Time slot / *Course İnformation*

Fields	DersId	SubeId	DersAdi	DersKredi	DersKodu	DersSaat	DersMe	HocaAdi	HocaSoyadi	HocaId	SubeAdi	Bolumu	Bolum	DerslikKap	DerslikAdi
1	69498	368817	'YÖNETİM VE ORGAN...	3	'İŞL206'	3	71	'BERİL'	'DÖNMEZ'	2652	'A Şubesi'	'ELEKTRİK - ELEKTR...	320	88	'MF301'
2	69498	368817	'YÖNETİM VE ORGAN...	3	'İŞL206'	3	71	'BERİL'	'DÖNMEZ'	2652	'A Şubesi'	'ELEKTRİK - ELEKTR...	320	88	'MF301'
3	69498	368817	'YÖNETİM VE ORGAN...	3	'İŞL206'	3	71	'BERİL'	'DÖNMEZ'	2652	'A Şubesi'	'ELEKTRİK - ELEKTR...	320	88	'MF301'
4	1838	352339	'TARİHSEL JEOLOJİ'	2	'JM228'	2	73	'NURDAN'	'İNAN'	371	'A Şubesi'	'JEOLOJİ MÜHENDİS...	620	88	'MF301'
5	1838	352339	'TARİHSEL JEOLOJİ'	2	'JM228'	2	73	'NURDAN'	'İNAN'	371	'A Şubesi'	'JEOLOJİ MÜHENDİS...	620	88	'MF301'
6	101359	366791	'BİLGİSAYAR DESTEKL...	2	'MM 102'	1	68	'KADER'	'METİN'	797	'A Şubesi'	'METALURJİ VE MAL...	910	88	'MF301'
7	83224	357195	'YAPI STATİĞİ II'	4	'İNŞ 322'	2	77	'SUAT'	'KAVAS'	731	'A Şubesi'	'İNŞAAT MÜHENDİS...	530	88	'MF301'
8	83224	357195	'YAPI STATİĞİ II'	4	'İNŞ 322'	2	77	'SUAT'	'KAVAS'	731	'A Şubesi'	'İNŞAAT MÜHENDİS...	530	88	'MF301'
9	102685	366818	'MATEMATİK 2'	4	'MAT106'	2	87	'ÖZGÜR'	'MIZRAK'	1182	'A Şubesi'	'METALURJİ VE MAL...	910	88	'MF301'
10	102685	366818	'MATEMATİK 2'	4	'MAT106'	2	87	'ÖZGÜR'	'MIZRAK'	1182	'A Şubesi'	'METALURJİ VE MAL...	910	88	'MF301'
11	101362	366793	'FİZİK II'	4	'FZK 122'	3	71.5000	'SERDAR'	'YILMAZ'	1057	'A Şubesi'	'METALURJİ VE MAL...	910	88	'MF301'
12	101362	366793	'FİZİK II'	4	'FZK 122'	3	71.5000	'SERDAR'	'YILMAZ'	1057	'A Şubesi'	'METALURJİ VE MAL...	910	88	'MF301'
13	101362	366793	'FİZİK II'	4	'FZK 122'	3	71.5000	'SERDAR'	'YILMAZ'	1057	'A Şubesi'	'METALURJİ VE MAL...	910	88	'MF301'
14	5776	352375	'JEOFİZİK'	2	'JM246'	2	60.5000	'MUSTAFA'	'ERYILMAZ'	382	'A Şubesi'	'JEOLOJİ MÜHENDİS...	620	88	'MF301'
15	5776	352375	'JEOFİZİK'	2	'JM246'	2	60.5000	'MUSTAFA'	'ERYILMAZ'	382	'A Şubesi'	'JEOLOJİ MÜHENDİS...	620	88	'MF301'
16	70648	352471	'FİZİK II'	4	'FZK122'	3	32.5000	'ŞEREF'	'KAPLAN'	246	'A Şubesi'	'JEOLOJİ MÜHENDİS...	610	88	'MF301'
17	70648	352471	'FİZİK II'	4	'FZK122'	3	32.5000	'ŞEREF'	'KAPLAN'	246	'A Şubesi'	'JEOLOJİ MÜHENDİS...	610	88	'MF301'
18	70648	352471	'FİZİK II'	4	'FZK122'	3	32.5000	'ŞEREF'	'KAPLAN'	246	'A Şubesi'	'JEOLOJİ MÜHENDİS...	610	88	'MF301'
19	133229	368835	'SAĞLIKLI YAŞAM'	2	'FEM384'	2	76	'EVREN'	'DEĞİRMENCİ'	1554	'A Şubesi'	'ELEKTRİK - ELEKTR...	320	88	'MF301'

Figure 5: Classroom course schedule

A sample of Engineering Faculty's course scheduling is seen in Figure 6. Course Identity is written on the figure.

Classroom / *Time Slot*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1329...	132...	5783	5783	5783	1038...	103...	103...	1038...	0	103566	103566	1035...	0	0	103893	1038...	70634	70634	0	10485	10485	10485	1599	0	5782	5782	5782	1005...	1005...	103828	103828	70640
3	1005...	100...	70625	70625	0	1863	1863	100...	1005...	100526	10494	10494	10494	1617	1617	19549	19549	19549	0	0	1332...	133226	103568	103568	103568	19518	19518	0	0	0	0	0	10474
4	98276	98276	98276	1035...	103...	0	0	0	83205	0	71952	71952	71952	0	0	83637	0	98273	98273	98273	70636	0	71946	71946	0	71934	71934	71934	1005...	1005...	1894	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	16040	16040	16040	0	0	71909	71909	71909	0	0	69245	69245	69245	0	0	71926	71926	1038...	1038...	0	71935	71935	71935	1890	1890	103...	103843	1709	1709	69399	103855	103855	103855
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	98275	98275	98275	1609	1609	98271	98271	98271	0	0	83225	83225	18276	182...	0	103574	69233	69233	69233	0	1026...	102685	6668	6668	6668	1709	1709	71948	1863	1863	0	0	19762
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	1013...	101...	101...	1038...	103...	1598	103...	103...	1038...	69414	103843	103843	1888	1888	0	71936	71936	71936	71927	71927	1850	1850	7132	7132	0	15901	15901	15901	0	0	100593	100593	100593
12	0	0	100...	1006...	0	69441	69441	69441	69251	0	69459	69459	69459	0	0	0	0	85355	85355	85355	1608	1608	1872	1872	1872	17531	70620	1034...	1034...	1034...	102597	83642	83642
13	69498	69498	69498	1838	1838	1013...	83224	83224	1026...	102685	101362	101362	1013...	5776	5776	70648	70648	70648	1332...	1332...	69396	69396	103894	103894	103894	83225	83225	83637	83637	83637	101361	101361	100535
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	83219	83219	2062	2062	0	11427	11427	83239	83239	0	71945	71945	71945	694...	694...	69250	69250	69250	10534	10534	83220	83220	83220	83219	83219	1706	1706	1706	69493	69493	0	0	71938
16	1652	1652	1652	0	0	1005...	100...	100...	69401	69401	83212	83212	83212	0	0	10495	10495	10495	69400	0	1005...	100533	100533	83214	83214	132...	16716	16716	0	0	1587	1587	1587
17	0	103...	0	0	0	0	0	70636	70636	70636	0	0	0	0	0	19782	19782	19782	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	1036...	100...	100...	1005...	0	1005...	100...	100...	1332...	133227	100588	100588	71956	719...	719...	103834	1038...	1038...	0	0	70630	70630	1910	1910	0	102...	102598	70635	70635	0	19503	19503	19503
19	71953	71953	100...	1005...	100...	1035...	103...	1601	1601	0	19777	19777	19777	719...	719...	71957	71957	71957	71943	71943	1035...	103575	103575	0	0	10534	10534	1611	1611	0	69398	69398	100583

Figure 6: A sample of Engineering faculty course schedule

VII. CONCLUSION

In this study, course scheduling problem of Mersin University Engineering Faculty is tried to be solved. ACO algorithm is developed for this aim. By this way, NP-hard scheduling problem is aimed to come up with a solution with the help of a computer in a short time and by taking into consideration maximum course, minimum classroom and desires of faculty members.

Developed ACO algorithm is developed in MATLAB software and the desired results are achieved. Also, the program is open to adding many parameters such as giving priority to classrooms that courses have been given in advance is case of knowing it. These parameters are open to reach desired conclusions or converge in the direction of the objective function.

ACKNOWLEDGMENT

We would like to thank Information Technologies Center of Mersin University for letting us use the data.

REFERENCES

[1] Y. Demir and C. Çelik , "*Müfredat bazlı akademik zaman çizelgeleme problem inin tam sayılı doğrusal programlama yaklaşımı*," vol. 31, no. 1, pp. 145–159,

[2] G. Özyandi, "*Ders çizelgeleme probleminin 0-1 tamsayili programlama tabanli uygulamasi*," 2010.

[3] M. Dorigo and L. Maria, "*Ant colonies for the travelling salesman problem*," vol. 43, pp. 73–81, 1997.

[4] R. Interdisciplinaires, "*Université Libre de Bruxelles Ant Colony Optimization*," no. September, 2006.

[5] C. Moler, "*MATLAB Users*' Guide," 1980.

Determine the Effect of Genetic Algorithm Performance Parameters in single-machine scheduling problem

Tuğba TUNACAN[1]; S. Büşra ORTOĞLU[2]
[1] Abant İzzet Baysal University, Bolu/Turkey, tugbatunacan@ibu.edu.tr
[2] Karabuk University, Karabuk/Turkey,

Abstract- This article will discuss the effect of performance variables in achieving optimal solution in single machine scheduling problem. First of all, To determine whether genetic algorithm solution is optimal, optimal solution values for different job size will be obtained by mathematical model. In this model, Primary performance measure is Tmax, while secondary performance measure is the number of tardy jobs (nt) and total tardiness (TT) values. Genetic algorithm performance variables are crossover and mutation ratio, generation and population size. However other variables are job size, strategies of scheduling. We will utilize statistical methods to understand the effect of performance variables. Application study is to schedule the bottleneck machine for a company in the textile industry.

Keywords: Single Machine Scheduling, Genetic Algorithm, Performance variable, Statistical Methods

I. INTRODUCTION

Scheduling covers various problem patterns according to data structure, the numbers of machine, and process type. Data structure can be deterministic or stochastic while a number of machine is single or multiple problem [1]. Single machine scheduling problems are significant for understanding scheduling theory and making an induction easily for decision makers. Therefore, each scheduling problem like parallel machines scheduling or serial machines scheduling problems is actually a set of single machine scheduling problems [2]. Each single scheduling problem has very different scheduling results belonging to different objective functions. For example, if objective function which is minimizing maximum tardiness is selected, problem is solved early due date rules, and result is "tardiness time".

In literature there are many problems which have different objective. Yazdani and etc. (2017) minimized the sum of maximum earliness and tardiness of jobs [3]. Hfaiedh etal (2015). aimed to minimize the maximum delivery time under the non-resumable scenario of jobs in a single machine scheduling problem [4]. Muştu and Eren (2015) tried to minimize total tardiness in order to solve single machine problem [5]. Ozçelik and Sarac (2011) intended to minimize both job completion time and total tardiness [6]. Choobineh and etc. (2006) considered minimizing completion time of the last job, the number of tardy jobs and total tardiness as objective function [7]. Eren and Guner (2006) took total completion time and total tardiness in consideration [8] while Eren ve Güner (2005) solved single machine scheduling problem by minimizing maximum tardiness [9]. Valente ve diğerleri (2006), solved early/tardy scheduling problem [10].

Although the purposes of the problem are different, the methods used in job scheduling in the literature in general are as follows [11];

1. Mathematical Programme
2. Sequencing according to priority rule,
3. Method based simulation,
4. Method based Artificial Intelligence,
5. Heuristic Models,
6. Multi-agent Paradigms,
7. Distibuted Artificial Intelligence),
8. Production smoothing and scheduling operator method

In last decades, most commonly used methods are mathematical methods and heuristic methods among these algorithms. Yuce and etc. (2017) presents a hybrid Genetic-Bees Algorithm based optimized solution for the single machine scheduling problem [12]. Yazdani and etc. (2017) developed in order to solve small size samples. Also, a knowledge-based variable neighborhood search algorithm is developed to solve large size samples, too [3]. Muştu ve Eren (2015) utilized four heuristic algorithm; genetic algorithm (GA), GA based solution combination method, kangaroo algorithm and hybrid algorithm based both GA and kangaroo algorithm. [5]. Hfadieh (2015) proposed a branch and bound algorithm, Jackson's algorithm and Schrage's sequence for large size intances with up to 1000 jobs [4].

Ozcelik ve Sarac (2011) tried to solve scheduling problem by using the shortest setup times, the shortest weighted setup times algorithm and the shortest setup time models improved with the local search method [6]

Choobineh vd. (2006) used tabu search and mixed integer programming to solve scheduling problem [7].

Eren ve Güner (2006) benefited from math model and two heuristic methods. The first method is a special purpose

E–G heuristic. The second method is a tabu search based heuristic method that the solution result of E–G heuristic is taken as an initial solution of tabu search method to improve the performance of tabu search [8].

Valente and etc. (2006) presented a hybrid genetic algorithm. The genetic algorithm was used to establish the order in which the jobs were initially scheduled, and a local search procedure was subsequently applied to detect possible improvements [10]

Eren ve Güner (2005) proposed math model with early due date rule. They used this mathematical model to solve problems up to 24 jobs. Tabu searched used to solve large scale problem [9]

It is known that GA among these models can be more as a hybrid rather than using alone. This is due to the fact that no selection rule exists for the crossover and mutation variables used to arrive at the solution in a short time, as indicated by Cicirello and Smith (2000). Touat and etc. (2017) used crossover probability P_c= {0.7, 0.8, 0.9} and mutation probability P_m= {0.01, 0.02, 0.05} [14].

Mustu and Eren (2015) used four different values for crossover probability; 0.6, 0.8, 0.95 and 1.00. Mutation propabilities were selected 0.2, 0.35 and 0.5. In the results of 36 experiments 0.95 for crossover probability and 0.5 for mutation probability generated the best solution [5].

Kodaganallur and etc. (2014) determined mutation probability which was fixed and 0.03. However they did not give any information about crossover probability while discussing about crossover methods [15].

Valente and etc. (2006) did not give any information about both mutation and crossover probability although methods were mentioned [10].

Engin ve Fığlalı (2002) performed 10250 experiments for 10 different problems by considering initial population, crossover method, mutation method and crossover and mutation probability parameters. In this study, mutation and crossover probabilities were selected respectively 0.01 and 0.9 [16].

The probability and method of Mutation and crossover or other variables used in each study were selected differently, but effects of those on scheduling success were not discussed. So, we try to discuss the effect of genetic algorithm parameters on results of scheduling. Parameters are selected the amount of job (NJ), objective function (FF) (fitness function for genetic algorithm) and job assignment strategy which are important for both math model and genetic algorithm. Only genetic algorithm parameters are considered as mutation probability, crossover probability, generation size and population size. Crosstab, chi-square and linear regression model are used to evaluate effects of parameters on experiments. We give information about parameters in second section. In third section, statistical tests are interpreted. Finally, conclusion and future studies are mentioned in fourth section.

II. APPLICATION

In this section, manufacturing system for single machine scheduling problem will be defined and information with respect to dependent and independent variables will be given.

A. Scheduling Problem

Data for single machine scheduling problem were collected from a textile company in Turkey. The production system of this company consists of three stages; cutting fabric, fabric printing and sewing. Sewing stage is bottleneck stage because they have only one machine for sewing up. Also, this machine is assembling parts which are processed in other two stages. So, our problem is non-zero ready times model. Process time, due-date and ready times belonging to jobs were generated for 10, 20 and 50 jobs by simulating according to times from random periods. Jobs were scheduled with math model before genetic algorithm for determining whether the results of genetic algorithm are optimal or not.

Ten replications were run using for each combination of parameters. If 8 optimal results of 10 replications were same or better than results of mat model, combination of those was "succeed (s)", otherwise "fail (f)". This decision was the dependent variable for statistical tests. If the combination between mutation and crossover probabilities and generation and population size was fail, then probabilities were fixed, but generation and population size were increased. Initial combination was {0.01;1;1000;500} as respectively mutation probability, crossover probability, generation and population size. All ranges of variables were defined in section 2.2. When all combination was run, we had 5788 data in total. The summary of this data will be given in frequency analysis in section 3.1.

B. The Independent Variables of The Estimation Model

We have six independent variables for solving single machine scheduling problem; mutation, crossover, fitness function, scheduling assignment strategies, population size and generation size. Both Mutation-crossover probability and generation-population size were determined random because we did not find precise value for all of them when searched literature. These variables can be explained as follow;

Fitness function: Objective function in math model can be shortly defined as fitness function. In this paper, three

different fitness functions is considered to solve problem. These are minimizing the number of tardy jobs, the maximum tardiness value and total tardiness in the presence of non-zero ready times.

Crossover operator: Crossover used here is one-cut point method, which randomly selects one-cut point and exchanges the right part of two parents to generate offspring. This procedure is applied by computing the probability of crossover. Crossover probability is selected [0,6 ; 1] because of the effect on estimation model.

Mutation *operator:* Mutation alters one or more genes with a probability equal to the mutation rate. This rate is the probability with which each gene in a new chromosome undergoes a random change after crossover operation. We use mutation probability: [0.01;0.1]

Scheduling strategies: To describe GA approach, it is important to define two other strategies related to machine scheduling and they are called delay (D) and non-delay (ND). While non-delay strategy does not allow a machine to be kept idle when there are jobs waiting to be processed, other strategy allows machine to be kept idle. Another strategy is defined as random strategy, which is the combination of those.

Population size: It shows collections of chromosomes in GA. Population size changes sometimes for each experiment as it tries to create best solution. So, this size can be decided with regard to the structure of the problem. In this study, initial value was 500. Experiments for 10 and 20 were succeed in this value, but when experiments for 50 jobs were failed, then this value was increased up to 3000.

Generation Size: This parameter is new collections of chromosomes which are generated by mutation and crossover processes. It is thought that as generation size increases, better results will be produced. So, while initial size was 1000, if results were failed, it was increased up to 3000.

III. RESULTS

In this study, frequency analysis was performed to get general information about data. Cross-tab and chi-square analyses were made to determine the relationship between NJ, FF, AS and decision (S or F) about the results of schedules. Linear and multiple regression model was used to generate an estimation model. In following subsections, the results obtained from these analyses were explained.

A. Frequency and Cross-Tab Analysis

In this section, the result of frequency analysis made according to job size was firstly interpreted. Later, chi-square and cross-tab analysis were explained whether there was a relationship between independent variables and dependent variable.

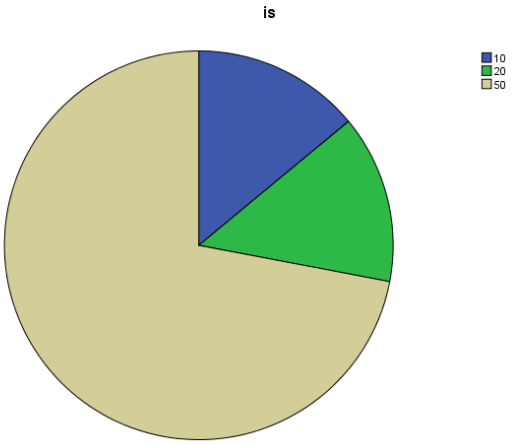


Figure 1: The result of frequency analysis belonging to job size

Figure 1 was represented the result of frequency analysis based on job size scheduled. When figure was viewed, it was determined that the number of experiments performed increases as job size to be scheduled increases. In the initial stage, the amount of experiment strategy was 810 (10 x 9 x 3 x 3) in total by depending on the combination of mutation, crossover, fitness function and assignment. If experiment was succeed at least 80 % of schedule results, generation-population sizes were not increased. For this reason, schedules of 10 and 20 jobs had successful results in initial combination sets (g.s.= 1000; p.s: 500) for all mutation and crossover probabilities, fitness function and assignment strategy, but experiments were repeated by increasing generation-population sizes because schedules for 50 jobs did not achieve success. In table 1, cross-tab and chi square analysis were presented in order to determine how many of these experiments are successful and whether they have changed depending on the job sizes.

Table 1: The results of crosstab analysis between job size and decision about schedule

		Decision		Total
		S	F	
JN	10	810	0	810
	20	810	0	810
	50	661	3507	4168
Total		2281	3507	5788
Chi-square: 3458,807 p=0.00				

When table 1 was examined, it was determined that 2281 experiments of 5788 were successful, 1620 of them were related to 10 and 20 jobs, but 661 schedules of them were for 50 jobs. At the same time, it was seen that 3507 schedule results were failed and all of them were for 50 jobs. In addition, when we assessed the result of chi-square, it was observed that success of scheduling varied depending on job size.

Crosstab analysis was verified to define relationship between job size and assignment strategy (table 2) and between job size and fitness function (Table 3). So, we decided factors that caused experiment to repeat.

Table 2: Analysis between Job Size and Assignment Strategy

		Assignment Strategy			Total
		delay	nondelay	random	
JN	10	270	270	270	810
	20	270	270	270	810
	50	1714	1003	1451	4168
Total		2254	1543	1991	5788
Chi-square: 56,515; p= 0.00					

In table 2, It was seen that experiments were repeated in the combination of 50 jobs-delay options, and the combination of 50 job-random followed. Table 3 showed that 271 experiments were conducted for the combination of 50 jobs - total tardiness (tt) and it was almost same situation like 10 jobs and 20 jobs (270 experiments).

Table 3: Analysis between Job Size and Fitness Function

		Fitness Function			Total
		nt	tmax	tt	
JN	10	270	270	270	810
	20	270	270	270	810
	50	1949	1948	271	4168
Total		2489	2488	811	5788
Chi-square: 697,060; p= 0.00					

When data was researched for extra one experiment, since mut: 0.1; generation size: 1000 and population size: 500 initial set did not acquire 80% success rate, this set was repeated by changing as generation size: 1000 and population size: 1000. The experiments of the number of tardy jobs and maximum tardiness options for 50 jobs were repeated until decision was successful.

Table 4 and Table 5 were represented relations respectively between fitness function and decision; between assignment strategy and decision.

Tablo 4: The results of crosstab between Fitness Function and decision about schedule

		Decision		Total
		S	F	
FF	NT	738	1751	2489
	T_{max}	733	1755	2488
	TT	810	1	811
Total		2281	3507	5788
Chi-square: 1444, 204; p= 0.00				

In table 4 and table 5, it was observed that both fitness function and assignment strategy had effects on decision (chi-square values and $p<0.05$). The experiments of tmax and nt options was repeated with same frequencies, while experiments of "total tardiness" option were successful except for one experiment. In table 5, it was determined that the frequencies of successful decision for assignment strategy options were almost similar, but these were very different in other situation. Also, most repetitive experiments were performed for "delay" assignment strategy option.

Table 5: The results of crosstab between assignment strategy and decision about schedule

		Decision		Total
		S	F	
AS	delay	743	1511	2254
	nondelay	783	760	1543
	random	755	1236	1991
Total		2281	3507	5788
Chi-square: 124,105; p= 0.00				

Table 6 and table 7 presented the experimental results to determine whether mutation and crossover probabilities affected on decision. In table 6, as mutation probability increased, schedule success decreased. In crossover probability, although there was no significant difference like mutation results, it was observed that as crossover probability was increased, the amount of failed decisions increased. When experiments conditions were viewed, we determined that "fail" decision did not depend only on crossover probability, but also on the combination of it with mutation probability.

Table 6: The results between mutation probability and decision

		Decision		Total
		S	F	
mut	0,01	243	1	244
	0,02	243	0	243
	0,03	243	0	243
	0,04	242	0	242
	0,05	244	8	252
	0,06	240	126	366
	0,07	228	490	718
	0,08	211	813	1024
	0,09	198	992	1190
	0,10	189	1077	1266
Total		2281	3507	5788
Chi-square: 2687,591; p=0.00				

Table 7: The results between crossover probability and decision

		Decision		Total
		S	F	
Crossover	0,60	258	323	581
	0,65	256	335	591
	0,70	256	352	608
	0,75	255	371	626
	0,80	252	389	641
	0,85	253	404	657
	0,90	251	421	672
	0,95	251	448	699
	1,00	249	464	713
Total		2281	3507	5788
Chi-square: 23,181; 0.003				

We made a decision that all the variables used to establish a linear multiple regression model because they affected on "decision about schedule" according to the results of crosstab and chi-square analyzes.

B. The results of Multiple Regression Model

In this section, the results of linear multiple regression model was presented. Dependent variable was "decision about schedule", while independent variables were six different performance parameters (defined in section 2.2).

Table 8: Model Summary

Model	R	R^2	Adjusted R^2	Std. Error of the Estimate	Durbin-Watson
1	0,868	0,754	0,754	0,24258	0,603

Table 9: Anova Results for Model

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	1041,944	7	148,849	2529,443	0,000
Residual	340,133	5780	0,059		
Total	1382,078	5787			

Table 10: Variable Coefficient in Model

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	0,353	0,027		12,881	0,000
JN	0,263	0,005	0,389	48,939	0,000
FF	-0,083	0,005	-0,119	-17,3	0,000
AS	-0,015	0,004	-0,027	-4,1	0,000
Mut p.	5,363	0,146	0,284	36,678	0,000
Cross. p.	0,111	0,025	0,029	4,466	0,000
GS	0,097	0,003	0,310	30,865	0,000
PS	0,013	0,003	0,036	4,034	0,000

Table 8 showed that there was very high significant relationship between independent parameters and dependent variable ($R= 0,868$ $R^2= 0,754$ $p < 0.05$). Also, this model established by all parameters explained %75.4 of variance of results.

First four parameters which allow us to comment according to the values of standardized coefficients were job size, generation size, mutation probability and fitness function (table 10). Also, the results of t-test presented that job size was very important interpreting parameter on the results of decision (t-value: 48,939). This was followed by mutation probability and generation size. However, population size, crossover probability and assignment strategy were less important parameters than the others (seen the results of both t-value and standardized coefficients)

IV. CONCLUSION AND FUTURE STUDIES

Our study included in discussion about parameters in genetic algorithm used by single machine scheduling problem. Job sizes were selected as 10, 20 and 50 and six parameters were also chosen. We made two decisions about each schedule result of each job size; "succeed" and "failed". Chi-square and crosstab analyzes were used to determine the effect of each parameters on decision. All variables were found to have an effect on the outcome performance.

A linear multiple regression model was established in order to generate the estimation model. All variables were used in consequence of the results of chi-square analysis. Three important parameters for model were identified as job size, mutation probability and generation size. It can be said that According to the results of crosstab analyzes, as job size and mutation probability increased, success decision about schedule decreased.

In future studies, a new estimation model will be established by using artificial neural network or other estimation models and these models will be compared with the results of regression model. Also, two-factor anova analysis will be performed between mutation*crossover probability, generation*population size, mutation probability*generation size and decision results. Thus, bilateral relations on decision can be discussed.

REFERENCES

[1] T. Eren, and E. Güner," Tek Ve Paralel Makinalı Problemlerde Çok Ölçütlü Çizelgeleme Problemleri İçin Bir Literatür Taraması", Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi, Cilt:17, No:4, 37-69, 2002

[2] M. D. Toksari and O. A. Arık, "Single machine scheduling problems under position-dependent fuzzy learning effect with fuzzy processing times", Journal of Manufacturing Systems, Vol: 45, 159–179, 2017

[3] M. Yazdani, S. M. Khalili, M. Babagolzadeh, and F. Jolai, "A single-machine scheduling problem with multiple unavailability constraints: A mathematical model and an enhanced variable neighborhood search approach", Journal of Computational Design and Engineering, Vol: 4, 46–59, 2017

[4] W. Hfaiedh, C. Sadfi, I. Kacem, and A. Hadj-Alouane, "A branch-and-bound method for the single-machine scheduling problem under a non-availability constraint for maximum delivery time minimization", Applied Mathematics and Computation, Vol:252, 496‑502, 2015

[5] S. Muştu, and T. Eren, " Geliş Zamanlarının Farklı Olduğu Öğrenme Etkili Tek Makine Çizelgelemede Toplam Gecikmenin Çözümü", Social Sciences Research Journal, Volume 4, Issue 3, 11-34, 2015

[6] F. Özçelik, and T. Saraç, "Sıra Bağımli Hazırlık Süreli İki Ölçütlü Tek Makine Çizelgeleme Problemi İçin Sezgisel Bir Çözüm Yöntemi", Endüstri Mühendisliği Dergisi ÜAS 2009 Özel Sayısı, Cilt: 22, Sayı: 4, 48-57, 2009

[7] F .F. Choobineh, E. Mohebbi, , and H. Khoo, H., "A Multi- Objective Tabu Search For A Single-Machine Scheduling Problem With Sequence-Dependent Setup Times," European Journal of Operational Research, 175, 318–337, 2006

[8] T. Eren, and E. Guner, "A Bicriteria Scheduling With Sequence Dependent Setup Times," Applied Mathematics and Computation, 179, 378–385, 2006

[9] T. Eren, and E. Guner, "Öğrenme Etkili Çizelgeleme Probleminde Maksimum Gecikmenin En küçüklenmesi İçin Çözüm Yaklaşimları", V. Ulusal Üretim Arastırmaları Sempozyumu, İstanbul Ticaret Üniversitesi, 25-27, 2005

[10] J. M. S. Valente, and J. E. Schaller, "Dispatching Heuristics For The Single Machine Weighted Quadratic Tardiness Scheduling Problem", Computers & Operations Research, Vol: 39, 2223–2231, 2012

[11] F. Maturana, P. Gu, A. Naumann, and D. H. Norrie, "Object-Oriented Job Shop Scheduling Using Genetic Algorithm" Computers In Industry, Vol: 32, 281-94, 1997

[12] B. Yuce, F. Fruggiero, M. S. Packianather, D.T. Pham, E. Mastrocinque, A. Lambiase, and M. Fera, "Hybrid Genetic Bees Algorithm Applied To Single Machine Scheduling With Earliness And Tardiness Penalties", Computers & Industrial Engineering, Vol: 113, 2017, 842-858, 2017

[13] V. A. Cicirello, and S. F. Smith, (2000), "Modeling GA Performance For Control Parameter Optimization", Proceedings. Genetic And Evolutionary Computation Conference , 2000

[14] M. Touat, S. Bouzidi-Hassini, F. Benbouzid-Sitayeb, B. Benhamou, "A Hybridization Of Genetic Algorithms And Fuzzy Logic For Thesingle-Machine Scheduling With Flexible Maintenance Problem Underhuman Resource Constraints", Applied Soft Computing, Vol: 59, 556–573, 2017

[15] V. Kodaganallur, A. K. Sen, and S. Mitra, "Application Of Graph Search And Genetic Algorithms For The Single Machine Scheduling Problem With Sequence-Dependent Setup Times And Quadratic Penalty Function Of Completion Times", Computers & Industrial Engineering, Vol:67, 10–19, 2014

[16] Orhan ENGİN, and Alpaslan FIĞLALI, "Genetik Algoritmalarla akış tipi çizelgelemede üreme yöntemi Optimizasyonu", ITU Dergisi/ d Mühendislik, Cilt:1 Sayı:1, 2002

Artificial Atom Algorithm in Image Processing Applications

B. AÇMA[1], E. ÇİFCİ[1] and B. SELÇUK[1]

[1]Karabuk University, Karabuk/Turkey, burakacma@gmail.com
[1]Karabuk University, Karabuk/Turkey, eminecifci@ogrenci.karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, bselcuk@karabuk.edu.tr

***Abstract* - In this study, Artificial Atom Algorithm, which is a new meta-heuristic algorithm approach, have been used to calculate the threshold value which is frequently performed in image processing applications. The entropy-based method has been adopted for the target function which is common in the meta-heuristic algorithms. For this aim, Shannon entropy and a new entropy method, Fractional Order Entropy, have been used. It is seen on standard test images that the Artificial Atom Algorithm can be applied to image processing applications. In addition, the results of the application were supported by comparing the values produced by certain parameters of fractional entropy with the values produced by Shannon Entropy.**

***Keywords* - Artificial Atom Algorithm, Image Processing, Entropy**

INTRODUCTION

Optimization is structures aiming to find optimal solutions on the solution of a problem. For this purpose, some algorithms have been used to solve problems. General-purpose metaheuristic algorithms are categorized into seven different groups: biological-based, physics-based, swarm-based, social-based, music-based, sports-based and chemistry-based as used in this article [1,3,4]. These algorithms have applications in different areas and one of them is image processing [2,5,6,7,8,11,12].

Motivated by [8,12], the aim of the study is to show that Artificial Atom Algorithm (A^3), a new optimization algorithm, can be used in image processing applications. In addition, Shannon Entropy formula and Fractional Order Entropy formula [10] defined by a new fractional derivative have been used for entropy calculations. The results of the application were supported by comparing the values produced by certain parameters of fractional entropy with the values produced by Shannon Entropy.

This paper is arranged as follows. A summary of A^3 is given in section II. In Section III, Fractional Order Entropy is to be mentioned. In Section IV, Entropy-based image thresholding is to be discussed. Section V includes the application results obtained by using Shannon Entropy and Fractional Order Entropy with A^3.

ARTIFICIAL ATOM ALGORITHM (A^3)

A^3 is a metaheuristic algorithm inspired by chemical compounds. As with many optimization problems, the formation of chemical compounds can also be used in optimization problems. Ionic bond and covalent bond events, occurring independently from each other, form the backbone of this algorithm [2,5,6,7,8,11].

In this algorithm, problems are assumed as atoms, and the parameters that contain a feature of the problem are assumed as electron, too. In the initial state, each of the atoms is a random solution of the problem, and processing will be started with more than one atom. The cluster formed by these solutions is called the Atomic cluster. The atomic cluster A, containing n-electrons, is given as $A = \{E_1, E_2, \dots, E_n\}$.

There are two situations while atoms exchange their electrons [7]. This method is based on covalent and ionic bonds which are effective in interatomic attraction. Atoms have nuclei and electrons around this nucleus. Some electrons in the outermost orbit are used jointly during bonding. The covalent bond is formed by electron partnership, and the ionic bond is formed by the attraction forces of oppositely charged ions. Effects of both types of bonds are present on atoms.

The operations performed in the heuristic calculation method can be divided into steps respectively as follows;

1) The atom cluster is generated randomly.
2) The electron effects of each atom are calculated, and the order of magnitude is determined. (It will be from small to large for this application)
3) Covalent bond treatment is applied.
4) Ionic bond treatment is applied.
5) The objective function value and electron effects of the new state of atoms are calculated.
6) This process continues until the end condition is provided.

The effects of each electron on the objective function are calculated. Then, electrons of each atom is sorted from large to small according to the effects of every electron on the objective function. But, in the current study, we sort them from small to large due to that we need the smallest one. After this operation, the number of electrons in an atom is multiplied with covalent ratio. With this operation, how many best electrons will be used for covalent bonding is determined. The field created by these electrons is called Covalent Field. The remainder electrons are used for ionic bonds. This field created by remainder electrons is called Ionic Field. The values of electrons in the ionic filed are canceled and instead of them new random values are produced. The electrons in the covalent field are part of solution.

A^3 is composed of two processes. Ionic and covalent bond represent these two processes. Covalent bonding determines which electrons are to be used jointly among at least two atoms, whereas ionic bonding determines which electrons' values are to be changed. Ionic bonding is the process of producing new values for electrons whose effect are too small to take into account. If a new electron is generated from these electrons with a better value than the electrons in the covalent field, this electron will find itself in the covalent field [8].

Covalent Bonding Process: Every electron has a value, and according to this value, each electron in the atom is sorted from small to large. Each atom has covalent and ionic regions. A covalent ratio is calculated depending on the aim (for example, 0.7), and electron numbers of atoms are multiplied with this ratio. The field created by these electrons is called as covalent field as in Figure 1, and covalent bonding process is applied this field. The ionic bonding process is applied to the remainder field called ionic field.

Covalent Area (CA) ← → İonic Area (İA)

A_1	$E_{1.1}$	$E_{1.2}$	$E_{1.3}$	...	...	$E_{1.m-1}$	$E_{1.m}$
A_2	$E_{2.1}$	$E_{2.2}$	$E_{2.3}$	...	...	$E_{2.m-1}$	$E_{2.m}$
A_3	$E_{3.1}$	$E_{3.2}$	$E_{3.3}$	...	...	$E_{3.m-1}$	$E_{3.m}$
A_4	$E_{4.1}$	$E_{4.2}$	$E_{4.3}$	...	...	$E_{4.m-1}$	$E_{4.m}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_{n-1}	$E_{n-1.1}$	$E_{n-1.2}$	$E_{n-1.3}$	...	...	$E_{n-1.m-1}$	$E_{n-1.m}$
A_n	$E_{n.1}$	$E_{n.2}$	$E_{n.3}$	...	...	$E_{n.m-1}$	$E_{n.m}$

Figure 1: Demonstration of the ionic and covalent fields

The normal representation is as in Figure 1, but for the current study we need to change the fields as in Figure 2. The main reason behind this is that we sort electrons from small to large, and small values in the ionic field are necessary for our application.

İonic Area (İA) ← → Covalent Area (CA)

A_1	$E_{1.1}$	$E_{1.2}$	$E_{1.3}$	...	...	$E_{1.m-1}$	$E_{1.m}$
A_2	$E_{2.1}$	$E_{2.2}$	$E_{2.3}$	...	...	$E_{2.m-1}$	$E_{2.m}$
A_3	$E_{3.1}$	$E_{3.2}$	$E_{3.3}$	...	...	$E_{3.m-1}$	$E_{3.m}$
A_4	$E_{4.1}$	$E_{4.2}$	$E_{4.3}$	...	...	$E_{4.m-1}$	$E_{4.m}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_{n-1}	$E_{n-1.1}$	$E_{n-1.2}$	$E_{n-1.3}$	...	...	$E_{n-1.m-1}$	$E_{n-1.m}$
A_n	$E_{n.1}$	$E_{n.2}$	$E_{n.3}$	...	...	$E_{n.m-1}$	$E_{n.m}$

Figure 2: The ionic and covalent field demonstration for this study.

Ionic bonding process: With ionic bonding process, the value of an electron, whose effect is not good, is canceled and instead of it a new one is chosen randomly from the data set. After ionic bonding process, the effects of electrons in each atom on the objective function are recalculated, and then, electrons of each atoms are resorted from good to bad according to their effects. The electrons whose effect are considered as good take part in covalent field and do not change. The electrons fallen within ionic field take new values again.

Here, the random data selection function chooses randomly an electron, which is in between the largest and the smallest value of the atom's electrons, from data set.

FRACTIONAL NEW ENTROPY

Entropy is the measure of irregularity of a system. It is very important notion. In so many disciplines, entropy functions are used. The most used and well-known entropy is the Shannon entropy [13].

The interval $[a, b]$ can be divided into very small interval such as

$$[a, b] = [a, a_1] \cup [a_1, a_2] \cup \ldots \cup [a_{n-2}, a_{n-1}] \cup [a_{n-1}, b]$$

and the corresponding probabilities for each interval can be $p_i, 1 \leq i \leq n$. The Shannon entropy is

$$H(A) = -\sum_{i=1}^{n} (p_i \, log_2 \, p_i). \qquad (1)$$

In [10], Karci give new entropy definition via fractional order derivation;

$$E_1 = D_t^{(\alpha)} \sum_i |p_i - t| = \sum_i |(-p_i)^{\alpha} \ln p_i|. \qquad (2)$$

Karci's claim, the new entropy is better than Shannon entropy. For this, we will use (1) and (2) formulas in this study. As [10], we select the following values of α;

$$\{0.01,0.02,0.03,0.04,0.05\},$$

$$\{0,46.0.47,0.48,0.49,0.5\},$$

$$\{0,96,0.97,0.98,0.99,1.00\}.$$

Thus, we calculate entropy results using each value of α.

ENTROPY BASED IMAGE THRESHOLDING

The problem of determining the threshold values on the image thresholding can be defined as the k-dimensional optimization problem. Determination of K optimum threshold values can be achieved by optimizing the entropy-based problem. As is known, optimum threshold values for the maximum entropy criterion on image thresholding are found by the method developed by Kapur [9]. In this method, given I image with M pixels is represented L-levels of grayscale. The objective function which is the base for optimization problem is obtained from the histogram of the image.

The mathematical equations used in this method is given in Eq. (3), (4) and (5);

$$P_i = h(i)/M \qquad (3)$$

where $h(i)$ represents the number of pixels at the i. level. M denotes the total number of pixels in the image. P_i represents the normalized probability at the i. level.

The main purpose on the entropy-based method is to find the threshold value ensuring the maximum of f-function.

$$f([t_1, t_1, \dots, t_k]) = H_0 + H_1 + H_2 + \cdots + H_k \quad (4)$$

$$H_0 = -\sum_{i=0}^{t_1-1} \frac{P_i}{w_0} ln \frac{P_i}{w_0}, w_0 = \sum_{i=0}^{t_1-1} P_i,$$

$$H_1 = -\sum_{i=t_1}^{t_2-1} \frac{P_i}{w_1} ln \frac{P_i}{w_1}, w_1 = \sum_{i=t_1}^{t_2-1} P_i, \quad (5)$$

$$H_2 = -\sum_{i=t_2}^{t_3-1} \frac{P_i}{w_2} ln \frac{P_i}{w_2}, w_2 = \sum_{i=t_2}^{t_3-1} P_i, \dots$$

$$H_k = -\sum_{i=t_k}^{L-1} \frac{P_i}{w_k} ln \frac{P_i}{w_k}, w_k = \sum_{i=t_k}^{L-1} P_i$$

APPLICATIONS AND RESULTS

The aim of the A^3 proposed for image thresholding is to supply finding of threshold values in the k-dimensional vector to be used on the obtaining of maximum values of f-function on the entropy-based method.

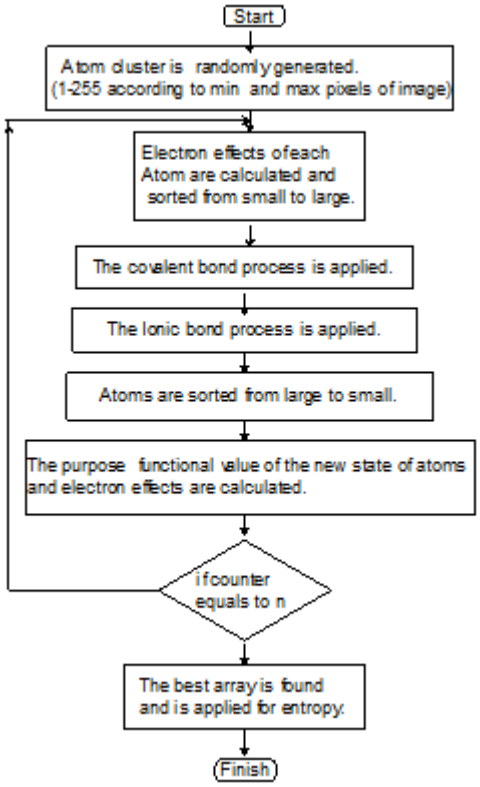


Figure 3: The flow chart for these applications

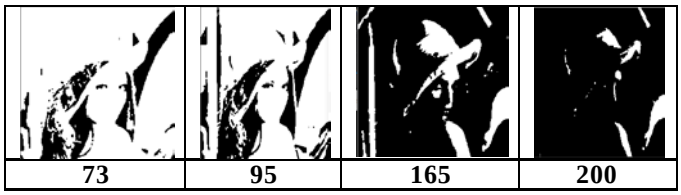
To this end, we sort the applied procedure step by step as follows:

1) First, by performing random solutions in the range of $[1, L]$, the initial values of the target function are generated.
2) By using entropy-based method, the fitness values are calculated for target function.
3) The maximum one among these fitness values is assigned as the best fitness value. The solutions providing this is assigned as the best solutions.
4) Covalent and ionic bonding process are applied.
5) By sending the obtained new solutions to fitness function again, algorithm is continued with the new fitness values [11].

The flow chart for these applications is given in Figure 3. Applications is prepared in C language on an Intel Core-i5 (4gb Ram) PC. By using Shannon entropy and new entropy formula, we studied with A^3. As examples Lena, Cameraman and Bird images in Figure 4 are used.

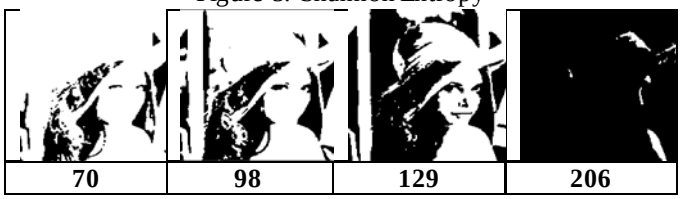
Figure 4: The images used in these applications

It is seen that Fractional entropy method was found good results from threshold for images. In Figure 5-12, the application results are shown.



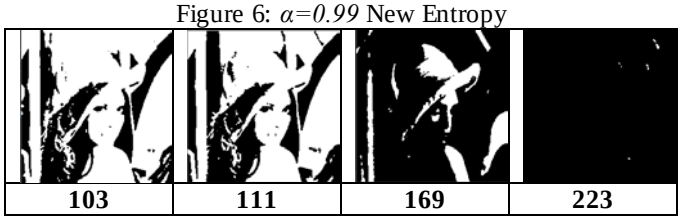
73 **95** **165** **200**

Figure 5: Shannon Entropy



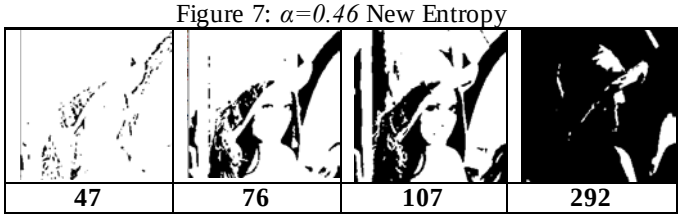
70 **98** **129** **206**

Figure 6: *α=0.99* New Entropy



103 **111** **169** **223**

Figure 7: *α=0.46* New Entropy



47 **76** **107** **292**

Figure 8: 6. *α=0.03* New Entropy

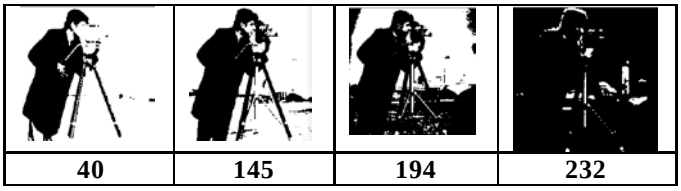


Figure 9: Shannon Entropy

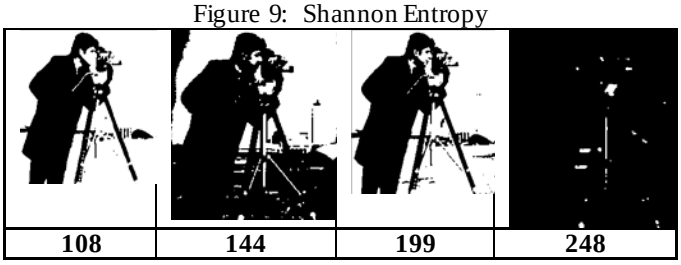


Figure 10: *α=0.97* New Entropy

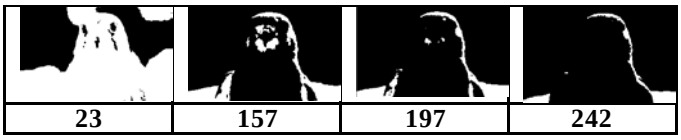


Figure 11. Shannon Entropy

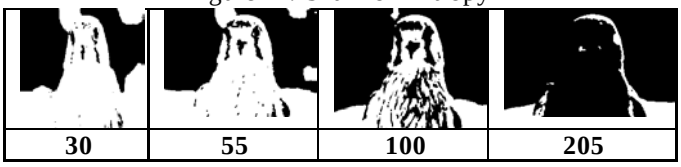


Figure 12. *α=0.98* New Entropy

The calculated threshold values are given in Table 1-2;

Images	K	Thresholding Values	Target Function	Time
Cameraman	1	227	7.026086	0.243897
	2	102,212	11.650915	0.180416
	3	41, 142, 239	17.575764	0.201734
	4	24 ,78,147,245	22.292700	0.287478
Lena	1	211	7.317718	0.203767
	2	84,218	12.357946	0.179926
	3	92,141,222	17.155878	0.218453
	4	81,109,168,223	21.571177	0.231524
Bird	1	211	6.798947	0.312652
	2	76,235	12.203512	0.220673
	3	171,196,238	16.580679	0.208311
	4	13,59,133,235	21.361496	0.250271

Table 1. Shannon Entropy Results

Images		Thresholding Values	Target Function	Time
Cameraman	1	128	5.896718	0.128110
	2	104,195	11.929565	0.236362
	3	41, 142, 227	18.145861	0.336316
	4	24 ,78,147,245	23.201468	0.522192
Lena	1	170	7.299936	0.185058
	2	84,189	12.572029	0.258869
	3	92,141,182	17.155878	0.239397
	4	81,109,168,193	21.247978	0.362142
Bird	1	75	5.502429	0.097429
	2	76,234	12.764874	0.282258
	3	171,196,238	17.268362	0.296482
	4	32,54,154,204	21.424461	0.364550

Table 2. New Entropy Based Results

When the results are examined, it is seen that A^3 can be used in image processing applications, and the new entropy also yields significant results.

REFERENCES

1. J. Kennedy, RC. Eberhart, "Particle swarm optimization". IEEE International Conference on Neural Network; 27 Nov -01 Dec 1995; Perth, WA: IEEE pp. 1942-1948.

2. E. Cifci, B. Açma, B. Selcuk, "Yapay atom algoritmasi kullanarak bireye özgü öğün programlı beslenme çizelgesi hazirlanmasi", International Artificial Intelligence and Data Processing Symposium'17, Inonu University, Malatya, TURKEY (2017).

3. D. Karaboğa, Yapay zeka optimizasyon algoritmaları, Atlas Yayınları, 2004.

4. S. Akyol, B. Alataş, Güncel sürü zekâsi optimizasyon algoritmaları, Nevşehir Üniversitesi Fen Bilimleri Enstitü Dergisi 1 (2012) 36-50.

5. A. Karcı, A new meta-heuristic algorithm based on chemical process: atom algorithm. In: 1st International Eurasian Conference on Mathematical Sciences and Applications; 3-7 September 2012; Kosovo.

6. A. Karadoğan, A. Karcı, Artificial atom algorithm for reinforcement learning. In: 2nd International Eurasian Conference on Mathematical Sciences and Applications; 26-29 August 2013; Sarajevo, Bosnia and Herzegovina.

7. A. E. Yıldırım, A. Karcı, Bireye özgü optimum beslenme çizelgesinin yapay atom algoritması kullanılarak hazırlanması, Mustafa Kemal Üniversitesi Tıp Dergisi, 2015; 6 (24): 1-11.

8. M. Canayaz, M. Demir, Veri kümelemede yapay atom algoritması ve cırcır böceği algoritmasının karşılaştırılma analizi, In 4th International Symposium on Innovative Technologies in Engineering and Science 3-5 November 2016.

9. JN. Kapur, PK. Sahoo, AKC. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, Computer Vision Graphics Image Processing, 29, pp.273–285, 1985.

10.A Karci, Fractional order entropy: New perspectives, Optik-International Journal for Light and Electron Optics 127 (20), 9172-9177.

11.M. Canayaz, A. Karcı, İmge işleme uygulamalarında cırcır böceği algoritması, Akademik Bilişim Konferansı 2015.

12. G. Das, Bat algorithm based soft computing approach to perceive hairline bone fracture in medical x-ray images, Int J of Comp Sci & Eng Tech 2013; 4: 432-436.

13.C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379–423, 623–656.

Improved Vortex Search Algorithm for Single Objective Optimization Problems

G. UÇAR[1], U. YÜZGEÇ[2]

[1] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey

gokhan.ucar@bilecik.edu.tr

[2] Department of Computer Engineering, Bilecik Seyh Edebali University, Bilecik/Turkey

ugur.yuzgec@bilecik.edu.tr

Abstract- In daily life we may encounter many different optimization problems. There are many different algorithms developed for solving these problems. The abundance of optimization problems in every field, such as science, engineering, economics, health, communication, production, etc., have made these algorithms are more popular. These algorithms are divided into two main groups as mathematical based and meta-heuristic based algorithms. The meta-heuristic algorithms are inspired by the nature events in general. In this study, we deal with the Vortex Search (VS) algorithm, which is one of the meta-heuristic algorithms. This algorithm is inspired from the vortex pattern produced by the vortical flow of stirred fluids. We proposed some improvements on vortex search algorithm to increase the performance of the original VS algorithm. To show the performance of Improved Vortex Search (ImpVS) algorithm, we used well-known meta-heuristic algorithms, such as Differential Evolution (DE) algorithm, Particle Swarm Optimization (PSO) algorithm, Artificial Bee Colony (ABC) algorithm, Simulated Annealing (SA) algorithm and original Vortex Search (VS) algorithm. The obtained Benchmark results show that the proposed ImpVS algorithm has got the competitive performance for the optimization problems.

Keywords - Meta-heuristic, vortex pattern, Benchmark , optimization problems.

I. INTRODUCTION

Heuristic algorithms are widely used to solve challenging problems in the real world [1,2]. Heuristic algorithms are designed to solve a problem faster and more efficiently than traditional methods, sacrificing integrity for optimality, accuracy, precision or speed [3]. Approximate algorithms are divided into two classes: probabilistic heuristics and metaheuristics. Specific heuristic algorithms are designed solely to solve a specific problem. On the contrary, metaheuristic algorithms represent a family of approximate algorithms that are valid for more general problem solving and for solving a wide variety of optimization problems [18].

A number of challenging scientific and industrial applications can be formulated as follows: Optimization problems. Optimization can be defined as minimizing cost and risk, or maximizing parameters such as profit, quality and productivity. It is called multi-objective problems in. Many real engineering problems are actually multi-objective, i.e., minimize cost, maximize performance, maximize reliability, etc [4]. For example, there are many possible ways to design a network to optimize cost and service quality; There are many ways to plan a production to optimize time [5]. We will talk about the most commonly used ones from these algorithms: differential evolution (DE) [6,7,8], genetic algorithm (GA) [9,10], simulated annealing (SA) [11,12], touring ant colony optimization (TACO) [13], particle swarm optimization (PSO) [14,15], artificial bee colony algorithm (ABC) [16,17] etc. VS is an optimization algorithm that is implemented by executing the behavior of the vortex pattern[18,19].

The VS algorithm gives good results for solving optimization problems. In the VS algorithm, candidate solutions are generated around the best solution using gaussian distribution in each iteration. This leads to some problems as well as simplification of the algorithm. For example, selecting a single point for generating candidate solutions for functions with many local minimums may cause me to be stuck at one of the local minimums. Using adaptive step size, the locality of the candidate solutions generated is increased in each iteration pass. For this reason, the algorithm needs to get out of the local spot as quickly as possible. Otherwise, it becomes much more difficult to escape the algorithm the second time. In this work, some changes have been made in the VS algorithm to give better results and it has named IMPVS1 and IMPVS2.

II. VORTEX SEARCH ALGORITHM

A. Initial Solution

If we consider a two-dimensional optimization problem, a vortex pattern can be modeled by a number of nested circles. The largest) circle of the vortex is first one. μ_0 is the initial centre. For solving a two-dimensional optimization problem if a vortex pattern is modulated in 2D space, a structure in the form of nested circles will emerge. The radius is narrowed according to the determined step interval, and new circles are created, so that the search space is narrowed at every step.

Here the initial center μ_0 can be calculated using Eq. 1.

$$\mu_0 = \frac{upper\ limit + lower\ limit}{2} \quad (1)$$

The upper and lower limit values are vectors in dimension $dx1$ and define boundaries in d-dimensional space.

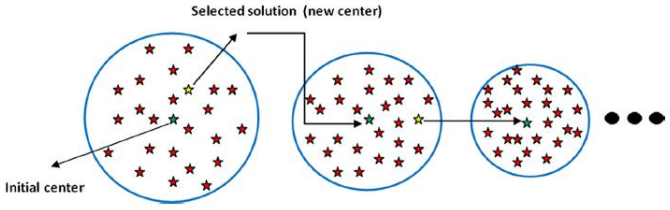


Figure 1: Search process[1].

B. Candidate Solutions

After the first center is calculated, a series of neighbour solutions ($C_t(s)$) around μ_0 are randomly generated using the d-dimensional space Gaussian distribution. Here t represents the iteration index and it is initially "0". Candidate solutions are represents as:

$C_0(s) = \{s_1, s_2, s_3, \ldots\ldots, s_k\}$, for $k = 1,2, \ldots\ldots n$.

n is the total number of candidate solutions. The general form of the multivariate Gaussian distribution is as in Eq.2.

$$p(x|\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\} \quad (2)$$

If the $\sum$'s variances of the values are equal and the covariance are zero (uncorrelated), then the resulting shape of the distribution will be spherical In this case, the value of $\sum$ can be calculated like Eq.3.

$$\Sigma = \sigma^2\,[I]_{dxd} \quad (3)$$

There, σ^2 is the variance of the distribution and I is the dxd matrix. The σ_0 can be calculate as:

$$\sigma_0 = \frac{\max(upperlimit) - \min(lowerlimit)}{2} \quad (4)$$

Here, σ_0 is equal to the initial radius (r_0) of the outer circle for optimization problem. In the beginning, the radius was chosen to get rid of the locality. Thus, in the first step, the entire search field is covered.

A. Change of Existing Solution

When it comes to the selection phase, the best solution to change the existing circle center μ_0 is to select $s' \epsilon C_0(s)$ and memorized it from $C_0(s)$. Remain within the boundaries of the candidate search solution is required. Therefore, prior to the selection process is shifted to overflowing candidate search out solutions to limit boundaries. This represents as in Eq. 5.

$$s_k^i = \begin{cases} rand\,(upperlimit^i - lowerlimit^i) + lowerlimit^i, & s_k^i < lowerlimit^i \\ s_k^i, & lowerlimit^i \le s_k^i \le upperlimit^i \\ rand\,(upperlimit^i - lowerlimit^i) + lowerlimit^i & s_k^i > upperlimit^i \end{cases} \quad (5)$$

Here , $rand$ is a random number, which is generated from uniformly distrubuted random number generator. In the next stage, the best solution is s', the second is assigned as the center. After which the new circle radius (r_1) is new solutions and $C_1(s)$ generated. $C_1(s)$ is evaluated to select a solution from $s' \in C_1(s)$.

If $C_1(s)$ is better than the best solution found so far, it is assigned as the best solution. These operations continue until reaching the termination condition.

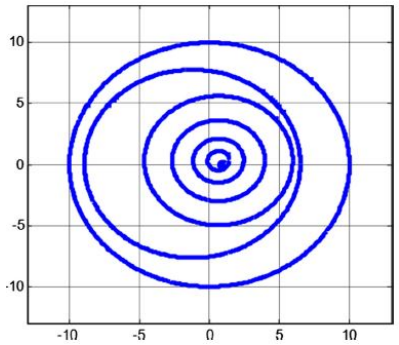


Figure 2: Pattern of the VS after a search process[1].

The radius decrement process given in Figure 3 can be considered as a type of adaptive step-size adjustment process which has critical importance on the performance of the VS algorithm. For this, radius must be tuned by decreasing in the search process. We used to inverse incomplete gamma to decrease the radius.

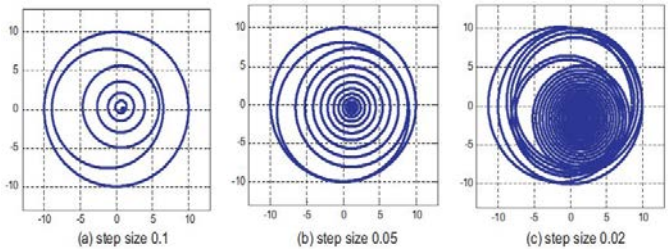


Figure 3: Vortex patterns due to resolution during search[18].

The incomplete gamma function is shown as:

$$\gamma(x,a) = \int_0^x e^{-t}t^{a-1}dt \;\; a > 0 \quad (6)$$

Incomplete gamma function's complementary Γ(a) is known as the gamma function (gamma function) and it is as Eq.7.

$$\Gamma(x,a) = \int_x^{\infty} e^{-t} t^{a-1} dt \quad a > 0 \qquad (7)$$

Different methods are proposed in the literature for the calculation of the gamma function. Inverse incomplete gamma function, the inverse of the incomplete gamma function, is represented by the $gammaincinv(x,a)$ and the invariant according to the limit value x in the integral form.

With certain step sizes, the call resolution can be adjusted by calculating a value from 0 to 1 in each step according to Eq.9.

$$a_t = a_0 - \frac{t}{Maxitr} \qquad (9)$$

Where a_0 must be selected as $a_0=1$ to search the entire search field. Lastly, the initial radius r_0 can be calculated with Eq.10. Because $a_0 = 1$, the resulting function value is $\left(\frac{1}{x}\right) * gammaincinv(x, a_0) \approx 1$, which means $r_0 = \sigma_0$ as indicated before.

$$r_0 = \sigma_0 * \left(\frac{1}{x}\right) * gammaincinv(x, a_0) \qquad (10)$$

The VS algorithm's pseudo code can be shows as follows:

Inputs: Initial center μ_0 (Eq.1)
Initial radius r_0 (Eq.10)
Fitness of best solution ($f(s_{best})$) =inf
t=0;

Repeat:
Generating candidate solutions $C_t(s)$ from μ_t and r_t
If $C_t(s)$ exceed values determined by Eq.5, shift the $C_t(s)$
s=select ($C_t(s)$);
if $f(s') < f(s')$
$s_{best} = s$;
$f(s_{best}) = f(s')$
else
Keep the best solution
end
$\mu_{t+1} = s_{best}$;
$r_{t+1} = decrease(r_t)$;
t=t+1;

Until: Maximum number of iteration
Output: Best solution s_{best}

III. IMPROVED VORTEX SEARCH ALGORITHM

Although the accuracy of the vortex algorithm is high, the processing time (CPU time) is much slower than some algorithms as DE and ABC . Two different enhancements have been tried to reduce the processing time without affecting the accuracy rate.

In the first case, the constant x coefficient is randomly generated in a certain range *[0.2-0.0001]* (IMPVS1).

```
y=1e-5;
z=0.15;
% x = 0.1  for gammaincinv(x,a)  function
 x = y + (z-y).*rand(1,1);  % random x value
ginv = (1/x)*gammaincinv(x,1);
 % initially  a = 1
```

The constant x, which is constant at the beginning, produces a value which is proportional to the specified maximum value and the minimum value range [0.2-0.0001] and decreasing proportionally starting from 0.2 (VSimp2).

```
y=1e-5;
z=0.15;
x = y+(z-(z./(tekrar-r+1)));  % reducing x value
 % x = 0.1 for gammaincinv(x,a)  function
ginv = (1/x)*gammaincinv(x,1);
 % initially  a = 1
```

The changes made in the calculation process of the algorithm are termed as IMPVS1 and IMPVS2. DE, ABC, PSO, Vortex and developed algorithms were tested with benchmark functions and performance comparisons were made.

IV. BENCHMARK TESTS

In this section, we will show results of benchmark tests for each algorithm with a table. To show the performance of the proposed IMPVS1 and IMPVS2 algorithms, some well-known benchmark functions were taken from the literature. In terms of some metrics, such as Mean Best & Std Deviation, number of function evaluations (NFE), CPU time, optimality and accuracy, both proposed algorithms provide the competitive results in comparison with the others.

Let's consider a problem as:

$$f: X \subseteq \mathbb{R}\text{^}n \rightarrow F \qquad (11)$$

is a function, and n dimension of search space, $x_0 \in X$ solution, $f(x_0) = f_0$ outcome of optimization problem and

$f(\hat{x}_0) = \hat{f}_0$ outcome of close solution, and metrics are defined as follows :

$$\alpha = 1 - \frac{\|f_0 - \hat{f}_0\|}{\|\bar{f} - \underline{f}\|} \in [0,1] \quad (12)$$

$$\beta = 1 - \frac{\|x_0 - \hat{x}_0\|}{\|\bar{x} - \underline{x}\|} \in [0,1] \quad (13)$$

$$mbs = \frac{1}{N}\sum_{i=1}^{N} \hat{f}_0 \quad (14)$$

$$std = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{f}_0 - mbs)^2} \quad (15)$$

Here, α optimality, β accuracy, mbs is mean best solution, std is standart deviation, , $\bar{f}$ and $\underline{f}$ lower and upper boundary of f , $\bar{x}$ and $\underline{x}$ lower and upper boundary of search space, and N is the run number.

We run all the algorithms for 50 times, 1000 iteration and 2-dimension. The obtained results are given in tables.

Table 1:Mean Best & Std Dev.

Number	Function	PSO	ABC	SA	DE	VS	IMPVS1	IMPVS2
1	Ackley	4,92E+02	6,21E-07	2,36E+03	8,08E-05	7,29E-05	4,61E-02	5,97E-05
2	Beale	1,60E+02	1,03E+00	2,92E+01	7,48E-06	3,65E-07	1,06E-04	3,50E-07
3	Bird	-1,04E+05	-1,07E+05	-1,05E+05	-1,07E+05	-1,07E+05	-1,07E+05	-1,07E+05
4	Booth	1,33E+01	1,10E+00	4,29E+02	5,79E-06	8,96E-07	3,77E-04	1,20E-06
5	Carromtable	-1,33E+04	-2,42E+04	-2,41E+04	-2,42E+04	-2,42E+04	-2,42E+04	-2,42E+04
6	Crossintray	-2,06E+03	-2,06E+03	-2,06E+03	-2,06E+03	-2,06E+03	-2,06E+03	-2,06E+03
7	Crosslegtable	-1,62E+02	-3,19E-01	-2,78E-01	-1,02E+02	-7,75E+00	-1,52E+00	-9,09E+00
8	Cube	3,62E+02	1,63E+01	2,86E+02	8,71E-06	1,14E-05	1,71E-02	3,31E-06
9	Giunta	6,45E+01	6,45E+01	6,47E+01	6,45E+01	6,45E+01	6,45E+01	6,45E+01
10	Griewank	4,48E+01	8,99E+00	5,99E+01	1,63E+00	3,01E+00	2,96E+00	3,35E+00
11	Himmelblau	1,32E+01	6,62E-01	5,61E+01	8,03E-06	2,16E-06	2,93E-04	2,29E-06
12	Leon	1,89E+01	7,75E-01	1,50E+01	1,14E-05	1,09E-07	2,07E-04	1,99E-07
13	Matyas	1,54E-02	1,44E-01	1,47E+01	9,04E-06	7,31E-07	1,30E-04	5,30E-07
14	Mccormick	-1,85E+03	-1,91E+03	-1,90E+03	-1,91E+03	-1,91E+03	-1,91E+03	-1,91E+03
15	Rastrigin	8,03E+02	5,17E-01	7,79E+02	6,48E-06	2,31E-06	1,63E-03	1,73E-06
16	Rosenbrock	4,14E+01	1,90E+00	3,87E+01	1,18E-05	1,39E-07	4,88E-04	1,56E-07
17	Schweffel	1,20E+05	7,53E+00	2,50E+04	2,55E-02	2,55E-02	2,72E-02	2,55E-02
18	Sinenvsin	-1,49E+03	-1,49E+03	-1,48E+03	-1,49E+03	-1,49E+03	-1,49E+03	-1,49E+03
19	Sixhumpcamel	-1,03E+03	-1,03E+03	-1,02E+03	-1,03E+03	-1,03E+03	-1,03E+03	-1,03E+03
20	Threehumpcamel	9,39E-02	2,76E-02	1,57E+01	8,46E-06	1,43E-06	1,18E-04	1,52E-06

Table 2. Number of function evaluation and CPU time.

Number	Function	PSO	ABC	SA	DE	VS	IMPVS1	IMPVS2
1	Ackley	2776.80 (0.1210)	14028.52 (0.7122)	20000.00 (1.0369)	2955.20 (0.1298)	17840.40 (0.7379)	15347.20 (0.6418)	18008.40 (0.7312)
2	Beale	2160.00 (0.0646)	21002.62 (0.7627)	20000.00 (0.7430)	3078.00 (0.0899)	16337.20 (0.4510)	12951.60 (0.3703)	16624.40 (0.4616)
3	Bird	2436.00 (0.0751)	21023.82 (0.7698)	20000.00 (0.7441)	3001.60 (0.0906)	16726.80 (0.4656)	13722.00 (0.3848)	16984.00 (0.4849)
4	Booth	2315.20 (0.0664)	21006.02 (0.7377)	20000.00 (0.7105)	2675.20 (0.0772)	16434.40 (0.4257)	12959.60 (0.3379)	16712.40 (0.4318)
5	Carromtable	2303.20 (0.0710)	21012.50 (0.7866)	20000.00 (0.7665)	2966.00 (0.0900)	16633.60 (0.4717)	13534.80 (0.3894)	16900.80 (0.4784)
6	Crossintray	1717.60 (0.0605)	20996.02 (0.8949)	20000.00 (0.8658)	2213.20 (0.0775)	15785.60 (0.5227)	12058.80 (0.4078)	16111.60 (0.5299)
7	Crosslegtable	10858.40 (0.3911)	21002.26 (0.8987)	20000.00 (132.6947)	8728.40 (0.3052)	18878.00 (0.6411)	16446.80 (0.5686)	18984.00 (0.6430)
8	Cube	2577.20 (0.0863)	21002.72 (0.7755)	20000.00 (0.7700)	6000.40 (0.1833)	16995.20 (0.4916)	13940.00 (0.3963)	17225.20 (0.4859)
9	Giunta	1536.40 (0.0501)	4485.82 (0.1768)	20000.00 (0.8003)	1202.00 (0.0398)	15253.20 (0.4658)	11482.00 (0.3594)	15631.20 (0.4872)
10	Griewank	2440.00(0.1200)	20993.82	20000.00	4920.00	16677.60	13551.60	16946.80

			(1.1168)	(1.0856)	(0.2249)	(0.7400)	(0.6002)	(0.7520)
11	Himmelblau	2318.00 (0.0681)	20997.94 (0.7319)	20000.00 (0.7264)	3461.60 (0.0983)	16541.60 (0.4453)	13474.80 (0.3683)	16818.00 (0.4494)
12	Leon	2131.60 (0.0610)	21006.94 (0.7255)	20000.00 (0.7152)	3830.80 (0.1081)	16396.40 (0.4309)	12884.80 (0.3412)	16680.40 (0.4391)
13	Matyas	1645.20 (0.0618)	20995.54 (0.9351)	20000.00 (0.9147)	2635.60 (0.0993)	15951.20 (0.5618)	12293.20 (0.4367)	16279.20 (0.5783)
14	Mccormick	1812.80 (0.0572)	20777.00 (0.7740)	20000.00 (0.7978)	1760.00 (0.0555)	18822.80 (0.5534)	14896.80 (0.4371)	18917.20 (0.5535)
15	Rastrigin	2493.20 (0.1030)	21003.78 (1.0134)	20000.00 (0.9936)	2260.40 (0.0922)	16718.00 (0.6425)	13753.60 (0.5450)	16980.00 (0.6546)
16	Rosenbrock	2273.20 (0.0882)	21002.50 (0.8762)	20000.00 (0.8670)	4316.80 (0.1702)	16570.40 (0.5575)	13202.40 (0.4475)	16838.40 (0.5781)
17	Schweffel	2716.80 (0.1585)	21000.18 (1.4456)	20000.00 (1.3641)	2279.20 (0.1358)	16925.20 (0.8996)	14105.60 (0.7442)	17164.00 (0.8981)
18	Sinenvsin	1989.60 (0.1039)	21049.04 (1.3067)	20000.00 (1.2889)	13318.00 (0.6950)	16307.60 (0.7978)	12696.00 (0.6280)	16598.40 (0.8132)
19	Sixhumpcamel	2041.20 (0.0895)	21138.02 (1.1100)	20000.00 (1.0807)	2367.20 (0.1033)	16234.80 (0.6571)	12736.80 (0.5209)	16542.00 (0.6672)
20	Threehumpcamel	1930.80 (0.0897)	21002.28 (1.1595)	20000.00 (1.1357)	1780.80 (0.0834)	16010.00 (0.6899)	12596.00 (0.5541)	16322.00 (0.7084)

Table 3. Optimality results of meta-heuristic algorithms.

Number	Function	PSO	ABC	SA	DE	VS	IMPVS1	IMPVS2
1	Ackley	0,9779493	1	0,8944504	1	1	0,999997933	0,999999997
2	Beale	0,9999991	1	0,9999998	1	1	1	1
3	Bird	0,9889028	0,9999993	0,994998	1	1	0,999999996	1
4	Booth	0,9999949	0,9999996	0,9998347	1	1	1	1
5	Carromtable	0,5497388	0,9999975	0,9990237	1	1	0,999999977	0,999999999
6	Crossintray	0,9999926	1	0,9991703	1	1	0,999999912	0,999999999
7	Crosslegtable	0,1623085	0,0002701	0,0002296	0,1023614	0,0077046	0,001466634	0,009038829
8	Cube	1	1	1	1	1	1	1
9	Giunta	0,9999977	1	0,999617	1	1	0,999999974	0,999999997
10	Griewank	0,9932231	0,9986392	0,9909409	0,9997538	0,9995448	0,999552191	0,999492542
11	Himmelblau	0,999994	0,9999997	0,9999743	1	1	1	1
12	Leon	0,9999731	0,9999989	0,9999786	1	1	1	1
13	Matyas	0,9999998	0,9999986	0,999853	1	1	0,999999999	1
14	Mccormick	0,9986341	0,9999999	0,9996716	1	0,999941	0,999940958	0,999940979
15	Rastrigin	0,9900416	0,9999936	0,9903328	1	1	0,99999998	1
16	Rosenbrock	0,9999928	0,9999997	0,9999933	1	1	1	1
17	Schweffel	0,9284108	0,9999955	0,985105	1	1	0,999999984	0,999999985
18	Sinenvsin	0,9964748	0,9999916	0,9838475	0,9999952	0,9999952	0,999995242	0,999995242
19	Sixhumpcamel	1	1	0,999998	1	1	1	1
20	Threehumpcamel	1	1	0,9999923	1	1	1	1

Table 4. Accuracy results of meta-heuristic algorithms

Number	Function	PSO	ABC	SA	DE	VS	IMPVS1	IMPVS2
1	Ackley	0,998900	0,846717	0,994733	1,000000	1,000000	1,000000	1,000000
2	Beale	0,944447	0,713023	0,975558	0,999991	0,999995	0,999967	0,999996
3	Bird	0,999149	0,863554	0,996384	0,999999	0,999999	0,999998	0,999999
4	Booth	0,998185	0,713284	0,984632	0,999998	0,999998	0,999990	0,999998
5	Carromtable	0,947343	0,792363	0,999246	0,999999	0,999999	0,999996	0,999999

6	Crossintray	0,999749	0,828812	0,996600	0,999995	0,999989	0,999977	0,999989
7	Crosslegtable	0,741319	0,757991	0,611545	0,571173	0,658680	0,656101	0,646709
8	Cube	0,975287	0,749513	0,953502	0,999994	0,999996	0,999762	0,999997
9	Giunta	0,999740	0,941186	0,996223	0,999982	0,999964	0,999967	0,999953
10	Griewank	0,965012	0,750228	0,962157	0,995832	0,992486	0,992994	0,992367
11	Himmelblau	0,999335	0,915523	0,998318	0,999999	0,999999	0,999996	0,999999
12	Leon	0,940512	0,565894	0,957122	0,999961	0,999993	0,999833	0,999993
13	Matyas	0,999919	0,749528	0,979386	0,999986	0,999989	0,999957	0,999989
14	Mccormick	0,989658	0,654685	0,990113	0,999993	0,994961	0,994959	0,994968
15	Rastrigin	0,962949	0,746924	0,975253	1,000000	0,999999	0,999997	1,000000
16	Rosenbrock	0,970028	0,691160	0,958804	0,999983	0,999997	0,999862	0,999996
17	Schweffel	0,729166	0,562322	0,952725	1,000000	1,000000	0,999999	1,000000
18	Sinenvsin	0,962218	0,757433	0,955214	0,968010	0,967424	0,967121	0,966518
19	Sixhumpcamel	0,999708	0,793095	0,997980	0,999998	0,999997	0,999991	0,999997
20	Threehumpcamel	0,999569	0,751394	0,994395	0,999996	0,999994	0,999987	0,999992

V. CONCLUSION

In this study , improved vortex search algorithm is evaluated with some metrics as, optimality, accuracy, standart deviation, mean best solution and CPU time. In general, when the results are considered, IMPVS1 and IMPVS2, which are obtained as a result of the proposed changes in the Vs algorithm, yield better results. However, the results obtained with IMPVS1 seem to be better.

These simulations shows that, vortex is good for finding best value and accuracy but CPU time is not best of one.

REFERENCES

[1] Doğan, Berat, " A modified vortex search algorithm for numerical function optimization, International Journal of Artificial Intelligence and Applications (IJAIA), Vol. 7, No. 3, May 2016.

[2] Kokash, N. (2005). An introduction to heuristic algorithms. *Department of Informatics and Telecommunications*, 1-8 . Avaliable: https://www.researchgate.net/publication/228573156_An_introduction_to_heuristic_algorithms.

[3] Reeves, C. R. (1993). *Modern heuristic techniques for combinatorial problems*. John Wiley & Sons, Inc.

[4] Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, *91*(9), 992-1007.

[5] E.-G. Talbi, Metaheuristics: From Design to Implementation, Wiley & Sons, Hoboken, New Jersey, 2009.

[6] Storn R., Price K., Differential evolution–a simple and efficient adaptive scheme for global optimization over continuous spaces, Technical report, International Computer Science Institute, Berkley, 1995.

[7] Qin, A. Kai, and Ponnuthurai N. Suganthan. "Self-adaptive differential evolution algorithm for numerical optimization." *Evolutionary Computation, 2005. The 2005 IEEE Congress on*. Vol. 2. IEEE, 2005.

[8] Qin, A. Kai, Vicky Ling Huang, and Ponnuthurai N. Suganthan. "Differential evolution algorithm with strategy adaptation for global numerical optimization." *IEEE transactions on Evolutionary Computation* 13.2 (2009): 398-417.

[9] Davis, L. (1991). Handbook of genetic algorithms.

[10] , M. (1998). *An introduction to genetic algorithms*. MIT press.

[11] Davis, L. (1987). Genetic algorithms and simulated annealing.

[12] Van Laarhoven, P. J., & Aarts, E. H. (1987). Simulated annealing. In Simulated annealing: Theory and applications (pp. 7-15). Springer, Dordrecht.

[13] Karaboga, D., Guney, K., & Akdagli, A. (2004). Antenna array pattern nulling by controlling both amplitude and phase using modified touring ant colony optimization algorithm. *International journal of electronics*, *91*(4), 241-251.

[14] Kennedy, J. (2011). Particle swarm optimization. In *Encyclopedia of machine learning* (pp. 760-766). Springer US.

[15] Xie, X., Zhang, W., & Yang, L. (2003). Particle swarm optimization. *Control and Decision*, *18*, 129-134.

[16] Karaboga, D., & Basturk, B. (2008). On the performance of artificial bee colony (ABC) algorithm. *Applied soft computing*, *8*(1), 687-697.

[17] Karaboga, D., & Basturk, B. (2007, June). Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. In *International fuzzy systems association world congress* (pp. 789-798). Springer, Berlin, Heidelberg.

[18] Doğan, B., & Ölmez, T. (2015). A new metaheuristic for numerical function optimization: Vortex Search algorithm. *Information Sciences*, *293*, 125-145.

[19] Doğan, B., & Ölmez, T. (2015). Vortex search algorithm for the analog active filter component selection problem. *AEU-International Journal of Electronics and Communications*, Volume 69, Issue 9, September 2015, Pages 1243-1253, ISSN 1434-8411.

Car Model Categorization with Different Kind of Deep Learning Convolutional Neural Network Models

K. UYAR[1] and E. ÜLKER[1]

[1] Selcuk University, Konya/Turkey, kubrauyar@selcuk.edu.tr
[1]Selcuk University, Konya/Turkey, eulker@selcuk.edu.tr

***Abstract* - With the increasing of data size in recent years, processing and analyzing of the big data has become difficult. Therefore, interest in this area has also gone up. In 2006, Geoffrey Hinton, the pioneers in the field of machine learning, developed a deep learning model at the University of Toronto. The basis of the deep learning model is based on the use of a large number of hidden layers in artificial neural networks. Previously, increasing the number of hidden layers in artificial neural networks, deepening the network, was not preferred because it caused complexity in calculation processes. Deep learning that provides a nonlinear transformation of the big data can model complex relationships with a multi-layered structure. As this model provides more comprehensive attribute learning, the classification becomes more successful. For this study, Convolutional Neural Network (CNN) model is represented for the large-scale image classification. CNN is the most widely used deep learning model in feature learning, recognition, and classification. In order to obtain a powerful image classifier, there should be large amount training data. Because of that, data augmentation techniques are used to boost the performance of the deep networks. In this study, data was increased using image augmentation methods that create artificial training examples through different ways of processing techniques such as adding noise, rotating, shifting, shearing, and resizing. The data that consists of car photos taken from the rear and are provided from the literature are separated into 40 categories. Finally, the classification performances using different deep learning CNN models such as AlexNet, VGG16, and VGG19 are evaluated and experimental results are reported.**

***Keywords* – Alex NET, big data, convolutional neural network, deep learning, VGG NET.**

I. INTRODUCTION

Deep learning is a kind of machine learning which is one of the artificial intelligence methods[2]. Machine learning is an approach that estimates and deduces from existing data. In order to establish a machine learning system, the feature vector must be extracted. In the deep learning methods, the data is directly given to the network. It obtains the necessary information about the data in the representations that creates in the different layers.The concept of "deep" refers to multi-layer artificial neural networks. Deep learning has come to the forefront with the 2012 ImageNet competition. There are several reasons for the increase in the popularity of deep learning. In particular, thanks to the internet, huge amounts of data have been produced and stored in digital media. Deep learning systems have benefited from using this big data. In addition, graphics processors are specialized in parallel computing. In this way, some operations that CPU is slow can do much faster. The result is an increase in processing power, allowing deeper models to be used in practice. Today, deep learning has many different applications in the field such as speech recognition, image recognition, and natural language processing.

In literature, there are lots of studies deep learning approach was used. Eunsuk et al.[3] used deep learning networks for stock market analysis and prediction. Joseph et. al[4] developed real-time object detection system using R-CNN deep learning model with training ImageNET dataset. Yannis et. al[5] presented LipNet which maps a variable-length sequence of video frames to text with integrating spatiotemporal convolutions and recurrent networks. In addition, there are numerous studies on video images to analyze it [6-9]. There are also studies on classification using medical images and extraction of semantic information from the image. [10-11]. Finally, Guoxing et. al[12] proposed a robust and efficient visual tracking system using regional deep learning model.

The rest of this paper is organized as follows. Section 2 presents an overview of deep learning. Section 3 provides comprehensive experimental results and Section 4 draws conclusion.

II. RELATED WORK

Deep Learning Methods

Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Auto-Encoder(AE) are among the popular deep learning neural network architectures. Different deep learning architectures can be used depending on the structure of the problem. For example, while CNN is preferred for image-based studies, RNN is preferred for time-dependent studies.

Convolutional Neural Networks (CNN)

CNN, one of the deep learning methods, can be thought of as a multi-layered artificial neural network. It is frequently used in computer vision applications. It is made up of some basic layers. These layers are input layer, convolution layer, rectified linear unit layer (ReLu), polling layer, fully connected layer, drop-out layer, and classification layer. The general structure is shown in Fig 1:

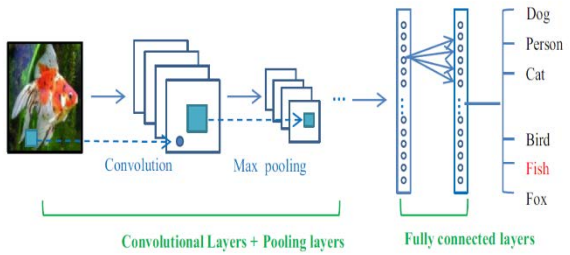


Figure 1: The general structure of the CNN

In the input layer, the data is given raw to the network. The size of the data in this layer is important. A suitable input image size should be selected for hardware computation cost and network performance.

In the convolution layer, feature map is created by convolution process. Filters can be in different sizes such as 2x2, 3x3, and 5x5 etc.

Rectified Linear Unit Layer (ReLu) comes after the convolution layer. It is known as the activation layer. Other activation functions such as tangent hyperbolic and sigmoid may also be used. The graphical representation of the ReLu is illustrated in Fig 2:

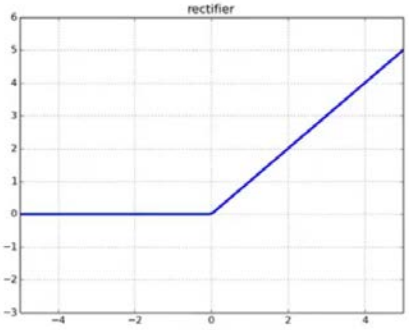


Figure 2: Graphical representation of ReLu

Also, the mathematical definiton of the ReLu is defined as follows:

output_pixel_value=max(0, input_pixel_value)

The pooling layer is usually applied after the ReLu layer. The main goal is to reduce the input size for the next convolution layer. Reduction in size causes loss of information
but creates less computational overhead for the next network layer. In addition, the system's tendency to overfit is avoided. In general, max pooling is preferred. However, there are a number of varieties such as average pooling and minimum pooling.

The fully connected layer transforms the two-dimensional feature map that results from the last convolution layer into a one-dimensional vector for classification. This layer itself depends on all areas of the previous layer.

CNN Models

Deep learning CNN models have been developed rapidly since 2012. All models mentioned below are trained on the ImageNET dataset. The most important ones are given below:

AlexNet (2012): This model was developed by Krizhevsky, Sutskever, and Hinton. It is the first model in the ImageNET competition. This model consists of 5 convolution layers, pooling layers, dropout layer and 3 fully connected layers. The architecture is designed to classify 1000 objects[1]. Training was done on ImageNet data and ReLU was used as a nonlinearity function. With this CNN model, the error rate of object recognition has been reduced to 16.4% from 25.8%.

ZFNet(2013): This model was developed by Matthew Zeiler and Rob Fergus. Apart from a few minor changes, the architecture resembles AlexNet, but it is more developed. In the ZFNet architecture, smaller size filters (7x7) and reduced steps (stride) were used in the first layer. The logic behind this change is that a smaller filter size in the first convolution layer helps to preserve many original pixel information in the input volume. ReLU is used as a nonlinearity function. While AlexNet has trained around 15 million images, ZFNet has been trained on 1.3 million images[1].This model won the ImageNet contest in 2013 with an error rate of 11.7%.

VGGNet(2014): This model was developed by Karen Simonyan and Andrew Zisserman. Due to the use of 3x3 size filters, it is quite different than AlexNet (11x11) and ZFNet (7x7) filters. ReLU is used as a nonlinearity function. In ImageNet competition, 7.3% error was achieved[1].

GoogLeNet(2014): GoogLeNet is a complex architecture due to its Inception modules. GoogleNet has 22 layers and has won ImageNet 2014 competition with a 6.7% error rate[1].

Microsoft ResNet(2015): It is an architecture that is designed deeper than all architects. Microsoft ResNet consists of 152 layers. This model won the ImageNet 2015 competition with 3.57% error rate[1].

The success of the CNN models according to the ImageNet competition results are illustrated in Fig. 3.

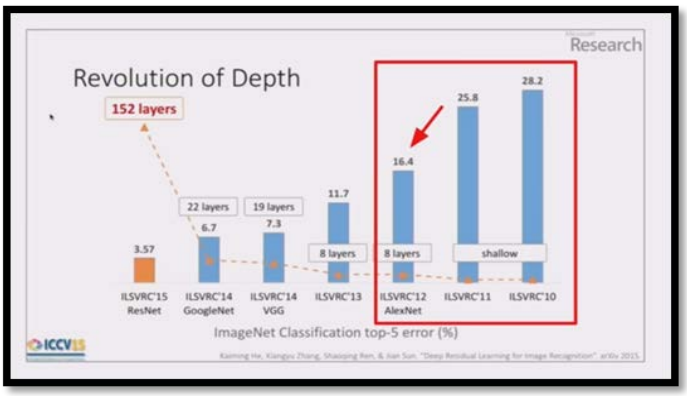


Figure 3: ImageNET competition results

R-CNN(2013): Image classification is usually based on the prediction of an object. Object identification is the process of determining where the object is located and the boundaries it covers. The R-CNN structure generally consists of 4 phases.
In the first stage, the picture is taken. In the second step, the areas are determined by Selective Search.In the third stage, each selector field is sent to a CNN that is designed similar to

AlexNet. In the last stage, if the CNN output is an object, an arrangement is made on the frame to produce the final result[1].

III. EXPERIMENTAL RESULTS

For this study, it is aimed to classify car models with convolutional neural networks, one of the deep learning methods. Color images (RGB images) have been worked on. Classification performances were tested using the existing AlexNet, Vgg-16 and Vgg-19 CNN deep learning models in the literature. Experimental results are presented using 5-fold cross-validation.

Data Set

The dataset originally consists of 526 car images taken from the rear of the cars. 40 categories were created. With data augmentation method, the number of data is increased to 2000. Data augmentation methods used: image rotation, adding noise, histogram equalization and translating (shifting).

Training Strategy of the CNN

The training phase consists of four basic steps which are the forward pass, loss function calculation, backward pass, and weight updating. The forward stage allows each layer to be represented by the weight and bias parameters of the input image. The prediction output is used to calculate the loss cost. The backward stage calculates the gradient of each parameter with the chain rule, depending on the cost of the loss. Depending on the gradient, all parameters are updated and prepared for the next forward calculation. After a certain number of iterations, network training is stopped. There are learning algorithms such as Sgd, Sgdm, Adam, Adagrad, and RMSProp. In order to prevent overfitting, dropout, data augmentation and fine-tuning can be used.

Classification Strategy of the CNN

In order to use machine learning methods, the feature vector to represent the data in the best way must be specified. In order to determine these feature vector, a preprocessing is required before the classification process. However, deep learning approaches can perform the classification process without any preprocessing. Data are directly given to the deep learning network, and features appear in the convolution layers.Finally, classification is performed in the classification layer. The learning algorithm is also of importance to training network. For this study, Stochastic Gradient Descent with Momentum (SGDM) approach is used.

Results

The classification performances of the trained models which are AlexNet, VGG-16 and VGG-19 were tested with the dataset we modified. The pipelines of the mentioned models are shown in Fig. 4.

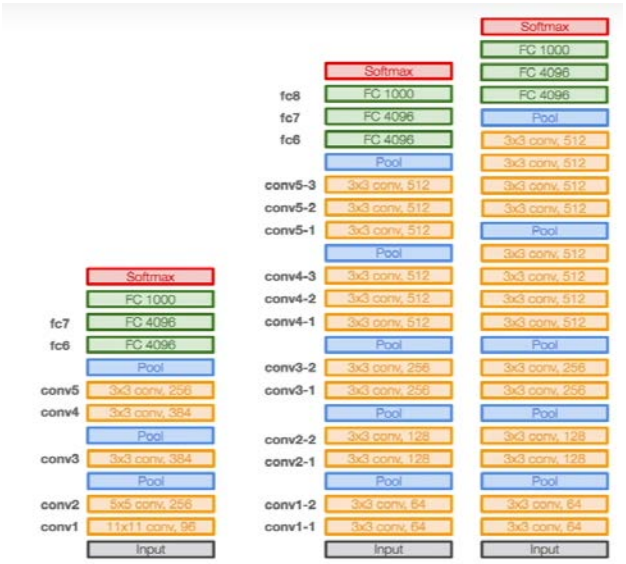


Figure 4: The pipeline of the AlexNET, VGG-16, and VGG-19

Since these models were trained on the Imagenet data, re-training was performed on the data. This process is known as fine-tuning. The following Fig.5, Fig.6. and Fig. 7 show the training success rates of the three models.

Some parameters such as learning rate, batch size for network training should also be determined. These parameters are given in Table 1, Table 2, and Table 3 for each model.

Table 1: Training parameters of the AlexNET

Parameter	Interval	Value
Batch size	(2)^n, n∈{1,2,…n}	128
Initial learning rate	[0,1]	0.01
Maximum epoch	Depends on user	30

Table 2: Training parameters of the VGG-16

Parameter	Interval	Value
Batch size	(2)^n, n∈{1,2,…n}	16
Initial learning rate	[0,1]	0.001
Maximum epoch	Depends on user	5

Table 3: Training parameters of the VGG-19

Parameter	Interval	Value
Batch size	(2)^n, n∈{1,2,…n}	8
Initial learning rate	[0,1]	0.001
Maximum epoch	Depends on user	30

Because of the hardware shortcomings, smaller batch sizes have been specified for VGG-16 and VGG-19 models.

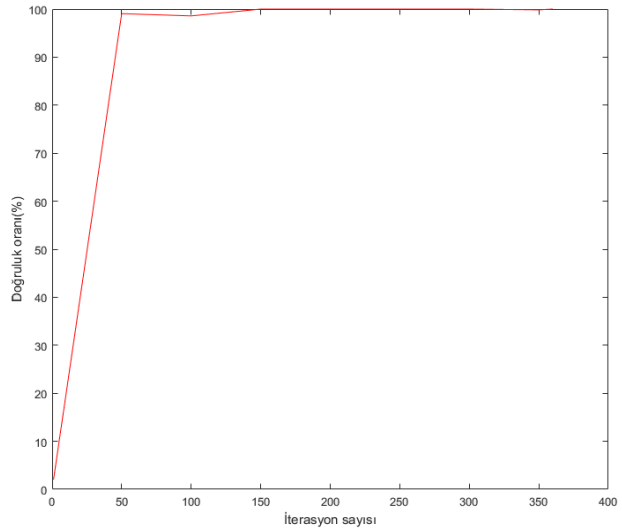


Figure 5: AlexNET training accuracy results

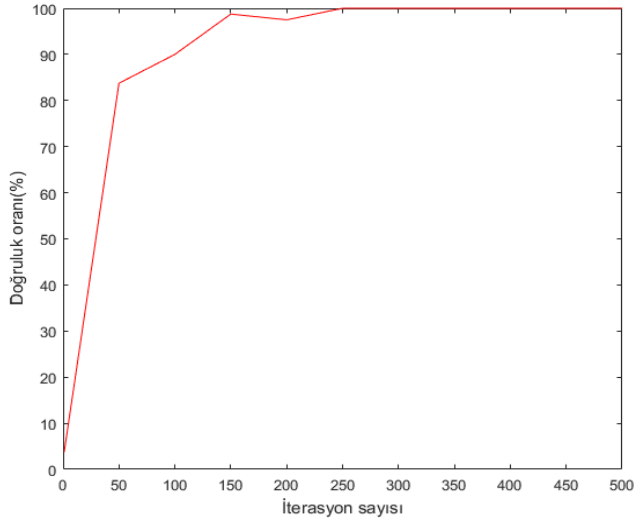


Figure 6: VGG-16 training accuracy results

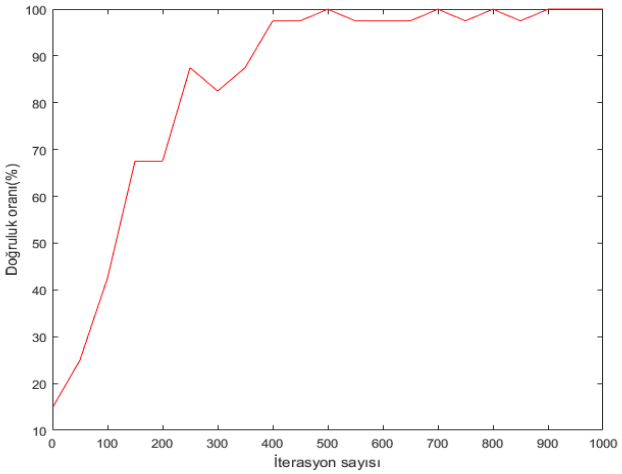


Figure 7: VGG-19 training accuracy results

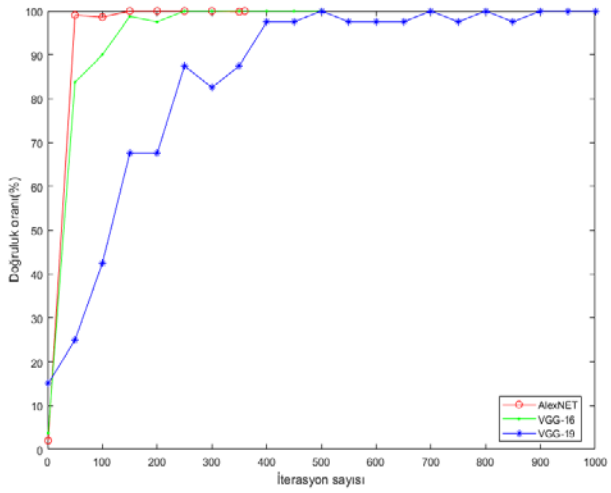


Figure 8: Comparison of training accuracy rates

Table 4: Test classification results of the CNN models

CNN Model	Input size	Classification Accuracy
AlexNet	227x227x3	% 100
Vgg-16	224x224x3	% 99.95
Vgg-19	224x224x3	% 99.95

IV. CONCLUSİON

The results of the study conducted to measure the classification success of CNN models are presented. Due to the simplicity of the dataset, AlexNet, a simple architecture compared to others, is better at classifying as seen in Table 4. Result from here, AlexNet model provides more successful results on simple data sets than other models. The more the data set is varied and the more available the data, the better the results are obtained with deep learning[13].

ACKNOWLEDGMENT

This study was funded by OYP with Project 2017-OYP-047.

REFERENCES

1. İnik Ö, Ülker E, Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri. Gaziosmanpasa Journal of Scientific Research, 2017. 6(3):p.85-104.
2. Guo Y, Liu Y., Oerlemans A., Lao S., Wu S., and Lew M.S, Deep Learning for visual understanding. Neurocomputing, 2015. 187(2016):p.27-48.
3. Chong E., Han C., and Park C.F., Deep Learning Networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications,2017. 83(2017):p.187-205.
4. Redmon J., Divvala S., Girshick R., and Farhadi A., You Only Look Once: Unified, Real-Time Object Detection.

5. Assael Y., Shillingford B., Whiteson S., and Freitas N., Lipnet: End-to-End Sentence-Level Lipreading.
6. Wang L., and Sng D., Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey.
7. Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R., and Fei-Fei L., Large Scale Video Classification with Convolutional Neural Networks.
8. Burney A., and Syed T.Q., Crowd Video Classification using Convolutional Neural Networks. 2016 International Conference on Frontiers of Information Technology,2016.
9. Olmos R., Tabik S., and Herrera F., Automatic handgun detection alarm in videos using deep learning, 2017. 275(2018): p.66-72.
10. Mohsen H., El-Dahshan E.A., El-Horbaty E.M, and Salem A.M., Classification using Deep Learning Neural Networks for Brain Tumors.
11. Vasconcelos C.N., and Vasconcelos B.N, Experiments using deep learning for dermoscopy image analysis. Pattern Recognition Letters, 2017. p.1-9.
12. Wu G., Lu W., Gao G., Zhao C., and Liu J., Regional deep learning model for visual tracking. Neurocomputing, 2015. 175(2016): 310-323.
13. Zhang Q., Yang L.T., Chen Z., and Li P., A survey on deep learning for big data. Information Fusion,2017.42(2018):p.146-157.

Increasing the Performance of SAR Image Despeckling Using Convolutional Neural Networks

Y.S. GUNAYDIN[1] and B. SEN[1]

[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, yusufsevkignydn@gmail.com
[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, baha.sen@gmail.com

***Abstract* - Using Synthetic Aperture Radar images become popular in many military or civilian applications such as algorithm design, geo-referencing and Automatic Target Recognition. One of the main reason is SAR images can be obtained in any weather condition like rainy or cloudy weather even without daylight. However, Synthetic Aperture Radar (SAR) images contain multiplicative noise called speckle which makes analyzing images difficult. Therefore, there are many algorithms developed about despeckling SAR images in last decades. Each algorithm has strengths and weaknesses such as some algorithms work great in texture areas and some can work fine about homogeneous regions. To achieve more efficient result in despeckling SAR images, we proposed a method which uses 3 despeckling algorithms (SSD, MSAR_BM3D and FANS) and apply those algorithms in the regions which they are powerful. The proposed method splits a SAR image into smaller images and use Convolutional Neural Networks to categorize the sub images to find which algorithm is the best for that region. Afterwards, sub images despeckled using the algorithm which CNN selected and sub images come together and create the final despeckled image. The proposed method aimed despeckling of noises from the Synthetic Aperture Radar images more effective than the available despeckling algorithms.**

***Keywords* – Convolutional Neural Networks, Deep Learning, Denoising, Despeckling, Synthetic Aperture Radar.**

I. INTRODUCTION

SYNTETIC aperture radar images are the high resolution images of specific locations or the targets, which can be taken in any weather condition such as rainy or windy weather or any time like at night or daylight. This power of SAR images come from remote sensors which uses low wavelength. Under favor of the power, SAR image usage becomes popular in many areas like geo-referencing, monitoring of an area and target recognition [1,2]. Besides, there are challenges about SAR images such as speckle in images and the complexity of despeckling that noise. The multiplicative noise on SAR images makes difficult to complete those tasks. Speckle makes difficult to obtain information from SAR images and reduces the reliability of applications which using SAR images. Therefore, making SAR images more meaningful by denoising the speckle becomes an intense preprocessing task [5].

Although despeckling synthetic aperture radar images is complex, there are many algorithms in this area such as Sparsity-Driven Despeckling (SDD), Fast Adaptive Nonlocal SAR Despeckling (FANS), Multitemporal SAR Image Despeckling Based on Block-Matching (MSAR-BM3D) and many others. Each algorithm has advantages and disadvantages on despeckling SAR images. For instance, an algorithm can perform better in homogeneous areas but worse in the heterogeneous areas of synthetic aperture radar images [3-5].

Inspired from these, we propose an approach which uses Convolutional Neural Networks to increase the performance of synthetic aperture radar despeckling algorithms. The main idea of this approach is detecting the areas which algorithms works well by training a Convolutional Neural Network. After that divide a speckled image into sub images and learning the best algorithm by testing sub images in CNN model which we create and apply the algorithm which CNN chose. In consequence of these, sub images despeckled with best algorithm according to them, create the denoised SAR image. In proposed approach, 3 despeckling algorithms are used to increase the performance:

1) Sparsity-Driven Despeckling Algorithm [3]
2) Fast Adaptive Nonlocal SAR Despeckling Algorithm [4]
3) Multitemporal SAR Image Despeckling Based on Block-Matching Algorithm [5]

These algorithms are chosen because of their success in despeckling images and their public accessibility. In the proposed method sub images are classified out of these 3 algorithms according to the Structural Similarity Index of despeckled images by these methods. Proposed despeckling algorithm improved performance of despeckling SAR images.

The next section gives brief information about 3 despeckling algorithms which proposed method used, Sparsity-Driven Despeckling Algorithm, Fast Adaptive Nonlocal SAR Despeckling Algorithm and Multitemporal SAR Image Despeckling Based on Block-Matching Algorithm. Section III defines proposed method and Section IV shows result of algorithms. Finally, conclusions are mentioned in Section V.

II. DESPECKLING ALGORITHMS

a) Sparsity-Driven Despeckling:

SDD is a variational approach and uses l_0-norm, l_1-norm and l_2-norm in Total Variation for despeckling. Sparsity Driven Despeckling Algorithm describes denoising synthetic aperture radar images as an optimization problem and SDD algorithm despeckles homogeneous areas very well [3].

b) Fast Adaptive Nonlocal SAR Despeckling:

FANS is a denoising technique which uses nonlocal approach. Nonlocal methods are complex but their results of despeckling images are very well. The algorithm offers less complexity, less time cost and excellent performance on despeckling SAR images [4].

c) Multitemporal SAR Image Despeckling Based on Block Matching:

MSAR-BM3D is a despeckling method which uses block matching and collaborative filtering techniques to denoising operation and this technique is based on nonlocal approach and improved version of SAR-BM3D [5].

III. PROPOSED METHOD

Simply, the purpose of this study is giving an image to Convolutional Neural Network as an input and then, CNN classifies and despeckles sub images which are created from input image into 3 classes which show best despeckling performance for that sub image out of the chosen algorithm.

The Convolutional Neural Network architecture of proposed method is shown in the Tab.1. The network consists of 4 convolutional (Conv) layers with 3 pooling (Pool), 2 dropout (Drop) steps and 2 Fully Connected (FC) layers.

Table 1: Convolutional Neural Network Architecture

	Layer	Filter Size	#Filters
Layer 1	Conv ReLU Drop Pool	3 x 3 x 1	32
Layer 2	Conv ReLU Drop Pool	3 x 3 x 32	64
Layer 3	Conv ReLU	3 x 3 x 64	128
Layer 4	Conv ReLU Pooling	3 x 3 x 128	256
Layer 5	FC ReLU FC ReLU	3 x 3 x 256	512 256

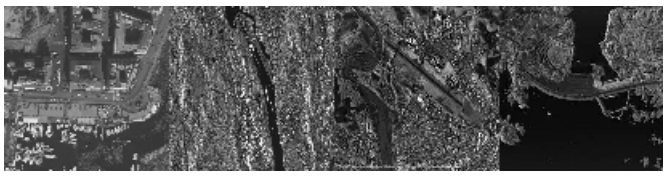


Figure 1: Some images used to train CNN

The method which we proposed for increasing the performance of despeckling algorithms works as follows.

SAR images are collected and Gaussian noise is added to the SAR images as shown in Fig.2. Synthetic aperture radar images are divided into 32 x 32 sub images after adding artificial speckle to them. Subsequently, these SAR images are despeckled by using the algorithms that we chose (SSD, FANS, MSAR-BM3D). Afterwards, the similarity between original and despeckled images is calculated by using Structural Similarity Index (SSIM) and the training set for the Convolutional Neural Network is created by this way. Therefore, CNN is trained by sub images and SSIM values of algorithms.

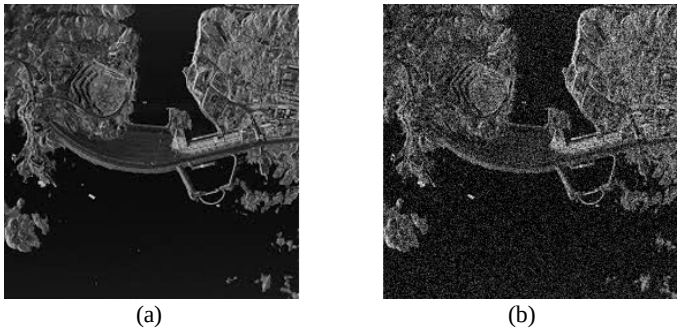
The Structural Similarity Index (SSIM) is a method which measures the similarity between two images based on degradation of structural information [7].

After Convolutional Neural Network is trained, input images are selected outside of the training set. Speckle noise is added to these images and 32 x 32 sub images are created from the speckled SAR image. CNN selects the best despeckling algorithm for these sub images.

32 x 32 sub images despeckled with best algorithm according to the result of Convolutional Neural Network and these images come together and create the output images.

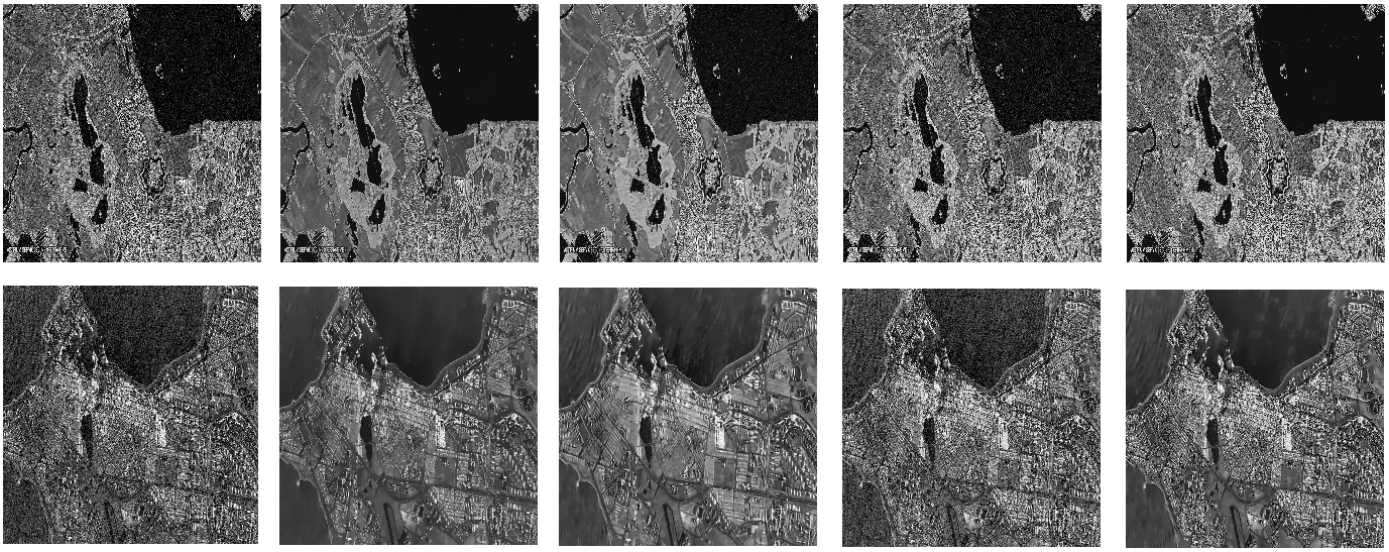
IV. EXPERIMENTAL RESULTS

The despeckled images of proposed method and the images which are despeckled by 3 algorithms (SDD, FANS and MSAR-BM3D) are shown in Fig.3 with speckled input images. These despeckled images are compared with noise free original image and SSIM values are found as shown in Tab.2. The experiments show that proposed method provides better results than 3 algorithms which are used in our method.



(a) (b)

Figure 2: Sample of adding Gaussian noise to SAR images. (a) Original SAR image and (b) Speckled image



(a) (b) (c) (d) (e)

Figure 3: Speckled images and results of despeckling algorithms
(a) Speckled SAR images, (b) SDD, (c) FANS, (d) MSAR-BM3D and (e) Proposed method

Table 2: Structural Similarity Index (SSIM) Results of Algorithms

	SDD	MSAR_BM3D	FANS	Proposed Method
Image1	0.58	0.59	0.47	**0.67**
Image2	0.66	0.56	0.66	**0.72**

MATLAB R2016a and Python 3.6 were used in experiments with an Intel i7-6700HQ CPU at 2.60 GHz. Convolutional Neural Network was created in Python using Keras library. Image processing tasks and setting up training set were done by MATLAB using Image Processing Toolbox.

V. CONCLUSION

In this paper, we proposed a method aiming increasing the despeckling performance of SAR images. Our method uses 3 existing despeckling algorithm which are SDD, FANS, MSAR-BM3D and a Convolutional Neural Network to compound their strengths about denoising. The experiments show that proposed method improved the results of available algorithms noteworthily.

As a future work, number of despeckling algorithms can be increased and also optimization of that algorithms can be integrated to our algorithm. For each image, the best variables of each algorithm can be found and then our proposed method can be applied.

REFERENCES

[1] Y. Maghsoudi, M. J. Collins, and D. Leckie, "On the use of feature selection for classifying multitemporal Radarsat-1 images for forest mapping," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, pp. 904–908,
Sep. 2011.

[2] G. Moser and S. B. Serpico, "Unsupervised change detection from multichannel SAR data by Markovian data fusion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, pp. 2114–2128, Jul. 2009.

[3] C. Ozcan, B. Sen and F. Nar, "Sparsity Driven Despeckling for SAR Images" *IEEE Geoscience and Remote Sensing Letters*, vol. 13, pp. 115–119, January 2016.

[4] D. Cozzolino, S. Parrilli, G. Scarpa, G. Poggi and L. Verdoliva "Fast Adaptive Nonlocal SAR Despeckling" *IEEE Geoscience and Remote Sensing Letters*, vol. 11, pp. 524–528, February 2014.

[5] G. Chierchia, M.E. Gheche, G. Scarpa and L. Verdoliva "Multitemporal SAR Image Despeckling Based on Block-Matching and Collaborative Filtering" *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, pp. 5467–5480, October 2017.

[6] P. Wang, H. Zhang, and V.M. Patel "SAR Image Despeckling Using a Convolutional Neural Network" *IEEE Signal Proccessing Letters*, vol. 24, pp. 1763–1767, December 2017.

[7] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity" *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, April 2004.

Evaluation of Object Tracking Performance of ADNet Method with Different Datasets and Color Spaces

H.ÜZEN[1*] and K. HANBAY[2*]

* Bingol University, Computer Engineering Department, Bingöl, Turkey,
[1] huseyin730@gmail.com, [2]khanbay@bingol.edu.tr

Abstract **- Recently, object tracking methods based on deep learning have shown great successes. Using deep neural network methods allows following the object even in highly complex scenarios by learning more details of the object in the video and the object motion model. The developed network architecture needs to be trained with a powerful dataset. The variety and size of the selected dataset affect the success of object tracking methods' results directly. Some object tracking studies have also used image datasets to obtain the diversity in the training set in addition to the video dataset. The object in the image has gained some kind of artificial motion/action between the two images by making the certain rotational or translational movements. But the uses of this technique brings out the question of whether the object has gained a right action or not. In this study, a new perspective is introduced to provide the data the diversity which is required by deep learning-based methods. In this paper, ADnet, which is a deep learning based object tracking method, is used to test our new perspective which is mentioned above. For the training of the ADNet method, color spaces such as HSV, L*a*b*, NTSC, YCbCr, inverted HSV (converted to HSV channeled by taking BGR instead of RGB) were analyzed. As a result of the analysis, HSV and inverted HSV (IHSV) color spaces have been found out to provide stronger and more varied training dataset. Moreover, in order to prepare a proper training set for the object tracking, two different datasets were created with the videos taken from Vot2015 and Vot2014 datasets. Six different training sets were created for each group by translating them into RGB, HSV and IHSV color spaces. A separate ADNet network architecture was trained for each training set. Tests were carried out for each method with the dataset, which included 61 videos. This test dataset includes some videos which are not used in training dataset. In conclusion, it was found that stronger training sets could be created by applying different color transformations such as HSV and IHSV to strengthen the training set.**

Key words: Object tracking, Deep learning, Color spaces, Training dataset

I. INTRODUCTION

Object tracking is a very important issue in image processing and machine learning. It is still a task on trial to develop efficient tracking because of difficulties from partial occlusion, illumination variation, motion blur and so on. Many classifiers in machine learning have been applied for tracking to develop a powerful object tracking method; such as Random Forest [1], Naive Bayes [2], Metric Learning [3], Multi-Instance learning [4] and Correlation Filters [5].

Moreover, the classifier updating is very important for providing strong success in object tracking. A method for updating of the classifier is using a positive instance (the current object position) and a few negative instances [6]. Generally, these methods are able to perform inaccurate updates by taking false positives when the object's appearance changes or disappearance problems occur. To solve this problem, Grabner et al. proposed an on-line semi-supervised learning method [7]. In this method, only the first sample is labeled. Another solution for update operations is the multi-sample learning framework provided by Babenko et al. [8]. In this method, samples of the images taken at the current frame at the time of testing were evaluated by collecting them in positive and negative bags.

Deep learning based object tracking methods have achieved great success with the developed multi-layer artificial neural network architectures. These methods learn the relationship between object and its motion throughout the training of network architectures. In this way, it has achieved strong success against challenging object tracking problems. But these methods need to be trained with a strong dataset. Bertinetto et al. [9] used image pairs from the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [10] study for the SiamFC method training. The network architecture trained in this way has been able to successfully cope with challenging object tracking problems. Held et al. [11] proposed GOTURN method which was trained with a 41 GB dataset to be able to track objects at high speeds.

The VGG [12] network architecture used by many methods is trained with a dataset which is in a quite large size and contains approximately 1.3 million pictures and 1000 classes. The ADNet method was proposed by Yun et al [13] was trained with 360 video from datasets VOT2015 [14], VOT2014 [15], VOT2013 [16] and ALOV300 [17]. Generally, the methods in the literature have used a training set in high dimensions to provide data diversity. We gained a new perspective to ensure the diversity of the data needed in this work. From this point of view, tests have been conducted on the ADNet method, which provides strong success in object tracking.

II. METHOD

1) *Action-Decision Networks (ADNET)*

Yun et al. [13] proposed the ADNet method of tracking objects through sequential actions. In this method, the network architecture learns 11 motions to move the bounding box (patch).The ADNET network architecture was trained offline for learning these movements and online updates were made for adaptation. An overview of the ADNet method is shown in figure 1. The working principle of ADNET is based on Markov Decision Process (MDP) method. The MDP method constitutes a decision making mechanism to move from one state to another. The agent responsible for the object movement changes its status in the direction of certain movements. At the end of the movement, the agent receives a response. Furthermore, if the response of the vehicle is correct it will be rewarded otherwise it will be punished. According to the response, the probability values of the situations that the system can pass are updated [18].

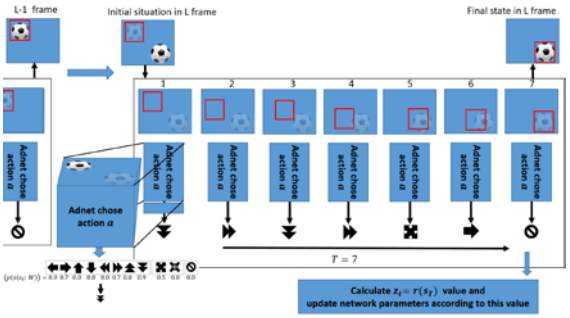


Figure 1. Flow of ADNet object tracking method

The agent in the ADNet method defines the movements that the object can make in a space. With the help of the defined movements, the patch that the rectangle bounds the object has moved to the new position of the crossing between the frames. These operations are repeated throughout all frames to aim at object tracking. In ADnet method, there are 4 definitions which are action, status, status transition function, and reward [13].

Action: It is defined as the form of movement that the object can do. A total of 11 motion patterns were created in the ADnet method as shifted and resized. These are 8 shifted movements, left, right, up, down and double of them respectively; Two-dimensional motion is defined as enlargement and reduction in size. The last movement is the stop-motion, indicating that the move is over. Each motion (action) is shown and coded with an 11-dimensional vector [13].

Status: $\{p_t, d_t\}$ represents the position and movement information of the identification box (defined as a part) representing the object with a description group. Here, $p_t \in R^{112x112x3}$ is the selected parted image from the frame. The positional values of the p_t image are expressed by the $b_t = [x^{(t)}, y^{(t)}, w^{(t)}, h^{(t)}]$ vector. In the b_t vector; t is the iteration, $[x\ y]$ the coordinates of the center of the selected view, $[w\ h]$ the width and height. $d_t \in R^{110}$ in the identification group represents a vector containing k previous movements [13].

State transition functions: a_t is defined as transition functions from S_t state to S_{t+1} state after the action decision is made. The transition function consists of two parts which are the transition function of the part image and the action dynamics function. The transition function of the part image is defined as $b_{t+1} = f_p(b_t, a_t)$. Here, the position of the part image is changed by a certain amount according to the action. The action dynamics function is defined by a function in the form of $b_{t+1} = f_p(d_t, a_t)$. When $a_t = 'dur'$ the action takes the last position in the frame. Finally, it gets the response according to whether it finds the object.

Reward ($r(s_T)$): The state transition function was assumed find to object in the frame after the stop motion. After that, it is calculated how accurate the object is by using the ground truth. To calculate the response value is used (1) function. If the precision value is greater than 0.7, it returns $r(s_T) = +1$ value otherwise $r(s_T) = -1$ [13]. Where b_t and G are the areas of the bounding box and the object, respectively.

$$r(s_T)\begin{cases} 1, & if\ IoU(b_t, G) > 0.7 \\ -1, & otherwise \end{cases}$$

$$IoU(b_t, G) = \frac{b_t \cap G}{b_t \cup G} \quad (1)$$

2) *ADNet Network Architecture and Training*

Table 1 shows the layers and functions of the ADNet network architecture. This architecture has 11 output units in the Fc6 layer. These outputs show the probabilities of choosing the actions that are probabilistically $(p\ (a \mid s_t; W))$ defined to pass from state s_t to s_{t+1} . The confidence layer (fc7) generates the probability that the calculated part belongs to the object $(p\ (\text{object} \mid s_t; W))$. The probability generated here is defined as the confidence score, which is used for online adaptation during the test. The ADnet method was trained in two stages: supervised learning (SL) and reinforcement learning (RL) using different datasets. In the SL phase, the network architecture was trained from scratch using samples from training videos. In the RL phase, the network architecture is strengthened by verifying over the trained network at SL stage [13].

Table 1: Layers of ADNet network architecture

	Name	**Explanation**
Input Layer	Conv1	Deep features are achieved.
	Conv2	
	Conv3	
Full Connected Layer	Fc4	ReLU and dropout functions are used. Motion estimation operations are performed here.
	Fc5	
Output Layers	Fc6	Determines the probability of selecting the motions in the d_t vector $(p\ (a \mid s_t; W))$.
	Fc7	The confidence layer yields the probability of correctness of the target and background class for the state s_t $(p\ (object \mid s_t; W))$.

3) Color Spaces

In order to provide data diversity, this study has analyzed YCbCr, HSV, inverted HSV (IHSV), CIE L*a*b*, CIE XYZ and NTSC color spaces. In addition to these color spaces, the IHSV color space is added because of the success achieved in the tests performed. The IHSV color space is transformed into the HSV color space by replacing the R and B color channel in the normal RGB color space (Fig. 3). Specific frames were taken from the various videos for analysis. These frames were then transformed into selected color spaces as shown in Fig. 2 and experiments were made following the object. As a result of the experiments, the following were obtained:

- HSV and IHSV color spaces have strong results for data diversity.
- RGB is more sensitive than HSV to light change
- Apart from the HSV color channel, the YCbCr, CIE XYZ and NTSC color spaces provided lower diversity across other color spaces.
- Using L*a*b* color space has reduced the success rate in object tracing

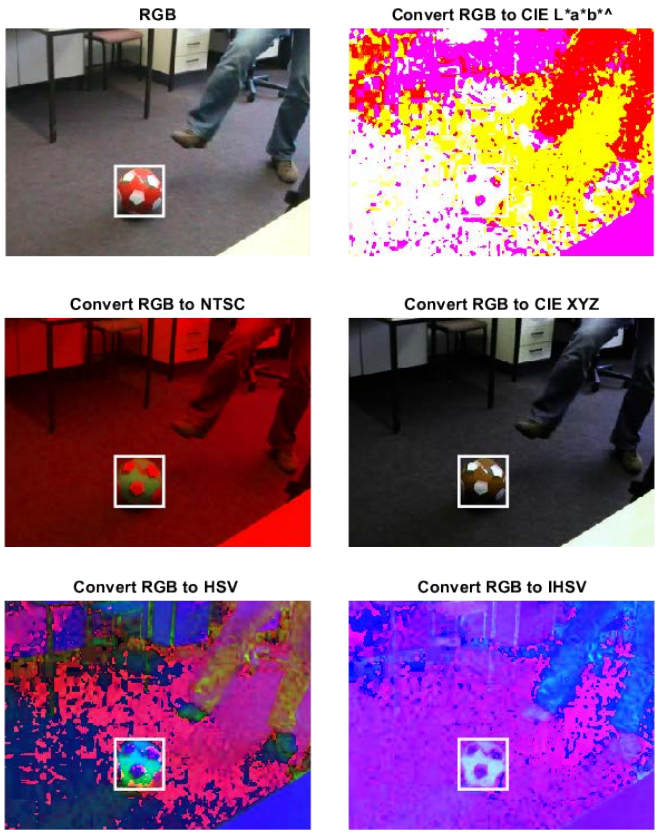


Figure 2. RGB color space to other color space

In the direction of the above results, the RGB, HSV and IHSV color spaces are applied separately to the two sets of data taken from the VOT2014 [15] and VOT2015 [14] datasets. The six datasets obtained are shown in Table 2. Then six different ADNet methods were trained using six different training sets. For Tests, the trained ADNet methods were conducted on VOT2013 [16], VOT2015 and VOT2016 which composed a total of 61 videos (including videos not used in training).

Table 2: Training sets for ADNet.

Datasets	Video count	Color Space	Reference
Training_Set_1	11	RGB	VOT2014[15]
Training_Set_2		HSV	
Training_Set_3		IHSV	
Training_Set_4	55	RGB	VOT2014 [15] VOT2015 [14]
Training_Set_5		HSV	
Training_Set_6		IHSV	
Test Set	61	Original Video	VOT2013 [16] VOT2014 [15] VOT2015 [14]

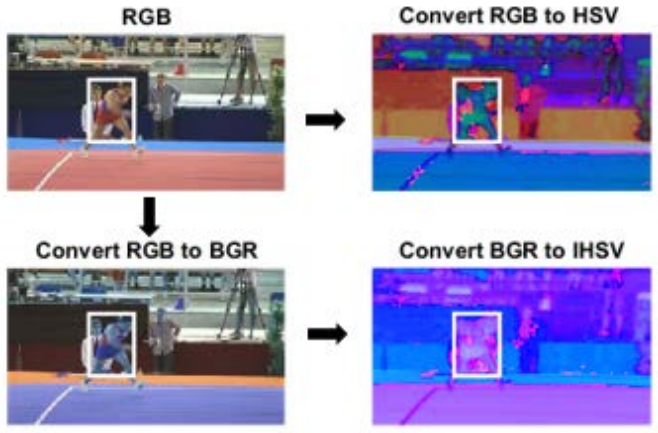


Figure 3. IHSV generate

III. EXPERIMENT

The tests were carried out with a test set containing 61 videos prepared by six ADNet methods. The tests were performed on an Intel i7 6700HQ processor, 16 GB RAM, and GeForce GTX 960M graphics card. The results of dinosaur, fish3, gymnastics4, graduate, handball1, handball2, singer2 and sphere in the dataset are very promising. The average results of these video are shown in Fig. 4. The result of the methods trained with HSV and IHSV color channel has got strong achievements. But the excessive increase in the number of videos on HSV and IHSV color channels has adversely affected success. The fact that the result dataset we extracted here is extremely difficult, has a negative effect on the tracer. Therefore, the limited use of dataset consisting of HSV and IHSV color blood provides a more accurate result.

The method of training the singer2 and handball1 videos in the test data with the RGB color channel has lost the object. But HSV and IHSV have achieved success in following the trained object with color channels. In Fig. 5, singer2 and handball1 precision results and the averages of all the results of the 61 videos are given. When we look at the general result in Fig. 4.(c), we have proven that a stronger dataset can be won by adding the IHSV and HSV color channels to the RGB color

channel. One of the important points here is that the IHSV and HSV color channel makes video data on the RGB channel difficult and adversely affects the outcome of overuse.

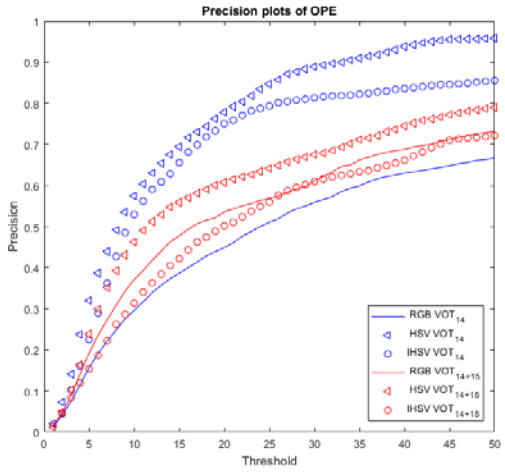
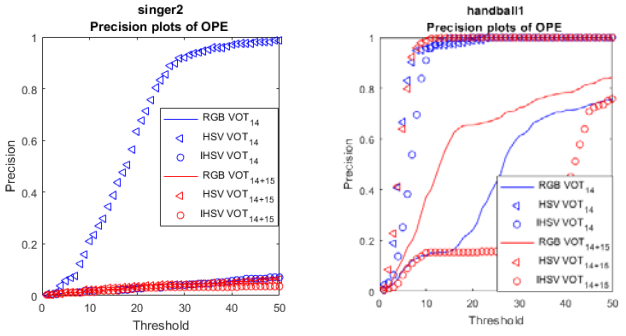
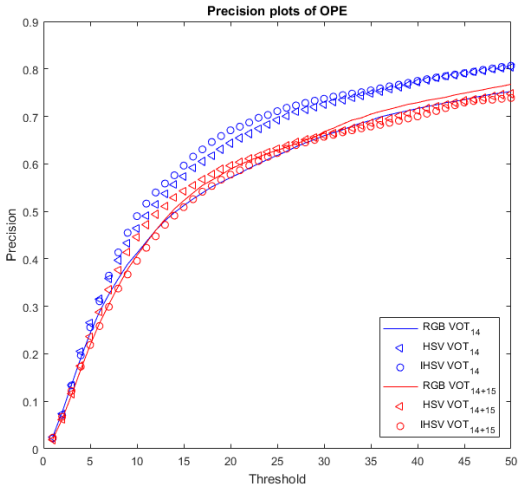


Figure 4. The average results of dinosaur, fish3, gymnastics4, graduate, handball1, handball2, singer2 and sphere



(a) *(b)*



(c)

Figure 5. (a) singer2 and (b) handball1 precision results and (c) the averages of all video results

IV. CONCLUSION

In this study, we acquired a different perspective for the training set used in deep learning network architectures. From this point of view, the ADNet method with its educational set consisting of RGB, HSV and IHSV color spaces has been trained. Then we performed extensive tests on these methods. As a result of the tests, we have proven that ADNet method is strengthened by using different color spaces. In the future work, we aim to combine these color channels in an adaptive way and to increase the success in this way.

V. REFERENCES

[1] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof, "On-line Random Forests," in *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, 2009, pp. 1393–1400.

[2] K. Zhang, L. Zhang, and M.-H. Yang, "Fast Compressive Tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2002–2015, Oct. 2014.

[3] N. Jiang, W. Liu, and Y. Wu, "Learning adaptive metric for robust visual tracking," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2288–2300, Aug. 2011.

[4] B. Baben, Ming-Hsuan Yang, and S. Belongie, "Visual tracking with online Multiple Instance Learning," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 983–990.

[5] Z. Chen, Z. Hong, and D. Tao, "An Experimental Survey on Correlation Filter-based Tracking," *arXiv Prepr. arXiv1509.05520*, Sep. 2015.

[6] S. Avidan, "Support vector tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 8, pp. 1064–1072, Aug. 2004.

[7] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised On-Line Boosting for Robust Tracking," Springer, Berlin, Heidelberg, 2008, pp. 234–247.

[8] B. Babenko, Ming-Hsuan Yang, and S. Belongie, "Robust Object Tracking with Online Multiple Instance Learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011.

[9] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-Convolutional Siamese Networks for Object Tracking," Jun. 2016.

[10] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," Sep. 2014.

[11] D. Held, S. Thrun, and S. Savarese, "Learning to Track at 100 FPS with Deep Regression Networks," Apr. 2016.

[12] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2014.

[13] S. Yun, J. Choi, Y. Yoo, K. Yun, and J. Y. Choi, "Action-Decision Networks for Visual Tracking with Deep Reinforcement Learning," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1349–1358.

[14] "The Visual Object Tracking VOT2015 Challenge Results," in *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2015, pp. 564–586.

[15] M. Kristan *et al.*, "The Visual Object Tracking VOT2014 Challenge Results," Springer, Cham, 2015, pp. 191–217.

[16] M. Kristan *et al.*, "The Visual Object Tracking VOT2013 Challenge Results," in *2013 IEEE International Conference on Computer Vision Workshops*, 2013, pp. 98–111.

[17] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual Tracking: An Experimental Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[18] G. E. Monahan, "State of the Art—A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms," *Manage. Sci.*, vol. 28, no. 1, pp. 1–16, Jan. 1982.

A Study on Prediction Success of Machine Learning Algorithms for Wart Treatment

K. AKYOL[1], A. KARACI[2], Y. GÜLTEPE[3]

[1] Kastamonu University, Kastamonu/Turkey, kakyol@kastamonu.edu.tr
[2] Kastamonu University, Kastamonu/Turkey, akaraci@kastamonu.edu.tr
[3] Kastamonu University, Kastamonu/Turkey, yasemingultepe@kastamonu.edu.tr

***Abstract* - Data mining and machine learning algorithms are utilized in order to discover meaningful information by thorough analysis of dataset. They are used in multi-disciplinary field. Wart is caused by the human papillomavirus. It inhibits body growth by activating ecdysone steroid production systematically. There are several treatment methods for this illness. These methods focused on offering a solution for people. In this framework, a study on the analysis of the best two wart treatment methods, Cryotherapy and Immunotherapy, is carried out. The first one of these datasets collected by applying the cryotherapy method consists of seven features. The second dataset collected by applying the immunotherapy method consists of eight features. Fuzzy Rule, Naive Bayes and Random Forest based models are designed in order to evaluate the effectiveness of these methods in wart treatment. The performances of these algorithms are judged within the frame of Accuracy and Sensitivity performance measures.**

***Keywords* - Wart treatment, Immunotherapy, Cryotherapy, Machine Learning, Naive Bayes, Fuzzy Rule, Random Forest.**

I. INTRODUCTION

The most common clinical sign of the human papilloma-virus infection in the skin and mucous membranes [1]. Warts restrict body growth by activating ecdysone steroid production systemically. There are several treatment methods for this illness. Immunotherapy is a therapy that uses certain parts of an individual's immune system to help fight various diseases like cancer [2-3]. Cryotherapy locally destroys lesional tissue and is used frequently for the treatment of Oral Leukoplakia and Oral Verrucous Hyperplasia lesions [4].

There are numerous investigations on the wart treatment in literature. Khozeimeh et al. proposed and presented a fuzzy logic rule-based system in order to predict the responses to the treatment method. According to their results, the accuracies of immunotherapy and cryotherapy methods were 83.33% and 80.7% respectively [5]. Redman et al. reviewed the latest developments in immunotherapy for prostate cancer treatment [6]. Donin et al. reviewed the biological mechanisms of action, clinical safety and efficacy of immunotherapies for urothelial carcinoma. They also described current research areas in immunotherapy and highlighted ongoing trials and promising and novel investigational agents [7]. Qian et al. examined the application of nanomaterials in cancer immunotherapy, including nanovaccines, tumor immune environment modulation, and other application [8]. Ye and Formenti discussed the mechanisms in which Radiation therapy can trigger an immune response for tumor rejection, and provide emerging preclinical and clinical data of combination immune radiotherapy, and its potency in breast cancer treatment [9]. Chen et al. examined the anti-tumor mechanisms of metronomic chemotherapy, which an antiangiogenic and multi-targeted therapy with immunological properties, and the preliminary research addressing the combination of immunotherapy and metronomic chemotherapy for cancer treatment in animal models and in the clinical setting [10]. In another study, Jung et al. tackled recent improvements in immunotherapy for the treatment of multiple myeloma [11].

II. MATERIALS AND METHOD

The best two wart treatment methods, Cryotherapy and Immunotherapy, are evaluated in this study. The datasets, which are taken from the University of California Irvine Machine Learning Repository, are collected from patients, with plantar and common warts, who had referred to the dermatology clinic of Ghaem Hospital in Mashhad, Iran from January 2013 to February 2015. One of these datasets collected by applying the cryotherapy method consists of seven features. The second dataset collected by applying the immunotherapy method consists of eight features. The *Result_of_Treatment* attribute in both datasets is the class attribute. These attributes are presented in Table 1.

Table 1: The information about the features in datasets.

Cryotherapy dataset		
Attribute no.	Attribute name	Data Type
1	Gender	Categorical
2	Age (year)	Continues
3	Time elapsed before treatment (month)	Numeric
4	The number of warts	Numeric
5	Types of ward (count)	Categorical
6	Surface area of the warts (mm^2)	Numeric
7	Class 0 and 1 (Response to treatment)	Yes / No

Immunotherapy dataset		
Attribute no.	Attribute name	Data Type
1	Gender	Categorical
2	Age (year)	Continues
3	Time elapsed before treatment (month)	Numeric
4	The number of warts	Numeric
5	Types of ward (count)	Categorical
6	Surface area of the warts (mm^2)	Numeric
7	Induration diameter of initial test (mm)	Numeric
8	Class 0 and 1 (Response to treatment)	Yes / No

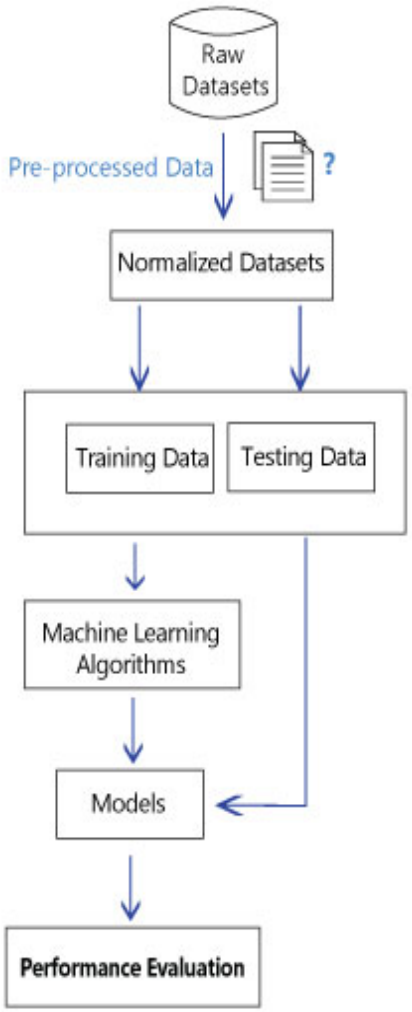


Figure 1: The block diagram of the study.

Firstly, the data are normalized via min-max normalization. The min-max normalization is a pre-process which is necessary for well-done machine learning. Original datum is converted into a datum between the minimum and maximum, i.e. zero and one [12]. The datasets normalized are shuffled and then divided into 70-30% training and testing data respectively. Hence, 62 train and 28 test instances are obtained for both Cryotherapy and Immunotherapy datasets. These datasets are sent as input data to the Naive Bayes, Fuzzy Rule and Random Forest algorithms which are introduced in [13-15] respectively in order to evaluate the effectiveness of these methods in wart treatment.

III. RESULTS

The experimental results and performance measures which are obtained on testing data are presented in Table 2.

Table 2. Experimental results.

Learning Algorithms	Cryotherapy dataset			Immunotherapy dataset		
		True	False		True	False
	Train	30	32	Train	48	14
	Test	18	10	Test	23	5
FR		No	Yes		No	Yes
	No	10	0	No	2	3
	Yes	6	12	Yes	2	21
	Acc: 78.57, Sen: 66.67			Acc: 82.14, Sen: 91.3		
NB		No	Yes		No	Yes
	No	10	0	No	1	4
	Yes	5	13	Yes	1	22
	Acc: 82.14, Sen: 72.22			Acc: 82.14, Sen: 95.65		
RF		No	Yes		No	Yes
	No	10	0	No	3	2
	Yes	1	17	Yes	1	22
	Acc: 96.43, Sen: 94.44			Acc: 89.29, Sen: 95.65		

According to these results;

a) The FR algorithm classified correctly 12 out of 18 instances which are considered as positive for the Cryotherapy dataset. Also, it classified correctly all of 10 instances which are considered as negative. The Acc and Sen measures of this model are 78.57% and 66.67% respectively for this dataset.

b) The NB algorithm classified correctly 13 out of 18 instances which are considered as positive for the Cryotherapy dataset. Also, it classified correctly all of 10 instances which are considered as negative. The Acc and Sen measures of this model are 82.14% and 72.22% respectively for this dataset.

c) The RF algorithm classified correctly 17 out of 18 for the Cryotherapy dataset. Also, it classified correctly all of 10 instances which are considered as negative. The Acc and Sen measures of this model are 96.43% and 94.44% respectively for this dataset.

d) The FR algorithm classified correctly 21 out of 23 instances which are considered as positive for the Immunotherapy dataset. Also, it classified correctly 2 out of 5 instances which are considered as negative. The Acc and Sen measures of this model are 82.14% and 91.3% respectively for this dataset.

e) The NB algorithm classified correctly 22 out of 23 instances which are considered as positive for the Immunotherapy dataset. Also, it classified correctly 1 out of

5 instances which are considered as negative. The Acc and Sen measures of this model are 82.14% and 95.65% respectively for this dataset.

f) The RF algorithm classified correctly 22 out of 23 instances which are considered as positive for the Immunotherapy dataset. Also, it classified correctly 3 out of 5 instances which are considered as negative. The Acc and Sen measures of this model are 89.29% and 95.65% respectively for this dataset.

IV. CONCLUSION

Wart is caused by the human papillomavirus. It inhibits body growth by activating ecdysone steroid production systemically. There are several treatment methods for this illness. In this framework, a study on analysis of the best two wart treatment methods, Cryotherapy and Immunotherapy, is carried out. Fuzzy Rule, Naive Bayes, and Random Forest based models are designed to evaluate the effectiveness of these methods in wart treatment. The experimental results clearly show that the Random Forest classifier outperforms other algorithms for both the Cryotherapy and Immunotherapy datasets. Acc and Sen measures are 96.43% and 94.44% respectively for Cryotherapy dataset, and 89.29% and 95.65% respectively for Immunotherapy dataset. This algorithm has the best performance in experimental studies.

ACKNOWLEDGMENT

The authors would like to thank the UCI Machine Learning Repository for providing the Cryotherapy and Immunotherapy datasets.

REFERENCES

[1] F. Khozeimeh, F.J. Azad, Y.M. Oskouei, M. Jafari, S. Tehranian, R. Alizadehsani, P. Layegh, "Intralesional immunotherapy compared to cryotherapy in the treatment of warts," *International Journal of Dermatology*, vol. 56, pp. 474-478, 2017.

[2] American Cancer Society, Cancer Immunotherapy, Available: https://www.cancer.org/treatment/treatments-and-side-effects/treatment-types/immunotherapy/what-is-immunotherapy.html

[3] Y. Shih, L. Elting, A. Pavluck, A. Stewart, and M. Halpern, "Immunotherapy in the Initial Treatment of Newly Diagnosed Cancer Patients: Utilization Trend and Cost Projections for Non-Hodgkin's Lymphoma, Metastatic Breast Cancer, and Metastatic Colorectal Cancer," *Cancer Investigation*, vol. 28, no. 1, pp.46-53, 2009.

[4] C.H. Yu, H.P. Lin, S.J. Cheng, A. Sun, H.M. Chen, "Cryotherapy for oral precancers and cancers," *Journal of the Formosan Medical Association*, vol. 113. no. 5, pp. 272-277, 2014.

[5] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layeghe, S. Nahavandi, "An expert system for selecting wart treatment method," *Computers in Biology and Medicine*, vol. 81, pp. 167-175, 2017.

[6] J.M. Redman, J.L. Gulley, R.A. Madan, "Combining immunotherapies for the treatment of prostate cancer," *Urol Oncol*, vol. 35, no. 12, pp. 694-700, 2017.

[7] N.M. Donin, A.T. Lenis, S. Holden, A. Drakaki, A. Pantuck, A. Belldegrun, K. Chamie, "Immunotherapy for the Treatment of Urothelial Carcinoma," *J Urol* vol. 197, no. 1, pp. 14-22, 2017.

[8] H. Qian, B. Liu, X. Jiang, "Application of nanomaterials in cancer immunotherapy," *Materials today Chemistry*, vol. 7, pp. 53-64, 2018.

[9] J.C. Ye and S.C. "Formenti, Integration of radiation and immunotherapy in breast cancer - Treatment implications," *Breast*, vol. 38, pp. 66-74, 2018.

[10] L. Chen, M.C. Chang, W.F. Cheng, "Metronomic chemotherapy and immunotherapy in cancer treatment," *Cancer Letters*, vol. 400, pp. 282-292, 2017.

[11] S.H. Jung, H.J. Lee, M.C. Vo, H.J. Kim, J.J. Lee, "Immunotherapy for the treatment of multiple myeloma," *Crit Rev Oncol Hematol*, vol. 111, pp. 87-93, 2017.

[12] Y.K. Jain, S.K. Bhandare, "Min Max Normalization Based Data Perturbation Method for Privacy Protection," *International Journal of Computer & communication Technology*, vol. 2, no.8, pp. 45-50, 2011.

[13] J.A. Roubos, M. Setnes, and J. Abonyi, "Learning fuzzy classification rules from labeled data," Information Sciences, vol. 150, no. 1-2, pp. 77-93, 2003.

[14] N. Friedman, "Bayesian Network Classifiers," Mach. Learn., vol. 163, no. 29, pp. 131–163, 1997.

[15] L. Breiman, "Random forests," *Mach Learn*, vol. 45, pp. 5-32, 2001.

Turkish Sign Language Alphabet Recognition with Leap Motion

A.KARACI[1], K. AKYOL[2] and Y. GÜLTEPE[3]

[1] Kastamonu University, Kastamonu/Turkey, akaraci@kastamonu.edu.tr
[2]Kastamonu University, Kastamonu/Turkey, kakyol@kastamonu.edu.tr
[3]Kastamonu University, Kastamonu/Turkey, yasemingultepe@kastamonu.edu.tr

Abstract **- Sign language recognition is used to help communicate effectively between normal hearing peoples and hearing-impaired. According to literature review, Turkish sign language recognition studies are very few. For this reason, this study has been performed on Turkish sign language recognition. Depth cameras, such as the Leap Motion controller, allows the researchers to exploit depth knowledge to better understand hand movements. In this study, data of 10 letters in Turkish sign language was taken from Leap Motion. Five of these data are composed of letters (I, C, L, V, O) that It can be expressed with one hand, while the other five are composed of letters (B, D, M, N, K) that It can be expressed with two hands. The dataset was taken by two different people. Each person made five trials for each letter. Ten samples were taken at each trial. In this study, Artificial Neural Network, Deep Learning and Decision Tree based models were designed and the effectiveness of these models in recognizing the Turkish sign language is evaluated. Regression (R), Mean Square Error (MSE) and Estimation Accuracy performance metrics are used to evaluate models' performance. The data set was randomly divided into 30% for training and 70% for testing. According to the experimental results, the most successful models for the data set with 120 features are decision tree and DNN models. For the data set with 390 features, DNN is the most successful model.**

Keywords **- Leap Motion, Machine Learning, Deep Learning, Artificial Neural Network, Decision Tree, Deep Neural Networks.**

I. INTRODUCTION

SIGN language is the most common and important way for deaf and hearing impaired in order to communicate and integrate with their society. It is a kind of visual language that consists of a sequence of grammatically structured human gestures. According to the most recent statistics of World Health Organization, there are 360 million persons in the world with disabling hearing loss i.e. (5.3% of the world's population). For this reason, the development of automatic systems that can translate sign languages into words and sentences becomes a necessity [1]. Hand gestures use the palms, finger positions, and shapes to create forms referring to different letters and phrases. Microsoft Kinect and Leap Motion Controller (LMC), for example, are two of the latest devices used to create natural user interfaces (NUIs), which enhance communication between the user and the devices, relying on hand and body gestures and movements. In 2013, LMC was released a device that a user can use to interact with objects on screens through hand gestures. This device is a combination of hardware and software that tracks the movement of hands and fingers and then converts them into a 3D input [2]. LMC was developed by an American company called leap motion company. It detects and tracks position and motion for hands, fingers and their joints with a rate of 200 frames per second approximately. The captured frames contain information about how many hands are being detected and also vectors that contain information about the position and rotation of hands and fingers based a skeletal model of the hand [1].

A standard neural network (NN) consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons [3]. Deep learning using neural networks have claimed state-of-the-art performances in a wide range of tasks [4]. Since 2006, deep structured learning, or more commonly called deep learning or hierarchical learning, has emerged as a new area of machine learning research. During the past several years, the techniques developed from deep learning research have already been impacting a wide range of signal and information processing work within the traditional and the new, widened scopes including key aspects of machine learning and artificial intelligence [5].

According to literature review, Turkish sign language recognition studies [6-9] are very few. Among these studies, there is only one study [6] directly related to our study. For this reason, this study has been performed on Turkish sign language recognition. There are several studies in the literature about recognizing finger and hand movements in the marking language with Leap Motion. Chuan et al (2014), applied k-nearest neighbor and SVM to classify the 26 letters of the English alphabet in American Sign Language using the derived features from the sensory data. According to the study results, classification rate of 72.78% and 79.83% was achieved by k-nearest neighbor and support vector machine respectively [10]. Mohandes et al (2014) applied Naive Bayes and Multilayer Perceptron (MLP) neural networks to classify the 28 letters of the Arab alphabet in Arabic Sign Language. According to the study results, classification rate of 98.3% and 99.1% was achieved by Naive Bayes and MLP neural networks respectively [11]. Che et al (2015) proposed to a rapid recognition for dynamic hand gestures using leap motion [12]. Elons et al (2014) proposed recognition system for Arabic sign language recognition based on ANN and leap motion sensor [13]. Tong et al (2016) proposed a novel feature vector which

is suitable for representing dynamic hand gestures, and presented a satisfactory solution to recognizing dynamic hand gestures with a LMC only [14]. Kumar et al (2017) presented a framework to recognize manual signs and finger spellings using Leap motion sensor [15]. Wang et al (2017) proposed a method for recognition of hand gestures representing Arabic numbers (0-9) using the information derived from Leap Motion [16].

Depth cameras, such as the Leap Motion controller, allows the researchers to exploit depth knowledge to better understand hand movements. In this study, data of 10 letters in Turkish sign language was taken from Leap Motion. Five of these data are composed of letters (I, C, L, V, O) expressed with one hand, while the other five are composed of letters (B, D, M, N, K) expressed with two hands. In this study, Artificial Neural Network, Deep Learning and Decision Tree based models were designed and the effectiveness of these models in recognizing the Turkish sign language is evaluated.

II. METHOD

In this study, data of 10 letters in Turkish sign language was taken from Leap Motion. Five of these data are composed of letters (I, C, L, V, O) that It can be expressed with one hand, while the other five are composed of letters (B, D, M, N, K) that It can be expressed with two hands. The dataset was taken by two different people. Each person made five trials for each letter. Ten samples were taken at each trial. Thus, a total of a thousand data was obtained. The data set has 195 features for one hand. This number is 390 for two hands.

Leap motion gives information about 3 axes (x, y, z). These axes are shown in Fig 1.

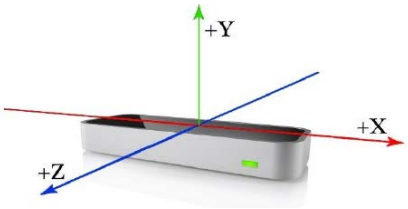


Figure 1: Leap motion coordinate axes.

Moreover, the information of these three axes (x, y, z) is produced for all the bones and positions shown in Fig 2. Information such as start and end positions of distal, intermediate, proximal and metacarpal bones, bone directions, palm position, arm direction, wrist position, elbow position are measured in 3 axes. As a result of this measurement, 195 features are obtained for one hand.

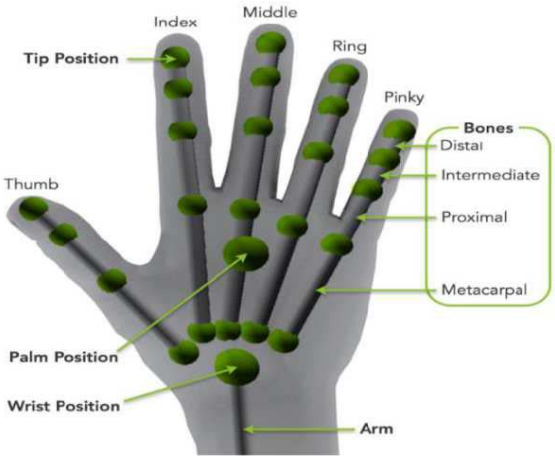


Figure 2: Bones and positions measured by Leap motion [17]

In order to reduce the number of features in this data obtained from this study, the end position of each bone was subtracted from the palm position (equation 1). This operation is repeated for 3 axes. With this method, the number of features are reduced to 120 for a double hand. In the study, artificial neural network, deep neural network and decision tree machine learning methods were used. These machine learning methods are trained for both the 390 and 120 feature data sets and their performance is tested. The data set is randomly divided into 70% and 30%. 30% of the data (300 data) were used for training and 70% (700 data) were used for the test.

$$\text{New Data}_{x,y,z} = \text{Palm\_Position}_{x,y,z} - \text{Bones\_End}_{x,y,z} \quad (1)$$

The model of deep neural network (DNN) modeled in the study consists of 3 hidden layers (50 neurons each), an input layer (120 or 390 neurons) and an output layer (1 neuron). The 3 hidden layers in the deep learning model by analyzing features of the input layer determined the most important twenty features and used them in the training phase. Deep learning is an approach that aims to find the ideal representation of a machine with non-linear methods using machine learning techniques and architectures [18]. Deep Learning is based on deep architecture. The number of hidden layers has been increased. Each layer is a set of artificial neural networks, in which a feature related to the problem is learned. In this architecture, a feature of problem is learned on each layer, and this learned feature creates an input to a top layer. Thus, from the bottom layer to the top layer, a structure becomes established where the simplest and most complex attribution is learned [19]. In deep neural networks, data is first processed in auto encoder layers which having close to one another of inputs and outputs, and preliminary processing and size reduction steps are performed. It has the ability to extract new features from raw data and reduce the size of the data set [20, 21]. A DNN consists of an auto encoder and a softmax classifier. The ability of the auto encoder (OE) to derive new features from raw data and effectively reduce the size of the data makes it the fundamental building block of the deep neural network. The Softmax layer enhances classification performance by using features from the OE. [21]. Today, deep learning is widely used

in areas such as image classification, video analysis, speech recognition, object identification and natural language learning [18, 22]. The great advantage of deep learning is that it does not require any pre-built mathematical model-based feature extractor. Thus, instead of deciding whether the feature type is appropriate, the network to be used in deep learning and the operations to be applied in each layer are gaining importance. The most commonly used deep learning method is the Convolutional Neural Network (CNN) [22].

The artificial neural network model has an input layer (120 or 390 neurons), 1 hidden layer (5 neurons) and an output layer (1 neuron). The training process used 1000 epochs. The performance of deep neural network, artificial neural network and decision tree are measured by MSE, R and accuracy parameters. Different hidden layers and neuron numbers were determined for deep neural network and artificial neural network, and experiments were carried out. In this study, the best results were presented. Some of the 10 Turkish sign language letters used in the study are shown in Fig. 3. The pictures in Fig. 3 are taken from tdk.gov.tr which is the official website of Turkish Language Association.

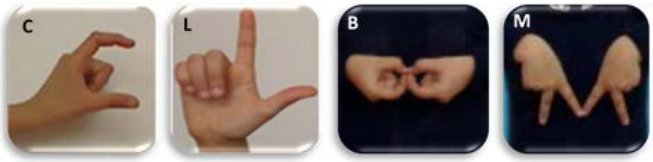


Figure 3: Examples of Turkish sign language

III. RESULTS AND DISCUSSION

DNN, ANN and decision tree models generated in the study were tested with test data (70%). The test operation was performed separately in the models trained with 120 and 390 features data sets. Test results were evaluated by means of R and MSE parameters for DNN, ANN and Decision Tree models. Test results are shown in table 1.

Table 1: DNN, ANN and Decision Tree models test results

MODEL	120 Features		390 Features	
	R	MSE	R	MSE
DNN	0.99	0.003	0.99	0.002
ANN	0.99	0.07	0.96	0.07
Decision Tree	1	0	1	0.011

if the R value is close to 1 and the MSE parameter is close to 0, the model performance is high. The most successful models for the data set with 120 features according to Table 1 are decision tree and DNN models. For the data set with 390 features, DNN is the most successful model. In addition, to determine the success rates, the round function is applied to the test outputs obtained from the models and rounded to the nearest integer. Estimation accuracy is calculated according to the outputs obtained from this. These rates are shown in table 2.

Table 2: Estimation accuracy of the models

MODEL	Accuracy rate	
	120 Features	390 Features
DNN	%100	%100
ANN	% 99	%93
Decision Tree	%100	%98,85

When the results obtained are evaluated in general, it is observed that the DNN model gives better results as the number of features increases. This is an expected situation. DNN's give better results on big data than other models. In other words, deep learning method is the most ideal method if the modeling process is performed without any process on the raw data. Because the most important features in the dataset are used by DNN in training.

REFERENCES

[1] B. Hisham and A. Hamouda, Arabic Static and Dynamic Gestures Recognition Using Leap Motion, *Journal of Computer Science*, 13 (8): 337-354, 2017.

[2] M. A. Almasre, H. Al-Nuaim, A Real-Time Letter Recognition Model for Arabic Sign Language Using Kinect and Leap Motion Controller v2, *International Journal of Advanced Engineering, Management and Science (IJAEMS)*, 2(5), 514-523, 2016.

[3] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61 pp. 85–117, 2015.

[4] Y. Tang, Deep learning using support vector machines, *in: ICML Workshop on Challenges in Representation Learning*, 2013.

[5] Deng, L. and Yu, D., "Deep Learning: Methods and Applications", *Foundation and Trends in Signal Processing* Vol. 7, pages 197-387, 2014

[6] B. Demircioğlu, G. Bülbül and H. Köse, "Turkish Sign Language recognition with Leap Motion," *2016 24th Signal Processing and Communication Application Conference (SIU)*, Zonguldak, 2016, pp. 589-592.
doi: 10.1109/SIU.2016.7495809

[7] A. Memiş& S. Albayrak. (2013). A Kinect Based Sign Language Recognition System Using Spatio-temporal Features. *Proceedings of SPIE - The International Society for Optical Engineering*. 9067. 10.1117/12.2051018.

[8] H. Haberdar, Saklı markov model kullanılarak görüntüden gerçek zamanlı Türk işaret dili tanıma sistemi, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 2005.

[9] B.S. Ertugrul, C. Gurpinar, H. Kivrak, H. KOSE, "Gesture Recognition for Humanoid Assisted Interactive Sign Language Tutoring", *The Sixth International Conference on Advances in Computer-Human Interactions (ACHI 2013)*, February 24 - March 1, 2013 - Nice, France.

[10] Ching-Hua Chuan, E. Regina, and C. Guardino. "American Sign Language Recognition Using Leap Motion Sensor". *In: Machine Learning and Applications (ICMLA)*, 2014 13th International Conference on. Dec. 2014, pp. 541–544. DOI: 10.1109/ICMLA.2014.110

[11] M. Mohandes, S. Aliyu, and M. Deriche. "Arabic sign language recognition using the leap motion controller". I*n: Industrial Electronics (ISIE), 2014 IEEE 23rd International Symposium on*. June 2014, pp. 960–965. DOI: 10.1109/ISIE.2014.6864742.

[12] Y. Chen, Z. Ding, Y-L. Chen, X. Wu, Rapid Recognition of Dynamic Hand Gestures using Leap Motion, *Proceeding of the 2015 IEEE International Conference on Information and Automation Lijiang*, China, August 2015.

[13] A. S. Elons, M. Ahmed, H. Shedid and M. F. Tolba, "Arabic sign language recognition using leap motion sensor," *2014 9th International Conference on Computer Engineering & Systems (ICCES)*, Cairo, 2014, pp. 368-373. doi: 10.1109/ICCES.2014.7030987

[14] Z. Tong, J. Chu, Dynamic Hand Gesture Recognition With Leap Motion Controller, *IEEE SIGNAL PROCESSING LETTERS*, VOL. 23, NO. 9, 1188-1192, 2016.

[15] P. Kumar, R. Saini, S. K. Behera and D. P. Dogra, Real-Time Recognition of Sign Language Gestures and Air-Writing using Leap Motion, *2017 Fifteenth*

IAPR International Conference on Machine Vision Applications (MVA) Nagoya University, Nagoya, Japan, May 8-12, 2017.

[16] Qinghui Wang, Ying Wang, Fenglin Liu, Wei Zeng, Hand Gesture Recognition of Arabic Numbers Using Leap Motion via Deterministic Learning, *Proceedings of the 36th Chinese Control Conference July 26-28, 2017*, Dalian, China.

[17] Y. Ji, C. Liu, S. Gong and W. Cheng, "3D Hand Gesture Coding for Sign Language Learning," *2016 International Conference on Virtual Reality and Visualization (ICVRV)*, Hangzhou, 2016, pp. 407-410. doi: 10.1109/ICVRV.2016.74

[18] Hatipoglu N., Bilgin G., "Histopatolojik Görüntülerde Fourier Özellikleri Kullanılarak Evrişim Yapay Sinir Ağı ile Bölütleme - Segmentation of Histopathological Images with Convolutional Neural Networks using Fourier Features", *IEEE 23rd Conf. on Signal Processing and Communications Applications*, MALATYA, TÜRKIYE, 16-19 Mayıs 2015, pp.455-458

[19] G. Işık and H. Artuner, "Recognition of radio signals with deep learning Neural Networks," *2016 24th Signal Processing and Communication Application Conference (SIU)*, Zonguldak, Turkey, 2016, pp. 837-840.

[20] M. H. Cılasun and H. Yalçın, "A deep learning approach to EEG based epilepsy seizure determination," *2016 24th Signal Processing and Communication Application Conference (SIU)*, Zonguldak, 2016, pp. 1573-1576.

[21] A. Basturk, M.E. Yuksel, A. Caliskan, H. Badem Deep Neural Network Classifier for Hand Movement Prediction, *Proceedings of the IEEE 25th Signal Processing and Communications Applications Conference SIU (2017)*

[22] Ş. Karahan and Y. S. Akgül, "Eye detection by using deep learning," *2016 24th Signal Processing and Communication Application Conference (SIU)*, Zonguldak, 2016, pp. 2145-2148. doi: 10.1109/SIU.2016.7496197

A Review on Web Crawlers and Ontology-Based Crawlers

Y. GÜLTEPE[1], A.B. ÖNCÜL[2], E. ALTINTAŞ[3] and F. UĞUR[4]

[1]Kastamonu University, Kastamonu/Turkey, yasemingultepe@kastamonu.edu.tr
[2]Kastamonu University, Kastamonu/Turkey, boncul@kastamonu.edu.tr
[3]Kastamonu University, Kastamonu/Turkey, emelaltintas@kastamonu.edu.tr
[4]Kastamonu University, Kastamonu/Turkey, fadikugur96@gmail.com

Abstract - **As known, web crawlers are programs that automatically browse on the web. Their purpose is to automatically navigate pages, saving source links that have target links, marking pages according to the words in those links, saving, indexing, collecting data to bring personalized ads, etc. Although the web crawlering algorithm is simple, it has various difficulties with respect to the existing pages on the web and the resulting amount of data. The semantic web works on generating computer readable data and is intended to overcome the quantity of data generated. Ontologies represent a pivoting source for semantic web applications. Ontology based crawlers scan the web by focusing on related web pages along with a specific ontology based on area ontology. The main advantage of the ontology based web crawlers over other crawlers is that no Conformance Feedback or Training Procedure is required to move wisely. In addition, both the number of documents and the more effective and efficient results will be obtained during the scanning process. As a result; The main advantage of an ontology based web crawler over other web crawlers is that it does not require intelligent, efficient operation and relevant feedback. In this study, traditional and ontology based web crawlers approaches and its infrastructure are examined. In addition, differences between ontology based web crawlers and traditional web crawlers have been investigated. A brief of literature summary on the subject has been included.**

Keywords **- Web crawlers, Challenges, Semantic, Ontology**

I. INTRODUCTION

THE World Wide Web (WWW) is an information system of hypertext documents linked to each other on the internet [1]. Web Crawler is a program that continuously scans the WWW to retrieve web pages, and finds the web pages needed for any topic. The most common approach used when developing Web Crawler is to go to the links in the pages and save these pages, then repeat the same process to find the relevant pages. Web Crawler's works are listed as follows: Automatically navigating pages, saving the links with links, marking pages according to the words in these links, saving, indexing, collecting data to bringing in personalized ads, and so on. [2]. Figure 1 shows the overall view of the web crawler which is the result of the works listed here.

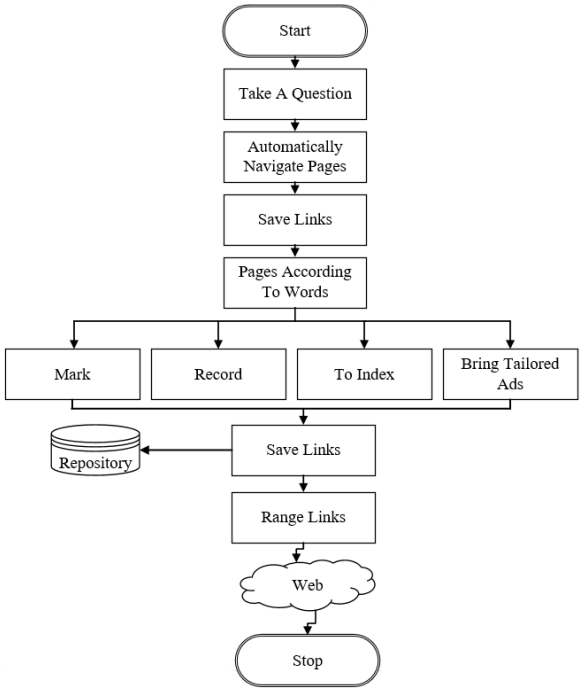


Figure 1: General of web crawler.

Search engines also find web pages they show in search results with the help of specially developed web crawlers [3]. In addition, crawlers can be adapted to work on deep web and secret records of web [4].

Web content mining defines techniques for automatic information search from web sources. Web content mining focuses on the content of your web sites. Web content is scanned with web crawler software. Web crawler, except for listing sites only from search engine results, images, videos, articles, files, music etc. It can scan information in many different formats and structures. This software scans the pages displayed by the links according to the given depth value starting from the main page and writes the page number and link name to the desired file [5].

In the digital world where computers, people and institutions are constantly communicating with each other, a safe, reliable, fast, compatible, collaborative and flexible data sharing system is needed to efficiently use information. Studies suggesting solutions to these problems have caused the beginning of semantic web research area.

In semantic web, web sites are related to each other by content, context, meaning [6]. Today, web crawlers are making keyword-based searches. However, in the semantic web, web sites are related to the subject matter and it will be much easier to find the searcher.

In this study, ontology-based web crawlers used for semantic web crawlers to make web content understandable, interpretable and usable by software have been examined in general. In this context, the development of semantic web, ontology and ontology based web crawlers are included in the topics. In the second part, information on the types of traditional web crawlers, strategies and constraints, semantic web and ontologies will be presented. In the third chapter, ontology based web crawlers are examined. Finally, the results are presented.

II. WEB CRAWLERS

In this section, different types of web crawlers and web crawling strategies are examined.

A. Types of Web Crawlers

It appears that various studies describing types of web crawlers have appeared. Research by Xiang, Yin and Han (20015) [7] web crawler types are divided into three types. These are respectively batch crawler, crawlers and focused crawlers. Similarly Udapuer et al. (2014) explained [8] it can be divided into four types: focused web crawlers are divided into incremental crawlers, distributed crawlers and parallel crawlers. These are summarized in the following articles, respectively.

Batch Crawler: Periodically scan a group of web pages. For example, a crawler is configured to scan 200 web pages every month. At the beginning of each month, crawler scans the pages that are set to update the web pages collected in the repository. After fulfilling his task, he will move on to the next task.

Incremental Crawler: A type of browser whose browsing never be stopped. It is thought that the web is constantly updated and will end at the end of fecthed pages. In addition to scanning a new web page, the incremental crawler has a built-in scheduler to schedule repeat visits of previously browsed websites to fetch web site content for any update. With this feature it is separated from other batch scanners.

Focused Browser or Topic Browser: Through Web page analysis, this type of browser is a method of selective identification with user defined topic pages. The focus of the research is the web page and topic relevance judge, the web analytics algorithm, and the ranking algorithm to access the URL co-ordination. To adapt to a complex web environment, the Web Analysis and the Sorting algorithms are used for the browser framework; enhances the scanner's ability to focus on itself. An example is the CINDI Robot [9]. CINDI Robot is used to find academic documents in computer science and software engineering.

Distributed Crawler: The basic research content of distributed browsers is to perform the scheduling process of the scanner to improve the scheduling strategy, load balancing and error recovery mechanism, scalability and fault tolerance of the scanner nodes.

Parallel Crawler: More than one browser is usually run in parallel, and these are called parallel browsers. A parallel scanner consists of multiple processes called C-procs that can work in the workstation network. Parallel browsers depend on page refresh and page selection. A parallel scanner can be in the local network and geographically dispersed locations. The parallelization of the scanning process is important in terms of downloading the documents within a reasonable period of time.

B. Limitations of Traditional Web Crawlers

Today, it is difficult to understand the information for the machines if web are not semantically structured. Therefore, web crawlers should be able to get the information that is made in key text. The semantic web is used to remove the following constraints of traditional web crawlers [10, 11, 12]:

- Synthesizer: Allows the identification and matching of synonymous tokens within the targeted data.
- Narrator: Source data creates new forms as a result of transformations using hidden semantics.
- Spiders/Crawler: Crawler allows the page to generate meaningful information when a page is first visited, assuming the starter cluster is empty. Then go to the next web page and apply these operations over and over again. The web search engine component crawler creates a local copy of the web pages. It will update this copy in a period of time. Variables that depend on the decision to update a web page vary with resource usage and refreshing of the page's content.
- Document Parser: Original documents automatically detected by Crawler; the document is subject to parsing by the parser. The main task of this phase is to prepare for indexing. This process is to filter the file information and make it ready to be indexed.
- Analyzer: The document divides document resources that are parsed by the parser, ie are ready to be indexed. By applying certain rules, the text is broken up into the smallest units that can index the information. Due to the repetition of indexing and searching, the same analysts are required.
- Indexer: The use of the document preprocessing method is to create a data structure for easy searching between documents. When the search engine needs to handle a large number of documents, the speed of the indexing process is to significantly increase the information search speed.
- Index Seeker: The index seeker retrieves the search requests of the user from the interface that the user has used. It sends its indexing strategy, which is determined by itself, from the index library to the sorting and filtering system, which is the result of searching and searching.
- Sequencer and Filterer: The query results of the document must be determined. It is a fact that these interrogations are located in the first order in the order of precedence. The closer the results are to the query, the upper the order

of precedence relative to the less relevant documents. First, a document contains multiple terms. Secondly, the importance of these terms is mutually variable. Determining the priority and importance of the terms that the current document contains while defining the stage of the association with the other documents. The weight of the term has an important role in the calculation of the relevance of documents to the query.

- User Interface: At this stage, it is the tool of human interaction with computer. It is the responsibility of the user to search for queries, to submit the query to the index seeker, to perform the search, to filter, sort, and to display the result list to the user.

C. Semantic Web

Nowadays, huge amounts of information, given the need for speeds and the production of information knowledge, the information means understanding what is important, and to be made by people that it is getting harder. Thus, the data itself will acquire within the meaning of the meaning by making sense of the relationship between the data to be hosted and it is believed that the problems can be solved.

Semantic web [6, 10], provides a common model for the reuse and sharing of information. Thanks to this model, it is easy to access the defined information again when requested and also to be used by other systems.

Semantic Web technology in the extraction of information through web sites, personalized and more effective content that delivers results-based search engines and through the websites of individual differences or features of the development of the portal is based on the information that is collected and evaluated.

D. Ontologies

Ontology plays a central role in the semantic web environment. Ontology is a formal and explicit presentation of shared conceptualization [13]. Conceptualization is the abstract modeling of ideas on concepts restricted to a specific domain. One of the other purposes of ontology is to create a reusable and shareable model in different applications.

III. ONTOLOGY BASED WEB CRAWLERS

Semantic focused crawler [14] is a software agent in the web environment that can download web information about specific topics using semantic technologies. The purpose of semantic focused crawler get information in a precise and efficient way and automatically collect and store relevant web information by automatically understanding the meaning behind predefined topics and web information.

The use of ontologies on the basis the semantic web the sharing of information is subject to. In the context of semantic focused approach; ontology makes explanations of managed information and can also perform queries on these explanations [15]. Ontology is significantly influencing the ontology based semantic crawlers' scanning performance. Because ontologies represent specific domain knowledge and updated created by domain experts and the ontology of the area of information, differences may occur between the perceptions of experts area. Another point is that information is constant and renewable. These contradictory situations; considering the differentiation of the issues of ontology and dynamism, has sometimes led to the idea that real-world knowledge can not be fully represented. The use of ontologies for semantic focused web crawlers as a result in [16] study has been found to negatively affect the performance evaluation.

General purpose crawlers are not enough to overcome the problem. They scan the internet blindly and in detail. Because of this, it is very difficult to find the information that is required when the current size of the web is taken into consideration, since the general purpose crawlers find very specific data in the web environment. For this reason, it is generally recommended to focus on specific areas of screening process [16, 17]. Such ontology-based crawlers are called preferential or heuristic-based crawlers. Heuristic approaches to ontology matching have been used. The semantic content is intuitively managed. The purpose of this research is to search for information on the content that has been separated according to meaning, using metadata to [15] semantically annotation the content, and search for information linked to URLs based on concepts and relationships on the ontology. Ontology based web crawler that was used to develop the algorithm, scanning the pages before solves the problem of finding the suitability of the most appropriate level. An ontology based crawler algorithm is embedded where the ontology is embedded to evaluate the level of interest of the related pages.

Managed by field ontology, Crawler performs browsing by focusing on pages related to a particular area ontology. The scan result list will be composed of a set of candidate matches with the ontologies found during the scan. Crawler, on the basis of domain ontologies for web pages more relevant crawled gives priority to the queue URL. Also, the ability to use the semantic structure of the URL obtained from the domain ontology to filter the URL queue is searched. The main advantage of the ontology based web crawler over the other crawler is that no conformance feedback or training procedure is required for wisely act. Also, both the number of documents will be reduced and more efficient and efficient results will be obtained in the scanning process. In the ontology based web crawler, the validity of the web pages is checked first. If the web page is valid, the web page is parsed and the parsed page content is matched with the ontology. If the web page is relevant, the word and web page are passed to the indexer.

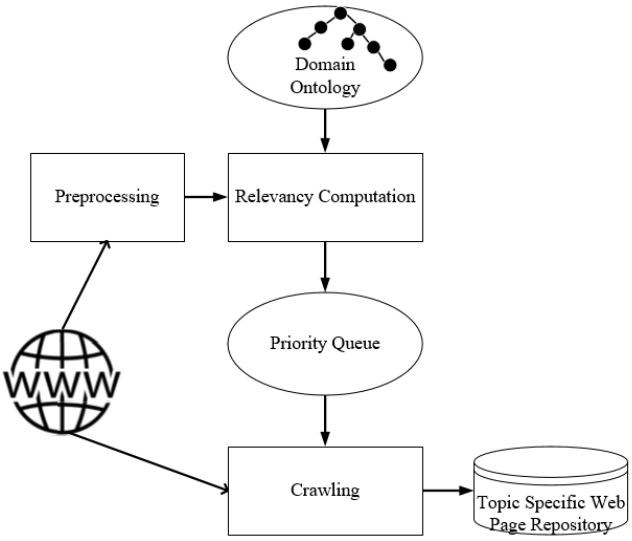


Figure 2: Ontology based web crawling framework.

The overall steps of the ontology based web crawler's general operations are shown in Figure 2. First is checked for the validity of the web pages. If so, it is parsed and the parsed content is matched to the ontology. If the web page is related, indexed otherwise is ignored. An algorithm for ontology based web crawler is proposed in [16] study. Basic steps of the algorithm are as follows:

- The seed get the URL.
- If the web page is the current type, add a queue.
- Parse the contents of the Web.
- Get answers from the server, then read the ontology file. Match the web page contents with the read ontology file.
- The web relevance point on account.
- Add the web page to index.
- Caches file to a folder.

This algorithm manually adjusts the weight of the ontology term in the interest level calculation algorithm. Actually, after you read the ontology and after visiting specific websites, the weight of the ontology can be designed in such a way that it can automatically provide. Thus, web crawler process time can be improved. The ontology developed in [16] study remains static; Ontology is created dynamically by visiting web pages and adding new concepts and relationships.

IV. CONCLUSION

The ontologies have many interesting features and a lot of usage area. Nowadays, the problem of increasing the amount of information garbage and its dispersion but will be able to recover with the use of such structures. The most important advantage of an ontology-based web crawler is that it does not need any relevance feedback or training procedure. In addition, the number of documents issued, along with the duration of the scan; The number of web page URLs extracted and the return time for the scan will also decrease. In this way, an effective and efficient search process is managed on the internet.

Ontology based web crawlers are becoming more common in recent years. Ontology based web crawlers, which will speed up semantic web applications, is still in its developmental stage.

REFERENCES

[1] S.S. Dhenakaran, and K.T. Sambanthan, "Web Crawler - An Overview", *International Journal of Computer Science and Communication*, Vol. 2 No.1, 2017.

[2] B. Pinkerton, "Finding what people want: Experiences with the WebCrawler", *Proceedings of the 2nd International World Wide Web Conference*, 1994.

[3] T. Berners-Lee, "Linked Data," The World Wide Web Consortium (W3C), 27.07.2006. Available: http://www.w3.org/DesignIssues/LinkedData.html.

[4] S.R. Bhor, S.B. Dumbre, S.A. Bakare, and M.V. Raut. "A Survey Paper on Intelligent Web Crawler". *International Journal of Advanced Research in Computer Engineering & Technology* ,Vol. 6, No. 10, pp. 1553-1556.

[5] Z. Guojun, J. Wenchao, S. Jihui, S. Fan, Z. Hao, and L. Jiang. "Design and Application of Intelligent Dynamic Crawler for Web Data Mining". 32nd Youth Academic Annual Conference of Chinese Association of Automation, pp.1098-1105.

[6] T. Berners-Lee, J. Hendler, and O. Lassila. *The Semantic Web*. Scientific American, Vol. 284, No. 5, pp. 34-43.

[7] L.C. Xiang, O.S. Yin, and P.Y. Han. "Intelligent Web Crawler for File Safety Inspection", *2015 IEEE International Conference on Signal and Image Processing Applications*.

[8] T.V. Udapure, R.D. Kale, and R.C. Dharmik, "Study of Web Crawler and its Different Types", IOSR Journal of Computer Engineering, Vol. 16, No. 1, VI, 2014, pp. 1-5.

[9] R. Chen, B.C. Desai, and C. Zhou, "CINDI Robot: An Intelligent Web Crawler Based on Multi-level Inspection", *11th International Database Engineering and Applications Symposium*, 2007.

[10] O. Menemencioğlu and İ.M. Orak "Semantik Web Arama Motoru Üzerine Bir İnceleme", *2nd Internatioanl Symposium on Innovative Technologies in Engineering adn Science*, 2014.

[11] C. Ma, M. Song, K. Xu and X. Zhang, "Web Service Discovery Research and Implementation Based On Semantic Search Engine", 2010 IEEE 2nd Symposium on Web Society Vol. 6, No. 10, pp. 672-677, 2010.

[12] D. Ezzat, M. Abdeen and M. F. Tolba, "A Memory Efficient Approach for Crawling Language Specific Web: The Arabic Web as a Case Study," *International Conference on Information Management and Engineering*, 2009.

[13] M.C. Daconta, L.J. Obrst, and K.T. Smith, "The Semantic Web: A Guide to the Future of XML, Web Services and Knowledge Management", *Wiley Publisher*, Indiana, 312p.

[14] L. Kozanidis, "An Ontology-Based Focused Crawler", International Conference on Application of Natural Language to Information Systems, pp. 376-379, 2008.

[15] P. Saxena, and M. Nosrati, "Slug: A Semantic Web Crawler", World Applied Programming, Vol. 2, No. 1, pp. 34-37.

[16] S. Ringe, N. Francis, P. Altaf H.S.A., 2012. "Ontology Based Web Crawler", *International Journal of Computer Applications in Engineering Sciences*, Vol. 2, No. 2, pp. 194-197.

[17] W.A. Gab-Allah, B.B.S. Tawfik and H.M. Nassar, "An Ontology Based Crawler for Retrieving Information Distributed on the Web", Journal of Engineering Research and Application, Vol. 6, No. 6, pp. 57-63, 2016.

Performance Comparison of Machine Learning Methods for Solving Handwriting Character Recognition Problem

Ş.G.KIVANÇ[1], A.E. BAKTIR[2] and B.ŞEN[3]

[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, gkivanc@ybu.edu.tr
[2] Ankara Yıldırım Beyazıt University, Ankara/Turkey, aebaktir@ybu.edu.tr
[3]Ankara Yıldırım Beyazıt University, Ankara/Turkey, bsen@ybu.edu.tr

Abstract **- Handwriting character recognition has been a popular problem among scientists for a few decades. United States Postal Service can be given as an example for a company that uses the recognition of digits in real life environment consistently. USPS uses digit recognition system to extract digits from pay checks and fastens the process of sending and receiving checks. Handwriting character recognition problem can be divided into two categories. Online character recognition and offline character recognition. A recognition pattern mainly based on angle of the strokes of stylus is called online recognition. A system is called offline when system takes images as inputs and tries to predict characters from given images by applying machine learning methods. We have worked on offline character recognition problem in this project. Many machine learning methods have been proposed over the years for solving this problem. In this paper we implemented 6 most popular machine learning methods to solve offline handwriting character recognition problem and compare the performance results to decide which method gives best accuracy results under pre-defined conditions. We have selected 92255 images from NIST Special 19 Database and used them as input images during the training phase of the selected machine learning methods. These methods are SVM, Decision Tree, Bag of Trees, Artificial Neural Networks (ANN), Deep learning network with autoencoders and Convolutional Neural Networks (CNN). We implemented all of these methods and compare the performance of the results according to accuracy metric. The results obtained from the comparison is going to help in deciding which ML method should be used to solve Offline Handwriting Character Recognition problem.**

Keywords **- Handwriting Character Recognition, Machine Learning, CNN, ANN, SVM, Performance Comparison**

I. INTRODUCTION

Handwriting character recognition (HCR) is a popular problem in computer science for a few decades. Although many works have been carried out about this subject, a feasible solution that can be implemented in a real world environment is still out of reach. Since every human has a distinctive handwriting, HCR is a significantly difficult problem to solve.

There are two subcategories of this problem called Online and Offline handwriting recognition problem. If a system tries to predict characters while a person writing that character, it is called online handwriting character recognition. If a system accepts images as inputs and tries to predict characters from input images, it is called offline handwriting character recognition [1]. We focused on offline handwriting character recognition problem in this study.

A lot of different techniques have been proposed to solve this problem by many researchers. Some gave more attention to prediction and tried to improve ML methods to improve their systems and some tried to extract more meaningful features from the images before prediction. In this paper we select some of the most popular Machine Learning(ML) techniques and implement each one of them to solve this problem. Later on we compare the resulting accuracies to decide which ML method can be considered to be most suitable one for this problem. The algorithms we select includes SVM, ANN, Decision Trees, Bag of Trees, DNN with Autoencoders and CNN.

There are three steps which are essential to design a system to tackle this problem. Preprocessing, feature extraction and prediction. Since all the images in NIST Special 19 Dataset have already been preprocessed, we only applied some resizing operations to the images. All images in the dataset have a size of 128*128. We resized these images to 28*28 for faster training time. Also we have selected 92255 images from NIST dataset and used 75% of this data for training and 25% for testing. Training set contains 70000 images and test set contains 22255 images. All 6 ML methods requires different features as inputs. We will talk about each feature extraction method in more details later on.

We are also going to talk about prediction processes of selected ML algorithms briefly on later sections of this paper.

The rest of the paper is organized as follows. On Section II we will look at selected ML algorithms and give details on required feature extraction techniques and prediction process. Also we will briefly look into previous works that have been carried out and share the accuracy results. Finally, we are going to compare the accuracy scores of each method for solving this problem on Section III.

II. METHODS

In this section we will talk about the implementation of given ML methods and present the accuracy results. We will also present previous works that have been carried out on this problem with given methods.

Support Vector Machine(SVM)

Support Vector Machine is considered one of the most useful ML methods. SVM have been used for object detection [2], regression [3], character recognition [4]. On many of these problems, SVM achieved comparable results with the state-of-the-art methods. SVMs are binary classifiers at their core however there are some strategies that can be applied to adapt the SVMs to multiclass classification. Two of the most popular ones are called One-Against-One and One-Against-All [5].

One-Against-All there is an SVM for each target class. Each SVM votes for either a sample falls into target class or falls outside of the target class. All the votes then will be counted and sample will be placed in the class with most votes. In One-Against-One strategy there is an SVM for each pair in target classes. If there are N classes, then N*(N-1) / 2 SVMs will be created. Each SVM casts a vote similar to One-Against-All but in this case it votes for either a given sample in class A or class B. Then all votes will be counted and sample will be placed in the class with the most vote. We used One-Against-One strategy in this paper. So we have created 29*(28/2) = 409 SVMs.

Like in many other field, on handwriting character recognition problem SVM is used extensively. Huang, Du et al. [4] used a hybrid model for handwriting character recognition problem. They used Hidden Markov Models(HMM) for feature extraction and SVM for prediction. Output of HMM has given to SVM. They achieved 97.48% accuracy on digits, 91.99% accuracy on uppercase letters and 91.74% accuracy on lowercase letters. Neves, Filho et al. [6] uses SVM as a digit recognizer. They construct 45 SVMs (1 SVM for each digit pair) and achieved 97.94% accuracy. We also built an SVM similar to suggested systemat [6]. Miyao,Maruyama, et al. [7] try to combine Online and Offline handwriting problem for a solution. They try to improve the performance of SVM by augmenting the data from online character database according to strokes of each user. They achieved different accuracy scores based on the number of samples. Highest accuracy is achieved on 950 samples with 97.4%. Naisen, Haron, Yuhaniz [8] used Freeman Chain Code(FCC) and SVM. FCC is used as feature extractor. SVM is used as predictor and received the output from FCC. They achieved 88% accuracy with their model.

We have used Histograms of Oriented Gradients (HoG) [9] features as input to our SVM. You can see examples of HoG features in ***Figure 1***. As a result we have achieved 93.27% accuracy score.

Decision Trees & Bag of Trees

Simplicity and fast training time make Decision Trees one of the most used ML methods. Decision Trees performs really well on simple classification problems. Stock trading [10], ,character recognition [11] problems can be given as an example of the usage areas of decision trees. Simple Decision Trees are not preferred on solving handwriting character recognition problem. Because of the problems complexity and having large amounts of features, decision trees tend to perform poorly on this problem. Liu and Huang [11] uses a decision tree method based on hierarchy decomposition. They built a feature extractor based on 4*4 mesh and 2*2 elastic mesh to extract vertical and horizontal lines separately. They combine these

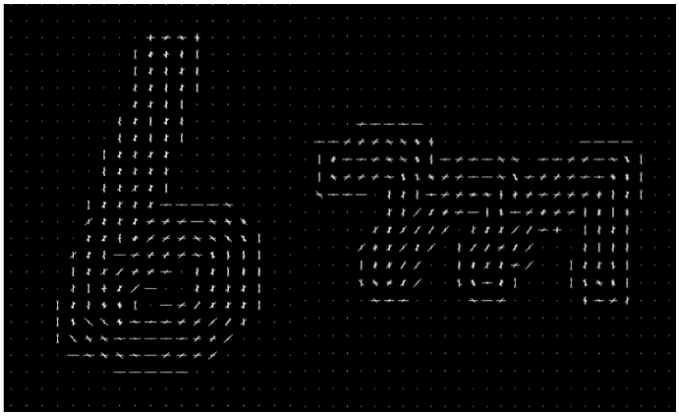


Figure 1 Two examples of HoG features. A random letter b on the left and random letter m on the right

two distinct features with the original set and used it as input. Hierarchy decomposition based decision tree is used as a predictor. It splits the training set by detecting different key attributes on different layers according to the correlation between classes and attributes. Designed system achieved 90.8% accuracy score.

We implemented a simple binary classification tree for multi-class classification. We convert all input images to a single vector and used it as input to train the decision tree. Resulting prediction accuracy from trained decision tree was 69.12%. Performance of the simple decision tree can be improved by some feature extraction tricks but we leave this part for the future studies. We try to avoid the maximization of performances of the selected ML methods because we like to see what these methods can achieve under same simple conditions.

Another technique we used to solve this problem was Bag of Trees. Bag of Trees is simply the combination of ***n*** trees. Single decision tree tends to overfit. By combining multiple trees we prevent the overfitting and increase the generalization. Cordella , Stefano et al. [12] used a Random Forest (RF) to build a reliable pre-classification model. Proposed model [12] based on a rejection method. Work mentioned in the paper investigates the relation between error rate of the system and rejection rate. System gives a probability for each class label. If the sum of all the class probabilities is below a certain threshold value then the sample is rejected. They use 300 trees to build RF and achieved 29.6% error rate for NIST dataset.

We ensemble with 100 bagged trees. Required input for the BoT is the same as simple decision tree. 28*28 images converted to a single vector. This way each pixel value became a feature. We achieved 84.43% accuracy with this method.

Artificial Neural Networks (ANN)

ANN has proved to be an efficient ML method for solving difficult problems such as handwriting character recognition. Architecture of ANN vaguely inspired by human brains and how it works. ANN has computation nodes called neurons. These neurons are fully connected with each other in the system. A connection between any two nodes has a property called weight. These weights are adjusted throughout the

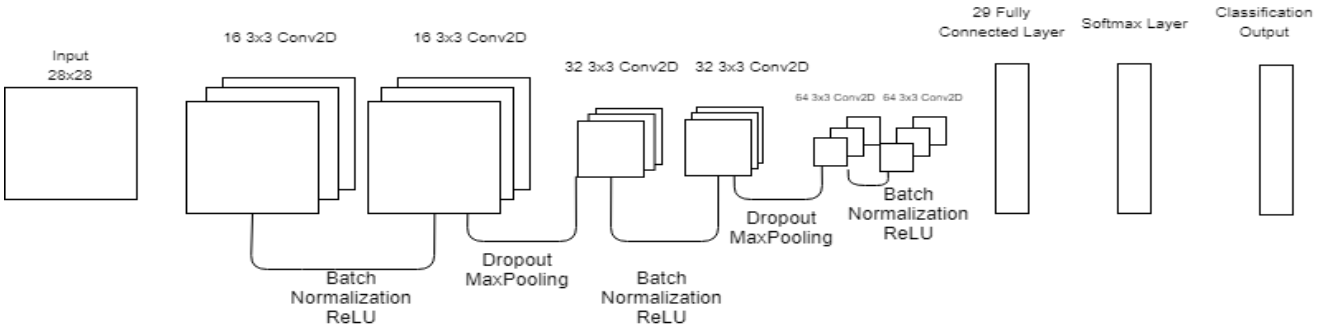


Figure 2 Architecture of our CNN

training process by comparing the simulated results with target classes and back-propagate error in the system.

ANN has been used in many different fields such as image classification, medical image segmentation, forecasting and handwriting character recognition. Rahaman, Hasan et al. [13] used a simple ANN with a row and column based feature extraction method. As a result of the feature extraction they get a 64x1 feature matrix in order to use as input to ANN. They achieved 94.3% accuracy. Pujari and Majhi [14] proposed work is similar to our work describe in this paper. They select 4 methods namely SVM, ANN, Decision Tree, Discriminant Analysis. They try to achieve the highest accuracy with all possible combinations of these methods. As a feature extraction they used a gradient and curvature based methods. The outputs of classifiers are combined by using confidential weighted scheme. They achieve best accuracy with SVM+DA and SVM+DA+C 5.0 with 97.5%.

We construct an ANN with three layers. Input layer, hidden layer and output layer. Input layer have 784 neurons, each neurons corresponds to a pixel value in an image. Hidden layer contains 100 neurons that are fully connected to input and output layers. Output layer have 29 neurons, each neuron corresponds to a target class. Sigmoid is used as and activation function on each neuron. We achieved 84.35% accuracy score with constructed ANN.

Convolutional Neural Network (CNN)

CNNs have become really popular for image classification tasks in the last decade. The reason why CNN has gotten popular recently is because image datasets have become large enough recently. CNN requires large image datasets to perform well, otherwise systemoverfits really quickly. The architecture of CNN is well suited for capturing fine details in images. It performs far better than any other ML method on image classification tasks [15]. CNN has been used in various fields such as tumor segmentation [16], object recognition [15] and character recognition [17]. Niu and Suen [18] proposed a CNN-SVM hybrid model. CNN is used for feature extraction and SVM is used for prediction. They train CNN with input images for few epochs. Then an SVM with Radial Basis Function(RBF) receives outputs of hidden layers of CNN as input. SVM is then trained with new input features. System achieves 0.19% error rate. Zong, Jin and Xie [19] used GoogleNET and Directional Feature Maps mainly Gabor Features, HoG features and gradient feature maps. They achieved 96.74% accuracy. Yuan, Bai et al. [20] uses a LeNet [21] architecture. They set the output of the CNN with Error Correcting code so CNN has the ability to reject in recognition process. They achieved 93.7% accuracy with uppercase letters and 90.2% accuracy with lowercase letters. Cireşan, Meier et al. [17] propose a CNN committee approach for solving handwriting recognition problem. They trained 5 differently initialized CNN with the same dataset with continuously distort each image slightly before every epoch. Distorting every image in the training set before every epoch is a useful technique to prevent overfitting and increase generalization capability of the system. According to their results, authors states that worst committee achieves better accuracy than best single CNN. The error scores they achieved is as follows, 1.91% for uppercase letters, 7.71% for lowercase letters.

We have constructed a 26 layer CNN. Network starts with an Input Layer accepting 32*32 gray scale images. Input layer is followed by a two sets of 3x3 16 Convolutional layer followed by Batch Normalization layer and ReLU (Rectifier Linear Unit) layer. After two sets of Convolutional layers we add a Dropout layer with 50% dropout chance. Then we have a 2x2 Max Pooling layer with [2 2] stride and 0 padding. These combination of layers again repeated with 3x3 32 Convolutional layer followed by normalization and ReLU layer and 3x3 64 Convolutional Layer. Followed by normalization and ReLU layer. Again these two sets of layers followed by Dropout layer and 2x2 Max pooling layers with [2 2] stride and 0 padding. Finally we have a 29 fully connected layer, softmax layer and classification layer. You can see the proposed architecture in ***Figure 2***. In this architecture convolutional layers extract features and max pooling layers downsize the features for simplicity and fast training time. We stacked two convolutional layers back to back to improve the feature extraction process. We have achieved 95.74% accuracy with this network.

Deep Neural Network with Stacked AutoEncoders

Another approach we tried is to build a deep neural network with two hidden layers. We increase the number of hidden layers to improve the accuracy of a simple ANN. With this approach, we used stacked auto-encoders for feature extraction at each hidden layer. Auto-encoder is a neural network which attempts to replicates its input as its output. If the neurons in hidden layer is less than the size of the input, auto-encoder compresses the input.

Autoencoders have been used in many fields. Some of those fields are emotion recognition [22], face recognition [23],

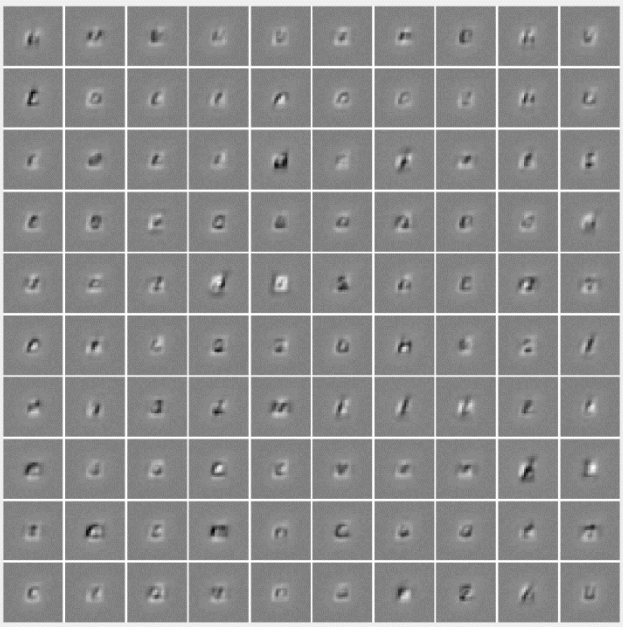


***Figure 3* Example of extracted features from Autoencoder on Hidden Layer 1**

medical image analysis [24], speech enhancement [25] and character recognition [26]. Wang,Chen et al. [26] proposed a system which is similar to ours. They try to predict handwritten characters in Chinese legal amounts. They built a network with two auto-encoders. They train each auto encoder separately and combine the outputs. They also used global overlapped elastic meshing algorithm. Elastic meshing algorithm intends to segment the image into pieces according to the foreground pixel position. They achieved 0.64% error rate with their network.

We also propose a neural network with two hidden layers. First we train hidden layers individually in an unsupervised fashion using auto-encoders. You can see the extracted features from autoencoder in Fi***gure 3***. Then we train a final sofmax layer , and stack all these layers to form a deep network. Newly formed deep network will be trained one final time in a supervised fashion. We have achieved 90.35% accuracy with this network.

III. CONCLUSION

In this paper we implement some of the most popular ML methods to solve handwriting character recognition problem and compare the performance results of each method. We have shown the comparison between results in **Table 1**. According to our findings we have shown that CNN is the most suitable ML method to solve this problem. We are aware that the results of we have found for each ML method can be improved. But it is not our main focus to achieve highest possible accuracy score for each ML method. Our aim was to implement each method on a given dataset and compare the performance results.

Table 1 Comparison of accuracy results between selected ML methods

Method	SVM	Decision Tree	Bag of Trees	DNN with Autoencoders	ANN	CNN
Accuracy	93.27	69.12	84.43	90.35	84.35	95.75

REFERENCES

[1] R. Plamındın and S. N. Srihari, "On-Line and Off-Line Handwriting Recognition:A Comprehensive Survey," *1EEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,* vol. 22, no. 1, pp. 63-84, 2000.

[2] P. F. Felzenszwalb, R. B. Grishick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 32, no. 9, pp. 1627-1645, 2009.

[3] Y. Bazi and F. Melgani, "Semisupervised PSO-SVM Regression for Biophysical Parameter Estimation," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 45, no. 6, pp. 1887-1895, 2007.

[4] B. Q. Huang, C. J. Du, Y. B. Zhang and M.-t. Kechadi, "A Hybrid HMM-SVM Method for Online Handwriting Symbol Recognition," in *Sixth International Conference on Intelligent Systems Design and Applications*, Jinan, 2006.

[5] G. Antony, H. Gregg and M. Tshilidzi, "Image Classification Using SVMs: One-against-One Vs One-against-All," in *28th Asian Conference on Remote Sensing*, Kuala Lumpur, 2007.

[6] R. F. Neves, A. N. G. Lopes Filho, C. A. B. Mello and C. Zanchettin, "A SVM based off-line handwritten digit recognizer," in *IEEE International Conference on Systems, Man, and Cybernetics*, Anchorage, AK, 2011.

[7] H. Miyao, M. Maruyama, Y. Nakano and T. Hananoi, "Off-line handwritten character recognition by SVM based on the virtual examples synthesized from on-line characters," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Seoul, 2005.

[8] D. Nasien, H. Haron and S. S. Yuhaniz, "Support Vector Machine (SVM) for English Handwritten Character Recognition," in *Second International Conference on Computer Engineering and Applications*, Bali Island, 2010.

[9] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005.

[10] M.-C. Wu, S.-Y. Lin and C.-H. Lin, "An effective application of decision tree to stock trading," *Expert Systems with Applications,* vol. 31, no. 2, pp. 270-274, 2006.

[11] D. Liu and X. Huang, "Recognition of Off-Line Handwritten Chinese Character by Using Decision Tree Based on Hiberarchy Decomposition," in *International Conference on Advanced Computer Control*, Singapore, 2009.

[12] L. P. Cordella, C. De Stefano, F. Fontanella and A. Scotto di Freca, "Random Forest for Reliable Pre-classification of Handwritten Characters," in *22nd International Conference on Pattern Recognition*, Stockholm, 2014.

[13] A. Rahaman, M. Hasan, F. Shuvo, A. S. Ovi and M. Rahman, "Analysis on handwritten Bangla character recognition using ANN," in *International Conference on Informatics, Electronics & Vision*, Dhaka, 2014.

[14] P. Pujari and B. Majhi, "Performance comparison of ensemble models for recognition of offline handwritten Odia numerals," in *IEEE Power, Communication and Information Technology Conference*, Bhubaneswar, 2015.

[15] A. Krizhevsky, G. E. Hinton and I. Sutskever, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems 25*, Nevada, 2012.

[16] S. Pereira, A. Pinto, V. Alves and C. A. Silva, "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images," *IEEE Transactions on Medical Imaging,* vol. 35, no. 5, pp. 1240-1251, 2016.

[17] D. C. Ciresan, U. Meier, L. M. Gambardella and J. Schmidhuber, "Convolutional Neural Network Committees for Handwritten Character Classification," in *International Conference on Document Analysis and Recognition*, Beijing, 2011.

[18] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN–SVM classifier for recognizing handwritten digits," *Pattern Recognition,* vol. 45, no. 4, pp. 1318-1325, 2012.

[19] Z. Zhong, L. Jin and Z. Xie, "High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps," in *13th International Conference on Document Analysis and Recognition*, Nancy, 2015.

[20] A. Yuan, G. Bai, L. Jiao and Y. Liu, "Offline handwritten English character recognition based on convolutional neural network," in *10th IAPR International Workshop on Document Analysis Systems*, Gold Cost, QLD, 2012.

[21] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[22] A. R. Garcia, M. Elshaw, A. Altahhan and V. Palade, "Stacked deep convolutional auto-encoders for emotion recognition from facial expressions," in *2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage,AK, 2017.

[23] M. Kan, S. Shan, H. Chang and X. Chen, "Stacked Progressive Auto-Encoders (SPAE) for Face Recognition Across Poses," in *IEEE Conference on Computer Vision and Pattern Recognition*, Colombus, OH, 2014.

[24] H.-C. Shin, M. R. Orton, D. J. Collins, S. J. Doran and M. O. Leach, "Stacked Autoencoders for Unsupervised Feature Learning and Multiple Organ Detection in a Pilot Study Using 4D Patient Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 35, no. 8, pp. 1930-1943, 2013.

[25] X. Lu, L. Tsao, S. Matsuda and C. Hori, "Speech Enhancement Based on Deep Denoising Autoencoder," in *INTERSPEECH*, Lyon, 2013.

[26] W. Meng, C. Youbin and W. Xingjun, "Recognition of Handwritten Characters in Chinese Legal Amounts by Stacked Autoencoders," in *22nd International Conference on Pattern Recognition*, Stockholm, 2014.

Artificial Neural Networks Approach to Greenhouse Heating Requirement Estimation

Özlem ALPAY[1] and Ebubekir ERDEM[2]

[1,2] Firat University, Elazig/Turkey, [1]oalpay@firat.edu.tr, [2]aberdem@firat.edu.tr

Abstract **- New and less energy consuming methods are developed to reduce the increasing heating costs day by day. Heat transfer method is one of the most commonly used methods for heating an environment. The amount of heat required to heat an environment in a heat transfer technique is found by the amount of heat lost from the environment. In this study, artificial neural networks were used for estimating the monthly heat demand for the heating needs of a greenhouse in Elazığ province with 2017 meteorological and spatial data. The amount of heating has been tried to be estimated using the MATLAB program with the lowest error. Heating Degree-Day (HDD) values and latitude, longitude and altitude data were used for analyzes to be made in the artificial neural network model. It has been estimated that the heat requirement for the heating of the greenhouse is lower than the heat requirement for the heat transfer method in the designed artificial neural network model.The artificial neural networks model has been found to be a useful method for studying the heating of greenhouses.**

Keywords **– Artifical Neural Network, Heat loss-gain, Greenhouse Systems, Heat Energy.**

I. INTRODUCTION

Increasing the quality and quantity of agricultural crops is classified as an important technology that provides a microclimate condition suitable for greenhouse vegetation development and plant growth. Recently, expert researchers have developed a number of models to understand the energy change events that govern the interior environment of greenhouses [1].

Manonmani and etc. [2]controlled the temperature and humidity values of the sera by developing artificial neural networks (ANN) model in order to produce efficient and high quality pepper cultivation in sera environment. For the winter conditions, the temperature of the greenhouse was accepted at about 32ºC-35ºC range and the humidity value at 12g / kg-8g / kg range. According to these assumptions, it was observed that the quality and efficiency of the peppers grown in the greenhouse controlled by the ANN they designed using the nonlinear autoregressive time series increased.

Critical environmental parameters affecting the growth of Salazar and other greenhouse [3] were determined as temperature, relative humidity, carbon dioxide, nutrition and irrigation. They admitted that the primary factor for the greenhouse grower is temperature and that temperature indirectly affects the relative humidity directly and the carbondioxide level.The most suitable average temperature for tomato products is between 20ºC-25ºC and the optimum night temperature is between 15ºC-20ºC. In addition, the optimal relative humidity is between 50% and 60%. Trying to keep the environmental variables at optimal levels causes an increase in the energy consumption of greenhouses. For this reason, it is important to have a precise control for heating, cooling and humidification. In this study, a sensitivity analysis was applied to estimate the internal and external environmental factors of the sera by using an artificial neural network model based on enthalpy. It has been identified that the variables are more successfully controlled by the generated ANN model.

Baytorun and the other[4], the greenhouse is commonly done in Turkey and greenhouse hours using the provinces where the potential heating- degree-day (HDD, High Degree Day) value with the ISIGER-SERA expert systems, they estimate greenhouse heat needs according to some of the climate value.

Bayram and Yeşilata [5] have introduced a simple approach to integrating (CDD , Cold Degree Day) with HDD data cooling rate day data to ensure the adequacy of insulation thicknesses according to TSE 825 and the reorganization of climate regions in our country.

In Cemek study [6], he estimated the heat requirements for unit surface and base areas of greenhouses by using ANN model, taking into consideration the latitude, longitude, altitude, average temperature data and months of Samsun provinces and districts. In the created model, the months, latitude, longitude, altitude and mean temperature input data are determined, and the heat requirement is determined as the output data. Samsun province and 8 districts of 7 districts training data, Samsun center and Bafra districts were used as test data. In different network constructions, approximately 99% predicted (R^2) values were obtained from the tested data using the Levenbergh-Marquardt (LM) training algorithm . As a result, in this study, using the artificial neural network model, the heating needs of the greenhouses were successfully estimated.

Yelmen and Çakır [7], analyzed the heating requirements of the unit area in Mersin provinces and districts using the artificial neural networks, one of the soft calculation methods, by using the LM algorithm. They have shown that the use of ANN is an appropriate method for heating the undergrounds.

Eredics and Dobrowiecki [8], have studied three methods from low complexity tables to high-complexity neural network structures to minimize consumption because of heating and cooling costs of greenhouses. The best solution is a neural network for the heating process, a separate neural network for the cooling process, and a combination of these two neural

networks. In these models, fewer hidden neurons were used and more reliable (errors in the range of 1 ° C) estimates were made for the heating temperature up to 12 hours before.

Jurenoks and Novickis [9] in their study describes a central heating system control method implemented by using the fuzzy control system designed.

In the first part of this work, it has been informed about the studies on the heating of greenhouses and greenhouses. In the second part, ANN and heat loss-gain methods were explained and they were used to estimate the heat requirement of a greenhouse 7mx25mx7m sized, plastic covered in Elazığ province. In the third chapter, the findings of the methods are given. In the last part, the conclusion of the study and the suggestion about the study were presented.

II. METHOD

In this study, the heating requirement of a greenhouse in Elazığ province was estimated using ANN and heat loss-gain methods.

A. Artifical Neural Network

ANN is a computational method neuroscientist-inspired and trained to achieve the desired result using input and output data. *ANN has the ability to learn, to supply output according to the inputs [10]. An artificial nerve cell general structure is shown in Figure 1.

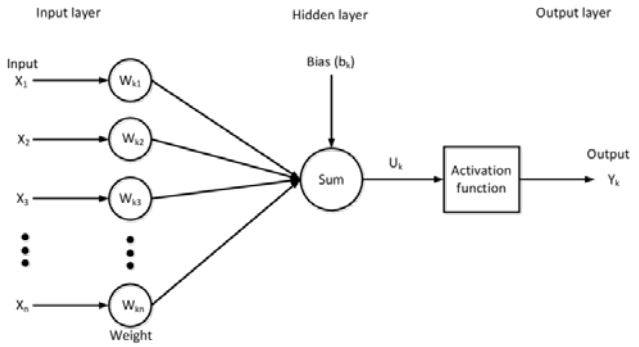


Figure 1. Artificial neural cell general structure

In this study, a feedforward ANN model was used to estimate the heat requirement. The LM training algorithm is preferred because of fast and stable in ANN training.

ANN model has two kinds of data sets including training data and testing data. The first data set to train the model and the second data set to test the model was used. The performance of the ANN model has completed learning 100 iterations while the performance of the model has been reduced to less than 100 iterations, the output value is repeated in more than 100 iterations.

Weight and target variables are used in the training of ANNs. The total input in the hidden layer is calculated by multiplying the input values by weights (wi). The ANN model have created the input data are 4 in the input layer, the weights of the inputs, the output data are 1 in the output layer,the number of neurons is 10 in the single layer and the single layer as shown in Figure 2. A total of 87 data were used, of which 75 training data and 12 test data.

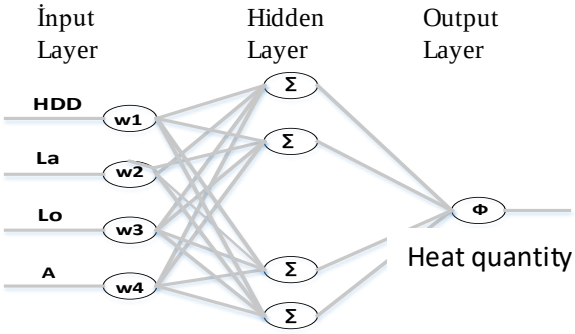


Figure 2. The designed ANN model

The HDD values of the province, latitude (La), longitude (Lo) and altitude (A) values were determined as input data. The designed model were trained by the dataset that Baytorun et al. [4] in the study created with the HDD values of the provinces.

HDD concept is expressed as high degree day in the international literature. The lowest and highest HDD values of the province and the HDD value of Elazığ province are used in training the model as shown in Table 1.In the calculation of the HDD value is used Equation 1. Th is threshold temperature and To is measured temperature.

$$HDD(Th) = \sum_{k=1}^{n}(Th - To) \qquad (1)$$

Table 1. Province and HDD values

Province	HDD values
Ardahan	4469
Erzurum	4205
Elazığ	2211
Ankara	2199
İstanbul	1433
İzmir	845
Adana	579
Mersin	552

There are many statistical test methods to examine the relationship between the variables for the intelligibility and comparability of the calculations. Correlation Coefficient (R^2) and Error Square Average (MSE) were chosen as the most accepted model performance criterion in the literature [11]. The highest values of R^2 and the lowest value of MSE are considered. In this case R^2 value is 0.996 and MSE value is 0,0191.

B. Heat Loss-Gain Method

Another method for estimating the heat requirement is the heat loss-gain method. The heat requirement of the greenhouse is found with amount of heat lost from the greenhouse. The type of cover material is important in the calculation of the heat capacity because the greenhouse is covered with a certain

					Heat Loss Calculation											Page	1
					ÖZLEM ALPAY											Floor	Ground
																Date	29.03.2018
Building Components			Field Measurement					Heat Loss Calculation				Raises					
Symbol	**Direction**	**Thickness**	**Length**	**Height or Width**	**Total Field**	**Amount**	**Extracted Field**	**Calculated Field**	**Total Heat Transfer Coefficient**	**Temperature Difference**	**Heat loss non raise**	**Operating**	**Floor Height**	**Direction**	**Total**		**Total**
									Kcal			Zd	Zr	Zh	Z		
		m	m	m	m²	Ad	m²	m²	m²h°	°C	KCal/h	%	%	%	1+%		KCal/h
Greenhouse 25 °C																	
B.K	K		1,00	2,10	2,1	1		2,10	2,6	35	191						
PVC	K	0,2	7,00	5,00	35	1	2,1	32,90	0,2	35	287						
PVC	D	0,2	25,00	5,00	125	1		125,00	0,2	35	1089						
PVC	G	0,2	7,00	5,00	35	1		35,00	0,2	35	305						
PVC	B	0,2	25,00	5,00	125	1		125,00	0,2	35	1089						
TAV	D	0,2			43,98	1		43,98	0,2	35	383						
T.DÖ		0,2			147	1		147,00	0,5	29	2132						
											5476	7		5	1,12		6134
	Q_k	=	15,00	x	5,46	x	0,90	x	0,8	x	36	x	1,00				2229
																	8363

Figure 3. Heat loss chart

covering material. The heat transfer coefficient of the used cover material is the most important effect used in calculating the amount of heat lost. In the calculation of the conduction heat losses is used the heat loss chart as shown in Figure 3. The values are written and calculated in this chart.

The heat loss calculator is created using the following steps.

Step 1. In the first column is symbols of building components, second column is direction, 3th column is, thickness, 4th column is lengths, and 5th column is height (or width).

Step 2. The total area in the 6th column is determined by multiplying the 4th and 5th columns.

Step 3. Tın the 7th column is determined with the number of fields calculated in column 6th.

Step 4. Window and door areas that it used to find the net wall area is written in the 8th column .

Step 5. The calculated area is written in the 9th column,

Step 6. K total heat transfer coefficient that it used in the calculation of building components is written in the10th column

Step 7. Temperature difference between outdoor and indoor environment is written in the 11th column.

Step 8. In the 12th column is written the heat loss resulting from the multiplication of the values in columns 9th, 10th and 11th.

Step 9. The combined incremental coefficient (Zd) is written In the 13th column

Step 10. Floor height increase (Zr) is written in the 14th column. It did not account because it was greenhouse.

Step 11. Direction increase (Zh) is written in the 15th column. This value should be taken into consideration for the direction of the greenhouse.

Step 12. Sum of increases is written in the 16th column. The total amount of increase (Z) is calculated using Equation 2.

$$Z = (1 + \%\ Zd + \%\ Zr + \%\ Zh) \quad (2)$$

Step13. Total heat requirement (Qh) is written in the 17th column.

Step 14. The process is terminated [12].

The features of the greenhouse are determined that comfort temperature 25 ° C, measured temperature -10 ° C, soil temperature -4 ° C, dimensions of the greenhouse 7mx25mx7m, greenhouse outer door iron, dimensions of the door 1mx2,1m direction north and greenhouse plastic (pvc). In this case, the untimely heat loss is 5476 KCal/h for the doors, walls and ceiling, Zd value is 7%, Zr value is not processed because of the greenhouse (single storey) and the Zh value is 5% and the total increase amount (Z) is 1,12%. The heat energy is calculated 6134 Kcal/h by using the untimely heat loss and the total amount of increase and it is calculated 2229 Kcal/h by using the field and the total heat transfer coefficients. The total heat energy is 8363 Kcal / h. Accordingly, the amount of heat to be given to the greenhouse is calculated as 9.72 W using the table in Figure 3.

III. FINDINGS AND DISCUSSION

In this study, the heating requirement of a greenhouse in Elazığ province was estimated using heat loss-gain method and ANN model for January, February, March, April, October, November, December. A total 5065 measurements were made in 2017. The measured values were evaluated in three classes: night, day and average because the temperature difference between day and night is higher in Elazığ province.

The heat loss-gain method predicted the heating needs by writing the greenhouse interior temperature, comfort temperature and the properties of the greenhouse on the heat loss calculator chart shown in Figure 3.

Table 2. Average temperature (AT, ℃) and relative humidity (RH %) values

	Night	Daytime	Average
	AT /RH	AT /RH	AT /RH
January	-5.5 / 84.6	3.9 / 76	-0.8 / 80.3
February	-7.5 / 59.3	7.03 / 54.8	-0.2 / 57
March	0.4 / 77.7	13.5 / 61,6	6.9 / 69.7
April	2.8 / 80.3	18.4 / 65.4	10.6 / 72.8
October	6.9 / 59.7	22.7 / 49.9	14.8 / 54.8
November	1.3 / 73.6	13.3 / 65.9	7.3 / 69.7
December	-1.3 / 84.1	7 / 83.1	2.8 / 83.6

In the ANN model, a total of 87 data were used; 75 training data and 12 test data. The dataset is Baytorun et al. [4] in the study created with the HDD values of the provinces as shown in Table 1. These data were evaluated using the LM training algorithm. Test data were tested on different network structures and network structure that the best predict result is selected. The temperature and relative humidity values of the night, daytime, average classes in Elazığ province are shown in Table 2. The type and characteristics of the greenhouse and its appearance as shown in Figure 4 and 5.

Figure 4. Types and characteristics of used greenhouse

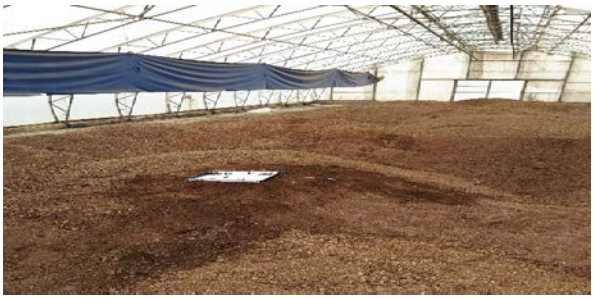


Figure 5. View of the greenhouse

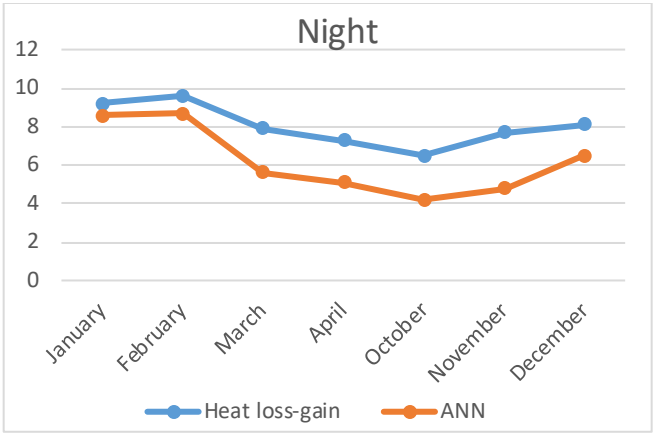


Figure 6. Amount of heat consumed according to nightly values

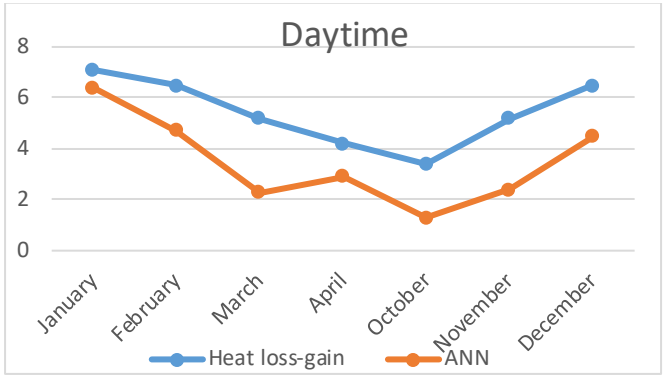


Figure 7. Amount of heat consumed according to daytime values

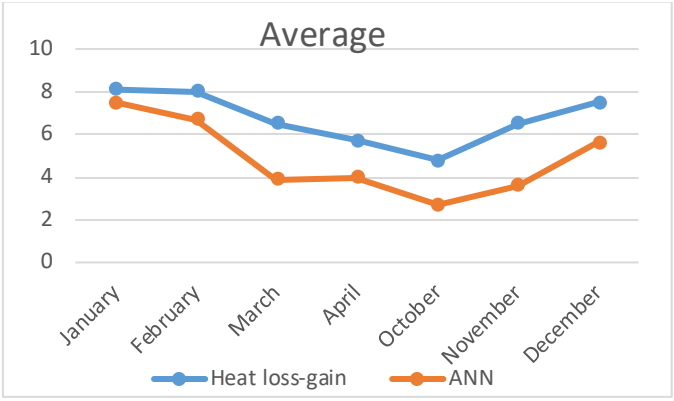


Figure 8. The amount of heat consumed by average values

The three classes of the output are night heat requirement, daytime heat requirement and average heat requirement, as shown in Figures 6, 7 and 8.

The methods are compared in Figure 6,7 and 8. the ANN model according to heat loss-gain method it operates with about 18%, 19% and 18% less heat energy respectively.

CONCLUSION

In this study, an ANN model was developed to estimate the heat requirement of the greenhouse. Temperature and relative humidity meteorological data of the Elazığ province were taken from Turkish State Meteorological Service. This study compare the ANN model with heat loss-gain method. The ANN model uses less about 18% to 19% heat energy than the heat loss-gain method. In study [9], there was an 8% reduction in the amount of heat energy used to heat an environment. The developed ANN model consumes less 10% heat energy than the study [9] The manufacturer contributes economical because the ANN model finds that the amount of heat required to heat the Greenhouse is lower than that of the heat loss-gain method. ANN are an algorithm that learns and recalls in the computer environment, which makes decisions, produces conclusions, has the ability to reach the best result with less data usage, can draw conclusions from the available data in case of insufficient data.It is suggested that the ANN model can be used to estimate the heat need for any region by meteorological and geographical parameters.

REFERENCES

[1] Labidi, A., Chouchaime, A. and Mami, A. "Control of relative humidity inside an agricultural greenhouse," 18th International Conference on Sciences and Techniques of Automatic control & computer engineering – STA, pp. 109-114,2017.

[2] A.Manonmani, T.Thyagarajan and S.Sutha "ANN based modeling and control of GHS for winter climate,"Trends in Industrial Measurement and Automation (TIMA), pp. 1 – 7, 2017.

[3] Salazar, R., Rojano, A., López, I. and Schmidt, U." A Model for the combine description of the temperature and relative humidity regime in the greenhouse," Ninth Mexican International Conference on Artificial Intelligence, pp. 113-117, 2010.

[4] Baytorun, A. N., Üstün, S. ve Akyüz, A. "Farklı Isıtma-Derece-Gün (HDD) değerlerine bağlı olarak seralarda ısı enerjisi gereksiniminin belirlenmesi," Çukurova Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi 31(2), pp.119-128 ,2016.

[5] Bayram, M. ve Yeşilata, B. "Isıtma ve soğutma derece gün sayılarının entegrasyonu,"IX. Ulusal Tesisat Mühendisliği Kongresi, 425-432, 2009

[6] Cemek, B. Samsun il ve ilçelerinde seraların iklimsel ihtiyaçlarının belirlenmesi. O.M.Ü. Ziraat Fakültesi Dergisi 20(3), 34-43, 2005.

[7] Yelmen, B. ve Çakır, M.T."Yapay sinir ağları kullanılarak sera ısıtma ihtiyacının tahmini," Politeknik Dergisi Cilt:14 Sayı:4, 235-241,201.1

[8] P. Eredics and T.P. Dobrowiecki, "Neural models for an intelligent greenhouse - the heating," Computational Intelligence and Informatics (CINTI), 11th International Symposium on 2010 pp: 63-68,2010.

[9] Jurenoks, A. And Novickis, L. "Fuzzy logic control method for autonomous heating system in energy efficient homes", 2nd International Conference on Integrated Circuits and Microsystems,2017.

[10] Sriharipriya, K.C. and Sanju, R. "Artifical neural network based multi dimensional spectrum sensing in full duplex cognitive radio networks", International Conference on Computing Methodologies and Communication (ICCMC), pp 307 – 312,2017.

[11] Şahan, M. and Okur, Y. "Akdeniz bölgesine ait meteorolojik veriler kullanılarak yapay sinir ağları yardımıyla güneş enerjisinin tahmini" SDU Journal of Science (E-Journal), 11 (1): 61-71,2016.

[12] Isı kaybı hesabı örneği - http://deneysan.com /content/images/docments /isitma2_65672174.pdf

Speech and Eye-Gaze-Based PC Control System for Disabled People

H. KAYA[1], F.F. İÇEL[1], F. ÖZATAK[1], S. KARAKOÇ[1] and O. SARIYER[1]

[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, hilalkaya@ybu.edu.tr
[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, fatihfurkanicel@gmail.com
[1]Ankara Yıldırım Beyazıt University, Ankara/Turkey, ferhat.ozatak.26@gmail.com
[1]Ankara Yıldırım Beyazıt University, Ankara/Turkey, snnkrkc1917@gmail.com
[1]Ankara Yıldırım Beyazıt University, Ankara/Turkey, osmannsariyer@gmail.com

***Abstract* - The field of human computer interaction (HCI) involves the creation of interactive computing systems for humans to enhance the quality of life of people especially with disabilities all over the world. This study proposes a multimodal system to give the opportunity to carry out all daily works with a personal computer (PC) for disabled people that cannot use their hands. In this study, it's aimed to create an interaction between the user and a machine that is performed by user's voice and eye movements. Turkish Speech Recognition was performed by using mel-frequency cepstral coefficient (MFCC) extraction, hidden markov model (HMM) and artificial neural networks (ANN). As a joint part of the software, an efficient eye tracking system with a Tobii 4C eye tracker, was developed having a feature of eye blink detection for controlling an interface that provides an alternate way of communication. This multimodal system was developed by the authors using Java language and Matlab library and the system performed promising results for Turkish training words. To increase the system's performance, usage of natural language processing methods is planned as a future work.**

***Keywords* - Human Computer Interaction, Speech Recognition, Eye-Gaze Tracking, PC control system, disabled people.**

I. INTRODUCTION

Many people are unable to operate PC by means of standard computer mouse or keyboard because of disabilities of their hands or arms. One possible alternative choice for these people is a multimodal system, which allows controlling PC without mouse and keyboard but using: (1) head movements to control the mouse cursor position on screen [1]; (2) speech for giving the control commands and (3) eye-gaze tracking to control the screen keyboard or mouse cursor position on the screen. For instance, a human can have problems with activity of neck and hence reduced ability to move the head in one or more directions. In such cases the eye tracking system can be more successful than a head tracking system [2]. Of course, speech input is only one acceptable alternative to keyboard for motor-disabled users [3].

The multimodal system is aimed for the disabled people, which need other kinds of interfaces than ordinary people [4]. Most of the multimodal systems combine two modalities only: speech and head movements. It is concerned with specific application area for hand-disabled people, also there are some applications using such modalities as gestures, haptics, handwriting. On the other side, systems using emotion recognition, facial moves, eye detection etc. can be faced in the literature [5-7]. Telemaque is a system which employs a force feedback ready pen to help children learning to write. According to its authors this system achieved positive results regarding the adoption of a proactive strategy to control handwriting movements by the children who participated in the evaluation [8].

Researchers have studied human speech interaction with computers for many years. Much of the focus in this area has been on creating better technical speech recognition (SR) systems, and almost all the testing has centered on accuracy and productivity gains [9]. [10] aims to develop an isolated-word automatic speech recognition (IWASR) system based on vector quantization (VQ). This system receives, analyzes, searches and matches an input speech signal with the trained set of speech signals which are stored in the database/codebook, and returns matching results to users. In [11], a new approach is introduced to develop a real time isolated word speech recognition system for human computer interaction. The system is a speaker dependent system and the main task is to recognize list of words in which the speaker says through the microphone. The features used are the mel-frequency cepstral coefficients (MFCC) which gives the good discrimination of the speech signal. The Dynamic Programming algorithm is used in the system measures the similarity between the stored template and the test template for the speech recognition which gives the optimum distance.

With a growing number of computer devices, and the increasing time spent for interacting with such devices, people are strongly interested in finding new interaction methods which ease the use of computers or increase interaction efficiency. Eye tracking seems to be a promising technology to achieve this goal [12]. Eye tracking is one of the most natural ways for people with amyotrophic lateral sclerosis (ALS) and other locked-in and paralysis diseases to communicate. The majority of existing eye-tracking computer input systems use cameras to capture images of the user's eyes to track pupil movements. Most camera-based systems are expensive and not

user-friendly. This paper proposes an eye-tracking system called EyeLive that uses discrete infrared sensors and emitters as the input device [13]. Patients with motor neuron disease and most terminal patients cannot use their hands or arms, and so they need another person for their all needs, only they can control their eyes. Using an eye-gaze tracking technique, authors have realized a real-time system for such patients. The system controls a motorized electrical hospital bed (EHB) by eye gaze with 4 degrees of freedom, using a low-cost webcam [14]. In [15], authors present the development of an augmentative system for people with movement disabilities (mostly cerebral palsy people) to communicate with the people that surround them, through a human-computer interaction mechanism. They developed an assistive technology application based on gaze tracking to select symbols in communication boards, which represent words or ideas, so that they could easily create phrases for the patient's daily needs.

In this study, the interaction between a user and a computer is performed by speech recognition and eye-gaze tracking techniques. A multimodal system is proposed to give the opportunity to carry out all daily works with a PC for disabled people that cannot use their hands. Turkish Speech Recognition was performed by using MFCC extraction, HMM and ANN. As a joint part of the software, an efficient eye tracking system with a Tobii eye tracker, was developed having a feature of eye blink detection for controlling an interface that provides an alternate way of communication.

II. PROPOSED METHOD

In this study, an isolated word Turkish speech recognition system has been developed. Speech recognition is used to convert speech signals into text after signal processing steps. System is trained by a group of words and controlled by these words. Although a lot of speech recognition studies have been made for English, it can be said that there are comparatively limited number of Turkish speech recognition studies.

HMM and ANN are known to be effective in the same way as for the voice signal processing and classification of the words. Fast fourier transformation (FFT) is a method for converting a change in the value of the time to the frequency domain. In contrast, the inverse Fourier transform is a method of converting the frequency values to the time domain. MFCC is the most popular method for extracting speech features from the speech recognition field. HMM, which is one of the state-of-the-art speech recognition methods, is used in the medium scale speech recognition studies [16]. HMMs provide a simple and effective framework for modelling time-varying spectral vector sequences. Therefore, almost all present day large vocabulary continuous speech recognition (LVCSR) systems are based on HMMs [17].

Steps of speech recognition process is shown in Fig. 1.

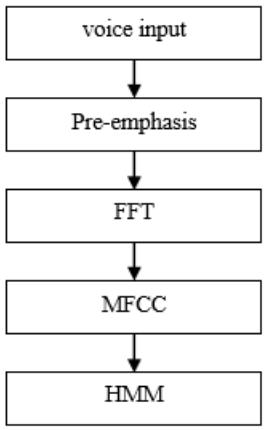


Figure 1. Speech recognition process [16]

Fig. 2 represents FFT spectrum and speech signal.

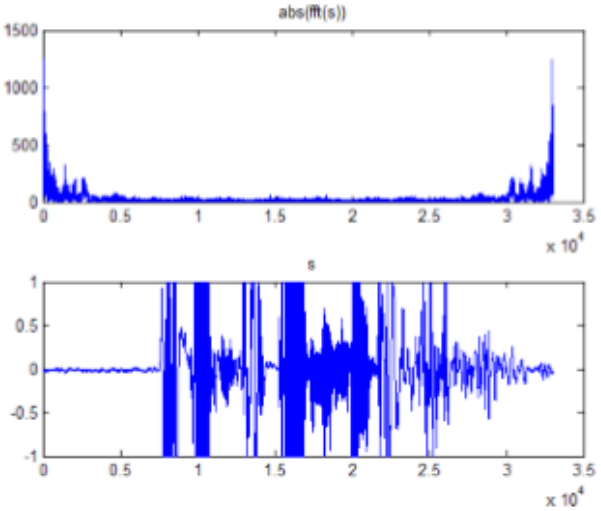


Figure 2. FFT spectrum and speech signal [16]

Steps of pre-processing process on the recorded word's signal as using noise removal method to achieve the speech without silence can be seen in Fig. 3.

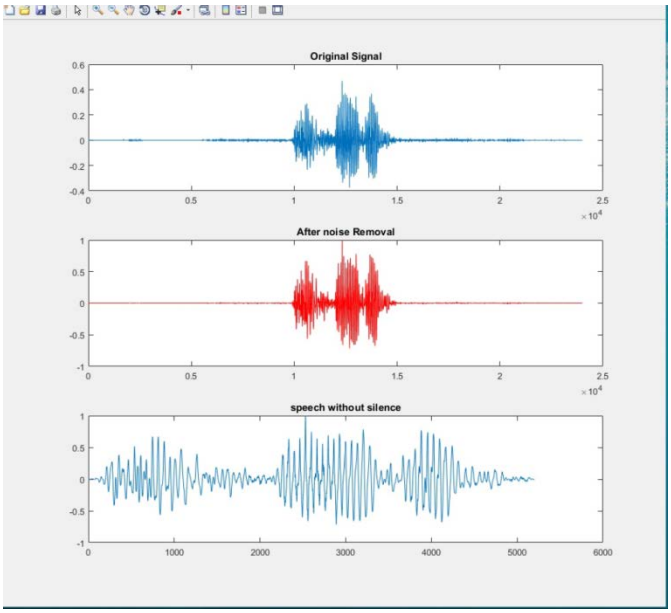


Figure 3. Noise removal step of signal pre-processing

After training the system with isolated words, system converts the signal to the text and uses it as a command to control the system. Interface can be seen in Fig. 4.

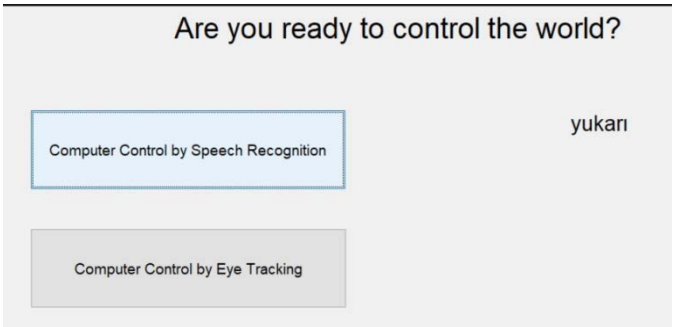


Figure 4. Graphical User Interface (GUI) of the system

How to extract the features from the signal using MFCC is represented in Fig. 5.

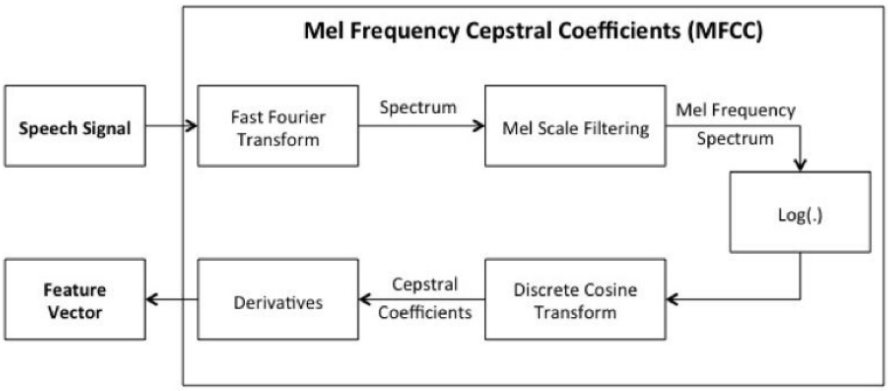


Figure 5. Study architecture of MFCC [18]

ANNs are nothing but the crude electronic models based on neural structure of brain. Artificial neurons are the basic unit of ANN which simulates the four-basic function of biological neuron. It is a mathematical function conceived as a model of natural neuron. Fig. 6 shows the basic artificial neuron architecture [19].

In this figure, various inputs are shown by the mathematical symbol, i(n). Each of these inputs are multiplied by connecting weights w(n). Generally, these products are simply summed and fed to the transfer function to generate the output results. The applications like text recognition and speech recognition are required to turn these real-world inputs into discrete values after MFCC, HMM steps [19].

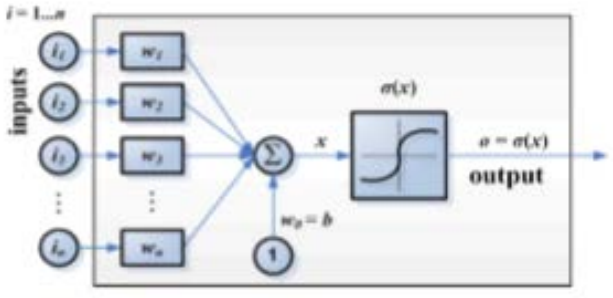


Figure 6. Basic artificial neuron [19]

In addition to the speech recognition phase, this paper also presents an efficient eye tracking system having a feature of eye blink detection for controlling an interface that provides an alternate way of communication for the people who are suffering from severe physical disabilities.

In this eye-tracking study, Tobii 4C eye tracker was used and on the area in Fig. 7, user's eye blink is tracked. Mouse cursor positioning on the screen is based on this eye-tracking process and activating the process is started with eye-blinking on the related area. Fig. 7 represents the eye-tracking area of the system.

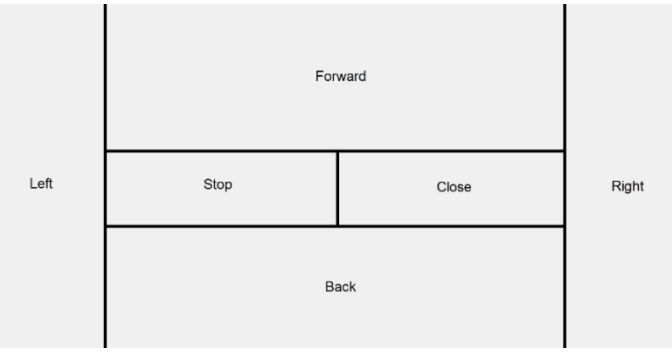


Figure 7. Eye-tracking interface of the system

The eye tracker – which consists of sensors (camera + projectors) and algorithms. The custom-designed sensors are the hardware designed to be a high-performance sensor consisting of custom components and advanced optics. With this technology, user can point, select, zoom and do actions in faster, easier, more natural ways than it would be done with the mouse or touchpad [20].

III. CONCLUSION

This study proposes a multimodal system to give the opportunity to carry out all daily works with a personal computer (PC) for disabled people that cannot use their hands. In this study, it's aimed to create an interaction between the user and a machine that is performed by user's voice and eye movements. Turkish Speech Recognition was performed by using mel-frequency cepstral coefficient (MFCC) extraction, hidden markov model (HMM) and artificial neural networks (ANN). As a joint part of the software, an efficient eye tracking system was developed having a feature of eye blink detection for controlling an interface that provides an alternate way of communication. This system performed promising results for Turkish training words but for an isolated training set. To increase the system's performance, adapting the natural language processing methods is planned as a future work.

REFERENCES

[1] K. Toyama, "`Look, Ma --- No Hands!' hands-free cursor control with real-time 3d face tracking", In Proc. of Workshop on Perceptual User Interfaces PUI'98, San Francisco, USA, 1998, pp. 49-54.

[2] R. Bates, H.O. Istance, "Why are eye mice unpopular? A detailed comparison of head and eye controlled assistive technology pointing devices", In Proc. of the 1st Cambridge Workshop on Universal Access and Assistive Technology CWUAAT, USA, 2002.

[3] Karpov, Alexey et al. Multimodal system for hands-free PC control. *2005 13th European Signal Processing Conference* (2005): 1-4.

[4] D. Tzovaras, G. Nikolakis, G. Fergadis, S. Malasiotis and M. Stavrakis. Design and Implementation of Haptic Virtual Environments for the Training of Visually Impaired, IEEE Trans. on Neural Systems and Rehabilitation Engineering, Vol. 12, No. 2, pp.266-278, June 2004.

[5] K. Kim, J. Kim, J. Choi, J. Kim, S. Lee. Depth Camera-Based 3D Hand Gesture Controls with Immersive Tactile Feedback for Natural Mid-Air Gesture Interactions. Sensors (Basel, Switzerland). 2015;15(1):1022-1046. doi:10.3390/s150101022.

[6] M. Kim, C. Jeon, J. Kim. A Study on Immersion and Presence of a Portable Hand Haptic System for Immersive Virtual Reality. Valle M, ed. Sensors (Basel, Switzerland). 2017;17(5):1141. doi:10.3390/s17051141.

[7] K. Kwangtaek, K. Joongrock, C. Jaesung, K. Junghyun, L. Sangyoun. Sensors 2015, 15, 1022-1046; doi:10.3390/s150101022.

[8] R. Palluel-Germain, F.Bara, A. Hillairet de Boisferon, b. Hennion, P. Gouagout, E. Gentaz. A visuo-haptic device- Telemaque- increases kindergarten children's handwriting acquisition. In: Proceedings of the Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC 2007), pp. 72-77. IEEE Computer Society, Los Alamitos 2007.

[9] Carl M. Rebman, Jr., Milam W. Aiken, and Casey G. Cegielski. 2003. Speech recognition in the human-computer interface. Inf. Manage. 40, 6 (July 2003), 509-519; doi=http://dx.doi.org/10.1016/S0378-7206(02)00067-8.

[10] M. Abushariah, R. N. Ainon, R. Zainuddin, O. Khalifa. (2007). Human computer interaction using isolated-words speech recognition technology. 1173 - 1178. 10.1109/ICIAS.2007.4658569.

[11] S. P. Nandyala and Dr.T.K. Kumar. Real Time Isolated Word Speech Recognition System for Human Computer Interaction. International Journal of Computer Applications 12(2):1–7, December 2010. Published by Foundation of Computer Science.

[12] H. Drewes (2010): Eye Gaze Tracking for Human Computer Interaction. Dissertation, LMU München: Fakultät für Mathematik, Informatik und Statistik.

[13] S. Liu, A. Rawicz, S. Rezaei, T. Ma, C. Zhang, K. Lin, E. Wu. (2012). An Eye-Gaze Tracking and Human Computer Interface System for People with ALS and Other Locked-in Diseases. Journal of Medical and Biological Engineering. 32. 10.5405/jmbe.836.

[14] N. A. Atasoy, A. Çavuşoğlu, F. Atasoy. Real-time motorized electrical hospital bed control with eye-gaze tracking, Turkish Journal of Electrical Engineering & Computer Sciences, Turk J Elec Eng & Comp Sci (2016) 24: 5162 – 5172; doi:10.3906/elk-1503-132.

[15] A. Galante, P. Menezes. A gaze-based interaction system for people with cerebral palsy, CENTERIS 2012 - Conference on ENTERprise Information Systems / HCIST 2012 - International Conference on Health and Social Care Information Systems and Technologies, Procedia Technology 5 (2012) 895 – 902.

[16] K. Dong-Ill, K. Byung-Cheol. Speech Recognition using Hidden Markov Models in Embedded Platform, Indian Journal of Science and Technology, Vol 8(34), DOI: 10.17485/ijst/2015/v8i34/85039, December 2015.

[17] M. Gales, S. Young. The Application of Hidden Markov Models in Speech Recognition, Foundations and Trends in Signal Processing Vol. 1, No. 3 (2007) 195–304.

[18] Mel-Frequency Cepstral Coefficients available at http://recognize-speech.com/feature-extraction/mfcc

[19] B.C. Kamble. Speech Recognition Using Artificial Neural Network – A Review, Int'l Journal of Computing, Communications & Instrumentation Eng. (IJCCIE) Vol. 3, Issue 1 (2016) ISSN 2349-1469 EISSN 2349-1477.

[20] Eye tracking technology available at https://www.tobii.com/tech/technology/what-is-eye-tracking/.

A Simple Heuristic Approach to Improve Performance of Extreme Learning Machine

C. KARAKUZU[1] and U. YÜZGEÇ[2]

[1] Bilecik Şeyh Edebali University, Bilecik/Turkey, cihan.karakuzu@bilecik.edu.tr
[2] Bilecik Şeyh Edebali University, Bilecik/Turkey, ugur.yuzgec@bilecik.edu.tr

Abstract **- Neural networks (NNs) is used to solve many engineering and science problem. Generally, feedforward architecture is preferred and gradient-based learning algorithms are extensively operated to tune all parameters of NN iteratively. This training method is a conventional one, but training process takes a long time due to the slowness of gradient-based learning algorithms. This slowness has been an important drawback in their applications. To overcome this disadvantage, extreme learning machine (ELM) concept introduced to science community in near past. Essentially, ELM is a data-driven learning algorithm for single-hidden layer feedforward neural networks (SLFNs). This algorithm provides extremely fast learning speed. In this study, performance of SLFNs learned by ELM algorithm is investigated on the problem of highly nonlinear dynamic system identification. As a result of studies on selected benchmark problems in the literature, it has been seen that ELM may not provide a good generalization success due to randomly chosen the number of hidden nodes and weight parameters for inputs in SLFN. For both the training and the test data set, very poor results have been obtained and observed surprisingly during the above-mentioned studies. Here, a simple heuristic approach has been proposed in this study in order to eliminate this bad situation and the findings obtained with this approach are discussed. Based on the obtained experimental results, it has been shown that the proposed approach determines the optimal the number of hidden nodes and a reasonable random selection of input weights required for a good generalization performance.**

Keywords **- Extreme learning machine, feedforward network, system identification, dynamic systems, heuristic approach.**

I. INTRODUCTION

THE half of past century was time period that the concept of learning has been put through artificial neural networks (ANNs) and it has become a general computation method applied in many problems. Especially multi-layer neural networks (MLP) have been extensively used to solve many science and engineering problems. The most important factor in the popularity of ANNs is to provide complex nonlinear mapping from input-output samples of systems. ANNs have given a so much valuable facility point of view to handle highly nonlinear and complex systems according to handling using classical numerical techniques to scientists and appliers.

On the other hand, the handicap of this convenient technique is training/learning process. Sometimes, it can be caused of so much time consuming. Gradient-based learning algorithms, such as back-propagation (BP) and it variants are conventional algorithms used in the training of MLPs. The most significant drawbacks of the gradient-based learning algorithms are slowness and getting stuck in a local minimum during training or learning phase. To remove slowness of gradient-based algorithms, in some studies such as [1, 2], training process implemented on specific hardware environment providing fast computation by parallel data processing like FPGA. Another current training method is the usage of both fast computing hardware environment and heuristic algorithms not required gradient like in [3-5]. This method can be operated by limited designer due to its design and implementation complexity. For this reason, training of neural and neuro-fuzzy systems may take considerable time using the conventional gradient-based and heuristic algorithms.

In the last decade, extremely fast and not iterative but data driven a novel learning algorithm was proposed [6,7] for single-hidden-layer feedforward neural networks (SLFNs). This algorithm is called extreme learning machine (ELM). In this algorithm, input weights and biases of a SLFN are randomly chosen, and its output weights are determined by a formulation based on generalized inverse.

In this study, performance of SLFNs learned by ELM algorithm is investigated on the problem of highly nonlinear dynamic system identification. For this purpose, benchmark problems selected from the literature are used. As a result of studies on them, it has been seen that ELM can not provide a good generalization success due to randomly chosen the number of hidden nodes and weight parameters for inputs in SLFN. To overcome this drawback, a simple heuristic approach has been proposed in this study and the findings obtained with this approach are discussed.

II. EXTREME LEARNING MACHINE (ELM)

ELM was proposed for learning of SLFNs by Huang, et al. [6]. Structure of a SLFN having n external inputs and single output with Nh hidden layer neurons can be given like in Fig. 1. As can be seen from this figure, input connection weights (ω_{ij}) and biases (b_k) are hidden layer parameters, and connection weights (θ_k) are output parameters of SLFN. In ELM algorithm, hidden layer parameters values and the number of hidden layer neurons (Nh) are randomly assigned and then output parameters are analytically determined by following transactions.

Suppose we are training SLFNs with Nh hidden neurons and activation function $\varphi(\cdot)$ to learn N samples. The number of hidden neurons is randomly chosen according to $Nh \leq N$ rule. Let

we assume inputs-outputs learning data pairs are given as $\{x_i, yd_i\}$. Signals on the input and output nodes in Fig. 1 can be expressed by a vector notation given in Eqs (1) and (2) respectively. In these equations, n is dimension size of input data, q is dimension size of output data, and i is index of data (i=1, 2, 3, …, N)

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, \cdots, x_{in}\} \in \mathbb{R}^n \quad (1)$$

$$yd_i = \{yd_{i1}, yd_{i2}, yd_{i3}, \cdots, yd_{ip}\} \in \mathbb{R}^p \quad (2)$$

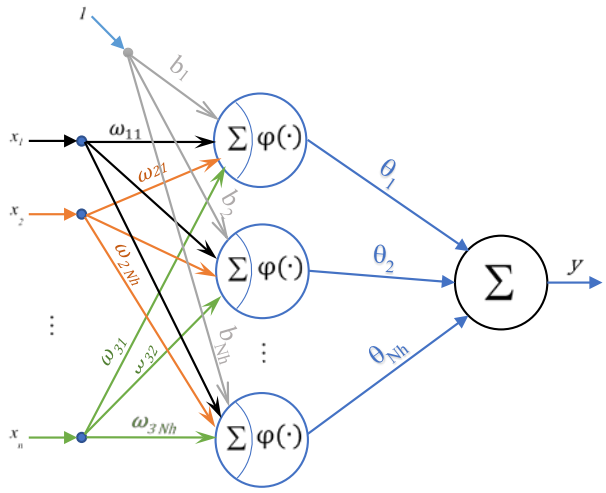


Figure 1: Structure of a SLFN with single output.

Output of a SLFN for the *ith* input is computed as following equation

$$y_i = \sum_{j=1}^{Nh} \theta_j \varphi(x_i w_j + b_j) \quad (3)$$

where input connection weights, $w_j = \{\omega_{1j}, \omega_{2j}, \ldots, \omega_{nj}\}^T$, and neuron biases, b_j, are randomly assigned and their values are kept fixed. In the equation, $\varphi(\cdot)$ represents the activation function of the neurons in SLFN and θ_j is jth connection weight of output.

There are N equations of the kind in Eq. (3), since there are N training samples. These equations can be written in a matrix, H, called hidden layer output matrix of SLFN [6,7].

$$H = \begin{bmatrix} \varphi(x_1 w_1 + b_1) & \cdots & \varphi(x_1 w_{Nh} + b_{Nh}) \\ \vdots & \vdots & \vdots \\ \varphi(x_N w_1 + b_1) & \cdots & \varphi(x_N w_{Nh} + b_{Nh}) \end{bmatrix}_{N \times Nh} \quad (4)$$

If we collect the output weights of each output in a matrix, it can be written as Eq. (5).

$$\vartheta = \begin{bmatrix} \theta_{1,1} & \cdots & \theta_{1,p} \\ \vdots & \vdots & \vdots \\ \theta_{Nh,1} & \cdots & \theta_{Nh,p} \end{bmatrix}_{Nh \times p} \quad (5)$$

Similarly, target of each output can be collected as in Eq. (6). And then, we can write N equations compactly as in Eq. (7).

$$Y_d = \begin{bmatrix} yd_1 \\ \vdots \\ yd_N \end{bmatrix}_{N \times p} \quad (6)$$

$$H\vartheta = Y_d \quad (7)$$

From the Eq. (7), a least square estimation of output connection weights, $\hat{\vartheta}$, can be obtained using generalized inverse of hidden layer output matrix (H^+) as in Eq. (8).

$$\hat{\vartheta} = H^+ Y_d \quad (8)$$

III. System Identification Using ELM

In this study, performance of ELM on dynamic system identification problem is investigated. For this purpose, three benchmark dynamic system (BDS) were chosen from the literature. These BDSs are given in Table 1.

Table 1: Benchmark dynamic systems (BDS) used for system identification based on SLFNs learned by ELM.

#	System definition
BDS#1	$y(k+1) = \frac{y(k)}{1+y(k)^2} + 1 + u(k)^3$ [8]
BDS#2	$y(k+1) = \frac{24+y(k)}{30} + y(k) - 0.8\frac{u(k)^2}{1+u(k)^2} y(k-1) + 0.5u(k)$ [9]
BDS#3	$y(k+1) = 0.5\left(\frac{y(k)}{1+y(k)^2} + (1+u(k))\,u(k)\,(1-u(k))\right)$ [10]

For each system given in Table 1, training and testing data sets were prepared as detailed seen in [11,12]. BDS#1, BDS#2 and BDS#3 are modelled by a SLFN with {u(k), y(k), y(k-1)}, {u(k), y(k), y(k-1)} and {u(k), y(k)} input compositions respectively using ELM defined the previous section. Logarithmic sigmoid transfer function (*logsig*) is used as activation function ($\varphi(\cdot)$) of hidden layer neurons in this study.

Essentially ELM is a data-driven learning algorithm for SLFNs and similar structures. In this study, a desktop computer with Intel (R) Core (TM) i7-2600 CPU @ 3.40 GHz processor, 8 GB RAM and 64 bit Windows 10 operating system was used for the experimental system identification. Mean, the best and the worst elapsed times of 100 separate runs for each BDS are given Table 2. As can be seen from this table, ELM algorithm provides extremely fast learning speed.

Table 2: Elapsed time metrics of BDS identification based on SLFN learned by ELM for 100 runs and *Nh*=50.

#	Mean [s]	Best [s]	Worst [s]	Standard Deviation
BDS#1	5.1727×10^{-04}	3.3718×10^{-04}	0.0071	6.6881×10^{-04}
BDS#2	4.0713×10^{-04}	3.1363×10^{-04}	0.0060	5.7225×10^{-04}
BDS#3	5.7799×10^{-04}	3.5046×10^{-04}	0.0063	5.9821×10^{-04}

As a result of studies on the selected BDS identification problems in here, it is not possible say that ELM always provides a good generalization success. As can be seen in Fig. 2, sometimes a good identification may not be occurred for both training and testing data set. Moreover, in some conditions, it is also possible to encounter a catastrophic result from the kind given in Fig. 3. The main cause of this kind of situations is due to randomly chosen *Nh* and weight parameters for inputs in SLFN. This is a handicap of ELM algorithm. For this reason, obtaining a model with a good generalization ability by trial and error may a lot of time consuming. Hence, ELM's fast learning feature disappears. To overcome of this drawback, we propose a heuristic approach in the next section.

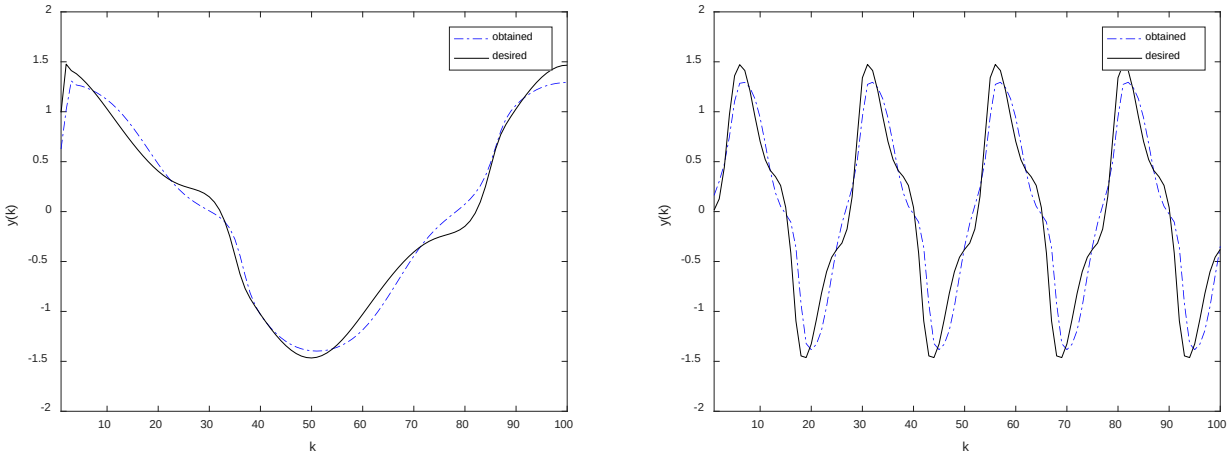


Figure 2: An SLFN model performance for BDS#1 with *Nh*=5 for training (left) and testing (right) data sets.

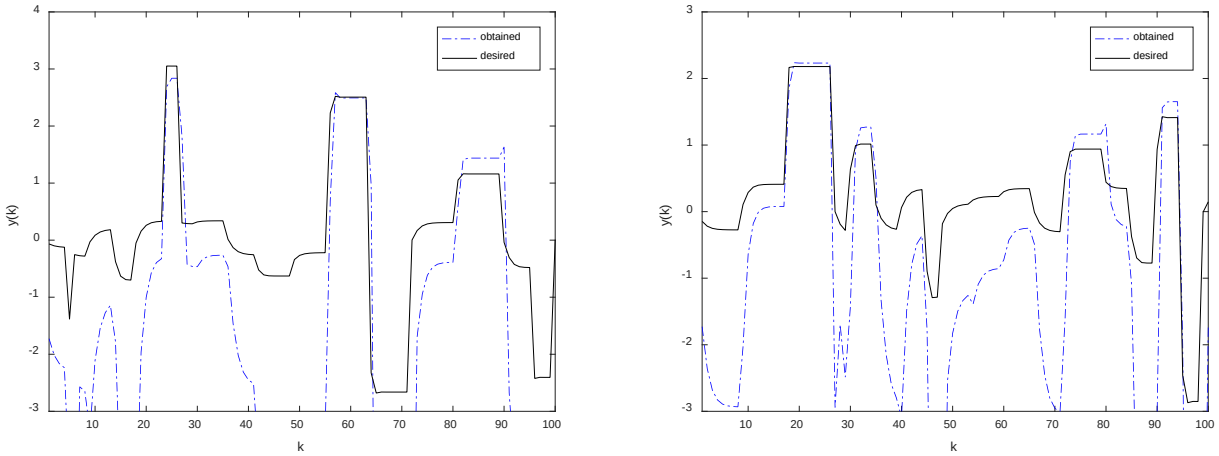


Figure 3: An SLFN model performance for BDS#3 with *Nh*=50 for training (top) and testing (bottom) data sets.

IV. A HEURISTIC APPROACH TO IMPROVE ELM ALGORITHM

In this section, a heuristic approach to solve the problem explained in the previous section is introduced and its results are given. The proposed approach firstly determines the most appropriate number of hidden layer neurons (*Nh*) based on training data set, and then inputs connection weights and biases parameters are chosen among *Trial_Number* candidates to obtain a good generalization performance based on testing data set as seen in Algorithm 1.

The proposed algorithm has been run 100 times for each BDS. Mean, the best and the worst elapsed times of 100 separate runs for each BDS are given in Table 3. Figures 4-6 show the identification performances of SLFN models for each BDS.

Table 3: Elapsed time metrics of BDS identification based on SLFN learned by the proposed approach for 100 runs (*Trial_Number*=50)

#	Mean [s]	Best [s]	Worst [s]	Standard Deviation
BDS#1 *Nh=33*	0.0399	0.0298	0.6974	0.0664
BDS#2 *Nh=74*	0.0419	0.0381	0.0627	0.0031
BDS#3 *Nh=37*	0.0371	0.0320	0.0586	0.0032

Algorithm 1: Pseudo-code of the proposed heuristic approach

Prepare training (n inputs, p outputs, N samples) and testing data sets
Select activation function $\varphi(\cdot)$ {logsig, tansig, cos, sin, etc.)
While *(i≤N)*
  Assign neuron number as *Nh=n*
  Assign randomly weights $W_{n\times Nh}$ and bias $B_{1\times Nh}$ matrixes
  Compute $H_{N\times Nh}$ matrix from Eq.(4)
  Determine an estimation of $\vartheta_{Nh\times p}$ matrix from Eq.(8)
  Find actual output using $Ya=H\hat{\vartheta}$
  Compute RMSE value based on e=Yd-Ya
    Cost (i)=RMSE value
End while
Determine optimal Nh_opt number as follows
  [value, index] =min(Cost)
  Nh_opt=index
For *j=1 to Trial_Number*
  Assign neuron number as *Nh=Nh_opt*
  Assign randomly weights $W_{n\times Nh}$ and bias $B_{1\times Nh}$ matrixes
  Compute $H_{N\times Nh}$ matrix from Eq.(4)
  Determine $\hat{\vartheta}$ matrix from Eq.(8)
  Compute Htest $_{N\times Nh}$ matrix using Eq.(4) using testing data
  Find actual testing output of candidate SLFN model
    $Ya\_test=Htest\,\hat{\vartheta}$
  Compute RMSE value based on e=Yd_test-Ya_test
    Cost_test=RMSE value
  If *j=1* ***then*** *Cost_opt=Cost_test, Wopt=W, Bopt=B,*
    ϑ_opt=ϑ;
  Else if *Cost_test<Cost_opt*
  then update *Cost_op,Wopt, Bopt, Beta_opt*
  End for
Result: *SLFN model with Wopt, Bopt, Beta_opt*

V. CONCLUSION AND REMARKS

In this study, firstly ELM algorithm is introduced and then its drawbacks are uncovered. This arises from randomly chosen the number of neurons in hidden layer, input connection weights and biases of the neurons. And lastly, to eliminate this disadvantage of the standard ELM algorithm, a heuristic approach is proposed, and its merits have been highlighted.

As can be seen from Table 2 and Table 3, the proposed approach is naturally slower roughly 80 times than that of ELM algorithm in terms of elapsed time metrics. But, as can be clearly seen from Tables 4 and 5, a roughly 27 times and 73 times better system identification performance was obtained for training and test sets respectively. This performance improvement especially for the test set is worth the expense in terms of elapsed time metrics. As a result, it has been shown that the proposed approach allows modeling a dynamic system with much better generalization capability.

Table 4: Performance metrics of SLFN model for BDS#1 learned by ELM for 100 runs (*Nh* is randomly chosen for each run).

	Average RMSE	Best RMSE	Worst RMSE	Standard Deviation
Training Set	1.9957	1.0384×10^{-04}	79.9528	8.6079
Testing Set	2.8080	0.0607	101.266	11.9375

Table 5: Performance metrics of SLFN model for BDS#1 learned by the proposed approach for 100 runs (*Nh=33*).

	Average RMSE	Best RMSE	Worst RMSE	Standard Deviation
Training Set	7.2466×10^{-05}	3.2101×10^{-05}	1.7692×10^{-04}	2.5818×10^{-05}
Testing Set	0.0383	0.0173	0.0589	0.0087

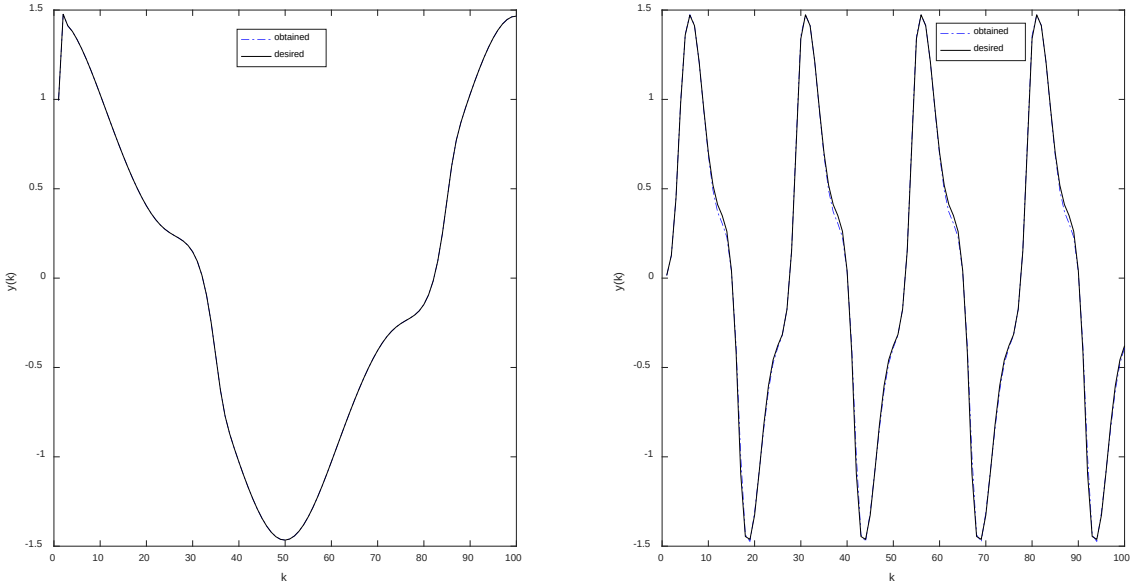


Figure 4: Performance of an SLFN model for BDS#1: training (left), testing (right).

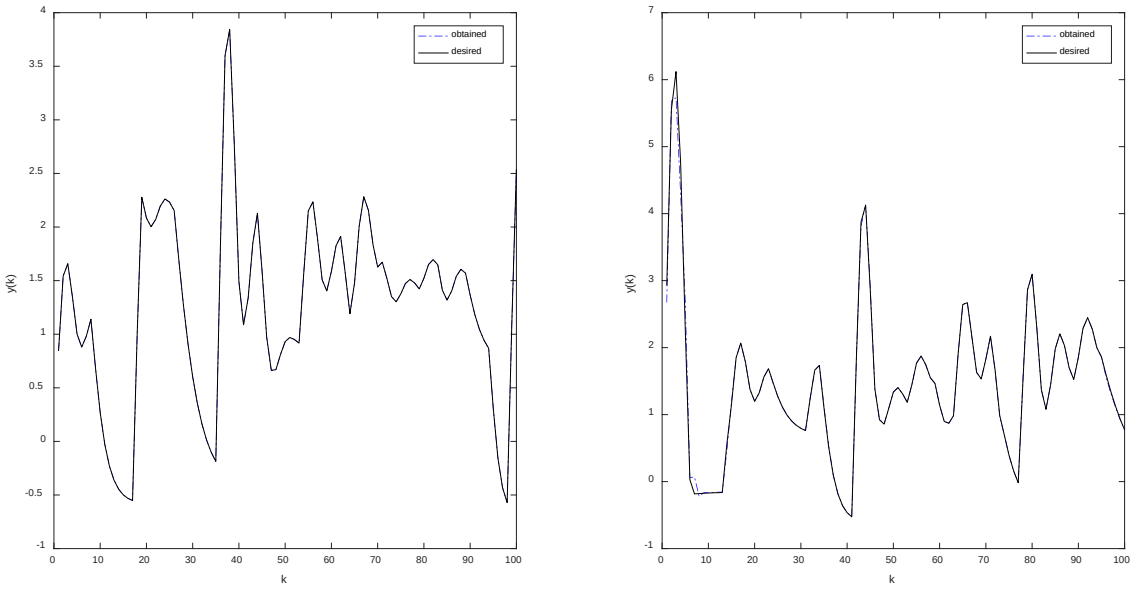


Figure 5: Performance of an SLFN model for BDS#2: training (left), testing (right).

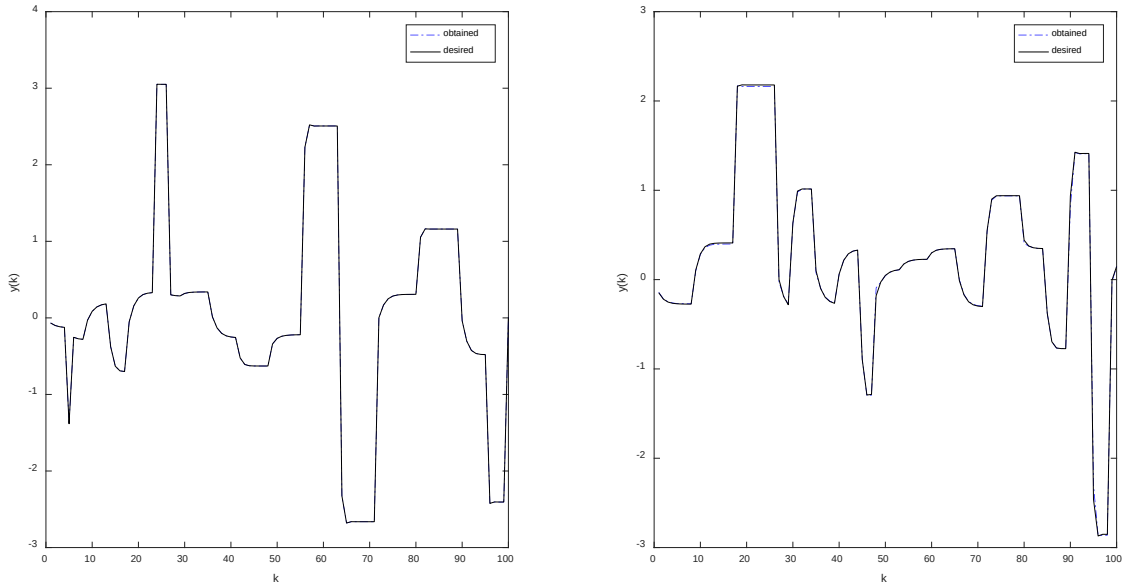


Figure 6: Performance of an SLFN model for BDS#3: training (left), testing (right).

REFERENCES

[1] M. A. Çavuşlu, C. Karakuzu, S. Şahin and M.Yakut, "Neural network training based on FPGA with floating point number format and it's performance," *Neural Computing and Applications*, vol. 20 (2), pp. 195–202, 2011.

[2] M. A. Çavuşlu, C. Karakuzu, S. Şahin and F. Karakaya, "Yapay sinir ağı eğitiminin IEEE 754 kayan noktalı sayı formatı ile FPGA tabanlı gerçeklenmesi," in *Conf. GömSis 2008 Gömülü Sistemler ve Uygulamaları Sempozyumu,* pp.49.

[3] C. Karakuzu, F. Karakaya, M. A. Çavuşlu, "FPGA implementation of neuro-fuzzy system with improved PSO learning" *Neural Networks*, vol. 79, pp. 128-140, 2016.

[4] N. Yalçın, G. Tezel and C. Karakuzu, "Epilepsy diagnosis using artificial neural network learned by PSO," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 23 (2), pp. 421-432, 2015. Available: http://journals.tubitak.gov.tr/elektrik/issues/elk-15-23-2/elk-23-2-7-1212-151.pdf

[5] S. Sahin and M. A. Cavuslu, "FPGA Implementation of Wavelet Neural Network Training with PSO/iPSO," *Journal of Circuits, Systems and Computers*, vol. 27 (6), 15 June 2018.

[6] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN2004)*, Budapest, Hungary, 25–29 July 2004.

[7] G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.

[8] K. S. Narendra, K. Parthasarathy, "Identification and control of dynamical systems using neural networks,". *IEEE T Neural Networks*, vol. 1(1), pp. 4-27, 1990.

[9] Y. Oussar, I. Rivals, and L. Dreyfus, "Training wavelet networks for nonlinear dynamic input output modeling," *Neurocomputing*, vol. 20, pp. 173-188, 1998.

[10] P. S. Sastry, G. Santharam, K. P. Unnikrishnan, "Memory neuron networks for identification and control of dynamical systems," *IEEE T Neural Networks*, vol. 5(2), pp. 306-319, 1994.

[11] C. Karakuzu, "On the performance of newsworthy meta-heuristic algorithms based on point of view fuzzy modelling,", *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 25, pp. 4706 – 4721, 2017. Available: http://journals.tubitak.gov.tr/elektrik/issues/elk-17-25-6/elk-25-6-22-1705-337.pdf

[12] C. Karakuzu, A. A. Makhaila, "Veriye Dayalı Bulanık Mantık Tabanlı Dinamik Sistem Modelleme," *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, vol. 4 (1), pp. 9-20, 2017. Available: http://dergipark.gov.tr/download/article-file/368181

PI and Self-Tuning PI Controller Design and Comparison for Speed Control of DC Motor

N. G. ADAR[1], M. EROĞLU[2] and R. KOZAN[3]

[1]Bursa Technical University, Bursa/Turkey, gokhan.adar@btu.edu.tr
[2]Sakarya University, Sakarya/Turkey, mustafaeroglu@sakarya.edu.tr
[3]Sakarya University, Sakarya/Turkey, kazan@sakarya.edu.tr

Abstract **- DC motors have a wide range of usage in the industry. So its control is one of the important topics. In literature, there are different control algorithms for the speed control of DC motor. This paper presents a comparative study of PI and Self-Tuning PI controller for the speed control of DC motor. DC motor is modeled and classic PI controller applies for the speed control of DC motor. Pole Placement method is used to get parameters of the PI controller which are K_P and K_I. Fuzzy Logic is used for Self-tuning PI controller design. In this controller, K_P and K_I controller gains are adjustable parameters and are updated depending on the speed error and change of error. Simulations of these two controllers are performed in the Matlab/Simulink. PI and Self-tuning PI controller are compared and results are given in graphs. The simulation results show that the Self-tuning PI controller has better efficiency than the classic PI controller.**

Keywords **- Dc motor, PI, Self-tuning PI, Fuzzy Logic.**

I. INTRODUCTION

DC motors have a wide range of applications due to their simple and easy controllability. Therefore, control of DC motors is an important issue. It is possible to control the position, speed etc. of the DC motors. In the literature there are different control algorithms for controlling DC motors. The most common of these is the classical PID control. In the PID, the error is obtained by taking the difference between the reference and the actual value then the control signal is obtained by using this error value. There are different methods to calculate PID control gains. These control gains are constant and do not be changeable. In the literature there are different methods to make these controller gains changeable.

Abdulrahman A.A. Emhemed et al. investigated how the different controllers are effective in controlling the dc motor in a simulation environment. They used PI and artificial intelligence algorithm in their studies. Artificial intelligence is Fuzzy Logic Control (FLC). The result of the study also compares PI and Fuzzy-PI controller. Fuzzy-PI controller is superior to all dynamic responses [1]. Ebrahim Abd al-Hamid Mohamed Ramadan et al. used FLC and PI control for dc motor control. Speed control according to the load change of the DC motor is done in real time. Performance comparison of both controllers was performed and it was determined that FLC is better than PI control [2]. Neeraj et al. worked on dc motor speed control with different controls such as PI, PID and fuzzy using matlab simulation. Fuzzy was used to obtain the direct control using the derivative of the error and error. As a result of the study, fuzzy control gave much better dynamic responses [3].

Akbari-Hasanjani used PID control for dc motor control, where he worked on the overshoot and settling time of K_p, K_i and K_d values. In addition, using the self-tuning fuzzy controller, they apply the control by taking the derivative of the error and the integral of the error. According to these values they have determined the amount of overshoot and settle time [4]. P. Tripura et al. used two different controllers to reduce the settling time and the amount of overrun of the dc motor. These are PI controller and ANFIS (Adaptive Neuro Fuzzy Inference System) controller. In their work he compared the answers given by the controllers in different situations. The ANFIS controller has shown that gives better results at 150 rad/sec in loaded and unloaded conditions of the dc motor [5].

M. Shadkam et al. have suggested Extended Kalman Filter (EKF) controller and compares this controller with other controllers which were PID and Fuzzy-PID. Simulation results obtained for both unloaded and load situations and EKF Fuzzy PID controller provided better results [6].

Y. Ma et al. have studied dc motor control with fuzzy PID. They compare conventional PID and Self-tuning fuzzy PID. As a result of the study, the dc motor gave accurate results in self-tuning fuzzy-PID despite the angular speed and load change in speed control [7].

Y.A. Almatheel et al. have used a fuzzy logic controller to control the dc motor. At the same time, they have studied PID control and compared PID with FLC. FLC gave more efficient results than PID [9]. H. Ahmed et al. have used four different controllers, Ziegler-Nichols, hand tuning, matlab simulink and fuzzy logic, for dc motor control and compare them to each other. Fuzzy logic control showed much better results than the other controls [10].

In this study, a simulation of speed control for different reference speeds of the DC motor was performed. PI and Self-tuning PI controller were used as speed controller. Simulation results were obtained for both controllers and their performances were compared.

II. DC MOTOR MODELLING

A simple DC motor use electricity and magnetic field for producing torque which rotate the motor. The electrical circuit of dc motor is shown in Figure 1.

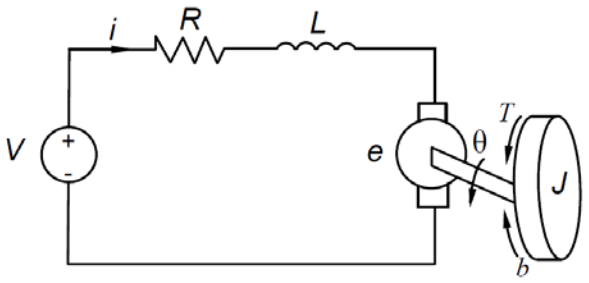


Figure 1: The electrical circuit of dc motor.

Figure shows a voltage source (*V*) across the coil of armature and the electrical equivalent of armature, inductance (*L*), resistance (*R*), back emf (*e*), motor viscous friction constant (*b*), torque (*T*), angular position (θ), angular velocity($\dot{\theta}$).

In this study voltage source (*V*) is the input of the system while rotational speed of the shaft ($\dot{\theta}$) is output.

For constant field current, the torque developed by the motor is

$$T = K_t i \tag{1}$$

where K_t is the motor torque constant and *i* is the armature current.

The back emf is proportional to the angular velocity for a constant flux

$$e = K_e \dot{\theta} \tag{2}$$

In SI units, the motor torque and back emf constants are equal, that is, *Kt = Ke*. So *K* represent for both of them.

Using Newton's second law and Kirchhoff's voltage law, following equations can be written

$$J\ddot{\theta} + b\dot{\theta} = Ki \tag{3}$$

$$L\frac{di}{dt} + Ri = V - K\dot{\theta} \tag{4}$$

If Laplace Transform is applied to the above equations, the following equations are obtained:

$$s(Js+b)\theta(s) = KI(s) \tag{5}$$

$$(Ls+R)I(s) = V(s) - Ks\theta(s) \tag{6}$$

The transfer function of a DC motor which input is voltage source while output is rotational speed can be developed by using equation 5 and 6.

$$p(s) = \frac{\dot{\theta}(s)}{V(s)} = \frac{K}{(Js+b)(Ls+R)+K} \quad [\frac{rad/\sec}{V}] \tag{7}$$

Pololu dc motor was selected. The parameters of the selected motor were given on the Table 1.

Table 1: The parameter of the motor.

(J)	moment of inertia of the rotor	0.01 kg.m^2
(b)	motor viscous friction constant	0.1 N.m.s
(Ke)	electromotive force constant	0.01 V/rad/sec
(Kt)	motor torque constant	0.01 N.m/Amp
(R)	electric resistance	1 Ohm
(L)	electric inductance	0.5 H

Using these parameters, transfer function of selected DC motor was given below:

$$p(s) = \frac{\dot{\theta}(s)}{V(s)} = \frac{0.01}{(0.01s+0.1)(0.5s+1)+0.01^2} \tag{8}$$

III. CONTROLLER DESIGN

A. PI controller

PID is a linear controller and the output of a PID controller, equal to the control input to the plant, in the time-domain is as follows:

$$u(t) = K_p e(t) + K_i \int_0^t e(t)dt + K_d \frac{de(t)}{dt} \tag{9}$$

Where K_p, K_i, and K_d are three parameters of the PID controller.

u(t) is the control signal and *e(t)* is error between the reference input and the process output.

$$error(t) = y_d(t) - y(t) \tag{10}$$

In this study, PI controller was used to control the velocity of DC motor. The expression of PI controller in the time-domain is as follows:

$$u(t) = K_p e(t) + K_i \int_0^t e(t)dt \tag{11}$$

Block diagram of system with PI controller was given in Figure 2.

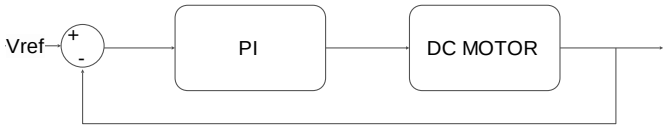


Figure 2: Block diagram of the system for PI.

To make the controller design simplicity, the first-order transfer function of the dc motor is calculated. The maximum speed of the motor is 160 rpm and the time corresponding to 63.2% in unit step response is 0.16 sec. Using these values, the first-order transfer function of the dc motor is as below.

$$p(s) = \frac{\dot{\theta}(s)}{V(s)} = \frac{160}{0.16s + 1} \qquad (12)$$

Pole Placement method is used to get parameters of the PI controller which are K_p and K_i.

Control gains can be chosen to achieve closed-loop pole locations that satisfy system requirements on settling time, peak time, and maximum overshoot.

In this method following expressions are taken to determine the control gains.

t_s=0.6 sec
t_p=0.4 sec
M_p=20%

Using these expressions σ and ω_d are calculated as follow:

$\sigma > 7.7$
$\omega_d > 7.85$

Based on these requirements, the closed-loop poles can be chosen as -9± 8j. Using close-loop poles, control gains can be achieved as follow:

K_p=0.011
K_i=0.14

After fine tuning, control gains are taken as follow:

K_p=0.035
K_i=0.3

B. *Self-Tuning PI Controller*

In this study FLC is used for Self-tuning PI controller design. Fuzzy logic is a branch that is based on thinking like human beings and transforms them into mathematical functions which are fuzzy cluster and sub-cluster.

FLC sets the PI controller parameters which are K_p and K_i gains. Input of the FLC are the speed error and change of error while the outputs are K_p and K_i gains .The block diagram of the whole system is given in Figure 3.

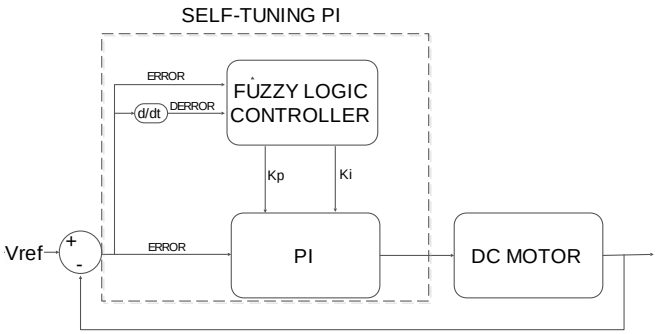


Figure 3: Block diagram of the system for Self-tuning PI.

In FLC the linguistic variables are the speed error and change of speed error, K_p and K_i. The linguistic variables for speed error have linguistic values which are NM (negative medium), NS (negative small), ZE (zero), PS (positive small), PM (positive medium), PL(positive large). The linguistic variables for change of speed error have linguistic values which are NL (negative large), NS (negative small), ZE (zero), PS (positive small), and PL (positive large). The linguistic variables for K_p and K_i have linguistic values which are S (Small), M (middle), L (Large), XL (X Large), XXL (XX Large). Each linguistic value is assigned a triangular membership function. The triangular membership functions are used for their simplicity.

The membership functions of error and change of error for K_p and error and change of error for K_i are shown in Figure 4-5-6-7 respectively.

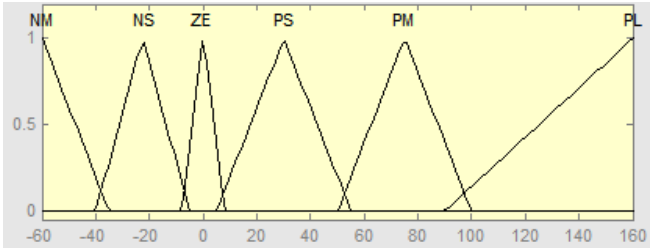


Figure 4: Membership function of error (K_p).

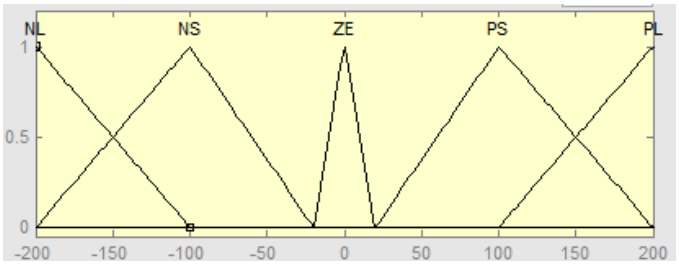


Figure 5: Membership function of change of error (K_p).

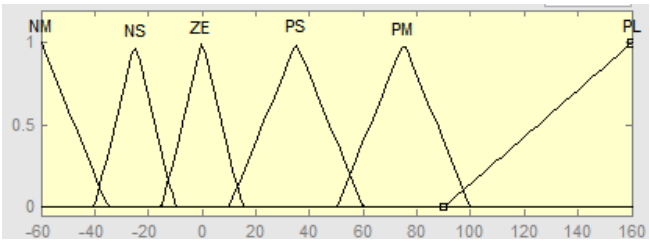


Figure 6: Membership function of error (K_i).

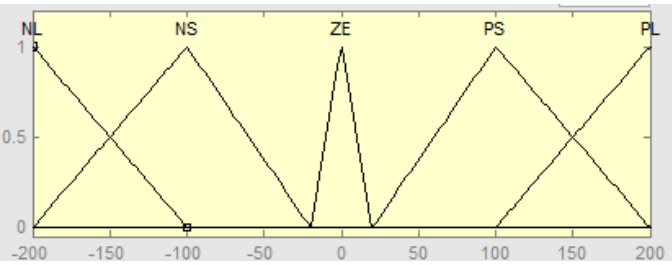


Figure 7: Membership function of change of error (K_i).

The membership functions of K_p and K_i are given in Figure 8-9 respectively.

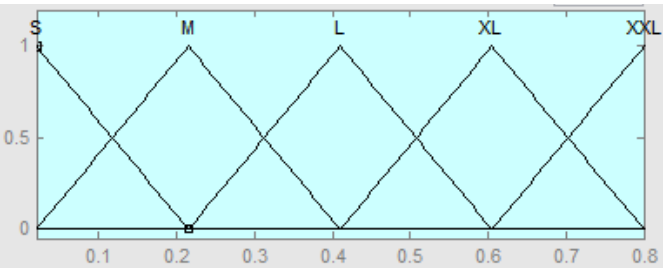


Figure 8: Membership function of K_i.

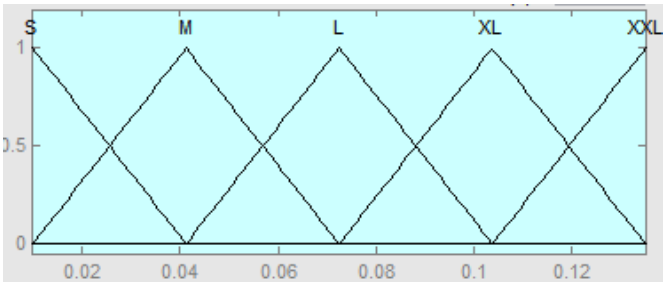


Figure 9: Membership function of K_p.

The maximum speed of the selected DC motor is 160 rpm. It is not desirable for the DC motor to reverse and not overshoot. Considering these, the error is selected from the range of -60 to +160. because the change of the error would be higher, it was chosen the range -200 to + 200.

The K_p value is selected in the range 0.01 to 0.135 while K_i is selected in the range 0.02 to 0.8.

Linguistic variables for change of speed error have been assigned five linguistic values while error assigned six linguistic values. So there are 30 total rules for K_p and K_i. The rule base for the FLC proposed in this study is listed in Table 2-3.

Table 2: Fuzzy sets of the K_p.

K_p		e					
		NM	NS	ZE	PS	PM	PL
de	NL	XL	S	XXL	L	XXL	XXL
	NS	XL	S	XL	M	XL	XXL
	ZE	XL	M	L	S	L	XXL
	PS	L	S	XL	M	XL	XXL
	PL	L	S	XXL	L	XXL	XXL

Table 3: Fuzzy sets of the K_i.

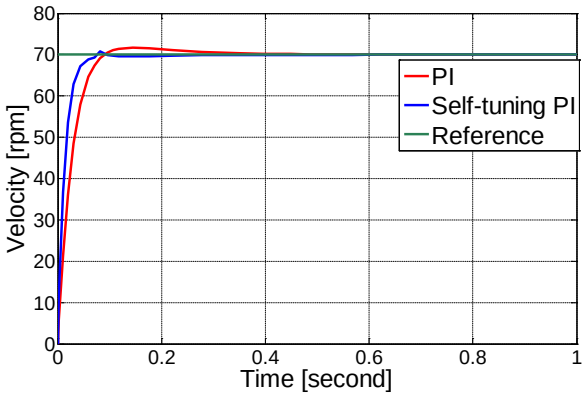
K_i		e					
		NM	NS	ZE	PS	PM	PL
de	NL	L	XL	XXL	L	XXL	XXL
	NS	L	L	XL	M	XL	XXL
	ZE	L	M	L	S	L	XL
	PS	L	S	XL	M	XL	XXL
	PL	L	S	XXL	L	XXL	XXL

IV. SIMULATIONS AND RESULTS

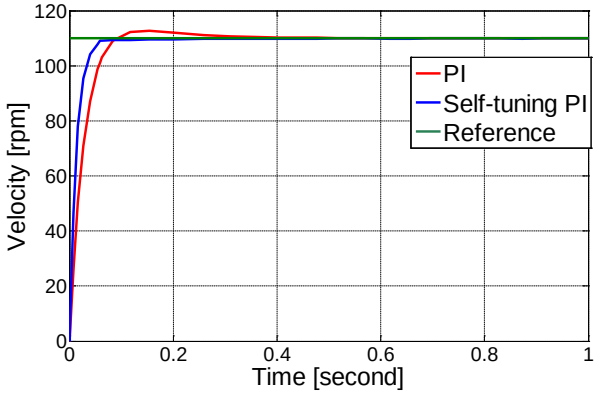
The mathematical modeling of the DC motor was obtained in Section II and PI controller's parameters in Section III Using these values model was developed in Matlab/Simulink. FLC proposed in this study was detailed in Section III. FLC was also built by using Fuzzy Logic Toolbox in Matlab.

First PI and Self-tuning PI controller were compared in different reference speed then Sine wave and Sawtooth wave forms were applied to the system and results were given in figures.

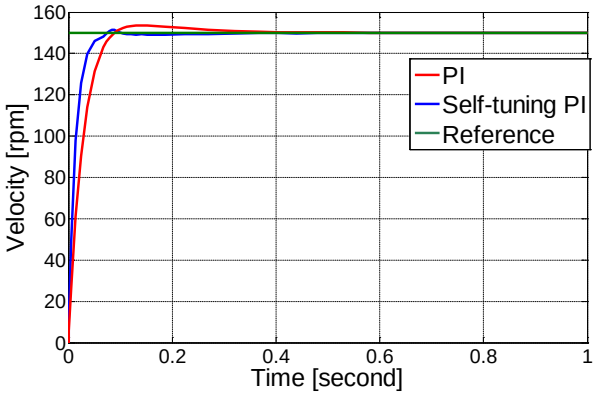
For the reference speed 70-110-150 rpm, comparative simulation results of PI and Self-tuning PI controller are given in Figure 10.



(a)



(b)



(c)

Figure 10: Simulation results of PI and Self-tuning PI controller (a) 70 rpm (b) 110 rpm (c) 150 rpm.

For the Sine wave reference speed, the simulation result was given in Figure 11.

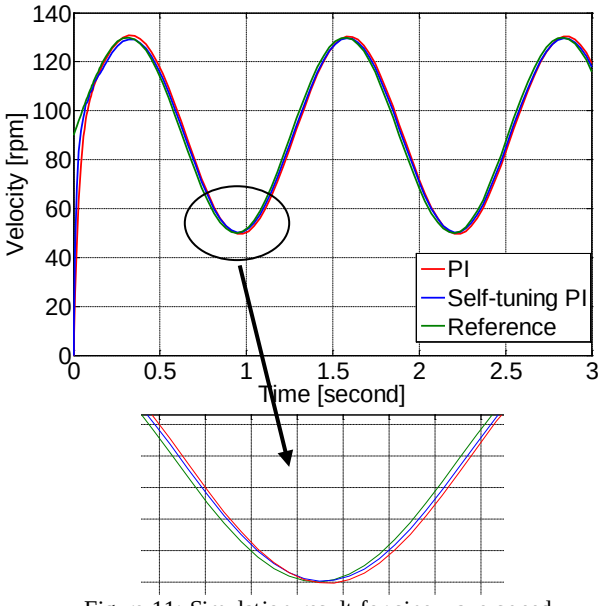


Figure 11: Simulation result for sine wave speed.

For the Sawtooth wave reference speed, the simulation result was given in Figure 12.

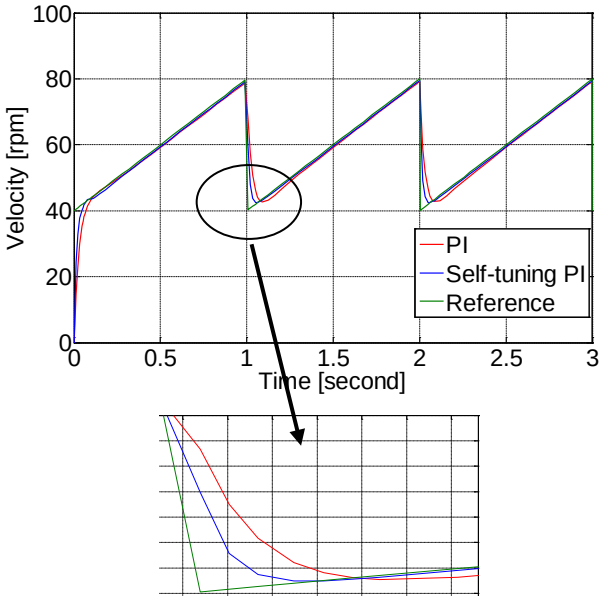


Figure 12: Simulation result for sawtooth wave speed.

For the performance indices Root-mean-square error (RMSE) was selected. The values of the RMS related to the controllers are given in Table 4.

Table 4: The values of the RMS.

Velocity (rpm)	PI	Self-tuning PI
70 rpm	29.6880	27.9982
110 rpm	48.3114	45.6078
150 rpm	61.2537	57.9195
Sine wave	26.2741	24.5833
Sawtooth wave	12.3865	11.0965

V. CONCLUSION

In this study, PI and Self-tuning PI controllers were simulated and compared for DC motor speed control. In PI controller, control gains were constant and do not be changeable. FLC is used for Self-tuning PI controller. Using FLC PI controller's gains (K_p and K_i) were made changeable. Thus, the disadvantage of the PI controller was eliminated. In constant speed simulations, Self-tuning PI controller's settling time was shorter and there was no overshoot. In Sine wave and Sawtooth wave reference speed, Self-tuning PI was more close to the reference than PI. The values of the RMSE for the Self-tuning PI controller were lower than PI controllers. As a result the Self-tuning PI had better dynamic compared to the conventional PI controller.

REFERENCES

[1] A. A. A. Emhemed, and R. B. Mamat, "Modelling and simulation for industrial DC motor using intelligent control," *Procedia Engineering*, 41 420-425, 2012.

[2] E. A. E. H. M. Ramadan, M. El-Bardini, N. M. El-Rabaie, and M. A. Fkirin, "Embedded system based on a real time fuzzy motor speed controller," *Ain Shams Engineering Journal*, 5.2 399-409, 2014.

[3] N. Kumar, H. Gupta, and R. Choudhary, "Analysis Fuzzy Self Tuning of PID Controller for DC Motor Drive," *IJITKM Special Issue*, pp. 148-152 (ISSN 0973-4414) May 2014.

[4] R. A. Hasanjani, S. Javadi, and R. S. Nadooshan. "DC motor speed control by self-tuning fuzzy PID algorithm," *Transactions of the Institute of Measurement and Control*, 37.2 164-176 2015.

[5] P. Tripura, and Y. S. K. Babu, "Intelligent speed control of DC motor using ANFIS," *Journal of Intelligent & Fuzzy Systems*, 26.1 223-227 2014.

[6] M. Shadkam, H. Mojallali, and Y. Bostani, "Speed Control of DC Motor Using Extended Kalman Filter Based Fuzzy PID," *International Journal of Information and Electronics Engineering*, 3.1 109, 2013.

[7] Y. Ma, Y. Liu, and C. Wang, "Design of parameters self-tuning fuzzy PID control for DC motor," *Industrial Mechatronics and Automation (ICIMA), 2nd International Conference on*. Vol. 2. IEEE, 2010.

[8] U. K. Bansal, and R. Narvey, "Speed control of DC motor using fuzzy PID controller," *Advance in Electronic and Electric Engineering* 3.9, 1209-1220, 2013.

[9] Y. A. Almatheel, and A. Abdelrahman, "Speed control of DC motor using Fuzzy Logic Controller," *International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE)*,. IEEE, 2017.

[10] H. Ahmed, G. Singh, V. Bhardwaj, S. Saurav, and S. Agarwal, "Controlling of DC Motor using Fuzzy logic controller," *In Conference on advances in communication and control systems*, (pp. 666-670), 2013.

[11] H.B. Kazemian, "Comparative study of a learning fuzzy PID controller and a self-tuning controller," *ISA transactions*, 40.3 245-253, 2001.

[12] M.M.F. Algreer, and Y. R.M. Kuraz, "Design fuzzy self-tuning of PID controller for chopper-fed dc motor drive," *Al-Rafidain Engineering*, 16 54-66, 2008.

[13] S. K. Gupta, and P. Varshney, "Fractional fuzzy pid controller for speed control of dc motor," *2013 Third International Conference on Advances in Computing and Communications (ICACC)*, IEEE, 2013.

[14] J. O. Salim, "Fuzzy based pid controller for speed control of dc motor using labview," *National Institute of Technology Kurukshetra* 10, 2015.

[15] L. A. Zadeh, "Fuzzy Sets," *Information and control*, 8, 338-353, 1965.

Cardiotocography Data Set Classification with Extreme Learning Machine

A. UZUN[1], E. CAPA KIZILTAS[1], E. YILMAZ[1]

[1]Uludağ University, Bursa/Turkey, aysenuruzun@uludag.edu.tr
[1]Uludağ University, Borçelik Celik Sanayii Tic. A.S.; Bursa/Turkey, capa.eda@gmail.com
[1]Uludağ University, Bursa/Turkey, ersen@uludag.edu.tr

***Abstract* – The purpose of the study is to efficient classification of Cardiotocography (CTG) Data Set from UCI Irvine Machine Learning Repository with Extreme Learning Machine (ELM) method. CTG Data Set has 2126 different fetal CTG signal recordings comprised of 23 real features. Data is two target class description that are based on fetal hearth rate and morphology pattern. The classification criteria based on morphology pattern (A-SUSP) is used in this study to serve better decision options to operators. Accuracy of ELM method will be compared with previous works in literature.**

***Keywords* – Cardiotocography, Extreme Learning Machine; Machine Learning, Classification**

I. INTRODUCTION

Fetus distress is common term used generally in third trimester of pregnancy because of oxygen inefficacy. Fetus distress can lead caesarian decision to protect fetus' health from fetal hypoxia and metabolic acidosis[1, 2]. Cardiotocography (CTG) is a vital monitoring tool for gynecologist in early detection of fetus distress. CTG allows graphics recordings of fetus hearth rate and contradiction of the uterus with two transducers [3]. In recordings, accepted baseline fetal heart rate is between 110 beats per minutes to 160 beats per minutes according to experts [4].

Due to emergency in interpretation of CTG recordings for fetus health, CTG data set is widely used in machine learning and classification papers [5]. Most works in literature are aimed to classify "fetus well-being" and used mostly two class (pathologic – healthy) and three class (normal – suspect – pathologic) methodology. In literature review, it is observed that 10 class (FHR) pattern code is preferred rarely due to adversity of classification.

Discriminant Analysis (DA), Decision Tree (DT) and Artificial Neural Network (ANN) is applied to classify CTG data set by Huang et al [6]. In 2012, accuracy results are 82.1 %, 86.36%, 97.78% respectively. In 2012, Sundar et all has applied ANN model to 3 class CTG data set, the accuracy result is 80% [7]. Yılmaz and Kılıkçıer (2013) use Least Square Support Vector Machine (LS- SVM) utilizing a Binary Decision Tree (BDT) for 3-class FHR, accuracy rate is improved to 91.62% [8]. Ocak and Ertunç (2013) has used two class CTG data with Adaptive Neuro Fuzzy Interface System (ANFIS) method, for pathological and normal class prediction accuracies are 96.6% and 97.2% [9]. In 2014, Karabulut et al has compared machine learning methods; Naïve Bayes (NB), Radial Basis Function Network (RBFN), Bayesian Network (BN), SVM, ANN and DT, without ensemble method AdaBoost and with AdaBoost [10]. In studies NB, RBFN, BN, DT methods' accuracy has improved to 87.39%, 87.67%, 92.61% and 95.01% respectively. By using K-Nearest Neighbors (k-NN) and Random Forest (RF), Şahin and Subasi has improved prediction accuracy to 98.4% and 99.18% respectively [4]. The ANN method performance has compared with ELM method by Cömert et al, it is stated that ELM is faster and more accurate than ANN. Classification error is stated as 93.42% for ELM [11]. Arif has used Random Forest classifier for 3 class CTG dataset and improved accuracy to 93.6% [12].

Kamath and Kamat has proposed to apply Random Forest model to 10 class CTG data set. Team has claimed to achieve accuracy over 87% for 600 number of tree in the forest with 5 partitioning [13].

II. FEATURE SELECTION

High dimensional data can be considered not only as detailed information to ease classification of targets but also correlated/redundant features to emphasize specific features and complicate classification. Therefore, analyzer should be aware of correlation between features before applying classification methods to dataset to prevent "curse of dimensionality" [14]. If dependent or irrelevant features exit, convenient subset of features should be selected by feature selection methods against over-fit [3]. In preprocessing stage of this work, dataset has examined through distribution and correlation of features. As a result, two main feature selection method is used: Principle Component Analysis (PCA) and Fisher Score (FS)

A. *Fisher Score (FS)*

Fisher Score is suitable feature selection method to sparse the most discriminative subset of features [15]. FS is a scoring algorithm over features which assigns a score for each feature and picks Z large score for subset [15].

For dataset with K samples, F features and C class label: $(x_k, y_k)_{k=1}^{K}$ for input samples $x_k \in \mathcal{R}^T$ and class target $y_k \in \{1,2,\dots.C\}$; score assignment can be calculated with equation (1).

$$FS(i) = \frac{\sum_{m=1}^{C} n_m (\mu_m^i - \mu_i)^2}{\sum_{m=1}^{C} n_m (\sigma_m^i)^2} \quad (1)$$

for $i = \{1,2,\dots..F\}$
where n_m number of instances in class m,
where μ_i is i[th] feature's mean,

where μ_m^i mean of m[th] class corresponding to i[th] feature
where σ_m^i standard deviation of m[th] class corresponding to i[th] feature [15].

After determining scores for each feature, top Z score can be selected for subset of features.

B. Principle Component Analysis (PCA)

Principle Component Analysis is a dimension-reduction tool that transforms original data into smaller set according to variance of data [16]. Main idea behind the PCA method is eliminating dependent variables with help of eigenvectors and eigenvalues of dataset.
PCA could be applied any dataset in four sequential steps [17]. First step is to obtain dataset with zero mean in each dimension. Each dimension's mean should be subtracted from each dimension of sample $x_{i,j}$ for $i \varepsilon \{1,2,\dots K\}$ and $j \varepsilon \{1,2,\dots . P\}$ where K is the sample number and P is the dimension number of dataset

$$\mu(j) = \frac{1}{K}\sum_{i=1}^{K} x_{i,j} \qquad (2)$$

where $\bar{X_j} = X_j - \mu(j)$

where $X_j = \begin{bmatrix} x_{1,j} \\ \vdots \\ x_{K,j} \end{bmatrix}$

Then covariance matrix of new dataset should be calculated.

$$\bar{\bar{X}} = cov(\bar{X}) \qquad (3)$$

where $\bar{X} = [\overline{X_1} \quad \dots \quad \overline{X_P}]$ and $\bar{\bar{X}}$ is $P \; x \; P$ matrix.
Eigenvalues and orthonormal eigenvectors of covariance matrix should be found.

$$Z = V^{-1}\,\bar{\bar{X}}\,V \qquad (4)$$

where V is eigenvectors and Z is the diagonal matrix of eigenvalues.
In the descending order of eigenvalues, top N eigenvalues with corresponding eigenvector can be chosen and new data set can be obtained.

$$Y = \bar{X}\,\tilde{V} \qquad (5)$$

where $\tilde{V} = \begin{bmatrix} V_1 \\ \vdots \\ V_N \end{bmatrix}$

III. EXTREME LEARNING MACHINE (ELM)

ELM is proposed as a new learning algorithm for Single-hidden Layer Feedforward Neural Network (SLFN) by Huang et al in 2006 [18]. The team has realized that traditional learning algorithms in ANN are slow due to iterative tuning of excessive parameter of network. Their proposed algorithm chooses weights of hidden nodes randomly but determines the output weights analytically [18]. As a result, learning in hidden nodes can be achieved without "iterative tuning". Due to reduce of gradient-descent based calculations and straightforward solution, ELM is a faster and less likely to converge in local minima [19]. Input-output weights and threshold values do not affect the performance of Single-hidden Layer Feedforward Neural Network (SLFN) [18]. Huang have improved analyses of ELM theory in 2014 [20]. He has determined three learning principle:

- In addition to linear activation functions and nonlinear activation functions, discrete and non-differential activation functions are applicable for hidden nodes of ELM [20].
- ELM has tendency for reaching both minimal training error and for minimal norm of output weights to achieve better generalization performance. ELM Learning algorithms provide capability of universal approach [20, 21]
- In SLFN point view, outputs of hidden layer need to be biased. In ELM approach, biasing is not a necessity. Huang (2014) claims that it could result in non-optimal solution [20, 21].

Huang and his team (2006) has summarized ELM as special version of single hidden layer feedforward neural network (SLFN). A typical SLFN scheme is shown in the figure 1.

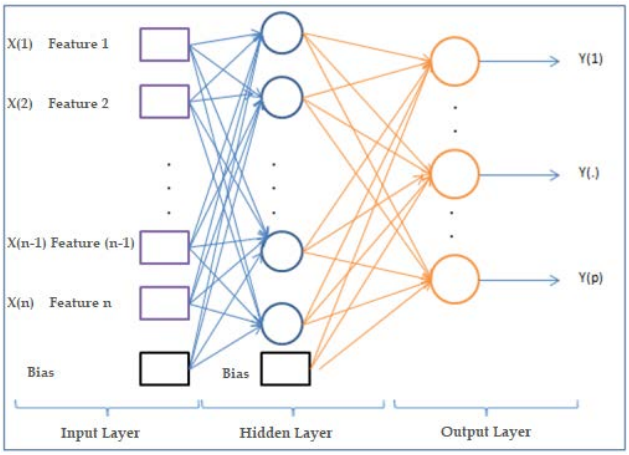


Figure 1: SLFN Model [22].

$X = (X_1, X_2, X_3 \dots . X_n)$, is the input vectors of system with n features
Y output (target) vectors of system.
SLFN's mathematical representation is as below for neuron number in hidden layer M and activation function g(x).

$$f_M(X) = \sum_{i=1}^{M} \beta_i g(W_i X_k + b_i), \quad k = 1,2,3,\dots n \qquad (6)$$

where $W = (W_1, W_2, W_3 \dots . W_n)$ is input layer weights.
where $\beta = (\beta_1, \beta_2, \beta_3 \dots . \beta_n)$ is output layer weights.
where b_i is hidden layer bias value

where g(W,X,b) is activation function.
Table 1 lists the most frequently used activation functions.

Table 1: Activation Functions.

Sigmoid function	$1/(1+e^{-(ax+b)})$
Hyperbolic tangent function	$(1-e^{-(ax+b)})/(1+e^{-(ax+b)})$
Gaussian function	$e^{(-b\|\|x-a\|\|)}$
Multiquadric function	$(\|\|x-a\|\|+b^2)^{1/2}$
Hard limit function	$\begin{cases}1, & if\ ax+b\le 0\\ 0, & otherwise\end{cases}$
Cosine function/Fourier basis	$\cos(ax+b)$

$g(W_iX_k+b_i)=h(x)$ transformation can lead the equation (6) into equation (7).

$$f_M(X)=\sum_{i=1}^{M}\beta_i h_i(x)=H\beta \quad (7)$$

where H matrix is the output of hidden layer.
Error in ELM network can be expressed as equation (8)

$$E=\min\|H\beta-T\|^2 \quad (8)$$

where H is the output of hidden nodes,

$$H=\begin{bmatrix}h(x_1)\\ \vdots\\ h(x_N)\end{bmatrix}=\begin{bmatrix}h_1(x_1) & \dots & h_M(x_1)\\ \vdots & \vdots & \vdots\\ h_1(x_N) & \dots & h_M(x_N)\end{bmatrix}$$

where T is the target matrix.

$$T=\begin{bmatrix}t_1^T\\ \vdots\\ t_N^T\end{bmatrix}=\begin{bmatrix}t_{11} & \dots & t_{m1}\\ \vdots & \vdots & \vdots\\ t_{N1} & \dots & t_{mN}\end{bmatrix}$$

The aim of network is to reduce error minimum, optimally zero. By the way, result of ELM network has become straightforward equation as equation (9).

$$H\beta=T \quad (9)$$

The optimum solution equation can be expressed in equation (10) for output layer weight β^* which minimizes the error of ELM.

$$\beta^*=H^{\dagger}T \quad (10)$$

$H^{\dagger}$, is the "Moore-Penrose" generalized inverse matrix form of H. In the detailed steps for $H^{\dagger}$ matrix calculation can be found [18].
The important part of ELM is to find the optimum solution for β^* output layer weight. This part is similar with least square problem and ridge regression, the more over-regularization the less accurate prediction [21]. The proposed approach for enhancement stability is high quality feature mapping for efficient input weights rather than random selection.
As a summary, For N input ELM with M neuron in hidden layer and g(x) activation function, Huang and team works can be summarized in 3 step [23]:

1. W_i input layer weights and b_i hidden layer neuron bias value is to assign randomly. i={1,..., N}
2. Hidden layer output matrix H is to calculate.
3. Output layer weight β is to calculate for optimal solution.

IV. CARDIOTOCOGRAPHY DATA (CTG)

CTG data set in UCI Irvine Machine Learning Repository is origin from requirement of CTG result conditioning (http://archive.ics.uci.edu/ml/datasets/Cardiotocography) [5]. The data set is comprised from 2126 real cases' recording with 21 feature:

Table 2: Summary of all CTG features used in classification

Symbol	Attribute information
LB	FHR baseline (beats per minute)
AC	# of accelerations per second
FM	# of fetal movements per second
UC	# of uterine contractions per second
DL	# of light decelerations per second
DS	# of severe decelerations per second
DP	# of prolonged decelerations per second
ASTV	Percentage of time with abnormal short-term variability
MSTV	Mean value of short-term variability
ALTV	Percentage of time with abnormal long-term variability
MLTV	Mean value of long-term variability
Width	Width of FHR histogram
Min	Minimum of FHR histogram
Max	Maximum of FHR histogram
Nmax	# of histogram peaks
Nzeros	# of histogram zeros
Mode	Histogram mode
Mean	Histogram mean
Median	Histogram median
Variance	Histogram variance
Tendency	Histogram tendency

In this work the classifications listed below are tried to predicted.

V. PERFORMANCE EVOLUTION

The performance of ELM system stated in this paper is measured with help of cross-validation(CV). Performance criteria are based on four outputs of system; prediction accuracy, confusion matrix, sensitivity and specificity [15].

A. Cross Validation (CV)

Main data set are divided as training and testing parts to properly construct system through training set and measure performance of system with testing set. Randomness of splitting is crucial to evaluate system performance and test system through all consequences. CV is widely used statistical tool for splitting data set. CV let us to rotate specified number of sets as training and testing in order. By the way all samples have chance both to train system and to validate classifiers either [15]. In this work, 10-fold-cross-validation is used. Firstly, CTG data set is divided into 10 random set which includes all classifier equally. Nine sets are used for training and remaining set is used for testing in each iteration. Iteration is repeated 10 times by changing testing set. The overall performance criteria is the average of 10 iterations [5].

B. Prediction Accuracy

Classifier accuracy or prediction accuracy is the strength of system for accurate prediction. Accuracy is the number of correct predictions over the entire data set number.

C. Confusion Matrix

Confusion matrix is a visual tool for distribution of predicted class and actual class. Confusion matrix is square matrix with n rows and n columns which are indexed as actual class of samples and predicted class of samples respectively. In Figure 2 it is shown a basic scheme of confusion matrix for n classifier. Cij represents the number of samples that belongs to class Ai but predicted as class Aj [24]. In this work, the confusion matrix is 10x10 due to 10 classifier problems.

$$\begin{array}{cc} & \text{Predicted} \\ & \begin{array}{ccccccc} & A_1 & \dots & A_j & \dots & A_n \end{array} \\ \text{Actual} & \begin{array}{c} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_n \end{array} & \begin{pmatrix} C_{11} & \dots & C_{1j} & \dots & C_{1n} \\ \vdots & & \vdots & & \\ C_{i1} & \dots & C_{ij} & \dots & C_{in} \\ \vdots & & \vdots & & \\ C_{n1} & & C_{nj} & & C_{nn} \end{pmatrix} \end{array}$$

Figure 2: Confusion Matrix [24].

D. Sensitivity and Specificity

Sensitivity is the positive predictive rate and specificity is negative predictive rate [15]. In multiclass prediction, sensitivity and specificity should be calculated by for each classifier with help of true positive(TP), true negative(TN), false positive(FP) and false negative(FN).

TP is true predicted classifier that belongs to class X and predicted as class X.

TN is number of true prediction of non-X classes that classed other than X class.

FP is false prediction of non-X classes that classed as X class.

FN is false prediction of X-class that classed as non-X class.

Then, sensitivity and specificity can be calculated as;

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (1)$$

$$Specificity = \frac{TN}{(TN+FP)} \quad (2)$$

VI. EXPERIMENTAL RESULTS

In this study, the CTG data set in UCI Irvine Machine Learning Repository is used. The dataset contains 21 classifications features according to 10-class morphological pattern and 3-class fetal heart rate. As can be understood from the literature review given in the introduction, researches are usually focused on 2-class or 3-class classification. In this study, 10-class morphological pattern is studied. Table3 shows the distribution of sample numbers in the CTG data by class for morphological pattern classification.

Table 3: Ten class of CTG

Abbreviation	Class Details	No. of Observations
A	calm sleep	384
B	REM sleep	579
C	calm vigilance	53
D	active vigilance	81
SH	shift pattern (A or SUSP with shifts)	72
AD	accelerative/decelerative pattern (stress situation)	332
DE	decelerative pattern (vagal stimulation)	252
LD	largely decelerative pattern	107
FS	flat-sinusoidal pattern (pathological state)	69
SUSP	suspect pattern	197

CTG Data set is processed by ELM machine learning method. In study, wide array of activation function is tried. The most valid 6 types of activation functions, Sigmoid, Sinus, Hardlim, Tribas, Radbas, Tansig; are applied to ELM algorithm because of high accuracies. Moreover, the hidden layer neuron numbers are important for machine learning performance. In this study, 100,500,1000,1500,2000 number of hidden layer neuron is applied to ELM. However, the sensible accuracies are come across over 1000 neurons. As a result, over 1000 neurons are focused in this study.

Data set is evaluated by cross validation method. İn this stage, ten-fold, seven-fold and five-fold cross validation performance analysis are used separately. In all analysis, folds have forced to contain equal size of examples from all 10 types classification. However, except from ten-fold validation

performance, the accuracy values are decreased to 73% drastically. As a result, in this result chapter, we have focused only in 10-fold cross validation.

As mentioned in Feature Selection section, PCA and FS methods are used to reduce the size of the feature set from CTG data.
In PCA analysis the total variance of features is examined. In Figure 3, the total variance / feature output of CTG data is given. For CTG data, feature 13 showed a total variance of 100. In this study, 14,11,7 & 5 features are selected to examine, however; 5 features results are not satisfactory. As a result, only PCA-144,11,7 features' results are represented in this paper.

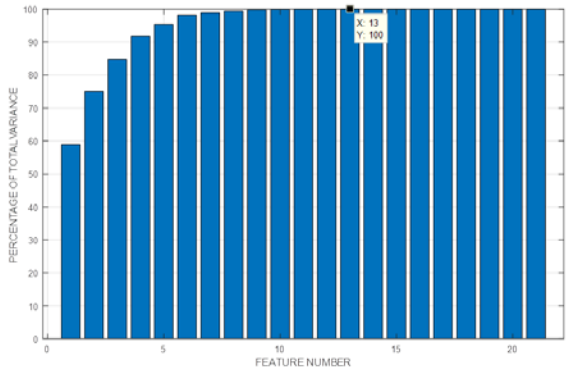


Figure 3: PCA output

Table 4: ELM results by PCA 14-11-7 method

Act. Func.	Hid. Num	PCA 14		PCA 11		PCA 7	
		Train	Test	Train	Test	Train	Test
Sig	1000	%99,66	%80,10	%99,69	%83,00	%99,58	%80,13
Sin		%99,28	%68,34	%99,15	%66,07	%99,23	%68,98
Hardlim		%99,71	%81,28	%99,85	%80,29	%99,74	%79,29
Tribas		%98,97	%69,00	%98,92	%69,08	%98,82	%71,58
RadBas		%99,29	%71,12	%99,31	%72,10	%99,32	%71,28
Tansig		%99,63	%81,38	%99,78	%81,43	%99,69	%80,89
Sig	1500	%99,95	%78,33	%100	%77,46	%100	%75,72
Sin		%99,95	%68,54	%100	%66,59	%100	%69,46
Hardlim		%99,95	%77,19	%100	%75,34	%100	%77,49
Tribas		%99,95	%69,98	%100	%69,42	%100	%70,69
RadBas		%99,95	%72,39	%100	%70,86	%100	%71,94
Tansig		%99,95	%78,17	%100	%77,69	%100	%77,75
Sig	2000	%99,95	%82,97	%100	%83,25	%100	%82,21
Sin		%99,95	%70,05	%100	%67,49	%100	%70,21
Hardlim		%99,95	**%84,29**	%100	%83,38	%100	%82,91
Tribas		%99,95	%73,87	%100	%70,69	%100	%73,79
RadBas		%99,95	%74,26	%100	%74,79	%100	%74,86
Tansig		%99,95	%83,44	%100	**%84,15**	%100	**%83,61**
Max. Accuracy			**%84,29**		**%84,15**		**%83,61**

In Table 4, the results for PCA method are demonstrated. The highest accuracies are found for PCA-14 features with Hard limit activation function (Hardlim) activation function and 2000 hidden layer neuron, PCA-11 features with Tansig activation function and 2000 hidden layer neuron and PCA-7 features with Tansig activation function and 2000 hidden layer neuron in order; 84,3%, 84.1%,83.6%.

In this study, FS score is used to order feature according to importance score. The features scores are 5,2,10,7,18,20, 12, 13, 9, 19, 17, 8, 14, 15, 4, 11, 1, 21, 16, 6, 3 in highest order. FS has used for 14,11,7 & 5 features, however; 7 features accuracy results are under %75. As a result, only FS-14,11,5 features' results are represented in this paper.

Table5 shows the results obtained for FS 14, 11 and 5 according to the ELM activation functions and hidden layer numbers used.

Table 5: ELM results by FS 14-11-7 method

Act. Func.	Hid. Num	FS 14		FS 11		FS 5	
		Train	Test	Train	Test	Train	Test
Sig	1000	%83,78	%71,00	%80,58	%69,15	%82,44	%75,90
Sin		%99,14	%67,44	%98,96	%69,43	%95,84	%75,69
Hardlim		%81,64	%69,46	%77,83	%67,07	%44,20	%40,84
Tribas		%66,12	%50,43	%68,30	%53,65	%56,22	%48,44
RadBas		%72,84	%55,65	%73,68	%58,54	%68,15	%60,17
Tansig		%82,74	%69,24	%79,78	%68,54	%77,28	%70,70
Sig	1500	%90,95	%74,45	%88,28	%72,20	%85,47	%76,62
Sin		%100,00	%67,60	%99,91	%71,40	%95,99	%75,59
Hardlim		%88,88	%73,28	%86,09	%70,54	%46,88	%42,29
Tribas		%77,59	%58,01	%80,73	%61,33	%66,35	%59,22
RadBas		%83,71	%64,30	%83,26	%63,93	%76,75	%68,33
Tansig		%89,56	%73,57	%87,37	%71,79	%83,35	%75,34
Sig	2000	%94,82	%75,21	%93,58	**%76,01**	%87,16	**%77,98**
Sin		%100,00	%67,49	%99,91	%71,87	%95,93	%74,97
Hardlim		%93,42	**%76,00**	%91,86	%73,23	%48,49	%42,15
Tribas		%87,26	%64,09	%88,61	%65,47	%71,73	%63,27
RadBas		%90,71	%68,30	%90,63	%69,05	%81,52	%72,67
Tansig		%94,44	%75,69	%92,98	%72,49	%85,78	%77,28
Max. Accuracy			**%76,00**		**%76,01**		**%77,98**

In Table 5, the results for FS method are demonstrated. The highest accuracies are found for FS-5 features with sigmoid activation function and 2000 hidden layer neuron, FS-11 features with Sigmoid activation function and 2000 hidden layer neuron and FS-14 features with Hardlim activation function and 2000 hidden layer neuron in order; 77,98%, 76,01%, 76,00%.

As mentioned in Performance Evaluation, the confusion matrix and sensitivity, specificity values are mentioned for top 3 accuracy values.sss

Table 6 and 7 indicate confusion matrix and performance values for PCA-14 feature application with Hardlim & 2000 neuron.

Table 6: Confusion Matrix for PCA-14 with Hardlim & 2000 neuron

Class	A	B	C	D	SH	AD	DE	LD	FS	SUSP
A	34	2	0	0	1	0	0	0	0	1
B	3	53	0	0	0	0	1	0	0	1
C	1	0	4	0	0	0	0	0	0	0
D	0	1	0	7	0	0	0	0	0	0
SH	1	0	0	0	4	0	0	0	0	2
AD	1	3	0	0	0	25	1	1	1	1
DE	1	2	0	0	0	2	19	0	0	1
LD	0	0	0	0	0	1	1	9	0	0
FS	0	0	0	0	0	0	0	0	5	2
SUSP	1	0	0	0	0	0	0	0	1	18

Table 7: Sensitivity & Specificity Values for PCA-14 with Hardlim & 2000 neuron

Class	Specificity	Sensitivity
A	0,954	0,895
B	0,918	0,914
C	0,995	0,800
D	0,995	0,875
SH	0,981	**0,571**
AD	0,941	0,758
DE	0,953	0,760
LD	0,985	0,818
FS	0,981	**0,714**
SUSP	0,948	0,900

Table 8 and 9 indicate confusion matrix and performance values for PCA-11 feature application with tansig & 2000 neuron.

Table 8: Confusion Matrix for PCA-11 with Tansig & 2000 neuron

Class	A	B	C	D	SH	AD	DE	LD	FS	SUSP
A	32	3	2	0	0	0	0	0	1	1
B	1	54	0	0	1	0	0	0	1	1
C	1	0	5	0	0	0	0	0	0	0
D	2	0	0	5	0	1	0	0	0	0
SH	1	0	0	0	5	0	0	0	0	1
AD	0	0	0	0	0	32	1	0	0	0
DE	0	1	1	0	0	0	20	0	1	2
LD	0	0	0	0	0	0	0	11	0	0
FS	0	0	0	0	0	0	0	0	5	2
SUSP	0	0	0	0	1	0	0	0	1	18

Table 9: Sensitivity & Specificity Values for PCA-11 with Tansig & 2000 neuron

Class	Specificity	Sensitivity
A	0,971	0,821
B	0,950	0,931
C	0,981	0,833
D	0,986	**0,625**
SH	0,981	**0,714**
AD	0,989	0,970
DE	0,969	0,800
LD	1,000	1,000
FS	0,971	**0,714**
SUSP	0,954	0,900

Table 10 and 11 indicate confusion matrix and performance values for PCA-7 feature application with tansig & 2000 neuron.

Table 10: Confusion Matrix for PCA-7 with Tansig & 2000 neuron

Class	A	B	C	D	SH	AD	DE	LD	FS	SUSP
A	35	3	0	0	0	0	1	0	0	0
B	4	50	0	2	0	1	0	0	0	1
C	0	0	5	0	0	0	0	0	0	0
D	1	0	0	6	0	1	0	0	0	0
SH	0	0	0	0	7	0	0	0	0	0
AD	2	2	0	0	0	27	2	0	0	0
DE	0	1	1	0	0	0	22	1	0	0
LD	0	1	0	0	0	2	0	8	0	0
FS	0	0	0	0	0	0	0	0	6	1
SUSP	3	2	0	0	0	0	1	0	0	14

Table 11: Sensitivity & Specificity Values for PCA-7 with Tansig & 2000 neuron

Class	Specificity	Sensitivity
A	0,943	0,897
B	0,896	0,862
C	0,995	1,000
D	0,981	0,750
SH	1,000	1,000
AD	0,946	0,818
DE	0,963	0,880
LD	0,980	**0,727**
FS	0,995	0,857
SUSP	0,960	**0,700**

VII. DISCUSSION AND CONCLUSION

In the experimental results, the top accuracy result belongs to PCA-14 ELM machine learning algorithm with Hard limit activation function (Hardlim) and 2000 hidden layer neuron number, 84.3%. CTG data set have been used frequently in the literature. The results of the research in the literature with the preferred number of classifications and the machine learning method are mentioned in Table 12.

Table 12: Some CTG Classification Accuracy Results in the Literature

Class.	**Method**	**Accuracy**
2-class	Adaptive Neuro Fuzzy Interface System (ANFIS)	%96.6-%97.2
3-class	Discriminant Analysis(DA)	%82,10
	Decision Tree (DT)	%86,36
	Artificial Neural Network (ANN)	%97,78
	Artificial Neural Network (ANN)	%80,00
	Least Square Support Vector Machine (LS-SVM)	%91,62
	Naive Bayes	%87,39
	Radial Basis Function Network	%87,67
	Bayesian Network	%92,61
	Support Vector Machine	%88,66
	Neural Network	%92,05
	Decision Tree (C4.5)	%95,01
	K-Nearest Neighbors (k-NN)	%98,40
	Random Forest (RF)	%99,18
10-class	Random Forest	%87,00
	Extreme Learning Machine (ELM)	%84,29

As can be seen from Table 12, the accuracy values for the 2-class and the 3-class can reach 99%, while the 10-class has not yet passed this 88%.

The reasons for the lower outcomes for the 10-class could also be a distinctive research topic that needs to be considered. For hypothetical 10-class classification, the reason for the lower results could be inadequacy and inequality of each class samples (for example, 579 from 2126 records belong to class B, whereas the number of records belonging to class C is only 53). From another point of view, it is also possible that the classes are so close to each other that the discriminability is low.

REFERENCES

[1] Z. Alfirevic, *et al.*, "Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour," *Cochrane Database Syst Rev*, vol. 3, 2006.
[2] S. G. Jacob and R. G. Ramani, "Evolving efficient classification rules from cardiotocography data through data mining methods and techniques," *European Journal of Scientific Research*, vol. 78, pp. 468-480, 2012.
[3] Q. Gu, *et al.*, "Generalized fisher score for feature selection," *arXiv preprint arXiv:1202.3725*, 2012.
[4] H. Sahin and A. Subasi, "Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques," *Applied Soft Computing*, vol. 33, pp. 231-238, 2015.
[5] E. Yılmaz, "Fetal State Assessment from Cardiotocogram Data Using Artificial Neural Networks," *Journal of Medical and Biological Engineering*, vol. 36, pp. 820-832, 2016.
[6] M.-L. Huang and Y.-Y. Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network," *Journal of Biomedical Science and Engineering*, vol. 5, p. 526, 2012.
[7] C. Sundar, *et al.*, "Classification of cardiotocogram data using neural network based machine learning technique," *International Journal of Computer Applications*, vol. 47, 2012.
[8] E. Yılmaz and Ç. Kılıkçıer, "Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree," *Computational and mathematical methods in medicine*, vol. 2013, 2013.
[9] H. Ocak and H. M. Ertunc, "Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems," *Neural Computing and Applications*, vol. 23, pp. 1583-1589, 2013.
[10] E. M. Karabulut and T. Ibrikci, "Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach," *Journal of Computer and Communications*, vol. 2, p. 32, 2014.
[11] Z. Cömert, *et al.*, "Cardiotocography signals with artificial neural network and extreme learning machine," in *Signal Processing and Communication Application Conference (SIU), 2016 24th*, 2016, pp. 1493-1496.
[12] M. Arif, "Classification of cardiotocograms using random forest classifier and selection of important features from cardiotocogram signal," *Biomaterials and Biomechanics in Bioengineering*, vol. 2, pp. 173-183, 2015.
[13] R. Kamath and R. Kamat, "Modeling fetal morphologic patterns through cardiotocography data: Decision tree-based approach," *Journal of Pharmacy Research| Vol*, vol. 12, p. 10, 2018.
[14] D. PEHRO and D. Stork, "Pattern classification," *D Wiley-Interscience Publication*, 2001.
[15] E. Yılmaz, "An expert system based on Fisher score and LS-SVM for cardiac arrhythmia diagnosis," *Computational and mathematical methods in medicine*, vol. 2013, 2013.
[16] L. L. C. Kasun, *et al.*, "Dimension reduction with extreme learning machine," *IEEE Transactions on Image Processing*, vol. 25, pp. 3906-3918, 2016.
[17] H. Zhou, *et al.*, "Stacked extreme learning machines," *IEEE transactions on cybernetics*, vol. 45, pp. 2013-2025, 2015.
[18] G.-B. Huang, *et al.*, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489-501, 2006.
[19] G.-B. Huang, *et al.*, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, pp. 513-529, 2012.
[20] G.-B. Huang, "An insight into extreme learning machines: random neurons, random features and kernels," *Cognitive Computation*, vol. 6, pp. 376-390, 2014.
[21] G. Huang, *et al.*, "Trends in extreme learning machines: A review," *Neural Networks*, vol. 61, pp. 32-48, 2015.
[22] Ö. F. Ertuğrul, *et al.*, "EMG signal classification by extreme learning machine," in *Signal Processing and Communications Applications Conference (SIU), 2013 21st*, 2013, pp. 1-4.
[23] G.-B. Huang, *et al.*, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Trans. Neural Networks*, vol. 17, pp. 879-892, 2006.
[24] X. Deng, *et al.*, "An improved method to construct basic probability assignment based on the confusion matrix for classification problem," *Information Sciences*, vol. 340, pp. 250-261, 2016.

Movie Rating Prediction with Machine Learning Algorithms on IMDB Data Set

D. ABİDİN, C. BOSTANCI, A. SİTE

Manisa Celal Bayar University, Manisa/Turkey, didem.abidin@cbu.edu.tr
Manisa Celal Bayar University, Manisa /Turkey, cse.canerbostanci@gmail.com
Manisa Celal Bayar University, Manisa /Turkey, atakansite77@gmail.com

***Abstract* – Predicting movie success with machine learning algorithms has become a very popular research area. There are many algorithms which can be applied on a data set to make movie success prediction if the data set is prepared and represented properly. In this study, we explained how IMDB movie data was used for movie rating prediction. The data set extracted from IMDB was formatted and prepared for datamining algorithms. These algorithms were executed on WEKA application environment and the performances in movie ratings and confusion matrices were obtained. The seven machine learning algorithms used have performed well on the data set with varying performance ratings of 73.5% to 92.7%. Random Forest algorithm had the best performance of 92.7%. This is the highest score obtained among similar studies.**

***Keywords* - Machine learning, WEKA, movie prediction, IMDB.**

I. INTRODUCTION

MOVIE industry has been expanded worldwide and now people, wherever on earth, have chance to watch a movie in the day it is released. There is a huge sector behind the preparation phases of each film and lots of directors and movie stars have burst. Every year, hundreds of films are produced. These movies have different genres, varying from comedy to romance or war to science fiction. To keep track of every movie produced, an online platform was needed. Internet Movie Database (IMDB) is the most popular platform to reach information about a rich collection of movies [1]. IMDB web site contains downloadable raw data about the movies, including data like cast, directors, genres, crew, scriptwriters, summaries, gross and even user ratings. This data is used for data mining on the movies for making prediction on user ratings of the movies. The data used in this study was extracted from Internet Movie Database (IMBD) web site.

There are some studies in literature which also make movie prediction. In some of these studies, similar data mining algorithms were used. Table 1 shows the studies in literature about movie prediction. The datamining algorithms applied in these studies are also given in the table. According to the table, Latif and Afzal [2] have studied the similar datamining algorithms as in this study. They used IMDB data set and obtained a movie prediction performance of 82.42% with J48 algorithm. They also performed a prediction of 79.07% with MLP and 79.52% with PART algorithms.

Butler et al. who worked on movie prediction [3] on IMDB data set have used different algorithms than we used and their performances are not as successful as [2] did.

Among these movie prediction studies, Lee et al. [4] tried many datamining algorithms on IMDB data and have obtained the best movie prediction results with Random Forest algorithm with 86.4% and MLP with 84% of performances.

In another study, Yu used J48 algorithm for movie prediction on IMDB data [5] and got only a performance of 49%.

In the study of Nithin et al., there are no common datamining algorithms used but they also worked on movie success prediction [6]. Their most successful algorithm was linear regression and it had the performance of 50.7% only.

Not only papers are published on movie prediction, but also there are some thesis which makes datamining on movie data from different points of view. For example Tashman have examined the datamining techniques for analyzing film industry success [7]. This study concluded that the former careers of the actors and the actresses have the most predictive information. Person also made movie rating prediction in his bachelor thesis [8] with random forests and support vector machines. In the thesis he collected data from 3three different resources like IMDB, Rotten Tomatoes [9] and MovieLens [10].

Table 1 – Studies in literature.

Reference	Aim	Algorithm/Technique
[Our Study]	Movie prediction	J48, MLP, Random Forest, Bagging, BayesNet, LMT, PART
[2]	Movie prediction	Logic Regression, Simple Logistics, MLP, J48, Naïve Bayes, PART
[3]	Movie prediction	KNN, Decision Trees, Gaussian Naïve Bayes
[4]	Movie prediction	Gradient Tree Boosting, Random Forests, Logistic Regression, Linear Discriminant, MLP, Adaptive Tree Boosting, Supported Vector Classifier
[5]	Movie rating prediction	SMO, Naïve Bayes, Logistic Regression, J48
[6]	Movie success prediction	Linear Regression, Logistic Regression, SVM

In this study, we used seven machine learning algorithms to make a movie rating prediction. We chose the most popular algorithms used in prediction. As the result of our preliminary tests, the most successful seven of them are explained here.

The layout of the paper is as follows: In Section 2, the workflow of the system and the machine learning algorithms we used in the study are explained. In Section 3, the data set and preparation of the data set is introduced. In Section 4, the experimental results are given and Section 5 is the discussion part for this study.

II. MATERIAL AND METHOD

Classification is one of the most common used machine learning techniques. Classification is a supervised method that classifies the items into several groups. The data set is divided into two groups as training set and test set, one for building a learning model on it and the other one to test it. In this study, seven classification algorithms are applied using WEKA (Waikato Environment for Knowledge Analysis) environment [11].

To make a proper analysis on the data set, the data should be cleaned and formatted. This phase is called the "preprocessing". Then, machine learning algorithms will be executed and these algorithms will build a model on the data in the first place. When the algorithms learn from the model built, they will perform a classification for the movie ratings. The workflow about the study is given in Figure 1.

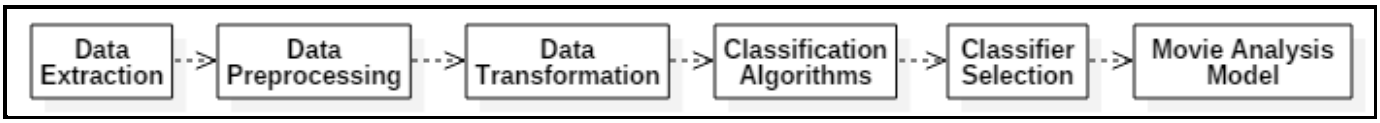


Figure 1: The workflow of this study.

Our workflow completely fits the CRISP-DM (CRoss Industry Standard Process for Data Mining) process model [12] which was stated in 1996. The steps of CRISP-DM for data mining applications are business understanding, data understanding, data preparation, modeling, evaluation and deployment [13]. We apply the data understanding, data preparation, modeling and evaluation steps exactly. The machine learning algorithms used in the study are explained briefly in the following subsections.

A. J48

Quinlan's first algorithm ID3 [14] was improved to C4.5 algorithm and J48 [15] is the open source Java implementation of C4.5 [16]. C4.5 is a classifier which accepts nominal classifiers. C4.5 can use both discrete and continuous attributes and training data with missing attribute values. The additional features of J48 are accounting for missing values, decision trees pruning, continuous attribute value ranges and derivation of rules. The WEKA tool provides a number of options associated with tree pruning. In case of potential over fitting pruning can be used as a tool for précising. In other algorithms the classification is performed recursively till every single leaf is pure, that is the classification of the data should be as perfect as possible [17].

B. MLP

MLP algorithm can be viewed as a logistic regression classifier where the input is first transformed using a learnt non-linear transformation. This transformation projects the input data into a space where it becomes linearly separable. This intermediate layer is referred to as a hidden layer. A single hidden layer is sufficient to make MLPs a universal approximator [18]. The power of the multilayer perceptron comes precisely from non-linear activation functions. Almost any non-linear function can be used for this purpose, except for polynomial functions [19].

C. Random Forest

Random forest is a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [20]. It is a classifier that builds many classification trees as a forest of random decision trees, each constructed using a random subset of the features [21]. Each tree gives a classification (vote) and the algorithm chooses the classification having most votes among all the trees in the forest [22].

D. Bagging

Bagging is an algorithm that is designed to improve the performance of machine learning algorithms. It also helps to reduce overfitting in the algorithm. It is a variance reduction technique for a given base procedure, such as decision tree and to grow each tree a random selection (without replacement) is made from the examples in the training set [23]. Bagging is a smoothing operation which turns out to be advantageous when aiming to improve the predictive performance of regression or classification trees [24].

E. BayesNet

BayesNet belongs to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain; each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. BNs correspond to another GM structure

known as a directed acyclic graph (DAG) [25] that is popular in the statistics, the machine learning and the artificial intelligence societies [26]. BayesNets improves the performance of naive Bayesian classifiers by avoiding unwarranted assumptions about independence [27].

F. Logistic Model Tree (LMT)

LMT basically consists of a standard decision tree structure with logistic regression functions at the leaves, much like a model tree is a regression tree with regression functions at the leaves. As in ordinary decision trees, a test on one of the attributes is associated with every inner node [28]. LMT is a supervised training algorithm that combines logistic regression and decision trees. [29].

G. PART

PART algorithm is used for generating rules in classification rule mining. PART stands for Projective Adaptive Resonance Theory. The input for PART algorithm is the vigilance and distance parameters [30]. It is a separate –and - conquer rule learner. The algorithm producing sets of rules called „decision lists" which are planned set of rules. A new data is compared to each rule in the list in turn, and the item is assigned the class of the first matching rule. PART builds a partial C4.5 decision tree in each iteration and makes the "best" leaf into a rule [31].

III. Data Set Description

A. The Movie Dataset

In this study, we used seven popular machine learning algorithms [32][33] for movie analysis on our data set. This data set was extracted from IMDB web site [1]. The data set given in the website is for about 270000 movies. Among these movies, the ones which have the gross data are used. The gross values were not included to the data set but that information was already given in the web pages of some of the movies. For this reason, a second data extraction was done through the movie web pages. This gross value is one of the most important attributes of the data set. We wrote a script to extract gross from the web site separately and it was merged with the former extracted data set. The data set prepared in three steps and these steps are described in the following subsections.

B. Data Extraction

The movie data set was extracted from IMBD web site, with one important information missing. The missing information was the "*gross*" field, which is already in IMDB web site. To include that column to the data set, a Python script is implemented and the gross data is added to the data set for the corresponding movies as an additional column with a join operation. The raw data set had 86 columns including "*gross*". We extracted 10843 movies in this way.

C. Data Preprocessing

Among the extracted movie data, some movies had same names. To eliminate the duplications, the movies were distinguished with respect to their movie identification numbers (*tconst*). After the elimination of the duplications and removing redundant columns, we had a data set of 6840 movies. Among these, animation movies are excluded from the data set. The remaining data have the type "*movie*" and finally we obtained a data set of 6548 movies

The given information about movies may vary in size of cast or directors. For example there are some movies with six directors, where most of the movies have one or at most two directors. The maximum number of directors in a film determine the number of columns that the directors occupy. Same case counts for the stars and writers of the movies. Including movies with so many directors, stars or writers mean so many null fields for movies with a few directors, stars or writers. For this reason, for the 6548 movies in the final data set, we did not include the columns of directors and stars. Including or excluding director and star data had no effect on the performances of the classification algorithms.

The preprocessed .csv file has 6 columns, which are explained below:

gross: Revenues for movie theaters in USA.

startYear: Release date of the movie in USA.

runtimeMinutes: The duration of the movie.

numVotes: Number of total votes taken from IMBD for a movie.

averageRating: Rating value in IMBD for a movie.

cluster: Rating clusters with values 0-9. This is the only column whose value is calculated by a computer program to obtain the 10 clusters of the movie ratings. The performance of the machine learning algorithms are measured according to this column.

D. Data Transformation

To prepare the data set for WEKA, a normalization step was needed. In this normalization step, the aim was to make all 6 columns' data have values in the same scale. These are the normalization steps:

- Data set was sorted by "gross" column and numbers were assigned to the instances from 1 to 6548 according to this sorting.
- Same sorting and enumerating operations were applied to the data set by "numVotes" column.
- Normalization of the values of an instance are done according to (1).

$$z = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

- Clusters of instances were obtained by a Python script given in Figure 2.

```
for each instance {
        value = (runtimeMinutes * 0.2 + gross * 0.2 + numVotes * 0.6) * averageRating
        for each value {
                normalizedValue = 10 * (value – minValue) / (maxValue – minValue)
        }
        assignCluster();     //cluster values from 0 to 9
}
```

Figure 2: Pseudocode for the data set normalization.

In the machine learning algorithms used, the "numVotes" column has the highest priority. For this reason, the weight of this column in the normalization algorithm has the highest value. The movies released before 2000 may have worse gross values although they have very high number of votes. This is the reason why we keep the weight of "gross" column low.

E. The .csv File

At the end of the data transformation step, the .csv file to be used in WEKA is obtained. Figure 3 shows the final status of the .csv file.

gross	startYear	runtimeMinutes	numVotes	averageRating	cluster
9287	2000	91	110	4	cluster0
1268760	2014	99	42587	7,2	cluster4
850	2001	85	1134	5,8	cluster0
230600	2009	96	1968	6,3	cluster0
71897215	2016	104	230049	7,2	cluster7
81416	2006	82	13425	6,7	cluster2
38176108	1999	97	250725	7,2	cluster6
203373	2011	100	21152	6,1	cluster2
16135	2012	92	3832	7,1	cluster1
7100000	1983	101	5115	6,3	cluster1
94784201	2008	109	115832	5,1	cluster4
59333	2015	113	4481	5,9	cluster1
6248	2012	91	3120	6	cluster0
136189294	1996	103	84853	5,7	cluster5
66957026	2000	100	29433	4,9	cluster3
53481	2006	107	5790	6,4	cluster1
125120	2007	159	12581	7,7	cluster2
12232937	2009	108	24685	5,6	cluster3
45020762	2018	130	8292	7	cluster3

Figure 3 - Definition of attributes in .csv file.

IV. EXPERIMENTAL RESULTS

We analyzed the performance of seven machine learning algorithms using WEKA environment. The classification operation is performed by choosing the "Classify" tab and the proper algorithm under this tab (weka/classifiers/). As the test option, "Cross Validation" with 10 folds is used. With this option, the data sets to be obtained can have different sizes and it does not need to divide the data into two sets as training and test sets. Cross validation option uses the whole data set for learning. To obtain the results, all algorithms were used with their default parameters.

In Table 2, the percentages of correctly classified instances for each algorithm are given. According to the results, Random Forest gives the best classification percentage. Bagging and LMT algorithms have remarkable successful results as Random Forest. Unexpectedly, MLP gives on of the lowest performance values.

Table 2 – Percentages of Correctly Classified Instances for Machine Learning Algorithms

Algorithm	% Correctly Classified
J48	87.9
MLP	76.1
RF	92.7
Bagging	90.2
BayesNet	73.5
LMT	90.1
PART	88

The confusion matrix for the results of Random Forest is given in Figure 4. According to the recall and precision values, rating cluster 0 is the best classified cluster; where the cluster with the movies having highest ratings (cluster 9) has the lowest recall and precision values. In cluster 9, there are less movies, which affects the precision and recall percentages.

	cluster0	cluster4	cluster7	cluster2	cluster6	cluster1	cluster5	cluster3	cluster8	cluster9	Recall %
cluster0	842	0	0	0	0	39	0	0	0	0	95.6
cluster4	0	817	0	0	0	0	32	30	0	0	92.9
cluster7	0	0	339	0	25	0	0	0	2	0	92.6
cluster2	0	0	0	930	0	27	0	45	0	0	92.8
cluster6	0	0	18	0	503	0	34	0	0	0	90.6
cluster1	18	0	0	42	0	953	0	2	0	0	93.9
cluster5	0	42	0	0	22	0	667	0	0	0	91.2
cluster3	0	32	0	43	0	0	0	860	0	0	92.0
cluster8	0	0	17	0	0	0	0	0	134	3	87.0
cluster9	0	0	0	0	0	0	0	0	4	26	86.7
Precision %	97.9	91.7	90.6	91.6	91.5	93.5	91.0	91.8	95.7	89.7	

Figure 4: Confusion matrix for Random Forest algorithm

The misclassified instances tend to be classified in the neighbor clusters of the expected cluster. For example the misclassified instances of cluster 8 are spread to cluster 7 and cluster 9. The number of instances misclassified in cluster 7 (17) is more than the number of instances misclassified in cluster 9 (3).

Kappa, precision and F-measure values for each machine learning algorithm are given in Figure 5.

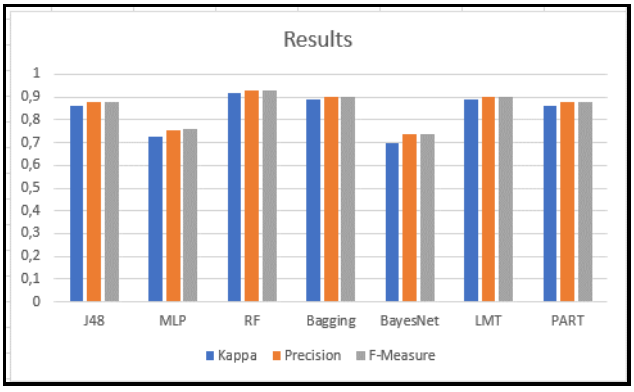


Figure 5: Kappa, precision and F-measure graphics

According to the results, the movies which are expected to be in cluster 5 tend to be misclassified more than the other misclassified movies. The reason for this is the *runtimeMinutes* value for the movies. There are very long movies expected to be classified in cluster 5 but because of the calculation of clusters in our application, they cannot be classified as expected. On the other hand, *runtimeMinutes* column has a remarkable effect on determining the cluster of the movie as happens in "The Green Mile".

MLP is one of the least successful machine learning algorithm, especially in determining the correct clusters of the movies having low rating values. Oppositely, MLP is the most successful algorithm to classify the best movies of cluster 9.

Therefore, these results show that machine learning techniques can be applied to movie score prediction effectively.

V. Discussion and Future Work

In this study, 6548 movies are used as the movie data set to make an analysis with machine learning algorithms and the performances of these algorithms were compared to similar studies. First, data was gathered and preprocessed to obtain a normalized data with no null fields. Later, the data set was used for executing the machine learning algorithms in WEKA. BayesNet was the least successful algorithm among the chosen ones. The misclassified instances are distributed to the neighboring clusters in cluster 5 and cluster 6 mostly. Three of the algorithms (Random Forest, MLP and Bagging) achieved the classification of movie ratings with a performance of over 90%. As the results show, the machine learning algorithms can be used for movie rating prediction.

As future work to this study, we plan to improve the data set by including more columns like genre and MPAA ratings. The Oscar awards will have a certain effect on the new data set to obtain more accurate classification results. The next study will be about exploring the genres of the movies and make a prediction on this column. Another future work will be studying on animation movies.

References

[1] Internet Movie Data Base. URL: https://www.imdb.com/

[2] M.H. Latif, H. Afzal, "Prediction of movies popularity using machine learning techniques", *International Journal of Computer Science and Nework Security*, vol. 16, pp. 127-131, August 2016.

[3] C.D. Butler et. al., "Predicting movie success using machine learning algorithms", Proceedings of The Fifteen Laccei International Multi-Conference For Engineering, Education Technology, Boca Raton, Florida – USA, July 19 - 21 2017.

[4] K. Lee, J. Park, I. Kim, Y. Choi, "Predicting movie success with machine learning techniques: ways to improve accuracy", *Information Systems Frontiers*, 2016. https://doi.org/10.1007/s10796-016-9689-z

[5] T. Yu., "On predicting the movie ratings", Carnegie Mellon University Human-Computer Interaction Institute, 2017.

[6] Nithin V.R., et. al., "Predicting Movie Success Based on IMDB Data", *International Journal of Data Mining and Techniques*, vol.3, pp. 365-368, June 2014.

[7] M. Tashman, "The Association Between Film Industry Success and Prior Career History: A Machine Learning Approach", Master's thesis, Harvard Extension School, 2015.

[8] K. Persson, "Predicting movie ratings: A comparative study on random forests and support vector machines", Bachelor Degree Project in Informatics, University of Skövde, 2015.

[9] Rotten Tomatoes. URL: https://www.rottentomatoes.com/ Date of access March 26, 2018.

[10] MovieLens. URL: https://movielens.org/ https://www.rottentomatoes.com/ Date of access March 26, 2018.

[11] E. Frank, I.H. Witten, *Data Mining*, Morgan Kaufmann Publishers, 2000.

[12] R. Wirth, H., Jochen Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining", (2000).

[13] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining", Journal of Data Warehousing, vol. 5, No:4, pp.13-22, 2000.

[14] J.R. Quinlan, "Induction of Decision Trees", Machine Learning vol.1, pp. 81-106, 1986.

[15] L.H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann Press, San Francisco, USA, 2005.

[16] J.R Quinlan, *C4.5 Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

[17] Kaur G., Chhabra A., "Improved J48 Classification Algorithm for the Prediction of Diabetes", *International Journal of Computer Applications* (0975 –8887), vol. 98, no:22, July 2014.

[18] Deep learning 0.1 Documentation. URL: http://deeplearning.net/tutorial/mlp.html Date of access March 20, 2018

[19] M.-C. Popescu, V.E. Balas, L. Perescu-Popescu, N. Mastorakis, "Multilayer Perceptron and Neural Networks", *WSEAS Transactions on Circuits and Systems*, Issue 7, vol. 8, pp. 579-588, July 2009.

[20] L. Breiman, "Random Forests", *Machine Learning*, pp. 5–32. doi: 10.1023/A:1010933404324, 2001.

[21] T.K. Ho, "Random Decision Forests", Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 278-282, 1995.

[22] F.A. Fontana, M.V. Mäntylä, M. Zanoni, A. Marino, "Comparing and Experimenting Machine Learning Techniques for Code Smell Detection", *Empirical Software Engineering*, vol. 21, pp. 1143-1191, 2016.

[23] L. Breiman, "Bagging Predictors", *Machine Learning*, vol. 24(2), pp. 123–140, 1996.

[24] Bühlmann, P., Yu, B., "Analyzing bagging", *Annals of Statistics*, vol. 30, pp. 927–961, 2002.

[25] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, San Francisco, CA, Morgan Kaufmann, 1988.

[26] I. Ben-Gal, Bayesian Networks, in Ruggeri F., Faltin F. & Kenett R., *Encyclopedia of Statistics in Quality & Reliability*, Wiley & Sons, 2007.

[27] N. Friedman, D. Geiger, M. Goldszmidt, M., "Bayesian network classifiers", *Machine Learning*, vol. 29, pp. 131–163, 1997.

[28] N. Landwehr, M. Hall, E. Frank, *Logistic Model Trees*, Kluwer Academic Publishers, 2006.

[29] N Landwehr, M. Hall, E. Frank, "Logistic Model Trees", *Machine Learning*, 59, 161, 2005.

[30] S.Vijayarani, M.Divya, "An Efficient Algorithm for Classification Rule Hiding", *International Journal of Computer Applications* (0975 – 8887), vol. 33, no.3, pp. 39-45, November 2011.

[31] V.S.Parsania, N.N. Jani, N.H. Bhalodiya, "Applying Naïve bayes, BayesNet, PART, JRip and OneR Algorithms on Hypothyroid Database for Comparative Analysis", *International Journal Of Darshan Institute On Engineering Research & Emerging Technologies*, vol. 3, no. 1, 2014.

[32] X. Wu, et al., "Top 10 Algorithms in Data Mining", *Knowledge and Information Systems*, vol.14, pp. 1-37, 2008.

[33] Meenakshi, Geetika, "Survey on Classification Methods using WEKA", *International Journal of Computer Applications*, vol. 86(18), pp. 16-19, 2014.

An algorithm of finding extreme points of curves

R.NAURYZBAYEVA[1]

Border Service Academy of National Security Committee of
The Republic of Kazakhstan, Almaty/Kazakhstan, rita.dali@mail.ru

Abstract. **General fundamentals of mathematics itself, geometric patterns, and tasks of mechanics field, physics, natural science and technology have the deep roots of developing of the theory of curves. The curves were studied in different ways and methods. At this stage of computerization, you can plot any curve. In this paper, an algorithm for finding the extreme points of a curve in explicit form in both polar and Cartesian coordinates is described.**

Key words: **function, extremum, polar coordinates, curves, algorithm.**

Curves are often found in life and nature, and are used not only in mathematics, physics, astronomy, but also in machinery engineering, in architecture. And their remarkable properties are widely used in mechanisms. The discovery and study of these curves have made a huge contribution to our life and science. For example, an important parameter on which the performance of a construction machine depends is the velocity of body movement. When the worker moves, the organ describes the hypotrochoid, then the velocity for the organ can be calculated by differential equation of the hypotrochoid at time. You can calculate the velocity of any point of the road construction machine. This makes it possible to determine the basic kinematic parameters of the working element and the planetary gears at the design stage of road-building machines [1].

In mathematics after the discovery of variables, the study of functions has a new context. The equations of curves are so complex that the study of each curve turned into a complex scientific work. For example, it took almost half a century to determine the field of curve equation, called the Folium of Descartes. If we look into the history, R. Descartes in 1638, defined the curve by the equation, known as the Folium of Descartes:

$$x^3 + y^3 - 3axy = 0 \cdot$$

But for many years mathematicians could not find the domain of definition. They thought that the curve is located in the first quarter of the Cartesian coordinate system. After that, Roberval strongly believed that that the curve is located in all quarters and had the form of a snowdrop. Only half a century later, in 1692, Huygens and Bernoulli found the domain of definition and called it the Folium of Descartes [2].

The study of curves has various approaches in the history. Besides of the analytical study, there is research, where the projective properties of geometrical figures are used. Desargues, investigating the projective properties of geometrical figures and using the concept of evolution, enriched the theory of the second-order curves with new discoveries. New methods for investigating the properties of second-order curves develop in the 19th century. Criticism of the analytical method for studying the shape and properties of curves was based on the circumstance of such method that could not provide a visual image of this curve and geometrical figures disappear. These views led to the emergence of algebraic geometry on the one hand, the foundations of which were laid by Hesse and Clebsch. The study of the properties of curves was defined by the study of algebraic invariants.

Defining the properties of curves with complex geometries, the elements of the vector functions and topology were used. However, in algebraic geometry, it was not possible to completely abandon the coordinate system as an extraneous element.

Functions are mainly specified in the Cartesian coordinate system. The classical study of functions in a Cartesian coordinate system consists of several algorithms: finding the domain of the definition of functions, the points of intersection of the graph with the coordinate axes, determining the form of the function, finding intervals of monotonicity and extremum points, convexity and concavity of the graph of the function and the inflection point, creating a function graph.

In this coordinate system, it is not always possible to investigate the curve by the indicated algorithm, since their equation cannot always be represented explicitly.

The equation of some curves in the polar coordinate system exposes the function.

This work represents an algorithm of finding the extreme points of a curve in explicit form in both polar and Cartesian coordinates.

Firstly, let us look at the pattern in order to understand the idea of investigating the curves on extremum points in both the polar and the Cartesian coordinate system. For this, we define the types of systems of the polar coordinates.

1. Two types of polar coordinates are specified:

1) Strictly polar coordinates, where, $\rho \geq 0$,

2) The generalized polar coordinate system, where both, $\rho \geq 0$ and $\rho \leq 0$.

For example, the graph of functions $\rho = \sin 2\varphi$ is in the first case a two-petalled rose, and in the second case, a four-petalled rose. Let us look at the polar coordinates of the first type. If the equation of the curve in the polar coordinate

system is $\rho = \rho(\varphi)$, so it can be viewed as a variable function φ, where $0 \le \varphi < \infty$.

Example. Examine the curve and sketch the graph.

$$x^2 + y^2 = x^4 + y^4$$

Solution: Polar equation

$$\rho = \frac{1}{\sqrt{\sin^4\varphi + \cos^4\varphi}}.$$

The finding of the domain of the function should be searched from the solution of non-equation $\rho \ge 0$. The domain of the function $[0, 2\pi]$. In order to find the critical values of the functions, we find the first derivative.

$$\rho' = \frac{4\sin^3\varphi\cos\varphi - 4\cos^3\varphi\sin\varphi}{-2\sqrt{\left(\sin^4\varphi + \cos^4\varphi\right)^3}} = \frac{\sin 2\varphi\cos 2\varphi}{\sqrt{\left(\sin^4\varphi + \cos^4\varphi\right)^3}};$$

Solving the equation $\sin 2\varphi = 0 \cup \cos 2\varphi = 0,$
we can find critical points of the function and appropriate value of the function:

$$\rho = \begin{cases} 1\ , & \text{если } \varphi = \dfrac{\pi(k-1)}{2};\ k = \overline{1,5} \\ \sqrt{2}, & \text{если } \varphi = \dfrac{\pi(2k-1)}{4};\ k = \overline{1,4} \end{cases}$$

Critical values of the argument correspond to the following points in the Cartesian coordinate system: in the vicinity of points $A(1,0), B(0,1), C(-1,0), D(0,-1)$ ρ -function attains its minimum, and un the vicinity of points $E(1,1), F(1,-1),$ $R(-1,-1),$ $Q(-1,1)$ - maximum.

The equation of a curve is an implicit function. If we assume that it's function from *x*, then $y = h = \rho\sin\varphi$ is a coordinate of the points located on the curve, i.e [3].

$$y = \rho\sin\varphi = \frac{\sin\varphi}{\sqrt{\sin^4\varphi + \cos^4\varphi}},$$

We determine

$$y' = \frac{\cos\varphi(1 - 2\sin^4\varphi)}{\sqrt{\left(\sin^4\varphi + \cos^4\varphi\right)^3}}.$$

$$\frac{\cos\varphi(1 - 2\sin^4\varphi)}{\sqrt{\left(\sin^4\varphi + \cos^4\varphi\right)^3}} = 0,$$

$\cos\varphi = 0,$ when $\varphi = k\pi/2,\ k = 1,3;$

$\sin^4\varphi = 1/2,$ $\varphi = (-1)^k \arcsin\dfrac{1}{\sqrt[4]{2}} + k\pi$

$\varphi = (-1)^{k+1}\arcsin\dfrac{1}{\sqrt[4]{2}} + k\pi.$

We find the value of the argument of the extreme points which belong to the domain of determining:

$\varphi_0 = \arcsin\dfrac{1}{\sqrt[4]{2}},$ $\varphi_1 = -\arcsin\dfrac{1}{\sqrt[4]{2}} + \pi,$ $\varphi_3 = -\arcsin\dfrac{1}{\sqrt[4]{2}},$

$\varphi_4 = \arcsin\dfrac{1}{\sqrt[4]{2}} + \pi.$

It can be verified, that the values φ_0, φ_3 correspond to the maximum of the function *y*, and the values φ_2, φ_3 correspond to the minimum of the function. We find the coordinates of the extreme points in the Cartesian system. For example, for: $\varphi_0 = \arcsin\dfrac{1}{\sqrt[4]{2}}$.

$$y(\varphi_0) = \rho\sin\varphi_0 = \frac{\sin\left(\arcsin\dfrac{1}{\sqrt[4]{2}}\right)}{\sqrt{\sin^4\left(\arcsin\dfrac{1}{\sqrt[4]{2}}\right) + \left(1 - \sin^2\left(\arcsin\dfrac{1}{\sqrt[4]{2}}\right)\right)^2}}$$

$$= \frac{\dfrac{1}{\sqrt[4]{2}}}{\sqrt{\dfrac{1}{2} + \left(1 - \dfrac{1}{\sqrt{2}}\right)^2}} = \frac{1}{\sqrt{2}\sqrt{\sqrt{2} - 1}}.$$

$$x(\varphi_0) = \rho\cos\varphi_0 =$$

$$= \frac{\cos\left(\arcsin\dfrac{1}{\sqrt[4]{2}}\right)}{\sqrt{\sin^4\left(\arcsin\dfrac{1}{\sqrt[4]{2}}\right) + \left(1 - \sin^2\left(\arcsin\dfrac{1}{\sqrt[4]{2}}\right)\right)^2}} =$$

$$= \frac{\sqrt{1 - \dfrac{1}{\sqrt{2}}}}{\sqrt{\dfrac{1}{2} + \left(1 - \dfrac{1}{\sqrt{2}}\right)^2}} = \frac{1}{\sqrt{2}}.$$

Thus, we found the coordinates of maximum point $M\left(\dfrac{1}{\sqrt{2}}, \dfrac{1}{\sqrt{2}\sqrt{\sqrt{2} - 1}}\right)$. This point lies on the curve.

Indeed, putting these coordinates into the equation of the curve, we obtain

$$\frac{1}{2} + \frac{1}{2(\sqrt{2} - 1)} = \frac{1}{4} + \frac{1}{4(\sqrt{2} - 1)^2}$$

After the transformation of the left and right sides we have, accordingly:

$$\frac{1}{2} + \frac{1}{2(\sqrt{2} - 1)} = \frac{1 + \sqrt{2} - 1}{2(\sqrt{2} - 1)} = \frac{1}{\sqrt{2}(\sqrt{2} - 1)};$$

$$\frac{1}{4}+\frac{1}{4(\sqrt{2}-1)^2}=\frac{1+2-2\sqrt{2}+1}{4(\sqrt{2}-1)^2}=\frac{1}{\sqrt{2}(\sqrt{2}-1)}.$$

Then, if the equation of the curve is an implicit function, we consider, that x function defined by y, then $x=h=\rho cos\varphi$ is the abcissa of the points located on the curve, i.e.

$$x=\rho cos\varphi=\frac{cos\varphi}{\sqrt{\sin^4\varphi+\cos^4\varphi}},$$

We find

$$x'=\frac{-\sin\varphi(1-2\cos^4\varphi)}{\sqrt{\left(\sin^4\varphi+\cos^4\varphi\right)^3}};$$

$\sin\varphi=0$ in $\varphi=\pi k;\quad k=0,1,3;$

$\cos^4\varphi=1/2$ in: $\varphi_0=\arccos\frac{1}{\sqrt[4]{2}},\ \varphi_1=-\arccos\frac{1}{\sqrt[4]{2}},$

$\varphi_3=\pi-\arcsin\frac{1}{\sqrt[4]{2}},\ \varphi_4=-\pi+\arcsin\frac{1}{\sqrt[4]{2}}.$

It is easy to verify that the values φ_0, φ_3 correspond to maximum, and the values φ_1, φ_4 correspond to minimum of the function x. Similarly, for these values of the argument of the function, you can find the coordinates in the Cartesian system. For example, for: $\varphi_0=\arcsin\frac{1}{\sqrt[4]{2}}$.

$$y(\varphi_0)=\rho\sin\varphi_0=\frac{\sin\left(\arccos\frac{1}{\sqrt[4]{2}}\right)}{\sqrt{\cos^4\left(\arccos\frac{1}{\sqrt[4]{2}}\right)+\left(1-\cos^2\left(\arccos\frac{1}{\sqrt[4]{2}}\right)\right)^2}}=$$

$$\frac{\sqrt{1-\frac{1}{\sqrt{2}}}}{\sqrt{\frac{1}{2}+\left(1-\frac{1}{\sqrt{2}}\right)^2}}=\frac{1}{\sqrt{2}}$$

$$x(\varphi_0)=\rho\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}}\cos\varphi_0=$$

$$=\frac{\cos\left(\arccos\frac{1}{\sqrt[4]{2}}\right)}{\sqrt{\cos^4\left(\arccos\frac{1}{\sqrt[4]{2}}\right)+\left(1-\cos^2\left(\arccos\frac{1}{\sqrt[4]{2}}\right)\right)^2}}=$$

$$=\frac{\frac{1}{\sqrt[4]{2}}}{\sqrt{\frac{1}{2}+\left(1-\frac{1}{\sqrt{2}}\right)^2}}=\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}}$$

Similarly, finding for the other extreme values φ coordinates in the Cartesian system, we obtain the points of the curve.

$M\left(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}}\right)$, $M_1\left(\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}},\frac{1}{\sqrt{2}}\right)$,

$N\left(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}}\right)$, $N_1\left(\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}},-\frac{1}{\sqrt{2}}\right)$,

$K\left(-\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}},\frac{1}{\sqrt{2}}\right)$, $K_1\left(\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}}\right)$,

$L\left(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}}\right)$, $L_1\left(-\frac{1}{\sqrt{2}\sqrt{\sqrt{2}-1}},-\frac{1}{\sqrt{2}}\right)$,

Representing all the points found on the plane, we obtain a graph of the functions (Fig. 2).

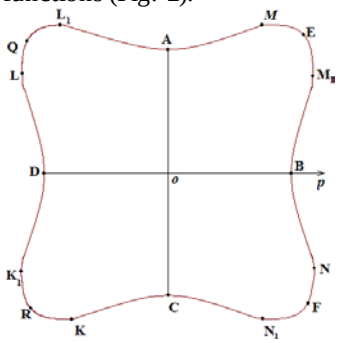


Figure 2: Graph of functions $x^2+y^2=x^4+y^4$

The function given in the polar coordinate system can be investigated for an extremum in both the polar and Cartesian coordinates simultaneously using the following algorithm:

1. Find the critical points of the radius of the function, check a sufficient sign of the extremum.
2. Find the coordinates of the extreme points of the radius of a function in a Cartesian coordinate system.
3. Find the critical points of the function $y=\rho\sin\varphi$ and check a sufficient sign of the extremum.
4. Find the coordinates of the extreme points of a function $y=\rho\sin\varphi$ in a Cartesian coordinate system.
5. Find the critical points of the function $y=\rho cos\varphi$ and check the necessary sign of the extremum.
6. Find the coordinates of the extremal points of a function $x=\rho cos\varphi$ in a Cartesian coordinate system.

Thus, we were convinced that if in the case when the equation of the curve can be represented as a function in a polar coordinate system, then it can be examined by the indicated algorithm.

REFERENCES

[1] S.Lee, B.Shin, Uravnenie *teoreticheskogo kontora rabochego organa mashinys planetarno-rotornom dvizheniem* // Vestnik KazATK, pp. 50-54, 2005.
[2] A. Savelov, *Ploskie krivye. Spravochnoe rukovodstvo.* Moscow: Fiz-mat. Literature, 1960.
[3] R.Nauruzbayeva, *Condition of bulge and concavity of a curve in the arctic system of coordinates in the extreme points of radius of functions* in *Conf. Rec.* 2011 Bonab, pp 239-244.

Choosing optimal truncation level on estimation of the nonparametric regression with Padé approximation

D. AYDIN[1] and E. YILMAZ[1]

[1] Mugla Sitki Kocman University, Mugla/Turkey, duaydin@hotmail.com
[1]Mugla Sitki Kocman University, Mugla/Turkey, yilmazersin13@hotmail.com

Abstract **– In this paper, Padé approximation based on truncated total least squares method (P-TTLS) is defined as a nonparametric method to estimate the nonparametric regression model and to obtain an optimal estimation, determination of the truncation level is illustrated. Here, three selection method are used for choosing the truncation level which are GCV, AICc and REML respectively. To compare and interpreting the performances of the criteria, a simulation study is carried out and results are presented.**

Keywords **–Nonparametric regression, Padé approximation, truncation level.**

I. INTRODUCTION

LET consider the nonparametric regression model will be as follows

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1,2,...,n, \quad a < x_1 < ... < x_n < b \qquad (1)$$

where y_i's are the values of the response variable, $f(.)$ is the smooth function to be estimated, x_i's are the values of explanatory variable and ε_i's are the error terms that have normal distribution with constant variance and zero mean. In this paper, the key idea is estimating the unknown function f in model (1).

In the literature, there are traditional methods to explain the relationship between variables in the model (1) such as kernel smoothing, smoothing splines, penalized splines. These methods are studied by several authors. Some of them could be ordered as; Green and Silverman (1994), Eubank (1999), Nadaraya-Watson (1964) and so on. Our purpose in this paper, to present an alternative method to mentioned methods at above. This method can be introduced as a rational function approximation which called Padé which connected to penalized spline method closely. This method is also studied by many authors such as Baker (1996), Cheney (1996). Also, Petrushev and Papov (1987) studied the relationships between splines and rational functions.

There are some important key points when estimating the nonparametric regression model with Padé approximation. Firstly, Pade-type estimator of $f(x)$ gives some overdetermined linear equation systems to be regularized. Also note that, this disordered linear system involves both response variables and explanatory variables and they are contaminated by error and noise. According to that, in order to obtain the accurate and stable results, we need to regularization method in here. Of course, there are some common regularization methods in literature but they assume that only response variable includes error. But in our method, error messes with both response and explanatory variables. Consequently, we need to overcome these two problems and to do that we used truncated total least squares method (TTLS) which is introduced by Golub and V. Loan (1980). In this paper, we defined Padé-type approximation based on truncated total least squares (P-TTLS) method. Note that, to obtain the P-TTLS solutions accurately, truncation parameter t must be chosen appropriately which also main purpose of this study.

As we mentioned above, in this study we considered the choosing truncation parameter optimally. To realize this purpose we used three selection methods for introduced P-TTLS method. Selection methods that used in this study are; GCV, AICc and REML methods respectively.

II. FUNDAMENTAL IDEAS

In this section, we introduced the P-TTLS method shortly and presented a algorithm for the estimating regression model. Unknown function $f(.)$ can be approximated by a rational function as follows;

$$y_i = f_{[p,q]}(x_i) = \frac{A(x)}{B(x)} = \frac{a_j x^j}{b_k x^k}, \; j = 0,2,...,p; \; k = 1,2,...,q \qquad (2)$$

where $A(x)$ and $B(x)$ are the polynomial functions with degrees p and q respectively. Here, $a_j \; (j = 0,1,...,p)$ and $b_k \; (k = 1,...,q)$ are the coefficients of the polynomial functions. Equation (2) can be rewritten as follows

$$a_0 + a_1 x_i + a_2 x_i^2 + ... + a_p x_i^p - b_1 y_i x_i - b_2 y_i x_i^2 - ... - b_q y_i x_i^q = y_i, \; 1 \le i \le n$$

and mentioned ill-posed and disordered linear system is obtained from this equation. From here, regression model will be

$$y_i = \left(f_{[p,q]}(x_i) = a_0 + a_1 x_i + ... + a_p x_i^p - b_1 y_i x_i - ... - b_q y_i x_i^q \right) + \varepsilon_i, \; 1 \le i \le n \quad (3)$$

And matrix form of the model (3) can be written as follows;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4)$$

where $\mathbf{X}$ is an $n \times (p+q+1)$ data matrix, $\mathbf{y}$ is the response vector and $\boldsymbol{\beta} = (a_j, b_k)'$ is a $(p+q+1)\times 1$ vector of unknown regression coefficients. In this case, our problem here turns into estimating vector of $\boldsymbol{\beta}$ now. In below, the algorithm of the estimating $\boldsymbol{\beta}$ is presented with P-TTLS method.

Algorithm P-TTTLS

Step 1. Calculate the SVD of the augmented matrix $(\mathbf{X}, \mathbf{y})$:

$$(\mathbf{X}, \mathbf{y}) = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}' \text{ with } \sigma_1 \ge \sigma_2 \ge ... \ge \sigma_{m+1} \ge 0$$

Step 2. Choose an appropriate truncation parameter t, $t \le min\left((p+q+1), rank\{(\mathbf{X}, \mathbf{y})\}\right)$

Step 3. Block-partition the $((p+q+1)+1)\times((p+q+1)+1)$ matrix $\mathbf{V}$ such that

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \text{ where } \mathbf{V}_{11} \in R^{(p+q+1)\times t} \text{ and}$$

$$\left\{ \mathbf{V}_{22} \equiv \left[v_{(p+q+1)+1,t+1}, ..., v_{(p+q+1)+1,(p+q+1)+1} \right] \neq 0 \right\} \in R^{1\times((p+q+1)+1-t)}$$

Step 4. Compute the P-TTLS solution as

$$\hat{\boldsymbol{\beta}}_{P\text{-}TTLS} = -\mathbf{V}_{12}(\mathbf{V}_{22})^{+} = -\mathbf{V}_{12}\mathbf{V}_{22}'\|\mathbf{V}_{22}\|_2^{-2} \quad (5)$$

where $(\mathbf{V}_{22})^{+} = \mathbf{V}_{22}'\|\mathbf{V}_{22}\|_2^{-2}$ denotes the pseudoinverse of $\mathbf{V}_{22}$.

Note that, there is a different wat to obtain of the estimation of the $\hat{\boldsymbol{\beta}}_{P\text{-}TTLS}$. Let $\mathbf{P}_t$ be the orthogonal projection onto a t dimensional subspace of $(\mathbf{X}, \mathbf{y})$, given by

$$\mathbf{P}_t = \mathbf{U}_t\mathbf{U}_t' = (\boldsymbol{u}_1, ..., \boldsymbol{u}_t)(\boldsymbol{u}_1, ..., \boldsymbol{u}_t)'$$

In this case, for the reduced rank system $\mathbf{P}_t(\mathbf{X}, \mathbf{y})$ we can obtain a useful alternative P-TTLS solution minimizes the $\left\| [\mathbf{X}, \mathbf{y}] - [\hat{\mathbf{X}}, \hat{\mathbf{y}}] \right\|_F$ such that

$$\hat{\boldsymbol{\beta}}_{P\text{-}TTLS} = (\mathbf{P}_t\mathbf{X})^{+}\mathbf{P}_t\mathbf{y} \quad (6)$$

Hence, the fitted values at training inputs are defined by

$$\hat{\mathbf{y}}_{P\text{-}TTLS} = \mathbf{X}\hat{\boldsymbol{\beta}}_{P\text{-}TTLS} = \mathbf{X}(\mathbf{P}_t\mathbf{X})^{+}\mathbf{P}_t\mathbf{y} = \mathbf{H}_t\mathbf{y} \quad (7)$$

where $\mathbf{H}_t = \mathbf{X}(\mathbf{P}_t\mathbf{X})^{+}\mathbf{P}_t$ is called as hat matrix.

III. SELECTION METHODS FOR THE TRUNCATION PARAMETER

Criteria that we used for selection of truncation parameter are introduced at below shortly.

***GCV Criterion*:** The generalized cross validation (GCV) score can be defined as (see Craven and Wahba, 1979)

$$\mathrm{GCV}(t) = \frac{n^{-1}\left\| (\mathbf{I} - \mathbf{H}_t)\mathbf{y} \right\|^2}{\left[n^{-1} tr(\mathbf{I} - \mathbf{H}_t) \right]^2} = \frac{RSS/n}{\left[1 - tr(\mathbf{H}_t)/n \right]^2}$$

where $\mathbf{H}_t$, as is defined after equation (7), is a hat matrix based on *t*.

***AICc*:** Note that the classic Akaike information criterion tends to overfit when the sample size is relatively small; therefore, Hurvich et al., (1998) suggested an improved version, called AIC_C, which is defined by

$$\mathrm{AIC}_c(t) = 1 + \log\frac{\left\| (\mathbf{H}_t - \mathbf{I})\mathbf{y} \right\|^2}{n} + \frac{2\{tr(\mathbf{H}_t)+1\}}{n - tr(\mathbf{H}_t) - 2} = 1 + \log\frac{RSS}{n} + \frac{2\{tr(\mathbf{H}_t)+1\}}{n - tr(\mathbf{H}_t) - 2}$$

***REML Criterion*:** The restricted maximum likelihood (REML) criterion treats λ as a variance parameter. The REML and GCV have a similar form and provide identical values. Moreover, the derivatives of both the REML and GCV with respect to can be determined naturally in a common form (see Reis et al. (2009)).

The REML score can be specified as

$$\mathrm{REML}(t) = \frac{\left\| (\mathrm{I} - \mathbf{H}_t)\mathbf{y} \right\|^2}{n - tr(\mathbf{H}_t)} = \frac{RSS}{n - tr(\mathbf{H}_t)}$$

In this study, we made a Monte-Carlo simulation study to see how the method works and also comparing the performances of the selection methods on estimating regression model. Empirical results are obtained.

REFERENCES

[1] P.J. Green, and B.W. Silverman, Nonparametric regression and generalized linear model, Chapman & Hall, 1994.

[2] R. Eubank, Nonparametric regression and spline smoothing. Marcel Dekker, New York, 1999.

[3] E.A. Nadaraya, On estimating regression, Theory of Probability and Its Applications, 10 (1964), pp.186-190.

[4] G.S. Watson, Smooth regression analysis. Sankhya, Series A, 26 (1964), pp.359-372.

[5] G. Baker, Essentials of padé approximants (1st Ed.), Academic Press, New York, 1975.

[6] E.W. Cheney, Introduction to approximation theory (2nd Ed.), Chelsea Publishing Company, 1996.

[7] P.P. Petrushev and V.A. Popov, Rational approximation of real functions. Cambridge University Press, New York, 1987.

[8] G.H. Golub and C.F. Van Loan, An analysis of the total least squares problem, SIAM Journal of Numerical Analysis, 17 (1980), pp.883-893.

[9] P. Craven and G. Wahba, Smoothing noisy data with spline functions, Numerische Mathematik, 31 (1979), pp.377-403.

[10] Hurvich, C., M., Simonoff, J. S., and Tasi, C. L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion., J.R. Statist. Soc. B., (1998), Vol. 60 pp.271-293.

[11] Reis, P. T., & Ogden, R. T., Smoothing parameter selection for a class of semiparametric linear models, J. R. Stat. Soc. B 71 (2009), pp. 505–523.

Elongation of Partially-Thermally Insulated on the Lateral Surface of the Rod Under the Influence of Temperature, Heat Flow, Heat Transfer and Tensile Force

K. AMIRTAEV[11]

[1]A. Yesevi International Turkish-Kazakh University, Turkistan/Kazakhstan, kanat.amirtayev@ayu.edu.kz

***Abstract*— In this paper, based on the laws of conservation and change of energy in the combination of the finite element method and minimization of the energy functional, is obtain a linear system of algebraic equations, the solutions of which allow us to find the values of the unknown variables.**

***Keywords* — tensile forces, temperature field, functional, thermo elastic state.**

INTRODUCTION

An investigation of the elongation of rods under simultaneous action of a tensile load and a heat flux, as well as compression of a rod of limited length clamped by two ends under the influence of a temperature field and a heat flux, is of interest for technology and technological processes of the strengths of elements of partially insulated structures of operating steady-state thermal fields.

With simultaneous consideration of the above factors, the analytical solution of the steady thermo elastic compression and stretching of the rods becomes very complicated. In this connection, the need naturally arises to develop appropriate universal computational algorithms oriented to modern computer facilities and object-oriented high-level programming languages that allow one to investigate the thermo elastic state of partially insulated one-dimensional structural elements, in the presence of heat exchange with the environment, the temperature field, thermal and axial tensile forces.

CALCULATION SCHEME OF THE TASK

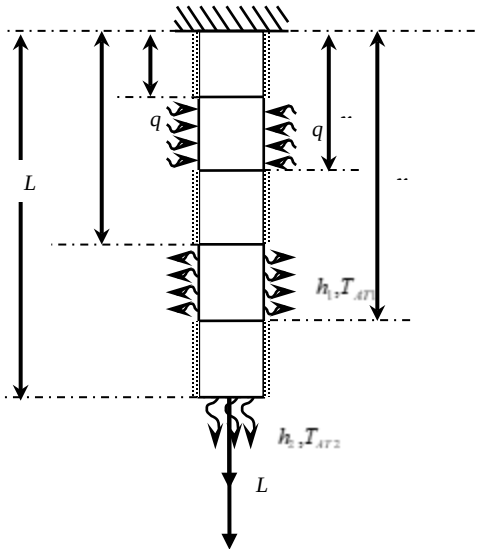


Figure-1 - Calculation scheme of rod extension due to temperature, heat flow and tensile force

It is necessary to investigate the elongations of the rod due to various kinds of thermal sources and tensile forces. Here, the lateral surfaces of the sections $0 \le x \le x_1$ and $x_4 \le x \le L$ the rod are thermally insulated.

It is necessary to find the elongation of the rod, taking into account the field of temperature distribution along the length of the rod and the tensile force L. To do this, the rod under consideration is discredited n into finite elements with three nodes. The length of each finite element will be $l = L/n$. And also within the finite element, that is, in any interval $L \cdot i/n \le x \le L(i+1)/n,\ i = 0 \div (n-1)$, we consider the temperature distribution field as a second-order curve that passes through three points, then in this interval the following expression will be appropriate

$$T(x) = \varphi_i(x)T_i + \varphi_j(x)T_j + \varphi_k(x)T_k, \qquad (1)$$

where $\varphi_i(x), \varphi_j(x), \varphi_k(x)$ - are the functions of the form of the finite element with three nodes. Now for each quadratic finite element we write the expression of thermal energies [1]. Due to the fact that the lateral surface of the sections $(0 \le x \le x_1), (x_2 \le x \le x_3)$ and $(x_4 \le x \le L)$ the rod is thermally insulated, the form of the corresponding functional for all elements of these sections will be as follows

$$I_i = \int_{V^{(i)}} \frac{K_{xx}}{2}\left(\frac{\partial T}{\partial x}\right)^2 dV,\ \left(i = 1 \div \frac{x_i}{l}\right),\ i = \left(\frac{x_2}{l} \div \frac{x_3}{l}\right),\ i = \left(\frac{x_4}{l} \div \frac{L}{l} - 1\right), \qquad (2)$$

here $V^{(i)}$ - are the volumes of finite elements.

The expression $\left(\frac{\partial T}{\partial x}\right)^2$ is found from (1), that is,

$$\left(\frac{\partial T}{\partial x}\right)^2 = \left[\frac{\partial \varphi_i}{\partial x}T_i + \frac{\partial \varphi_j}{\partial x}T_j + \frac{\partial \varphi_k}{\partial x}T_k\right]^2 = \left[\frac{4x-3l}{l^2}T_i + \frac{4l-8x}{l^2}T_j + \frac{4x-l}{l^2}T_k\right]^2 = \qquad (3)$$
$$= \frac{1}{l^4}\left[(16x^2 - 24lx + 9l^2)T_i^2 + 8(10lx - 8x^2 - 3l^2)T_iT_j + 2(16x^2 - 16lx + 3l^2)T_iT_k + \right.$$
$$\left. + 8(6lx - l^2 - 8x^2)T_jT_k + 16(l^2 - 4lx + 4x^2)T_j^2 + (16x^2 - 8lx + l^2)T_k^2\right]$$

Due to the fact that the heat flux is fed to the side surface $x_1 \le x \le x_2$ of the rod section, the expression of the corresponding functional for all finite elements in this section will be as follows

$$I_i = \int_{V(i)} \frac{K_{xx}}{2}\left(\frac{\partial T}{\partial x}\right)^2 dV + \int_{S_{n\delta n}^{(i)}} qT dS,\ \left(i = \frac{x_1}{l} \div \frac{x_2}{l}\right), \qquad (4)$$

where $S_{n\delta n}^{(i)}$ -is the surface of the lateral surface of that i -element. At the site $x_3 \le x \le x_4$ of the rod, heat exchange

takes place with the surrounding medium. Here the coefficient of thermal conductivity h_1, and the temperature of the environment T_{E1}. Therefore, on this section for all elements the expression of the corresponding functional has the following form

$$I_i = \int\limits_{V(i)} \frac{K_{xx}}{2}\left(\frac{\partial T}{\partial x}\right)^2 dV + \int\limits_{S^{(i)}_{n\delta n}} \frac{h_1}{2}(T - T_{OC1})^2 dS,$$
$$i = \left(\frac{x_3}{l} \div \frac{x_4}{l}\right). \qquad (5)$$

At a point $x = L$ corresponding to the area of the cross-section of the rod, heat exchange takes place with the surrounding medium [2]. Here the coefficient of thermal conductivity h_2, and the temperature of the environment T_{E2}. Then for the most finite n- element, the expression for the functional will be

$$I_n = \int\limits_{V(n)} \frac{K_{xx}}{2}\left(\frac{\partial T}{\partial x}\right)^2 dV + \int\limits_{S^{(n)}_{(2n+1)nnc}} \frac{h_2}{2}(T - T_{OC2})^2 dS. \qquad (6)$$

Then for the whole rod the form of the functional expressing the conservation and change of thermal energy will be as follows

$$I = \sum_{i=1}^{n} I_i. \qquad (7)$$

Because of the discretization of the rod on the n part, in the general case, the finite elements will have $(2n+1)$ nodal points [3]. The distance between any two points will be equal $L/n(cм)$. Now, in order to find the temperature $T_1, T_2, T_3, ..., T_{2n+1}$ at these nodal points $(2n+1)$, we minimize the functions (7) with respect to T_j $(j = 1 \div (2n+1))$ and construct a system of linear algebraic equations consisting of $(2n+1)$ equations

$$\frac{\partial I}{\partial T_j} = 0,\ j = 1 \div (2n+1). \qquad (8)$$

Solving the system of equations by the Gauss method, we find the temperature at the node points of the finite elements [4]. As a result, the temperature at any point of any finite element can be found by formula (1). Then the elongation of the rod L-length [5] by the temperature field $T = T(x)$ is determined as follows:

$$\Delta l_T = \int\limits_0^L \alpha T(x) dx, \qquad (9)$$

If the extension of the rod from the applied axial tensile force P is [1]

$$\Delta l_p = \frac{PL}{EF}, \qquad (10)$$

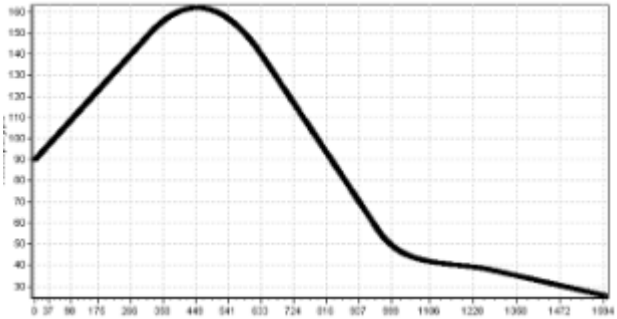
then, the total elongation of this rod will be

$$\Delta l = \Delta l_T + \Delta l_p = \int\limits_0^L \alpha T(x) dx + \frac{PL}{EF}. \qquad (11)$$

Here we investigate the effect of different values of the heat flux q_1, q_2, q_3 on the elongation of the rod.

Table 1 - The law of distribution of the temperature field at fixed points of the rod at $q = -50(W/cm^2)$

T1= 90,000	T350=154,3074	T700= 123,517	T1050= 44,493	T1400= 33,177
T50=99,110	T400=159,88779	T750= 110,794	T1100= 42,011	T1450= 31,267
T100=108,406	T450=162,02778	T800= 98,0721	T1150= 40,689	T1500= 29,357
T150=117,70208	T500=160,7273	T850= 85,3496	T1200= 39,764	T1550= 27,448
T200=126,99808	T550=155,9866	T900= 72,6271	T1250= 38,703	T1600= 25,538
T250=136,29408	T600=147,8054	T950= 59,9046	T1300= 36,997	T1601= 25,500
T300=145,59008	T650=136,2396	T1000= 49,564	T1350= 35,087	



Length of the rod

Figure-2 - The law of distribution of the temperature field along the length of the rod at $q = -50(W/cm^2)$

THE ANALYSIS OF THE EFFECT

So, after studying this problem, for a rod with a limited length insulated on the lateral surface of the rod, we can construct the below-mentioned final table (table 2) below. And also the dependence of the elongation of the rod on the temperature field on its length is given in Fig. 3.

Table 2-Relative elongation of the rod

Length of the rod $L(cm)$	Δl_T (cm)	Δl_P (cm)	$\Delta l = \Delta l_T + \Delta l_P$ (cm)	Δl_T is how many times less compared to when $L = 80cm$	Δl is how many times less compared to when $L = 80cm$
80	0,122083333	0,000666	0,12275	1	1
60	0,0759375	0,0005	0,0764375	1,6077	1,6059
40	0,040208333	0,000333	0,040541666	3,0363	3,0277
20	0,014895833	0,000166	0,0150625	8,1958	8,15

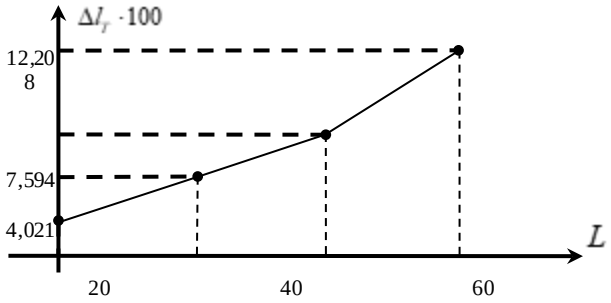


Figure 3-Dependence of the elongation of the rod on its length due to the temperature field

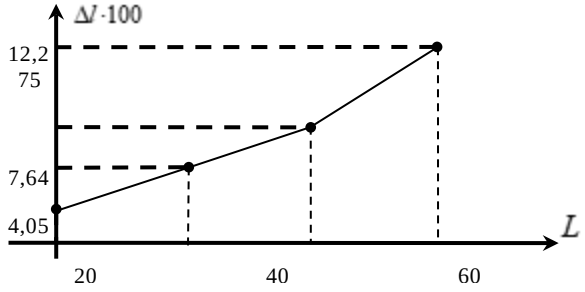


Figure 7 - Dependence of the elongation of the rod on its length counting tensile force and temperature field

CONCLUSION

1. In solving this problem, it was also found that with the rod extension from the field of distribution, the temperature will be greater than, compared with the elongation of the rod, because of the applied tensile force.

2. And also it is revealed, when the ambient temperature increases by 50%, 100%, 150%, 200%, the extension of the rod increases by 35.29%, 70.58%, 105.874%, 241.16%.

3. A corresponding computational algorithm for the numerical study of the dependence of the elongation of a rod on its length in the presence of a heat flux is constructed. It was found that under this law of heat flow, a decrease in the length of the rod by 25, 50, and 75% leads to a decrease in the value of its elongation from the field of temperature distribution by 37.8; 67.06 and 87.8% respectively. At the same time, with such a law of heat flow, a decrease in the length of the rod by 25, 50, and 75% leads to a decrease in the value of its elongation as a whole by 37.73; 77 and 87.7% respectively.

4. It was found that an increase in the value of the heat flux by 100 and 200%, leads to an increase in the value of the elongation of the rod by 139.4 and 178.8%, respectively.

REFERENCES

[1] Amirtayev K., Naizabayeva L., Ibadullayeva A. Development of the Complex of Software Applications to Control the State of Total Thermal Energy of an Elastic Rod // 8th IEEE International Conference on Application of Information and Communication Technologies (AICT), Astana, Kazakhstan, OCT 15-17, p. 115-119, 2014.

[2] Kudaykulov A., Amirtayev K., Utebaev U. Numerical study of strain-deformed state of the rod filled-rigidly at both ends, when exposed along the length to the parabolic law of the temperature field // Materialy IV mezinarodni vedecko-prakticka conference. – Sofia, – 2008. – T.28, – 21–24 p.

[3] Amirtayev K., Utebaev U., Tokkuliev B., Zhumadillaeva A. Determination of the law of temperature distribution in a partially insulated tube of limited length, with the heat flow on the inner bounded surface of closed mid-pipe. // Materialy IV mezinarodni vedecko-prakticka conference – Sofia, – 2008. – 66–69 p.

[4] Kudaykulov A, Arapov B, Kenzhegulov B, Amirtayev K, Utebaev U, Tokkuliev B. Numerical solution of the established thermal strained and deformed state of the heat exchanger in the presence of internal heat flow and external heat transfer of constant intensity // Materialy IV mezinarodni vedecko-prakticka conference «Veda a vznik-2008/2009». – Praha: Publishing house «Education and Science» s.r.o., – 2009. – 15–19 p.

[5] Amirtayev K.B., Ibadullaeva A.S., Akimhaze M. About one computing method of the study thermo-tense condition element to designs at presence of the sources of the heat and axial power // Abstracts of the third Congress of the World Mathematical Society of Turkic Countries. – Almaty, – 2009. V.2, – P. 188.

Impedance control of 2dof serial robot manipulator

A. ALSHAWI[1] and S. KİZİR[1]

[1] Kocaeli University, Kocaeli/Turkey, ali-sh90@hotmail.com
[1]Kocaeli University, Kocaeli/Turkey, selcuk.kizir@kocaeli.edu.tr

***Abstract*-This paper shows the implementation of an impedance control method to regulate the interaction forces between a robotic arm and the environment, when there is connect between them. A complete description of the system to model and control both an RR robot arm and its interaction with the environment is simulated and detailed by Matlab/Simulink; from the generation of a mechanical model in SimMechanics (Matlab) after export its 3D drawing. The description and setting of a dynamic model based on computed torque control, to cancel out the nonlinearities existing on the dynamic model of the robot. It is based on feedback linearization and computes the required arm torques using the nonlinear feedback control law, and lastly control of the reaction forces is done using the impedance control method after modeling the environment. This type of control modifies the dynamic behavior of the robot when there is contacting with the environment and it is widely used in industrial robots.**

***Keywords*- Impedance control, computed torque control, Robotic manipulator, Interaction, Simulation, Modelling.**

I. INTRODUCTION

APPLIYNG robot manipulators to a wider class of tasks, that will be necessary to control not only the position of a manipulator but also the force exerted by the end-effector to the environment. While many of the tasks performed by the robot manipulator r, that the robot interacts with its environment such as surgeries, pushing, touching, carrying sensitive objects and cutting, etc. The implementation of all these tasks require that the robot, besides realizing the expected position, provide the necessary force either to overcome resistance from the environment, or comply with the environment. Therefore, the robot manipulator to interact and contact safely and friendly to humans or to objects in unknown environments, it is necessary to include an interaction control method that adapts the forces exerted on the environment in order to avoid damages the manipulator and the environment. A force control methods can be used on those applications where the desired force or maximum force to exert is known in advance. However, in other case indirect force control methods can be used. That these methods try to make the manipulator compliant with the object connecting to, without any problem to control maximum or desired force. In the control loop, the position control loop takes the main role but when there is, an interaction is also being controlled for getting safe and clear contact. Impedance control [1] is one of these indirect force control methods. The main philosophy of impedance control, according to Hogan [2], is that the manipulator control system should be designed not to track a motion trajectory alone, but rather to regulate the mechanical impedance of the manipulator. Therefore, when the robot manipulator contacting the environment, impedance control aims to control the dynamic behavior of robot by controlling the properties of contact, that means controlling the damping and stiffness of the interaction. Similar to our work in this study, we find that the impedance control method is used in different fields, such as robots are able to move and carry sensitive things without any damage to them, industrial robots [3] used in vehicle assembly operations or surgical robots like Da Vinci surgical robot [4].

This paper presents a case study where impedance control is used to control the interaction forces of a simulated 2dof serial robot manipulator. In order to achieve this goal, the mechanical model of the robot will be created using SimMechanics after export its 3D drawing. A computed torque [5], [6] controller will be used to cancel out the nonlinearities present on the robot's dynamic model, that the mechanical system described in SimMechanics. It is depends on feedback linearization and computes the required arm torques using the nonlinear feedback control law. Finally, after the environment is modelled, a Cartesian impedance controller [7], [8] is used to control the forces exerted for the robot. To test the performance of the control system, the robot is given a desired trajectory as input that causes the robot to hit a wall, in order to observe the effectiveness of the controller and performance of the robot environment interaction.

II. MECHANICAL AND DYNAMIC MODEL

The first step in this paper is to create a model of a 2-dof robot arm to test our approach. Figure 1 shows the mechanical model of the 2-dof robot arm drew with 3D program (SolidWorks) and Figure 2 shows the mechanical model of the 2-dof robot arm described with SimMechanics after export its 3D drawing. The model is consisted of fixed part, two bodies tow and revolute joints. The module receives torques as input and the outputs are end-effector position in Cartesian Coordinates, joint angles and torques.

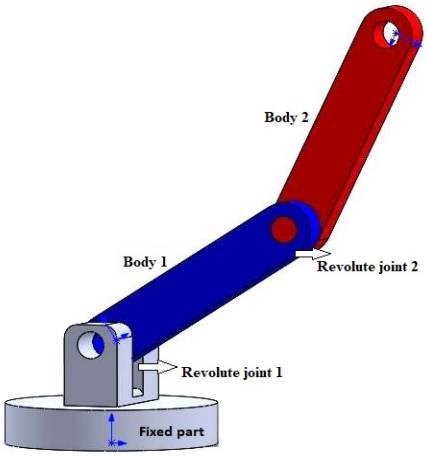


Figure 1: Model of 2-dof robot arm in SolidWorks.

The model of the system is non-linear and highly coupled. For the control system, there are two alternatives: either to use non-linear control techniques or to linearize the system and apply well-known linear control techniques. The second option will be used here like many of the studies in literature. The next section will describe the dynamical model of the mechanical manipulator.

The dynamics of a robot arm is explicitly derived based on the Lagrange-Euler formulation to explain the problems involved in dynamic modelling. That the dynamic model relates the forces acting on the mechanical structure with the resulting displacements, velocities, and accelerations. Where these forces can take different sources like, the torques generated by the motors, the friction forces, the inertia of the mechanical links, and the possible forces exerted from the environment on the robot. The mechanism of the arm was treated as a combination of open Kinematic-chains [9]. Through the formulation by Denavit Hartenberg convention, the forward kinematics of the system is derived. Then by using the Lagrange equation, we get the dynamics of the model.

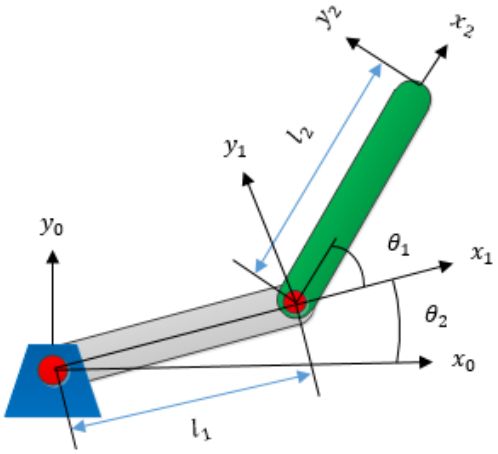


Figure 3: Coordinate systems of 2dof Robot manipulator.

The Lagrangian of a dynamic system is the difference between the kinetic and potential energy at an arbitrary instant [10]. The dynamic equations for the robot manipulator are usually represented by the coupled non-linear differential equations which was derived from lagrangian method

$$M(\theta)\ddot{\theta} + C(\theta,\dot{\theta}) + G(\theta) = u \qquad (1)$$

where θ is the joint variable and u is the vector of generalized forces acting on the robot manipulator. $M(\theta)$ n×n is the inertia matrix of manipulator. $C(\theta,\dot{\theta})$ n×1 is the vector of centripetal and Coriolis, and $G(\theta)$ n×1 is the gravity vector. Unlike the real robot manipulator the friction torques in our model will not be exist as we see in equation (1). Figure3 shows the Coordinate systems and details of 2-DOF robotic arm. The dynamic model following lagrangian formulation is:

$$M(\theta)\begin{bmatrix}\ddot{\theta}_1\\ \ddot{\theta}_2\end{bmatrix} + C(\theta,\dot{\theta}) + G(\theta) = \begin{bmatrix}u_1\\ u_2\end{bmatrix} \qquad (2)$$

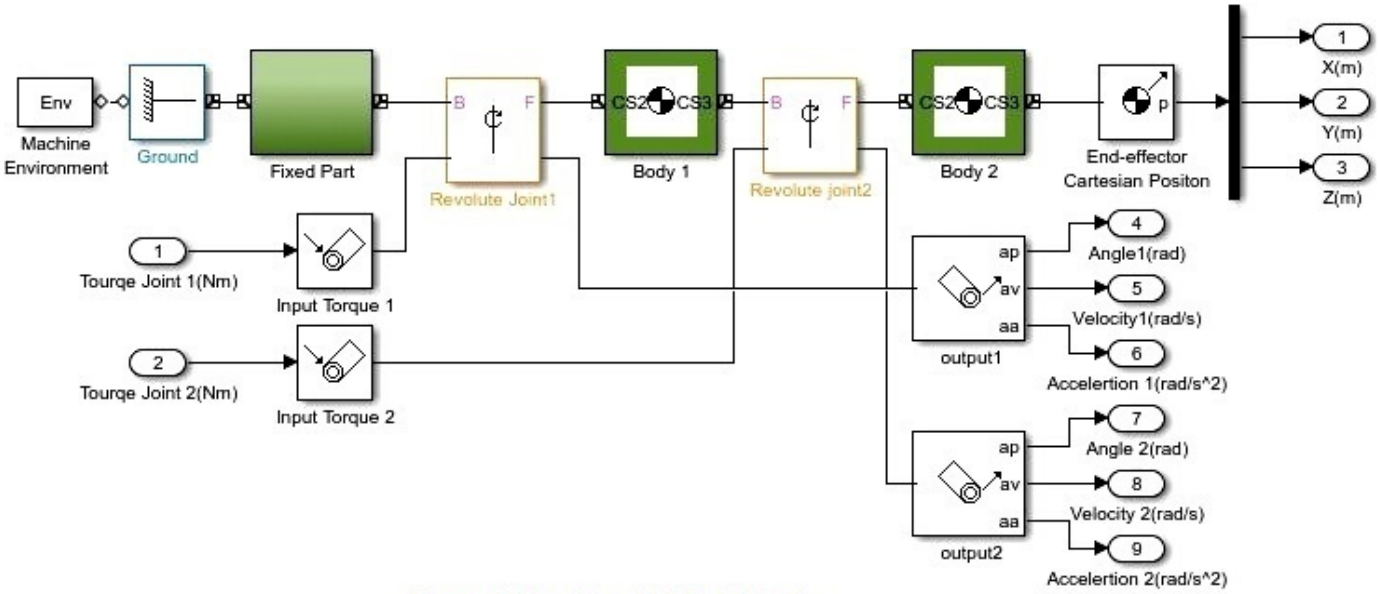


Figure2 : Mechanical model in SimMechanics

Where

$$C(\theta,\dot{\theta}) = \begin{bmatrix} -m_2 l_1 l_2 (2\dot{\theta}_1\dot{\theta}_2 + \dot{\theta}_1^{\,2}) sin\theta_2 \\ m_2 l_1 l_2 \dot{\theta}_1^{\,2} sin\theta_2 \end{bmatrix} \quad (3)$$

$$G(\theta) = \begin{bmatrix} (m_1+m_2) g l_1 cos\theta_1 + m_2 g l_2 cos(\theta_1+\theta_2) \\ m_2 g l_2 cos(\theta_1+\theta_2) \end{bmatrix} \quad (4)$$

$$M(\theta) = \begin{bmatrix} (m_1+m_2) l_1^{\,2} + m_2 l_2^{\,2} + 2 m_2 l_1 l_2 cos\theta_2 & m_2 l_2^{\,2} + m_2 l_1 l_2 cos \\ m_2 l_2^{\,2} + m_2 l_1 l_2 cos\theta_2 & m_2 l_2^{\,2} \end{bmatrix} \quad (5)$$

The terms l_1 and l_2 are the lengths of link1and 2, respectively and m_1, m_2 their masses. In this work $l_1 = l_2 = 0.16m$ and $m_1 = m_2 = 1kg$.

III. POSITION CONTROL

In this paper, a Computed torque controller will be used to control the position of the robot end-effector. The computed torque control is an effective motion control strategy for robotic manipulator systems, which the Computed torque controller is a significant nonlinear controller to certain systems, which it is depend on feedback linearization and computes the required arm torques using the nonlinear feedback control law. Let us describe now the technique to linearize and decouple the system. Figure 4 shows configuration of computed torque scheme.

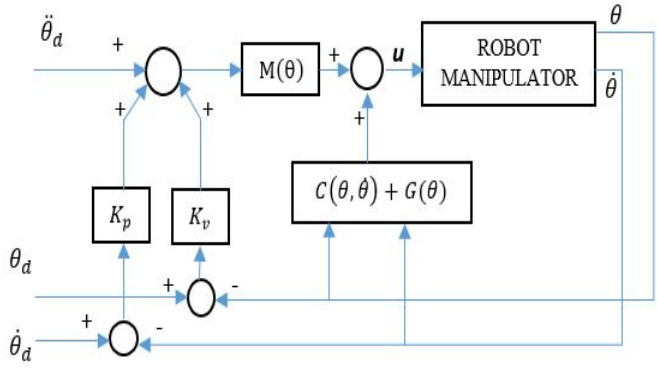


Figure 4: Conventional computed torque control scheme

The control law in (6) represents the standard structure of the compute torque control strategy.

$$u = M(\theta)\left[\ddot{\theta}_d + K_p\, \dot{e} + K_v\, e\right] + C\left(\theta,\dot{\theta}\right) + G(\theta) \quad (6)$$

where $\ddot{\theta}_d$, $\dot{\theta}_d$ and θ_d are re the vectors n×1 of desired acceleration, velocity, and position, respectively. The joint position error is denoted by the vector $e = \theta_d - \theta$while $\dot{e} = \dot{\theta}_d - \dot{\theta}$ the n×1 vector of velocity error. The computed torque control in (6) has two parameters, $K_p$and K_v which are the n×1 Proportional and Derivative gains, respectively.

In most of the industrial manipulators, the possibility to send torque commands directly to the robot is not available. To find solution and to simplify the tasks of the control engineer, these industrial robots include an internal joint controller together with the necessary inverse kinematics algorithms. In other words, the designer can choose either to send direct joint angle commands to the robot or an end-effector position or orientation command in Cartesian coordinates. Therefore, by using the Geometric solution approach [11] that is based on decomposing the spatial geometry of the manipulator into several plane geometry problems, the possible solutions for θ_1 and θ_2 angles of robotic arm can be written as in (7) and (8)

$$\theta_2 = \text{Atan2}\left(\mp\sqrt{1-\left[\frac{P_x^2+P_y^2-l_1^2-l_2^2}{2l_1l_2}\right]^2}, \frac{P_x^2+P_y^2-l_1^2-l_2^2}{2l_1l_2}\right) \quad (7)$$

$$\theta_2 = \text{Atan2}\left(P_y, P_x\right) + \text{Atan2}\left(\sqrt{P_y^2 + P_x^2 - (l_2 cos\theta_2 + l_1)^2}, l_2 cos\theta_2 + l_1\right) \quad (8)$$

Where P_y, P_x are the end-effector position, and l_1 and l_2 are the lengths of links 1 and 2 respectively. In this case and in order to be able to obtain the results of this work in an easy and useful manner to robot manipulator, Inverse kinematics of our module (7), (8) has been modeled in Simulink. In this way, the mechanical model of the manipulator, the computed torque controller, and the inverse kinematics module can be integrated into one block, this block represents the industrial robots in general. Then, the next step is to build up outer control loops that will be easily tested on a robot that explained above. In other words, focusing on the impedance controller in terms of design and selection of variables values, and study of the results obtained after application of the impedance controller on the robot rather than on internal joint controllers.

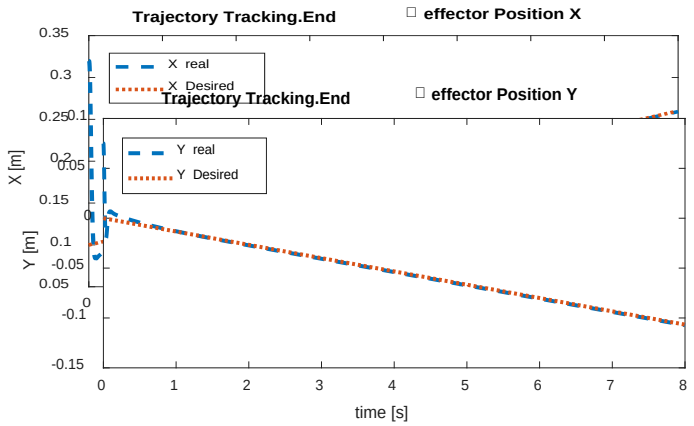


Figure 5: Trajectory tracking performance, top: End-effector Position X, right: End-effector Position Y.

Figure5 shows the trajectory tracking performance of the systemunder control for a given ramp trajectory as input signal in Cartesian space. The home position of our system when: $\theta_1 = \theta_2 = 0$ is $X = 0.32m$ and $Y = 0.075m$, as seen in figure5. The parameters $K_p$and K_v gains of computed torque controller were found empirically (K_p= 4000, K_v=2000) but it is not difficult to demonstrate how to choose them depending on the desired dynamical response. After canceling, the non-linarites of the mechanical model and tuned the gain parameters of the computed torque controller. It will be possible to obtain the desired performance for the control system.

IV. THE ENVIRONMENT AND CONTACT FORCES

Our design includes a manipulator model and its control system, a system that is able to track a desired trajectory in Cartesian space. Our dynamic model does not take into account possible external forces acting on the robot that would definitively change the dynamic behavior of our manipulator. To account for this situation, a model of the environment will be included in our system whose interaction force will act on our robot. Widely the environment's model is used as a linear spring with a spring constantK_e as in (9).

$$f = K_e(X - X_e) \tag{9}$$

Nevertheless, in our case, we will include a damping coefficient with the spring, as the environment is modelled in (10).

$$f = K_e(X - X_e) + B_e(\dot{X} - \dot{X}_e) \tag{10}$$

where K_e is the stiffness of the environment, B_e is the damping coefficient of the environment, f is the contact force, X_e is the static position of the environment, and X is the end-effector position at the contact point. Figure 6 appears the environment concept, where a manipulator of mass m contacts the environment at position X_e trying to reach the desired end-effector position X_d.

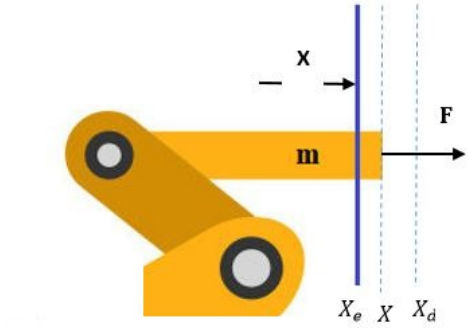


Figure 6: The contact between robot and environment

The dynamic equation of our system was defined in (1). That, *u* was the vector of generalized forces acting on the robot manipulator, also the external contact force. Therefore, to make our system closer to a real and clearer, we will modify (1) to show the effect of those forces and will represent for them in our model. The dynamic equation controlling the robot's behavior might be defined as in (11).

$$M(\theta)\ddot{\theta} + C(\theta,\dot{\theta}) + G(\theta) = u - J^T(\theta)f \tag{11}$$

The term $J^T(\theta)f$ translates the task-space forces to the joint. Then, equation (11) will include in Simulink model in order to take account the forces of contact on the dynamic response. As knowing, the relation between forces and torques is defined as in (12).

$$\tau_e = J^T(\theta)f \tag{12}$$

In this work, the case is 2dof robot manipulator, so the contact force can be written as in (13), (14).

$$\tau_{1e} = J_{11}f_X + J_{21}f_Y \tag{13}$$

$$\tau_{2e} = J_{12}f_X + J_{22}f_Y \tag{14}$$

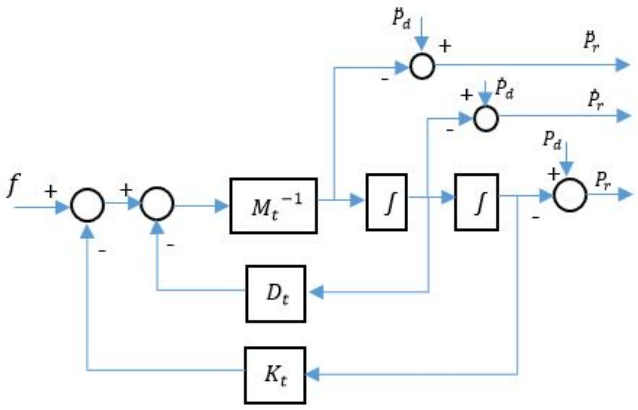
where τ_e are the contact torques, Jij are the elements of the transpose of the 2×2 Jacobian matrix, f_X and f_Y are the forces over the X and Y axes.

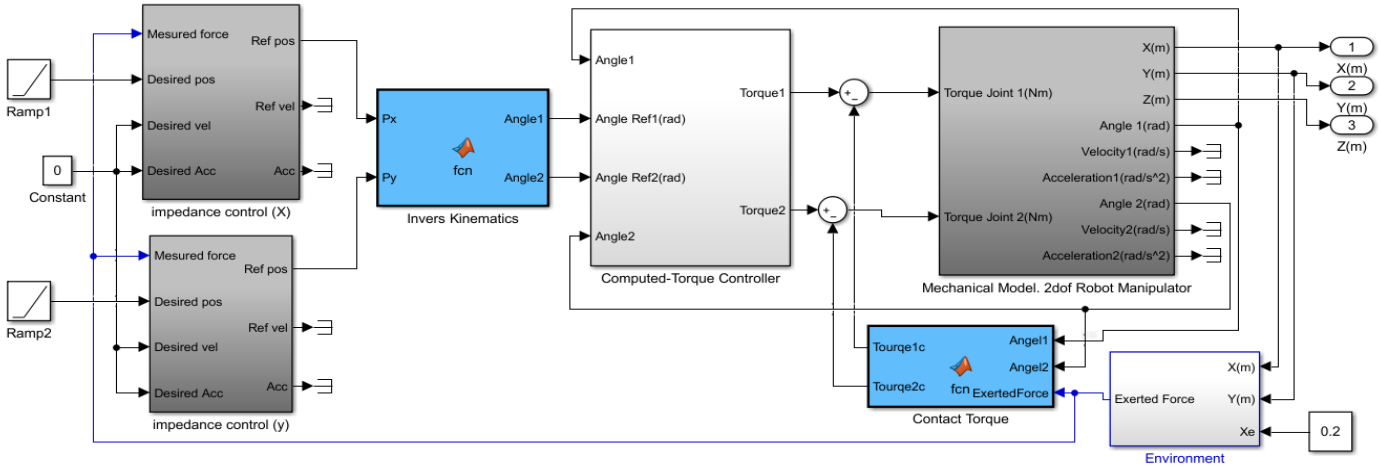
V. IMPEDANCE CONTROLLER

Finally, after we have modelled and controlled a 2dof robot manipulator and designed a simple model for the environment, the next step is to design a controller, which regulates the interaction when the robot contacts the environment. In the current situation, if suppose the robot follows a trajectory and suddenly an object or obstacle shows on its way, the robot will collide with it, trying to reach the final end position of the given trajectory, and exerting such a huge forces into the environment that would likely cause damages to a real robot or to both the object of collision and robot. To solve this situation and avoid it, an impedance controller will be designed. The impedance controller input is the desired trajectory at each time step, and the measured contact forces will be included on the controller for get a quick feedback of the contact state. The output of the controller will be a modified trajectory. That means if the forces are not sensed, the trajectory will be followed accurately. Otherwise, when forces are sensed, the trajectory will be modified in order to regulate the maximum forces. Equation (15) gives the control law of the impedance controller

Figure 7: Impedance controller structure.

Figure 8: Complete control system modelled in MATLAB/Simulink





$$M_t\ddot{e}_t + D_t\dot{e}_t + K_t e_t = f \qquad (15)$$

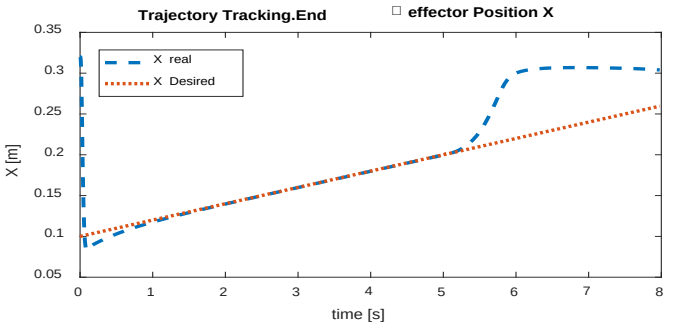
where M_t, D_t and K_t are the inertia, damping and the stiffness coefficients, respectively, e_t is the trajectory error.That, the error defined as $e_t = P_d - P_r$, where P_d is the desired input trajectory and P_r will be the modified trajectory, output of the impedance controller and the input for the inverse kinematics module (P_X,P_Y). Figure7 shows the structure of the impedance controller. The controller parameters have been optimized. $M_t = 2kg$, $D_t = 2\,Ns/m$, $K_t = 90\,N/m$.

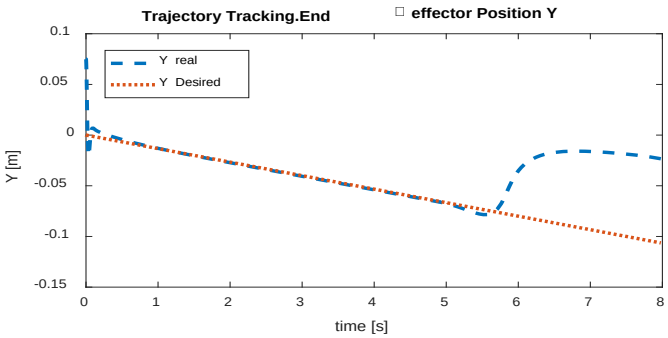
Figure8 shows the complete simulation model of the control system in this paper, after designing and modeling the all components of them as discussed above. It contains the mechanical model, the computed torque controller, the invers kinematics of robot manipulator, the model of the environment, the impedance controller and the desired input trajectories.

VI. SIMULATION RESULTS

In order to monitor and test the performance of the control sys-tem, the robot is given a desired input trajectory to follow. That, the X-axis input is a ramp with slop = 0.3/15 starting at *X=0.1*, and the Y-axis input is a ramp with slop = -0.2/15 starting at *Y=0*. In this experiment we put a wall modelled as defined in (11) like a barrier at X=0.2 with K_e=2000*N/m* and B_e =0.1*Ns/m*. Our goal from this experience, when the robot contacts to the surface of the wall, the robot remains at the point of contact as long as the wall exist, either in the case of removal, the robot follows the trajectory given to him to reach the target.

In Figure9, we can see the results of the experiment without using the Impedance controller. The manipulator tries to follow the given desired input trajectory and reaches to the target point, after contacting the wall surface, the contact forces increase exponentially as the robot travels "inside" the wall, and robot position changes because of a term $J^T(\theta)f$ as seen in (11) .In Figure10, we can see the results of the same experiment with using the impedance controller, that at the contact point (X=0.2), the contact force increases dramatically and the impedance controller reacts and works to redefine a new position trajectory.





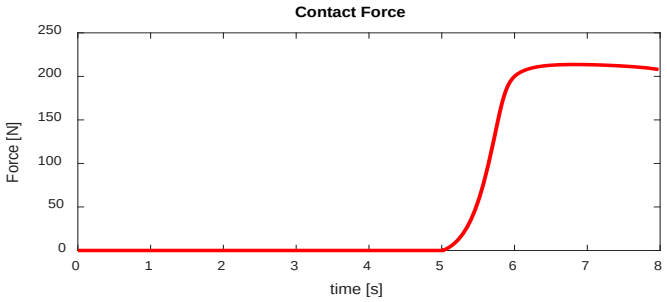
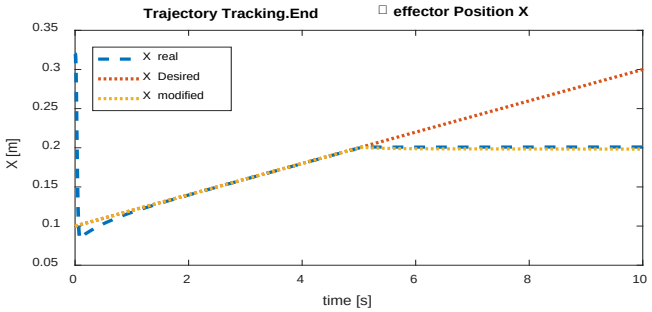


Figure 9: Response without Impedance Controller. top: End effector Position X and Y, bottom: Contact Force.

This new trajectory will be the reference position to value of contact point. The new trajectory regulates forces to a value of around 2N. X and Y real signals are end-effector position, X and Y Desired signals are desired input trajectory, X and Y modified signals are impedance controller outputs.



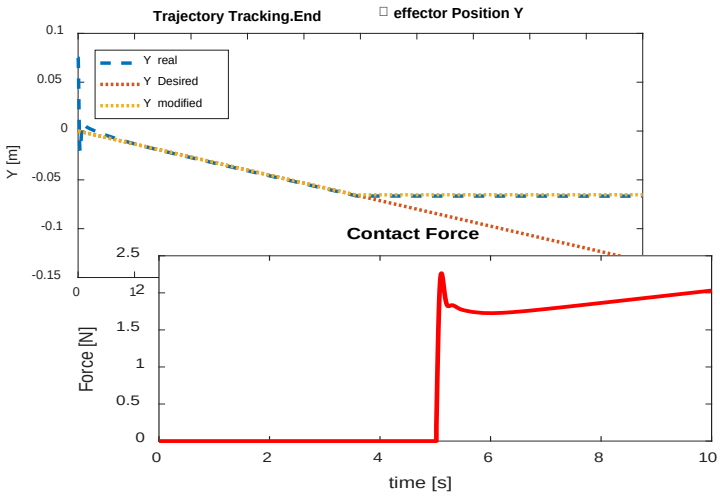


Figure 10: Response with Impedance Controller. top: End-effector Position X and Y, bottom: Contact Force.

VII. CONCLUSION

In this paper, the steps of modeling and simulation a 2dof robot manipulator, and the interaction between the robot and the environment was explained. The 2dof robot manipulator was modeled by using SimMechanics after export it's 3D drawing. That, the mechanics model of the robot gives a great potential ability to verify the control algorithms applied to the model. Then, a computed torque controller was applied and used to control the position of the robot end-effector. Therefore, an inverse kinematics was used to send direct joint angle commands to the robot or an end-effector position command in Cartesian coordinates. On the other hand, impedance controller was developed and applied to the system. When the robot contact with the environment that was modeled as (spring and damper), the contact forces increase exponentially and changes the robot position. By using the impedance controller, the impedance controller reacts and works to redefine a new position trajectory. This new trajectory regulates forces to a value of around2N. The parameters of the impedance controller were optimized. The use of impedance controller shows the ability to control the interaction between the environment and the robot, especially in those applications like Human-interaction robots and surgical robots.

REFERENCES

[1] Hogan N. (1984). *Impedance Control an Approach to Manipulation*, American Control Conference, pp. 304 – 314.

[2] Hogan N. (1985). *Impedance Control, an Approach to Manipulation*: Part I, II, Int. Journal of Robotics Res. 107, 1–24.

[3] Lange, Friedrich, Wieland Bertleff, and Michael Suppa. *"Force and trajectory control of industrial robots in stiff contact." Robotics and Automation (ICRA), 2013 IEEE International Conference on. IEEE, 2013.*

[4] Haidegger, Tamás, et al. "Force sensing and force control for surgical robots." IFAC Proceedings Volumes 42.12 (2009): 401-406.

[5] Middle tone, R. H., and G. C. Goodwin. "Adaptive computed torque control for rigid link manipulators." *Decision and Control, 1986 25th IEEE Conference on.* Vol. 25. IEEE, 1986.

[6] Gilbert, Elmer G., and In Joong Ha. "An approach to nonlinear feedback control with applications to robotics." *IEEE transactions on systems, man, and cybernetics* 6 (1984): 879-884

[7] Chiaverini, Stefano, Bruno Siciliano, and Luigi Villani. "A survey of robot interaction control schemes with experimental comparison." IEEE/ASME Transactions on mechatronics 4.3 (1999): 273-285.

[8] Caccavale, Fabrizio, et al. "Integration for the next generation: embedding force control into industrial robots." IEEE Robotics & Automation Magazine 12.3 (2005): 53-64.

[9] Xie, Ming. *Fundamentals of robotics: linking perception to action*. Vol. 54. World Scientific Publishing Company, 2003.

[10] Jazar, Reza N. *Theory of applied robotics: kinematics, dynamics, and control*. Springer Science & Business Media, 2010.

[11] *Welman, Chris. Inverse kinematics and geometric constraints for articulated figure manipulation. Simon Fraser University, 1994.*

A Review on Measurement Methods of Non-Invasive Blood Glucose Level

E. MENGUC[1] and S. HELHEL[1]

[1] Akdeniz University, Engineering Faculty, Antalya/Turkey, ercanmenguc@gmail.com
[1]Akdeniz University, Engineering Faculty, Antalya/Turkey, selcukhelhel@akdeniz.edu.tr

Abstract **- Today, approximately 200 million people around the world are required to regularly check blood sugar levels every day, and this number is increasing day by day. If hyperglycaemia and hypoglycemia occur as a result of blood sugar level abnormalities, serious tissue and organ damage and, most importantly, vital risks can occur. In generally, blood glucose level measurement methods are divided into three categories, which are invasive, minimally invasive and non-invasive. In invasive methods people need to measure blood glucose levels by drilling their fingers, squeezing blood droplets on test strips and treating the results with portable glucometers. The process can be uncomfortable and complicated and has to be repeated many times each day. By using minimally invasive methods, the glucose ratio is determined by the help of tissue fluid or very little blood. Actually this method is uncomfortable for human life too. For this reason, there is a growing need for a new generation of non-invasive glucose-level monitoring systems in which the glucose level can be reliably determined. The goal of the ongoing works in literature are to prevent or delay early diagnosis and complications, rather than to treat the diabetes. This review giving a comprehensive knowledge about non-invasive design methods used in the literature. General definitions are given for each of the design methods. Scientists, laboratories and universities have been working on the design of non-invasive glucometer with various methods, yet there is no product that can measure with high accuracy.**

Keywords **- non-invasive, polarimetry, glucose, diabetes, diabetes mellitus, blood glucose measurements.**

I. INTRODUCTION

Diabetes Mellitus (DM) is a life-long chronic disorder that is caused by the inability of the pancreas to produce sufficient insulin or not effectively use insulin produced by the body and is a continuing disease with a decrease in insulin-producing cells. It is estimated that the number of adults with diabetes of 415 million in 2015 in the world and will rise to 642 million by 2040. This means that in 2040 one person in every 10 people will have diabetes. In Turkey, it is stated that the Ministry of Health has 7 million diabetic patients according to 2015 data. So that in 2015, one adult (total 5 million) in every 6 seconds lost his life due to affiliated with diabetes. In addition, according to the International Diabetes Federation's 2015 data, 12% of global health spending is diabetes. This is equivalent to approximately 673 billion USD. In 2040, health expenditures of only diabetes-related diseases are estimated to exceed 802 million USD. As you can see, diabetes is both a serious threat to human health and a burden on countries in the fight against diabetes.

Diabetes mellitus is generally divided into two types and the definitions of Type-1 and Type-2 diabetes are used. Type-1 autoimmune system derived and include pancreatic beta cells (insulin-producing cells) damage. Type-2 represents the more prevalent diabetes mellitus. Type-2 diabetes mellitus is a group of diabetes forms in which pancreatic insulin secretion disorders and insulin resistance are both effective.

Type-1 diabetes is the type of diabetes that results in a series of eventual insulin-dependent diabetes that leads to the progressive beta cell depletion of the pancreas. It is known that the disease, which usually develops from autoimmunity, usually occurs in childhood and young adult ages. By measuring auto-anticore ratios with the type-1 diabetes process, it is possible to elicit early metabolic abnormalities before they begin. It is stated that around the world 50,000 new Type-1 diabetics are diagnosed every year. Type-1 diabetes accounts for about 5-10% of all diabetics.

Type-2 diabetes is the most common type of diabetes worldwide. Approximately 90% of all diabetics are Type 2 diabetes. Type-2 diabetes is seen, usually after age 40, of which chronic complications are common, with increased incidence of adulthood, increased or decreased diabetes symptoms, and sometimes no symptoms. The main causes of type 2 diabetes are saturated fat-rich diet, inactive lifestyle, obesity, inability to use insulin in tissues (insulin resistance), and impaired insulin secretion. The adverse effects of type-2 diabetes can be reduced by appropriate diet programs and exercise.

It is very troublesome to continuously monitor the blood sugar level with current invasive or minimally invasive methods. Portable blood glucose meter (glucometer), measuring sticks and finger punching device are required. These devices (needles and strips) need to be constantly renewed with certain periods in order to avoid the risk of infection. However, this situation causes both a financial burden and a painful feeling for the patients each time. In addition, results obtained by the invasive method should be noted regularly in order to be able to perform appropriate treatment. Conversely, continuous glucose level monitoring systems have no cost beyond the cost of acquisition, and glucose level monitoring systems do not affect patient comfort negatively.

Thanks to the high technology that can be produced today, it is now possible for existing invasive glucose meters to be developed as a wearable device with a non-invasive glucose level determination and continuous tracking capability. The success of these efforts has the potential to improve the quality of life of millions of people. For this reason, the importance of the subject is increasing even more. As a result of the when

design of a miniature non-invasive glucometer instrument suitable for hospital and home using; a continuous glucose level monitoring system can be set up so that the level of blood sugar can be continuously monitored and recorded, the patients can be pre-stimulated in order to prevent attacks of hypoglycemia and hyperglycaemia, glucose fluctuations can be minimized with appropriate doses and appropriate treatments can be made and the risk of developing complications can be reduced. The glucose level can be kept steady and thus the organ damage caused by diabetes can be avoided. Thus, diabetic patients will be provided with a more comfortable and safe life.

II. NON-INVASIVE BLOOD GLUCOSE LEVEL MEASUREMENT METHODS

In the 1800's, two French physicists, Dominique François Arago and Jean Baptiste Biot, were the first to investigate the branch of science known as stereo chemistry [1, 2]. The first documented use of the light after their work was first seen in applications towards the end of the 1800s, in which the concentration of sugar in industrial applications was determined or used to monitor the production of sugar [2]. After this date, although light is frequently used in industrial applications, the use of biomedical applications is first seen in the 1980s. In 1980, March and Rabinovic [2, 14] presented a method that could detect the level of glucose through the aqueous humor layer in the eye with the help of polarized light. March and Robinovic found that very sensitive and stable polarimetry is necessary to measure the millimeter-level rotation resulting from glucose in physiological levels. In the next decade, studies on the development of such a polarimetry have been carried out. Subsequently, Cameron and et al. presented a shaft-precision polarimeter based on the Faraday modulator technique and using a closed-loop feedback system [2, 33]. Since then, studies have been carried out by various groups to measure glucose intensities using different techniques.

Mostly optical techniques were used in 1990's studies. Most non-invasive methods are using optical techniques; which are spectroscopic based methods MIR(Mid-Infrared), NIR (Near-Infrared and Raman; transdermal method, electromagnetic technique, dielectric method, bio-impedance technique, fluorescence, photoacoustic technique, optical coherence and polarization techniques. Detailed information on these methods is given later in this review. Rather than discovering new things in these studies, priority has been given to areas such as increasing low signal-to-noise ratios, eliminating disruptive effects, reducing time delay, calibrations and minimizing the designs [14]. The non-invasive concept has only a 30-years history and it can be said that the majority of non-invasive technologies are still in development [21]. Currently, there is still a great demand for low-cost, easy-to-use non-invasive glucose-level monitoring systems capable of instant reliable measurements. This article contains extensive literature review to provide detailed and usable information on this topic.

Blood sugar determination methods have previously been described as invasive, minimally invasive and non-invasive techniques. Old and new technologies for tracking glucose level are tabulated in Table 1. In invasive methods, some blood is taken out of the body with the aid of a needle and the blood sugar level is determined by deteriorating the integrity of the body. This method is not suitable for continuous glucose level monitoring systems. In minimally invasive methods, the glucose ratio is determined by the help of tissue fluid or very little blood. In non-invasive methods, glucose levels are determined by direct contact from the body surface and measurements taken at a certain distance, without impairing body integrity, without pain, and not allowing any risk of infection. While the non-invasive method is suitable for use in continuous glucose level monitoring systems, other methods are not suitable.

Measurement of blood sugar by non-invasive methods have been done by different research groups and universities with different methods and technologies [1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,17,18,19,20,21,22,23,24,25,26,45]. There are a lot of publications and studies are continuing. Their number can be multiplied. The most commonly used methods in the literature for the measurement of blood sugar are given in Figure 1. Most of the methods based on optical techniques.

There are publications in the literature describing the advantages and disadvantages of the various methods used in determining the level of glucose in the blood by non-invasive methods. It has been shown in studies, determination of the blood glucose level non-invasively by optical, electrical or acoustic [9, 27] waves. Majority of methods used in the literature are optical techniques [10,11,43,45]; these methods are based on spectroscopic techniques (MIR [6, 14], NIR [13, 24], Raman technique [53, 54]) transdermal method [16], electromagnetic waves method [29,30,31,32], dielectric method [18,31,41,42], bio-impedance technique [15, 22], fluorescence, photoacoustic [9, 27], optical coherence tomography [4] and polarization [2,20,28,30,37,38,40,44,47,48,55]. In optical methods, analyzes are made about the amount of polarized light, wavelength and intensity of light passing through optically active substances. It is possible to that these parameters can be determined as a function of the amount of glucose.

Table 1: Old and new technologies for tracking glucose level

No	Method	Definition
1	Invasive	*Finger Drilling*
2	Min. Invasive	*Contact with the skin (puncture) sensors, probes*
	Min. Invasive	*Iontophoresis; Directing certain salt ions to body tissues through electrical current* *Micro dialysis; assessment of metabolic functions at cell level* *Sonoforez; A method of increasing the transdermal passage of molecules by ultrasonic waves*
3	Non-invasive [9, 27]	*Measurement of body fluids such as sweat, tear, uric acid*
	Non-invasive [53, 54]	*Raman Spectroscopy*

	Non-invasive [4]	*Optical Coherence Tomography*
	Non-invasive [2,20,28,30,37,38,40]	*Polarimetry*
	Non-invasive [9,27]	*Photoacoustic*
	Non-invasive [25, 30]	*Ultrasound Tech.*
	Non-invasive [15, 22]	*Bioimpedance Tech.*
	Non-invasive [6, 14]	*MIR*
	Non-invasive [13, 24]	*NIR*
	Non-invasive [16]	*Transdermal Tech.*
	Non-invasive [29,30,31,32]	*Electromagnetic Tech.*

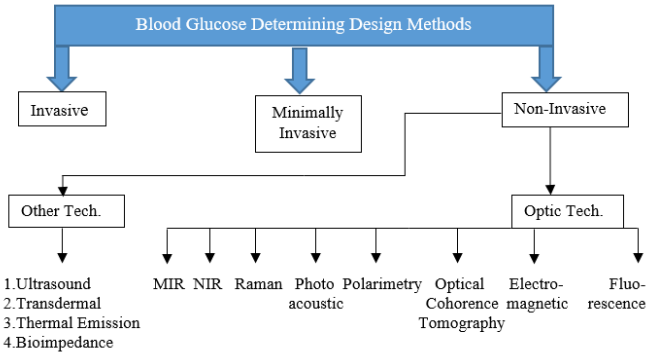


Figure 1: Design methods

MID-INFRARED SPECTROSCOPY

Mid-infrared spectroscopy (MIR) technique is similar to that of infrared spectroscopy. The prominent side of this technique is that the specific absorption band of glucose is in this region [6]. For this reason, it is expected that the glucose peaks will be more visible in this wavelength region than the other components in the blood. In various sources [6, 14] there are studies using this method wavelength range of 2500nm-25000nm. When determination of glucose, water which constitutes a large proportion of blood, appears to be a disruptive factor. However, penetration depth of substance is less than other methods, so it is an important feature that the effects of water and other agents below the skin are less disturbing. However, at the situation of less optical pathway, the light is less likely to interact with glucose molecules. Because the higher wavelengths are used in the MIR technique and the penetration depth is at a few micrometer levels so, higher absorption and lower scattering coefficients are obtained in measurements. Since the light transmitted by the tissue in this wavelength region is less powerful, only the amount of reflected or scattered light can be used.

NEAR-INFRARED SPECTROSCOPY

Near Infrared Spectroscopy (NIR) technique uses a wavelength range of 600nm-2000nm. The light that focuses on any tissue is partially scattered or absorpted due to the chemical components of the tissue has. Since light transmitted or reflected in this wavelength region is not at low levels, it is possible to obtain information on glucose ratios in both forms. It basically depends on the principle that each substance has an absorption coefficient in terms of its concentration. Measurement locations usually include ear lobe, fingertip, upper skin, forearm, mucosa layer. It has been shown in the literature [13, 24] that measurements can be made on the eye, finger and skin. It has been proposed that the difference between polarized light sent to optically active matters and reflected light from matters and the difference arising from angular rotation may provide an idea of the chemical and physical structure of the optically active substances [49]. With this method, high accuracy measurements can be made. It has also been shown that the signal used in this method is more tolerant to destructive effects due to its high energy [49]. Short wave length can be used and the penetration depth to the material is higher than the other methods. The depth of penetration is about 1-100 mm. The depth of penetration decreases with increasing wavelength. Because of the high depth of penetration, the infrared light used in this method is more interfered with the water, and therefore the water has a disruptive effect on measurements. Optical components using NIR wavelengths are less costly than MIR regions. In addition to all these; it is stated that there may be a time lag between increasing or decreasing level of glucose in the blood and determining this state by this method. This method requires a high variable analysis and high accuracy calibration, which is a negative feature. The main goal in the ongoing and continuing studies is to eliminate the destructive effect of non-glucose related parameters (blood pressure, temperature, blood flow rate, hemoglobin ratio, etc.).

RAMAN SPECTROSCOPY

Raman spectroscopy [53, 54] emerges as another method used in the literature. The method uses a wavelength range of 700nm-900nm. The glucose Raman spectrum can be distinguished from the Raman spectrum of other parameters in this wavelength region [21]. The Raman spectrum region describes a region that is sharper than the NIR spectral region and in which the signals do not overlap. Raman spectroscopy is based on the measurement of the slope of the light in the polarized laser beam. More specifically, in the Raman technique, the laser light causes the molecules to vibrate or rotate when they focus on the molecules in any solution. The emission of the laser light emitted from the molecules is determined by the amount of vibration and rotation. As a result, molecular densities are determined by utilizing the emission of laser light emitted by the amounts of vibration and rotation. In this method, water has less scattering parameters so has less disruptive effect on measurement. Sensitivity to temperature

changes is less, and weaker signals are detectable. Various Raman techniques have been applied on blood, water, serum, plasma, and eye, but many human cases have not been tested because of a lot of complications [11]. The simplest part of this technique is that it is easy and cheap to access the fixed wavelength laser beam. The negative side of this method is the low SNR ratio, the possibility that the high power laser beam used can be harmful to the human body and the data collection time is long. As in other methods, this method can also be referred to as the presence of interference. Raman technique is a method that can be used to distinguish different molecules (such as galactose and glucose) in the same solution instead of directly determining their molecular densities.

PHOTOACOUSTIC SPECTROSCOPY

The photoacoustic spectroscopy [9] is based on the fact that electromagnetic energy absorbed by a matter and causes heating. This heat generate sound wave or sound pressure due to thermal expansion. Briefly, this method based on the principle of light absorption of matter and principle of ultrasonic waves, which are caused by this absorbed light, are detected by ultrasonic sensors. The intensity of the glucose is calculated by means of these ultrasonic waves generated by the heat effect. Mostly NIR regions are used in the measurements. We have not found much application area in the literature.

OPTICAL COHERENCE TOMOGRAPHY

Another method used in the literature is optical coherence tomography. In this method, the polarized light wave is sent to the point to be measured while another light wave with the same characteristics is sent to another point of reference. For further information in [4] animal experiments were conducted. Densities can be determined by detecting and calculating the time difference between the two light waves obtained with the help of sensors [5]. As the concentration of glucose in the solution increases, the refraction index increases and a time delay appears between the reference point and the sample region. This time delay gives information about the densities directly. Although high resolution and high SNR ratio can be obtained with this method, the method is very sensitive to motion, which can affect the results [5]. The fact that technique penetrates deeply is a positive side. However, it is a negative aspect that the temperature differences that can occur on the skin can affect the accuracy of the measurements by changing the rotation of the polarized light.

ELECTROMAGNETIC SPECTROSCOPY

The electromagnetic method is based on the principle of determining the dielectric properties of the blood [29,30,31,32,41,42]. Inductor and electric current are used to determine these dielectric properties. An inductor is a kind of current source; a circuit element whose potential difference occurs between the two ends as the current passing through it changes. At different relative glucose ratios, the blood's total dielectric parameter will change, and the resulting current will change. Determination of the glucose level in the blood is done by analyzing the electromagnetic wave which is caused by this current. The frequency used is usually 2.4-2.9 MHz For better identification of glycosylated changes, the optimum frequency must be obtained [21]. The optimum frequency significantly affects the efficiency of the device being designed. This method is a reliable measurement technique and does not cause any ionization in tissue or blood. Negative aspects of the method are very sensitive to temperature changes. Because temperature changes the dielectric properties of the blood dramatically. Furthermore, the dielectric properties of total blood are not only determined by glucose. Therefore, more careful work is needed to calibrate.

FLUORESCENCE TECHNOLOGY

Another method is the fluorescence technique, in which short wavelengths of X-rays or ultraviolet rays are applied to optically active substances, based on the visible or invisible radiation emission of the materials. When human tissue is exposed to light at a wavelength of 308 nm, glucose molecules are stimulated and fluorescent emission of 340, 380, 400 nm is detected [14]. The concentration of NAD (P) H, which is the result of glucose oxidation in the cell, is monitored by fluorescent markers to measure glucose level. This method is very sensitive and visible light region can be used. In this method, there is no damage to the body surface. Negative aspects of the method are that they are greatly affected by time delay up to approximately one hour and strong scattering. With regard to this method, human experiments have not been done so far in the literature.

POLARIMETRIC METHOD

Polarimetry which is part of optical methods in the literature done by [20,28,30,33,37,38,44,47,51,56]. Polarimetry is the most useful method among non-invasive glucose measurement techniques [30]. Polarimetric systems measure the variation in intensity of an angular rotation and intensity of a transmitted light wave in any given sample. When a polarized light wave is sent to an optically active sample, the polarized light will undergo some changes (de-polarization) as it passes throughthe matter. Such changes can provide important information about the optically active substance's concentration, its chemical and physical structure, and the optical path that light follows in the substance. Due to the optical interaction, the plane of the polarized beam will deviate from the original plane by a positive or negative direction $\Delta\varphi$. If the optic sensors that are designed allow the deflection angle $\Delta\varphi$ can be determined depending on the concentration of the active substance, a variety of information can be obtained about the examined specimen. However, the study investigated is glucose, which is free in the blood of many active substances (amino acids, hormones, proteins and fat molecules, etc.). Although there are

high scattering parameters of human skin, there are many studies that have been done by the method of polarimetry [34,35,36,37,39,40,46,48,50]. While the visible light region or laser beam can be used in the polarimetry method and the optical components used in the detection can be present and miniaturized, the disadvantage is that the other components in the human skin and in blood the other components can distort the polarized light rotation.

The biggest problem in the method of polarimetry is how to influence the measurements of a wide variety of active substances (hemoglobin, water, hormones, oil molecules, etc.). Therefore, when measuring the glucose level in blood, these destructive effects will definitely have a positive or negative effect on the said deviation angle $\Delta\varphi$. Briefly, it is necessary to determine the highest accuracy of glucose present in blood by the very sensitive detectors, as well as to calibrate the detrimental effects by means of a strong calibration model in order to eliminate the undesirable effects.

BIO-IMPEDANCE TECHNIQUE

Another method is bio-impedance technique [15, 22], examines the electrical properties of biological tissues. With this feature it resembles an electromagnetic technique. Local increase in relative glucose concentration leads to decrease in plasma sodium level and increase in potassium level. This changes the dielectric properties of the plasma and changes the conduction and conductivity coefficients of the blood.

Dielectric properties of the glucose lead to changes in the total impedance of human blood [18]. This technique generally uses a frequency range of 0.1-100 MHz [21]. The dielectric properties of biological tissues are estimated from the voltage changes obtained by sending low currents to the region where the measurement is desired, and this is based on the technique of establishing a connection between glucose levels and the measurement of the dielectric spectrum of human blood. The method is practicable because it is low cost, simple, safe and have a fast response time. However, this technique is quite sensitive to changes in water rate and temperature changes in biological tissues, which require the patient to be calm for at least 60 minutes while resting [18,22].

THERMAL EMISSION SPECTROSCOPY

Another method, thermal emission spectroscopy [12], is to measure the infrared rays emitted by the human body, depending on the glucose concentration. When the local biological tissue illuminated by a light wave increases the temperature, the refractive index of the tissue also changes. This relationship between refractive index changes, light intensity and glucose ratio is attempted. The method reveals the relationship between light energy, spectral distribution, and the heat of the human body depending on Planck's rule. The most suitable measuring point in this method is the ear piece. In this method, it is an important feature that no calibration is needed and the method has high accuracy. However, IR (infrared) light emission of other human body components also appears to be a disruptive factor. There is not much application area in the literature.

TRANSDERMAL TECHNIQUE

Transdermal technique is based on the principle that the level of glucose at the side is derived from the intercostal fluid of glucose by various chemical components and the level of this glucose is measured [16]. The fact that the tissues and cells are not affected during this process causes this method to be regarded as a non-invasive method. By applying low DC voltage between two electrodes placed at the bottom, biological structures such as glucose to the interstitial fluid. The disadvantages are that the method needs at least 15 minutes delay time, 2 hours warm-up time and a strong calibration model.

ULTRASOUND TECHNIQUE

In the ultrasound technique, laser light in the low frequency region is used. Laser light is sent to the examined liquid to produce short waves of liquid. This method is still in the theory stage and there are no studies in which practical applications are made. In fact, it resembles a photoacoustic technique. In this method laser light is sent at short intervals to the region to be examined. The laser light causes the area to warm up. This warming event is also spreading as an ultrasonic wave. Ultimately, the ultrasonic wave is perceived by highly sensitive microphones. The best thing is that it can be used in any wavelength of the laser beam (UV-NIR). The method is very expensive and very sensitive to external factors. It is heavily influenced by pressure and temperature changes. There are also methods in which the glucose concentration changes the blood absorption and scattering parameters [30] and the glucose level is tried to be measured by measuring these parameters [25].

III. SUMMARY

Methods and technologies used in the determination of blood glucose level in non-invasive techniques with positive and negative aspects are compared and explained. A design that has been clinically proven to be reliable due to the presence of many other substances that absorb light or scattering materials and glucose in a heterogeneous environment has not yet been achieved. In experiments carried out in controlled laboratory conditions, the correlation between glucose concentrations and the control parameters and mathematical functions may be obtained. However, in real life, many uncontrolled parameters are introduced and appear to be detrimental to the determination of glucose concentrations. At this point it is important to have a strong calibration model in glucose detection techniques by using non-invasive method. The focus on concepts related to how to increase accuracy, how to eliminate distorting effects, how to set up a powerful calibration model, how to increase low SNR rates, and how to design the device in miniature

measurements, is much more than the discovery of new design methods in non-invasive studies and work continues in this direction.

REFERENCES

[1] Xudong Ge, Ph.D., Govind Rao, Ph.D., Yordan Kostov, Ph.D., Sunsanee Kanjananimmanont, B.S., Rose M. Viscardi, M.D., Hyung Woo, M.D., and Leah Tolosa, Ph.D., Detection of Trace Glucose on the Surface of a Semipermeable Membrane Using a Fluorescently Labeled Glucose-Binding Protein: A Promising Approach to Noninvasive Glucose Monitoring, Journal of Diabetes Science and Technology Volume 7, Issue 1, January 2013.

[2] Georgeanne Purvinis, Ph.D., Brent D. Cameron, Ph.D., and Douglas M. Altrogge, D.V.M., Noninvasive Polarimetric-Based Glucose Monitoring: An in Vivo Study, Journal of Diabetes Science and Technology, Volume 5, Issue 2, March 2011.

[3] Amos Mugweru, Ph.D., Becky L. Clark, B.S., and Michael V. Pishko, Ph.D., Electrochemical Sensor Array for Glucose Monitoring Fabricated by Rapid Immobilization of Active Glucose Oxidase within Photochemically Polymerized Hydrogels, Journal of Diabetes Science and Technology Volume 1, Issue 3, May 2007.

[4] Roman V. Kuranov, Ph.D., Veronika V. Sapozhnikova, Ph.D., Donald S. Prough, M.D., Inga Cicenaite, M.D., and Rinat O. Esenaliev, Ph.D., Prediction Capability of Optical Coherence Tomography for Blood Glucose Concentration Monitoring, Journal of Diabetes Science and Technology Volume 1, Issue 4, July 2007.

[5] Carl D. Malchoff, Kamal Shoukri, Julian I. Landau, Janusz M. Buchert, A Novel Noninvasive Blood Glucose Monitor, Diabetes Care, Volume 25, Number 12, December 2002.

[6] David C. Klonoff, Noninvasive Blood Glucose Monitoring, Diabetes Care, Volume 20, Number 3, March 1997.

[7] Vladimir L. Alexeev, Sasmita Das, David N. Finegold, and Sanford A. Asher, Photonic Crystal Glucose-Sensing Material for Noninvasive Monitoring of Glucose in Tear Fluid, Clinical Chemistry 50:122353–2360 (2004).

[8] Ok Kyung Cho, Yoon Ok Kim, Hiroshi Mitsumaki, and Katsuhiko Kuwa, Noninvasive Measurement of Glucose by Metabolic Heat Conformation Method, Oak Ridge Conference, Clinical Chemistry 50:10 1894–1898 (2004).

[9] Hugh A. MacKenzie, Helen S. Ashton, Stephen Spiers, Yaochun Shen, Scott S. Freeborn, John Hannigan, John Lindberg, and Peter Rae, Advances in Photoacoustic Noninvasive Glucose Testing, Clinical Chemistry 45:9, 1587–1595 (1999).

[10] Omar S. Khalil, Spectroscopic and Clinical Aspects of Noninvasive Glucose Measurements, Clinical Chemistry 45:2 165–177 (1999).

[11] R. W. Waynant, Ph.D. and V. M. Chenault, Ph.D., MT(ASCP), Overview of Non-Invasive Fluid Glucose Measurement Using Optical Techniques to Maintain Glucose Control in Diabetes Mellitus.

[12] Akesh Govada, Ch Renumadhavi, K B Ramesh, Non-Invasive Blood Glucose Measurement, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014.

[13] Jin Liu, Rong Liu, Kexin Xu, Accuracy of Noninvasive Glucose Sensing Based on Near-Infrared Spectroscopy, Applied Spectroscopy, Volume 69, Number 11, 2015, pp 1313-1318.

[14] Santhisagar Vaddiraju, Ph.D., Diane J. Burgess, Ph.D., Ioannis Tomazos, Ph.D., M.B.A., Faquir C. Jain, Ph.D., and Fotios Papadimitrakopoulos, Ph.D., Technologies for Continuous Glucose Monitoring: Current Problems and Future Promises, Journal of Diabetes Science and Technology Volume 4, Issue 6, November 2010.

[15] Andreas Caduff, Mark Talary, Lutz Heinemann, Yuri D Feldman, Non-invasive glucose monitoring in patients with diabetes: A novel system based on impedance spectroscopy, article in biosensors & bioelectronics January 2007.

[16] Losoya-Leal, Camacho-León, Dieck-Assad, Martínez-Chapa, State of the art and new perspectives in non-invasive glucose sensors, Revista Mexicana de Ingenieria Biomedica, Vol. XXXIII, Num. 1, Junio 2012, pp 41-52.

[17] Joseph Thomas Andrews, J. Solanki, Om P Choudhary, S. Chouksey, N. Malvia, P. Chaturvedi and P. Sen., Towards a Wearable Non-invasive Blood Glucose Monitoring Device, International Conference on Recent Trends in Physics (ICRTP 2012), Journal of Physics: Conference Series 365 (2012) 012004.

[18] Erdem Topsakal, Tutku Karacolak, and Elaine C. Moreland, Glucose-Dependent Dielectric Properties of Blood Plasma, 978-1-4244-6051-9/11/$26.00, 2011 IEEE.

[19] Orna Amir, Ph.D., Daphna Weinstein, M.D., Silviu Zilberman, Ph.D., Malka Less, M.Sc.,Daniele Perl-Treves, Ph.D., Harel Primack, Ph.D., Aharon Weinstein, M.Sc., Efi Gabis, B.Sc., Boris Fikhte, M.Sc., and Avraham Karasik, M.D., Continuous Noninvasive Glucose Monitoring Technology Based on Occlusion Spectroscopy, Journal of Diabetes Science and Technology Volume 1, Issue 4, July 2007.

[20] Roger J. McNichols, Brent D. Cameron, and Gerard L. Coté, Development of a Non-invasive Polarimetric Glucose Sensor, Biomedical Engineering Program, Texas A&M University.

[21] Chi-Fuk So, Kup-Sze Choi, Thomas KS Wong, Joanne WY Chung, Recent advances in noninvasive glucose monitoring, Medical Devices: Evidence and Research 2012:5 pp: 45–52.

[22] Anas Mazady, Non-invasive Glucose Meter, Electrical and Computer Engineering Department The University of Connecticut, Storrs, CT 06269-2157.

[23] Xinxin Guo, Andreas Mandelis, and Bernard Zinman, Noninvasive glucose detection in human skin using wavelength modulated differential laser photothermal radiometry, 1 November 2012 / Vol. 3, No. 11 / Biomedical Optics Express.

[24] Jyoti Yadava, Asha Rania, Vijander Singha, Bhaskar Mohan Murarib, Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy, Biomedical Signal Processing and Control 18 (2015) 214–227.

[25] Meixiu Sun, Nanguang Chen, Non-invasive measurement of blood glucose level by time-resolved transmission spectroscopy: A feasibility study, Optics Communications 285 (2012) 1608–1612.

[26] M.J. Tierney, J.A. Tamada, R.O. Potts, L. Jovanovic, S. Garg c, Cygnus Research Team, Clinical evaluation of the GlucoWatch® biographer: a continual, non-invasive glucose monitor for patients with diabetes, Biosensors & Bioelectronics 16 (2001) 621–629.

[27] Cerine Lal, Sujatha Narayanan, Correlation analysis of laser Doppler flowmetry signals: a potential non-invasive tool to assess microcirculatory changes in diabetes mellitus, Med Biol Eng Comput (2015) 53:557–566, DOI 10.1007/s11517-015-1266-y.

[28] L. Heinemann, G. Schmelzeisen-Redeker on behalf of the Non-invasive task force (NITF), Non-invasive continuous glucose monitoring in Type I diabetic patients with optical glucose sensors, Diabetologia (1998) 41: 848±854.

[29] Constantine A. Balanis, Advanced Engineering Electromagnetics, Arizona State University.

[30] K G Domnin, G A Cherevatenko, E T Aksenov, Angular measurements of light scattered by the glucose containing biological tissues and their phantoms, SPbOPEN2014, Journal of Physics: Conference Series 541 (2014) 012041.

[31] W. Kuang, S. O. Nelson, Low-Frequency Dielectric Properties of Biological Tissues: A Review With Some New Insights, Vol. 41(1):173-184, 1998 American Society of Agricultural Engineers.

[32] S. Gabriel, R.W. Lau and C. Gabriel, Dielectric Properties of Biological Tissues: II. Measurements in the Frequency Range 10 Hz to 20 Ghz, Phys. Med. Biol. 41 (1996) 2251-2269.

[33] Sunghoon Jang, Kenneth Markowitz, Hong Li, A New Approach to Present a Non-Invasive Optical Glucose Sensor Using Advanced Opto-Electronic Technology, New York City College of Technology of CUNY.

[34] Polarimetry, Chem 333L, Organic Chemistry Laboratory, Revision 1.3.

[35] Adler G. Santos, Quirino M. Sugon and Daniel J. McNamara, Polarization ellipse and Stokes parameters in geometric algebra, Vol. 29, No. 1 / January 2012 / J. Opt. Soc. Am. A.

[36] Dr. Theodore (Ted) Oakberg, Dr. Baoliang (Bob) Wang, Polarimetry: Optical Rotation, Hinds Instruments, Application Note.

[37] Shamaraz FIRDOUS, Masroor IKRAM, Polarized Mueller Matrix Analytical Model for Glucose Measurement in Vitro, Experimental / Laboratory Studies, Turk J Med Sci 35 (2005) 149-155, September 27, 2004.

[38] Sunghoon Jang, Hong Li, A New Method for A Non-Invasive Glucosesensing Polarimetry System, International Journal of Engineering Research & Innovation, Vol. 2, No. 1, Spring 2010.

[39] Russell A. Chipman, POLARIMETRY, Chapter 22, pp: 22.1-22.37.
[40] Stefan Böckle, Luigi Rovati and Rafat R. Ansari, Polarimetric glucose sensing using Brewster-reflection off of eye lens: Theoretical analysis, Proceedings of SPIE Vol. 4624 (2002) 2002 SPIE.
[41] Y.Ratna Kumar, Prof. G.S.N. Raju, Study of Characteristics for Dielectric Properties of Various Biological Tissues, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 1, January 2014.
[42] S Gabriely, R W Lau and C Gabriel, The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues, Phys. Med. Biol. 41 (1996) 2271–2293. Printed in the UK.
[43] Gerard L. Cote, Noninvasive and Minimally-Invasive Optical Monitoring Technologies, Biomedic Engineering Program, Texas A&M University, College Stat.ion, TX 77843-3120.
[44] Michael F. G. Wood, Nirmalya Ghosh, Xinxin Guo and I. Alex Vitkin, Towards noninvasive glucose sensing using polarization analysis of multiply scattered light, Division of Biophysics and Bioimaging, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto Toronto, Ontario, Canada.
[45] Md. Koushik Chowdhury, Anuj Srivastava, Dr. Neeraj Sharma, Dr. Shiru Sharma, Challenges & Countermeasures in Optical Noninvasive Blood Glucose Detection, International Journal of Innonative Research in Science, Engineering and Technology Vol. 2, Issue 1, January 2013.
[46] Darryl J. Bornhop and Stephen Dotson, Micro-scale polarimetry, Chapter-11, pp:343-360.
[47] Bilal H. Malik, Gerard L. Coté, Real-time, closed-loop dual-wavelength optical polarimetry for glucose monitoring, Journal of Biomedical Optics 15(1), 017002, January/February 2010.
[48] Michael F. G. Wood, Xinxin Guo, I. Alex Vitkin, Polarized light propagation in multiply scattering media exhibiting both linear birefringence and optical activity: Monte Carlo model and experimental methodology, Journal of Biomedical Optics 12(1), 014029, January/February 2007.
[49] A N Bashkatov, E A Genina, V. I. Kochubey and V V Tuchin, Optical properties of human skin, subcutaneous and mucous tissues in the wavelength range from 400 to 2000nm, journal of physics d: applied physics, J. Phys. D: Appl. Phys. 38 (2005) 2543–2555.
[50] Dmitry A. Markov, Kelly Swinney, Kristin Norville, David Lu, Darryl J. Bornhop, A Fourier analysis approach for capillary polarimetry, Electrophoresis 2002, 23, 809–812.
[51] Martin Hurtado and Arye Nehorai, Fellow, IEEE, Polarimetric Detection of Targets in Heavy Inhomogeneous Clutter, IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 56, NO. 4, APRIL 2008.
[52] Gerald L. CotC, Martin D. Fox, and Robert B. Northrop, Noninvasive Optical Polarimetric Glucose Sensing Using a True Phase Measurement Technique, IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, VOL. 39. NO. 7, JULY 1992.
[53] Narahara Chari Dingari, Ishan Barma, Gajendra P. Singh, Jeon Woong Kang, Ramachandra R. Dasari, Michael S. Feld, Investigation of the specificity of Raman spectroscopy in non-invasive blood glucose measurements.
[54] Omar S. Khalil, Non-Invasive Glucose Measurement Technologies: An Update from 1999 to the Dawn of the New Millennium, DIABETES TECHNOLOGY & THERAPEUTICS Volume 6, Number 5, 2004.
[55] Justın Shekwoga Baba, The Use of Polarızed Light for Biomedical Applications, Texas A&M University, August 2003.

A Survey on Predicting Survivability of Retinoblastoma on SEER Data

G. OZDEMIR OZDOGAN[1], H. KAYA[1], B. SEN[1] and I. CANKAYA[1]

[1] Ankara Yildirim Beyazit University, Ankara/Turkey, guldemiroz@gmail.com
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, hilalkaya@ybu.edu.tr
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, bsen@ybu.edu.tr
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, icankaya@ybu.edu.tr

***Abstract* - Retinoblastoma is a childhood cancer grows in retina. Although it could be treated in early stages, it can spread to nervous system and also other parts of the body and eventually may cause death in this situation. The prediction of survivability attracts a considerable interest and has been studied at different types of cancers, like breast, lung, colon and thyroid in literature by applying data mining methods. Data used in this study is obtained from The Surveillance, Epidemiology, and End Results (SEER) program which is an authorized data repository of cancer statistics. In our study, the survivability for retinoblastoma is predicted on SEER dataset consisting of 1258 patients by using data mining algorithms (support vector machines, logistic regression, multi-layer perceptron, naïve bayes, random forest and decision trees). Two strategies for imbalanced data which are over-sampling (synthetic minority over-sampling - SMOTE) and under-sampling are used. Results are analyzed and compared with the ones studied in other cancer types.**

***Keywords* - Retinoblastoma, SEER Data, Survival Prediction, Data Mining**

I. INTRODUCTION

RETINOBLASTOMA is an intraocular childhood cancer. It usually occurs in children before the age of five. It can affect one or both eyes. The most common sign of retinoblastoma is "leukocoria" which means a visible whiteness in the pupil. As leukocoria may be observed in clinical tests by doctors, it is generally noticed when parents take a photograph of their child with a flash. The second frequent symptom is lazy eye in which eyes appear to be looking in different directions. Furthermore, persistent eye pain, redness, irritation, blindness or poor vision can be seen. Early diagnosis is very important as in many other cancers. If retinoblastoma is not treated, the tumor may spread other parts of the body.

Retinoblastoma affects only one eye which is called *unilateral* or both of eyes called *bilateral* retinoblastoma. According to the research [1], 60% of the retinoblastomas is unilateral, while 40% of them is bilateral. Germinal forms of retinoblastoma may cause *trilateral* retinoblastoma (bilateral retinoblastoma with a brain tumor) and other cancers like lung, bladder or melanoma in later years. In treatment of retinoblastoma, chemotherapy, cryotherapy, thermotherapy, radiation therapy, enucleation and exenteration are used [2].

Since cancer causes death, patient's survival is very important after the diagnosis and treatment of cancer. Therefore, some historic information about the patient is collected to observe the survivability of patient. The information is needed to be kept for a period of time. Studies in the past have analyzed the survival time for ten or more years. However, due to the developments in the cancer technology, the survival time has been accepted as the five years, namely 60 months, after the time of diagnosis [3]. Some statistical methods such as Kaplan-Meier has been used to estimate survival rate. The topic of survival prediction has been studied by classification techniques in data mining. The aim is to predict the survivability best by using different classifiers.

Surveillance, Epidemiology, and End Results (SEER) Program is a large, comprehensive, reliable data repository on cancer incidence and survival. The data is accessible with submitting a request by signed SEER Research Data Agreement form over https://seer.cancer.gov/. SEER has different categories of data. Some of these are demographic (age, gender etc.), diagnosis (primary site, grade, tumor size, extension etc.), treatment (such as surgery, radiation), survival (vital status recode, survival time, cause of death) [4].

In this study, the survival prediction of retinoblastoma is discussed on SEER data. Data was preprocessed at first and then the survivability was predicted with popular algorithms used in survival prediction studies. Because we have used imbalanced data, under-sampling and oversampling were used before running algorithms to get balanced data. The results were evaluated by using performance metrics like accuracy, sensitivity, specificity and area under ROC curve. To the best of our knowledge, the topic of survival prediction on retinoblastoma has not been studied yet.

II. RELATED WORK

The most considerable work on prediction of survivability was studied by Delen et al. [3] on SEER breast cancer data. Three popular data mining methods which are multi-layer perceptron, decision tree(C5) and logistic regression were used with 10-fold cross-validation. They collected more than 400,000 breast cancer cases with 72 variables on SEER data set. They defined the survival attribute to determine if a

patient is alive after 60 months from the date of diagnosis. They pointed out that they spent more time on data preparation. In this process, they took into account the attribute of cause of death and whether the record was regularly followed throughout 60-months. At the end of data preparation process, their data had 202,932 cases with 17 variables including survival attribute. The best result was retrieved using the decision tree with accuracy 93.6%. Then, multi-layer perceptron and logistic regression followed decision tree, respectively. They also presented the sensitivity and specificity values of each algorithm. Similar work [5] was also studied on SEER breast cancer data from Public-Use data by Bellaachia and Guven. After data pre-processing, they retrieved 151,886 records with 16 variables. They studied the Naive Bayes, the back-propagated neural network (Multi-layer perceptron) and C4.5 decision tree algorithms. Their results pointed out that C4.5 algorithm was much better when analyzed in terms of accuracy, precision and recall. They worked with Weka by using 10-fold cross-validation and emphasized that they used survival time recode, vital status recode and cause of death while setting survival attribute.

Survival prediction has also been studied for other types of cancer in addition to breast cancer. For example, SEER prostate cancer data was studied by support vector machine as well as decision tree, artificial neural network (MLP) and logistic regression in [6]. The results indicated that support vector machine was better than others. Artificial neural networks and decision trees followed support vector machines. SEER thyroid cancer data for survival prediction was studied with artificial neural network and logistic regression in [7]. A newer research about prediction survivability was studied in comorbidity cancers [8]. In this study, random forest was used as well as neural networks, decision tree and logistic regression. In the end of the study, it was indicated that the performance of random forest was better than the other algorithms. Besides, survival prediction was studied using ensemble data mining in lung [4] and colon cancer [9].

Imbalanced data problem was also studied on survival prediction. Liu et al. [10] studied C5 decision tree classifier with under-sampling technique and bagging algorithm on imbalanced breast cancer data. They obtained better results with bagging and under-sampling in terms of area under ROC curve. In [9], the authors mentioned that they had imbalanced data and they used synthetic minority over-sampling technique (SMOTE) [11] to balance data. They also emphasized the importance of balancing data to get better results. In another study [12], SMOTE, cost-sensitive classifier (CSC), under-sampling, bagging and boosting techniques were used to overcome the negative effect of imbalanced breast cancer data from SEER. The authors pointed out that logistic regression and decision trees with SMOTE, CSC and under-sampling techniques gave better performance compared to ones without using these techniques.

To study survival prediction on retinoblastoma, some research about survival rates has been made. Survival of retinoblastoma in USA was studied in [13] and the 5-year survival time was found as 92-97% for the years 1975-2004. For Turkey, 141 retinoblastoma patients from Cerrahpasa Medical Faculty were retrieved between 1981 and 2004 in [14]. The authors found 3-year survival rate as 89.69%. In a newer study [15], the treatment trends and overall survival time were analyzed on SEER data between 1975 and 2010. The authors pointed out that there was increase in chemotherapy and decrease in radiation therapy in the specified time interval. Also, they calculated the 5-year survival time as 94-98% by Kaplan-Meier method.

III. DATA PREPARATION

In our study, SEER*Stat software [16] was used to retrieve SEER data. SEER*Stat is a software that was developed for providing the researchers to compute the frequency distributions, incidence rates, observed and relative survival rates [17]. Case listing session was used on Incidence-SEER 18 database [18]. Cases in database were between the years 1973 and 2013. Malignant cases were selected in research database. We selected as many categories of variables as possible, because we did not want to miss any important variables. According to our selections, case listing matrix with 1791 cases and 202 variables was created. Since there were many redundant and irrelevant information, data was analyzed carefully to eliminate all of these and the variables which had missing and constant values. In addition to elimination, some new variables were also created by combining the existing ones.

In this section, all of these steps in data preparation will be mentioned below:

Irrelevant variables: In this step, variables defined for other cancers and also those which were not in concept of the study such as insurance recode and health service area information were eliminated. Additionally, even though the variable named *marital status at diagnosis* is a property used in similar studies, it was not used in our study.

Variables with constant value: When analyzing data, due to our selection criteria, it was noticed that there were some variables with constant value like site recode, ICCC site recode, behavior code etc. Each of them was eliminated.

New variables created at specific years: New variables have been created at specific years. So, the records, which were created before that year (created retrospectively), consist no value for these variables. Because they contain the high rate of missing values, these variables were not used. On the other hand, some new variables were also created by redefining the old variables for different years. These were important and should be taken into consideration.

Redundant information: There were some redundant information. For example, although age is stored in three different variables, *age at diagnosis* was only used.

Derived variables: Besides the elimination process, four variables were derived. For the first three, the existing nine variables defined for different time intervals were used. When analyzing data and reading variable documentation, it was noticed that three new attributes for tumor size, extension and

lymph nodes could be created. These variables are defined after 1983. Since the data starts in 1973, for the cases between 1973 and 1982, there are blank values. Blanks were changed with suitable value that is defined for that variable. The fourth variable named "isSurvived" was created by using the variables named *vital status recode (VSR), survival months (SM)* and *SEER other cause of death classification (COD)*. Derived variables are summarized in Table 1. After this step, only derived attributes were used instead of the existing ones.

Table 1: Derived variables and their statements

Variable Name	**Statement**	
Tumor Size	EOD 4 - size (1983-1987), EOD 10 - size (1988-2003), CS tumor size (2004+)	
Extension	EOD 4 - extent (1983-1987), EOD 10 - extent (1988-2003), CS extension (2004+)	
Lymph Nodes	EOD 4 - nodes (1983-1987), EOD 10 - nodes (1988-2003), CS lymph nodes (2004+)	
isSurvived	Survived	SM>=60 AND VSR='Alive'
	Not Survived	SM < 60 AND VSR = 'Dead' AND COD='Dead due to cancer'

As well as eliminating the variables, some cases which are not be able to observe during 60 months or the cause of death is not this cancer were ignored. At the end of data preparation process, 1258 cases and 17 variables (listed in Table 2) were got. Since there were only 57 "not survived" cases, obtained data was imbalanced. The ratio of "not survived" class was 4.5%.

IV. METHOD

Our aim is to predict survival model on retinoblastoma data. To do it, six different algorithms that are popular in predicting the survival prediction studies such as decision trees (DT) (C4.5, [19]), logistic regression (LR) [20], support vector machines (SVM), random forest (RF) [21], multi-layer perceptron (MLP), naïve bayes (NB) were used. Artificial neural networks, decision trees, support vector machines and bayesian networks were also presented as widely used machine learning methods in cancer prediction in [22].

Table 2: All variables used in data and their types

Variable Name	**Type of Variable**
Sex	Nominal
Age at diagnosis	Numeric
Race/ethnicity	Nominal
Primary Site	Nominal
ICD-O-3 Hist/behav	Nominal
Grade	Nominal
SEER Historic Stage A	Nominal
Laterality	Nominal
Sequence Number	Nominal
Diagnostic Confirmation	Nominal
Tumor Size	Numeric
Extension	Nominal
Lymph Nodes	Nominal
Radiation sequence with surgery	Nominal
Reason no cancer-directed surgery	Nominal
Radiation	Nominal
isSurvived	Nominal

To develop a prediction model, the data needs to partitioning. As data is divided with holdout sampling, k-fold cross validation can also be used. In k-fold cross-validation, data is divided into k partitions. Each k-1 partition is used for training and the k^{th} partition is used for testing. This process is repeated for k times, and average of them is got as performance measure. The advantage of the method is that each case is able to be used as both training and testing data.

All predictions were made in Weka (3.8 version) [23] which is a tool for data mining. Data was divided by the approach of 10-fold cross validation. J48 for C4.5 decision tree and SMO for support vector machine were used in Weka.

Since data was imbalanced, two techniques called under-sampling and oversampling were used to overcome the problem of imbalance. For oversampling, SMOTE algorithm was used. In SMOTE, the aim is to increase the size of minority class by creating new synthetic samples. In our study, the percentage of over-sampling and nearest neighbors were set to 2000% and 5, respectively. For under-sampling, *SpreadSubsample* filtering was used in Weka and the class distribution was set to uniform distribution.

The results of the algorithms were evaluated in terms of accuracy (Eq. 1), sensitivity (Eq. 2), specificity (Eq. 3) and area under ROC curve.

$$Accuracy = \frac{TN+TP}{TN+FP+TP+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

where TP, FP, FN, TN denote true positive, false positive, false negative and true negative, respectively.

V. RESULTS

At first, all algorithms were run without using any imbalance strategies (in Table 3). In this case, they were not as successful as enough to classify the not survived class despite high accuracy values. However, it was noticed that naïve bayes gave a bit better results on not survived class.

Table 3: Results of algorithms on imbalanced data

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	ROC Curve
LR	94,44	98,25	14,04	0,67
C4.5	95,47	100,00	0,00	0,48
RF	94,99	99,25	5,26	0,82
SVM	95,55	99,67	8,77	0,54
MLP	93,56	97,67	7,02	0,79
NB	92,69	95,42	35,09	0,85

Then, same algorithms were run with oversampling (SMOTE) (in Table 4) and under-sampling (in Table 5) techniques.

Table 4: Results of algorithms after SMOTE

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	ROC Curve
LR	93,87	91,76	95,99	0,97
C4.5	96,41	96,17	96,66	0,98
RF	97,16	98,00	96,32	0,99
SVM	94,41	92,42	96,41	0,94
MLP	96,00	95,92	96,07	0,98
NB	91,24	85,68	96,83	0,96

By using oversampling with SMOTE, fairly better results were obtained. Random forest became the best one when analyzing accuracy, sensitivity, specificity and also area under ROC curve values. Decision tree and MLP methods followed Random forest.

Table 5: Results of algorithms after under-sampling

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)	ROC Curve
LR	65,79	77,19	54,39	0,67
C4.5	64,91	77,19	52,63	0,70
RF	76,32	78,95	73,68	0,84
SVM	76,32	77,19	75,44	0,76
MLP	78,95	80,70	77,19	0,82
NB	73,68	73,68	73,68	0,84

In under-sampling, MLP was the best one according to accuracy values. Despite long running time of MLP, it did not become a problem since data set was small. Random forest and SVM followed MLP. According to ROC curve values, random forest and naïve bayes, MLP, SVM gave three best results. When the results were compared with the ones of SMOTE, values of performance metrics decreased because of decreasing the size of survived class.

Finally, quite better results were retrieved when SMOTE was used instead of under-sampling. As decision tree results well in studies of survival prediction, it gave good results when it was run with SMOTE in our study.

VI. CONCLUSION

In our study, prediction of survival model of retinoblastoma was surveyed on SEER data. Some popular classifiers which are decision trees, logistic regression, support vector machines, random forest, multi-layer perceptron, naive bayes were studied. To overcome the imbalance, two strategies named oversampling (SMOTE) and under-sampling were used. The results indicated that classifiers with SMOTE technique gave the best results. On the other hand, random forest gave better results when it was compared with the other classifiers. As presented in literature, decision tree gave also quite good results when there was balanced data. As a future work, we would like to continue to study the problem of imbalance with genetic algorithms or deep learning.

REFERENCES

[1] G.J. Draper, B.M. Sanders, P.A. Brownbill and M.M. Hawkins, "Patterns of Risk of Hereditary Retinoblastoma and Applications to Genetic Counselling", British Journal of Cancer, 66(1), 211-219, 1992.

[2] S. Tuncer, "Retinoblastom", Klinik Gelişim Dergisi, 25, 56-65, 2012.

[3] D. Delen, G. Walker and A. Kadam, "Predicting breast cancer survivability: a comparison of three data mining methods", Artificial Intelligence in Medicine, 34(2), 113-127, 2005.

[4] A. Agrawal, S. Misra, R. Narayanan, L. Polepeddi and A. Choudhary, "Lung cancer survival prediction using ensemble data mining on SEER data", Scientific Programming, 20(1), 29-42, 2012.

[5] A. Bellaachia and E. Guven, "Predicting breast cancer survivability using data mining techniques", in 9th Workshop on Mining Scientific and Engineering Datasets in Conjunction with the 6th SIAM International Conference on Data Mining, 2006.

[6] D. Delen, "Analysis of cancer data: a data mining approach", Expert Systems, 26(1), 100-112, 2009.

[7] M. Jajroudi, T. Baniasadi, L. Kamkar, F. Arbabi, M. Sanei, M. Ahmadzade, "Prediction of Survival in Thyroid Cancer Using Data Mining Technique", Technology in Cancer Research & Treatment, 13(4), 353-359, 2014.

[8] H.M. Zolbanin, D. Delen and A.H. Zadeh, "Predicting overall survivability in comorbidity of cancers: A data mining approach", Decision Support Systems, 74, 150-161, 2015.

[9] R. Al-Bahrani, A. Agrawal and A. Choudhary, "Colon cancer survival prediction using ensemble data mining on SEER data", in 2013 IEEE International Conference on Big Data.

[10] Y-Q. Liu, W. Cheng and Z. Lu, "Decision tree based predictive models for breast cancer survivability on imbalanced data", in 2009 3rd International Conference on Bioinformatics and Biomedical Engineering, 312-315.

[11] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique", Journal of Artificial Intelligence Research, 16, 321-357, 2002.

[12] K-J. Wang, B. Makond and K-M. Wang, "An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data", BMC Medical Informatics and Decision Making, 13, 124, 2013.

[13] E. Broaddus, A. Topham, A.D. Singh, "*Survival with retinoblastoma in the USA: 1975–2004*", British Journal of Opthalmology, 93(1), 24-7, 2009.

[14] A. Ozkan, H. Pazarli, T. Celkan, S. Karaman, H. Apak, G. Kaner, O. Uzel and I. Yildiz, "Retinoblastoma in Turkey: survival and clinical characteristics 1981-2004", Pediatrics International, 48(4), 369-73, 2006.

[15] D. Tamboli, A. Topham, N. Singh and A.D. Singh, "Retinoblastoma: A SEER Dataset Evaluation for Treatment Patterns, Survival, and Second Malignant Neoplasms", American Journal of Ophthalmology, 160(5), 953-958, 2015.

[16] Surveillance Research Program, National Cancer Institute SEER*Stat software (www.seer.cancer.gov/seerstat) version 8.3.2.
[17] B.F. Hankey, L.A. Ries and B.K. Edwards, "The surveillance, epidemiology, and end results program: a national resource", Cancer Epidemiology, Biomarkers & Prevention, 8(12), 1117-1121, 1999.
[18] Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data + Hurricane Katrina Impacted Louisiana Cases, Nov 2015 Sub (1973-2013 varying) - Linked To County Attributes - Total U.S., 1969-2014 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2016, based on the November 2015 submission.
[19] J. Quinlan, *C4.5: programs for machine learning*. San Mateo, CA: Morgan Kaufmann, 1993.
[20] T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning*. New York, NY: Springer-Verlag, 2001.
[21] L. Breiman, "Random forests", Machine Learning, 45(1), 5-32, 2001.
[22] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis and D.I. Fotiadis, "Machine learning applications in cancer prognosis and prediction", Computational and Structural Biotechnology Journal, 13, 8-17, 2015.
[23] Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

Design Of A New Robot Manipulator For MALDI Sprayer

B.D.GÜNGÖR*, S. KÜÇÜK and M. KASAP

Kocaeli University, Kocaeli/Turkey, baris.gungor@kocaeli.edu.tr
Kocaeli University, Kocaeli/Turkey, skucuk@ kocaeli.edu.tr
Kocaeli University, Kocaeli/Turkey, mkasap@ kocaeli.edu.tr

***Abstract* – Various coating methods are used in the development of organic textures. The MALDI method is one of the most important one amongst the others. Three-axis robotic spraying devices are frequently used in tissue engineering. The inboard joints of the robot manipulator are designed with prismatic joints. The last two joints are designed with two revolute joints which are composed of a two-axes spherical wrist. With this design, complex surfaces can be reached with any orientation angle.**

***Keywords* – MALDI imaging, Tissue engineering, Coating.**

I. INTRODUCTION

SPRAYING devices are used to apply various biomaterials on TISSUE. These devices perform a process of spraying various biomaterials onto certain sized textures. There are various difficulties in the spraying process applied to the tissue in studies carried out in the field of tissue engineering. The roughness of the textured surface is the most fundamental problem encountered. The roughness of the tissue surface prevents the homogeneous application of the biomaterial to be applied to the tissue. Various methods of spraying are used in biochemistry, organic chemistry, polymer chemistry, microbiology, pharmacology, clinical chemistry and forensic medicine. There are major methods of spraying used in tissue engineering.

II. MATERIALS AND METHODS

A. Maldi Imaging Technique

The matrix-assisted laser desorption ionization (MALDI) imaging is based on the examination of various analysts in various materials that are intended for imaging mass spectrometry. It is a visualization method that visualizes the spatial distribution of peptides, pharmaceutical compounds and proteins obtained from fine tissue slices, usually from humans, animals or plants (Figure 1). Matrix-assisted laser desorption ionization mass spectrometry imaging techniques and applications have been highly developed and gained importance in recent years [1-3].

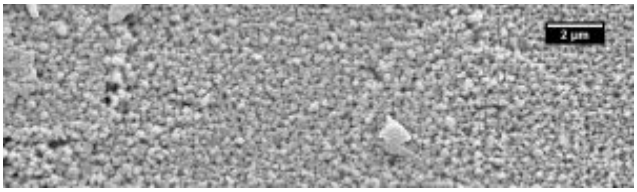


Figure 1: Tissue examination with MALDI imaging system.

Mass spectra of molecules are obtained by applying appropriate matrices to the tissues. Three-dimensional images of molecules are generated by MALDI imaging technique. Mass spectrometry makes it easier to identify the molecular diversity. This imaging system makes it easier to determine the boundaries of the tissue to be sprayed. It is simple to determine the amount of sputum to be applied to the tissue extracted by mass spectrometry [4-5].

B. TLC Sprayer Spraying Method

It is a technique where minimum homogeneity is targeted between matrix layers. When sprayed on the tissue, the matrix layer will also contain large droplets, since the droplets will not be uniform even if they are very small (Figure 2).

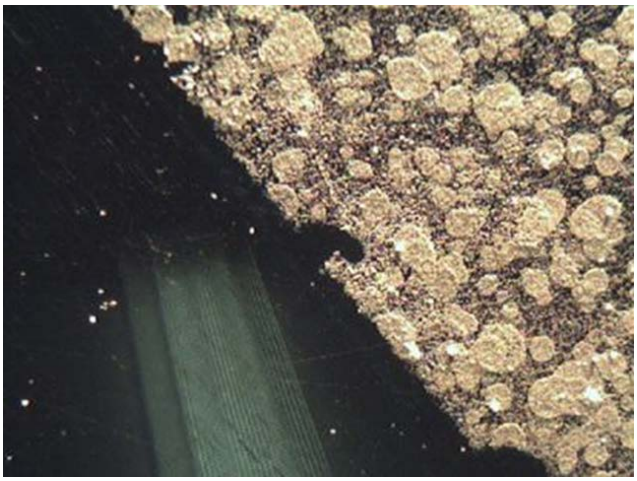


Figure 2: Matrix coating with TLC spraying.

This will result in an inhomogeneous coating. It is not a convenient method for single cell analysis because cell size is one micron (Figure 3) [6].

Figure 3: TLC sprayer spraying.

C. Micro Spotting Technique

Drips are sprayed on the textured surface in the size of picoliter. In this technique, the spraying nozzles are often clogged when using matrix solubilizes. Drop diameters are 150 micrometers and distances between droplets are between 20 and 250 micrometers (Figure 4). Therefore, this method is not suitable for single cell analysis [7].

Figure 4: Micro spotting [7]..

D. Ultrasonic Spraying

It forms droplets in different sizes from the sprayed matrix solution. Sprayed droplets of different size fall into different spots.

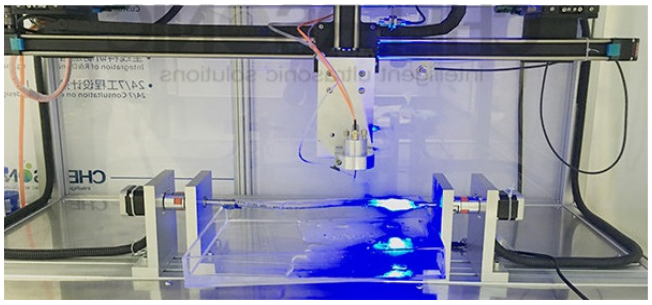


Figure 5: Ultrasonic spraying mechanism.

Big drips and small drips are different points. For this reason, the matrix concentration and droplet size are not homogeneous Sprayed layers are sprayed by drying with nitrogen gas (Figure 5). There is a detector that provides information about the structure and the thickness of the matrix layer [8].

E. Electrospray Technique

This technique is applied by spraying the matrix solution from the spray tip. During the spraying process, an electric field occurs between the sprayer and the target area (Figure 6) [9-10].

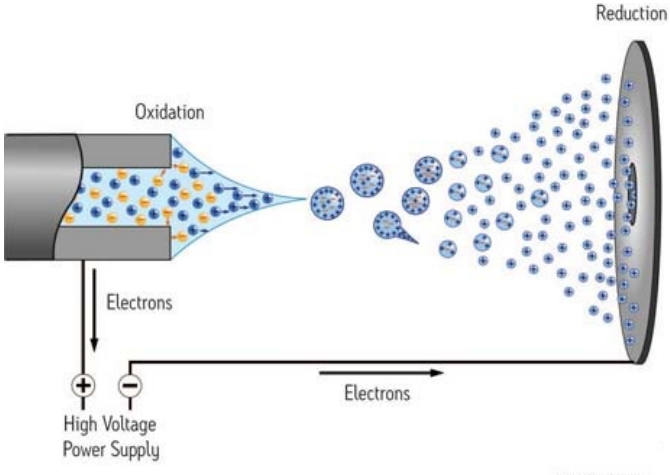


Figure 6: Electrospray method.

When a constant potential area is created, a homogeneous spray can be achieved (Figure 7). It can be applied to proteins and peptides [11-12].

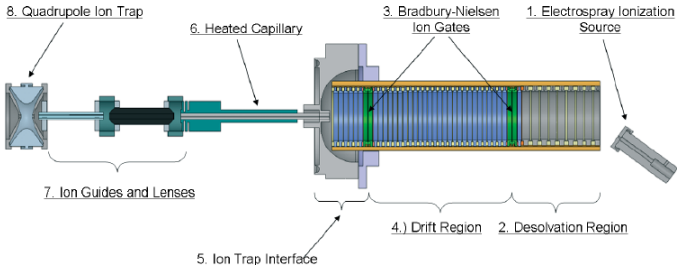


Figure 7: Electrospraying scheme.

F. Sublimation and Crystallization

It is a method that is applied with a coating of matrix components and many small crystals on the tissue. The matrix components crystallize from the gas phase on the tissue. This is not used for small molecules such as peptides and proteins as it will strain the extrusion. Instead, they are used for lipids and evaporable molecules [13]. The matrix solution is applied to the entire texture. For this reason, it is in the form of a layer on the tissue and the duration of the precipitation is short. The amount of solution may be too much as it is applied to the texture and the whole tissue is applied. Not all types of matrix solutions are suitable for use in this method (Figure 8) [6].

Figure 8: Tissue coating with sublimation method.

G. Sublimation and Recrystallization

It is the provision of the extrusion phenomenon by keeping the sublimated layer in a moist environment for a long time. The sublimated tissue is kept in a humid environment between 24-72 hours (Figure 9) [14].

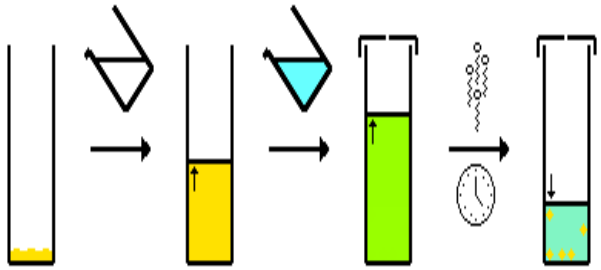


Figure 9: Recrystallization scheme.

III. CONCLUSION

The biomaterials or matrix solubles to be applied on the tissues in varying amounts will be sprayed onto the tissue using compressed air or nitrogen gas with the sprayer of the device. The end function of the new type of spraying device planned to be designed will be designed as a Cartesian robot arm that will provide movement in x, y, z coordinates. The internal connections of the robot manipulator are designed with prismatic connections. The last two joints are designed with two swivel joints which form a biaxial spherical joint. The end function will provide movement from the biaxial prismatic motion and the circular connection point. Three-axis motion will be provided on this vault. A design is planned to achieve complex surfaces with any orientation angle. After the coordinates of the taps are entered with the aid of an interface, they will be sprayed onto the spraying tissue via a cartesian robot [16-17].

REFERENCES

[1] S. C. Özen Karataylı, A. M. Bozdayı, "Proteomics and Gastroenterology", Current Gastroenterology, Hepatology Institute, Ankara.
[2] E. Gemperline, L. Li, "MALDI-Mass Spectrometric Imaging For The Investigation Of Metabolites In Medicago Truncatula Root Nodules", Department Of Chemistry, University Of Wisconsin-Madison, School Of Pharmacy, University Of Wisconsin-Madison.
[3] J. Hanrieder, A. Ljungdahl, M. Andersson, "MALDI Imaging Mass Spectrometry of Neuropeptides in Parkinson's Disease", Journal of Visualized Experiments, 2012.
[4] Y. Merit, K. K. Murray, "Multi-Component Matrix Sublimation Chamber For MALDI Tissue Imaging", Louisiana State University.
[5] A. Walch, S. Rauser, S. O. Deininger, ·H. Höxer, "MALDI Imaging Mass Spectrometry For Direct Tissue Analysis: A New Frontier For Molecular Histology", Histochem Cell Biology, 2008.
[6] M. Setou, "Imaging Mass Spectrometry: Protocols for Mass Microscopy", Hamamatsu University.
[7] T. H. Jupille, J. A. Perry, "Programmed Multiple Development: Independence Of Spot Placement And Size From Spotting Technique", Journal Of Chromatographic Science, Volume 13, Issue 4.
[8] R. Charbonneau, M. Tencer, N. Lahoud, P. Berini, "Demonstration Of Surface Sensing Using Long-Range Surface Plasmon waveguides On Silica", Sensors And Actuators, 2008, 455–461.
[9] S. J. Gaskell "Electrospray: Principles And Practice", Journal Of Mass Spectrometry, Vol. 32, 1997.
[10] C. Ho, C. Lam*, M. Chan, R. Cheung, L. Law, L. Lit, K. Ng, M. Suen, H. Tai, "Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications", Clin Biochem Rev., 2003.
[11] M. Wilm, "Principles of Electrospray Ionization", Molecular & Cellular Proteomics, Vol. 10, 2011.
[12] A. B. Kanu, P. Dwivedi, M. Tam, H. H. Hill, "Ion Mobility-Mass Spectrometry. J Mass Spectrom 43:1", Journal of Mass Spectrometry, Jan 2008.
[13] Joseph A. Hankin, Robert M. Barkley, and Robert C. Murphy, "Sublimation as a Method of Matrix Application for Mass Spectrometric Imaging", Journal of the American Society for Mass Spectrometry, Vol. 18, Issue 9, 2007.
[14] J. Yang, R. M. Caprioli, "Matrix Sublimation/Recrystallization for Imaging Proteins by Mass Spectrometry at High Spatial Resolution", Analytical Chemistry, 2011.
[15] R. C. Murphy, J. A. Hankin, R. M. Barkley, K. A. Z. Berry, "MALDI İmaging Of Lipids After Matrix Sublimation/Deposition", Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, Vol. 1811, Issue 11, 2011.
[16] http://www.chemeurope.com/en/whitepapers/126355/sample-preparation-for-maldi-imaging-with-suncollect-system.html
[17] https://sunchrom.de/

Investigation of Human Femoral Head by Wide Angle X-Ray Scattering (WAXS) Measurements

T.ÇAYIR*, [1], D.KARAARSLAN[2], K.MEMİŞOĞLU[1], S.İDE[2] and Ö.GÜNDOĞDU[1]

[1] Kocaeli University, Kocaeli/Turkey, tubacyr@hotmail.com
[2] Hacettepe University, Ankara /Turkey, karaarslandamla@gmail.com
[1] Kocaeli University, Kocaeli/Turkey, kayamemis@kocaeli.edu.tr
[2] Hacettepe University, Ankara /Turkey, semraide2015@gmail.com
[1] Kocaeli University, Kocaeli /Turkey, o.gundo@gmail.com

Abstract **- Osteoarthritis (OA) is the most common form of joint disease, and its impact is set to grow as the prevalence of obesity rises and our elderly population increases. As a result of this disease, it is necessary to make structural examinations for the determination of the deformations that occur. One of the methods used for structural analysis of biological samples is the wide-angle X-Ray scattering (WAXS) device. In this study, hip bone samples taken from humans during surgery were investigated using wide-angle X-ray scattering technique. Hip bone specimens were primarily compared based on regional density and later on based on sex. As a result of scatter patterns, the presence of crystal structures has been determined. The values of crystal structure obtained depending on low and high dense region and sex are different. As a result of gender related comparisons, it was found that crystal structures were better in males than females.**

Keywords **- Wide-Angle X-Ray Scattering, Osteoarthritis, Human Femoral Head.**

I. INTRODUCTION

The progress in structural studies with small-angle X-ray scattering (SAXS/WAXS) of biological macromolecules in solution is best reflected in the substantial increase of publications based on this technique over the last years. This gain in popularity can easily be explained by the synergistic improvement in hardware as well as software resulting in an automated data collection and in-depth analysis of the scattering data [1–3]. In recent years, the use of X-ray scattering methods in the analysis of size, shape and distribution of nanostructures in complex structures that are subject to biochemistry / biophysical researches has become widespread in recent years as the result of developing electro-mechanical device technology to experimental systems. SAXS / WAXS measurements, along with techniques such as XRD and NMR, also have an effective use area (SAXS. sources for bone tissue can be added). Optimization of the assay equipment for biological specimens to be examined and the highest performance improves the quality and sensitivity of the investigations. all experimental possibilities should be developed and used for the detailed examination of biological samples containing rich nano formation such as bone, cartilage, hair, tendon, dentin, retina. In particular, sample orientations should be determined by optimizing the absorbent and dispersive properties of the material. In the GISAXS mode for materials with high absorption characteristics, measurements are made in the Transmission mode for low scattering and good scattering samples. If the material is stable, the measurement period can be long while high-wise measurements can be made. If the long-term measurement and high flux will damage the biological sample to be examined (soft tissues such as the retina), these problems can be eliminated by using a capillary tube or appropriate specimen holders. The main purpose of the measurements is to optimize all the experimental parameters and to calculate the best quality. For this reason, each biological sample is passed through an intensive testing process before the measurement. The use of synchrotron radiation provides a high-flow, point collimated X-ray beam in the bone-cartilage in passages, facilitating the scanning of biological specimens with micrometric translocations starting from the bone region to the end of cartilage tissue. Since the line collimation X-ray is used in the HECUS SWAXS system, the sample is prepared as follows.

That is, the line interacts with all the different structural features of the collimation beam material. In some cases, the porous internal region is defined as less dense in terms of the tissue matrix, and the more dense structure in the cartilaginous region. The transmission mode, which gives the best scattering profile of the samples, is used for measurements. First, I (q) -q graph obtained by 3D View program is drawn and programs such as Igor Pro, PCG, Origin Pro and ATASAS Packet programs are used for detailed analysis [4-8].

Wide angle X-ray scattering (WAXS) was utilized on a decalcified human femoral head section affected by OA, direct measurements being made of spatial alterations of collagen fibres. Osteoarthritis is a joint disease that affects most people in the middle and elderly ages.

The cartilage structure of the joint due to aging changes. Being overweight, fractures in cartilage due to falling or some other mechanical trauma, followed by damage to the menus and ligaments, narrowing of the joint space and new bone formation.

Here, we present WAXS data obtained from the diseased bone-cartilage samples taken from male and female with the aim of extracting local and long-range structural information.

II. Materials and Method

Human femoral heads from total hip replacement procedures were used, due to surgical intervention in response to a degenerative joint disease. The femoral heads were sourced from the department of Orthopedics and Traumatology at Kocaeli University (Turkey). Using a water-cooled diamond saw (Isomet 1000 Precision) several sections at certain thicknesses were cut perpendicular to the articular surface from the superior aspect of the femoral head with thicknesses ranging from 200µm to 300µm. These sections were subsequently soaked in distilled water to remove any bone marrow and loose particulate matter.

Wide Angle X-ray scattering (WAXS) experiments were performed with a Kratky compact Hecus (Hecus X-ray systems, Graz, Austria) system equipped with a linear collimation system and X-ray tube Cu target (λ= 1.54 Å.)The generator was operated at a power of 2 kW (50 kV and 40 mA). Distances between channels and the sample-detector are 54 µm and 27.9 cm, respectively. All samples in the form of thin layer (size:5x3x1mm) were measured in transmission mode. The HECUS SWAXS test system is a system with Kratky Optics. Test system settings and calibration procedures must be performed before starting measurements. "Subilant" test samples were used for WAXS zone calibration. The detectors of the Hecus System3 device we use are line detectors and consist of 1024 channels. On the detector-related data screen, the magnitudes of the scattered vector of the photons falling on the horizontal axle channels and the intensity values in the vertical axis should be calibrated. The peak values of the silver stearate and subilantine test specimens are characteristic and are known in advance. Based on these values, the horizontal axle channels are calibrated.

III. Result

Figure 1and 2 shows a bone cartilage specimen taken from a female and male patient. 1 mm sections were cut from the bone-cartilage sample taken from the femur head and analyzed.

Figure 1: Sample of a bone-cartilage taken from a female patient.

Figure 2: Sample of a bone-cartilage taken from a male patient.

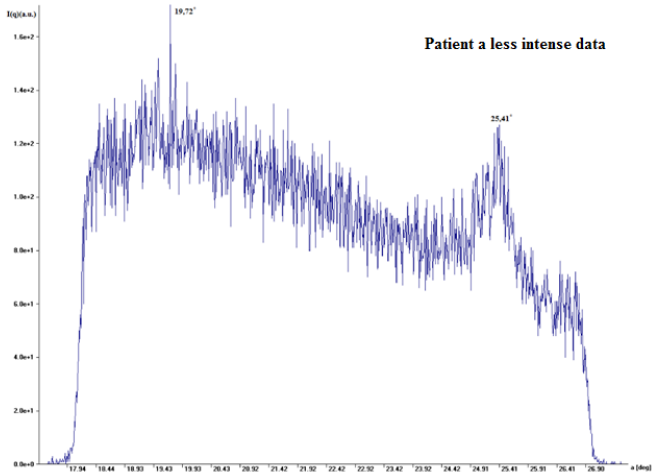


Figure 3: I(q)(a.u)-θ(°) chart for a less dense cartilaginous tissue specimen of a bone-cartilage sample from a female patient.

In Figure 3, the two-peak θ values for the female patient with less dense cartilaginous tissue samples were 19.72 ° and 25.41 °, respectively.

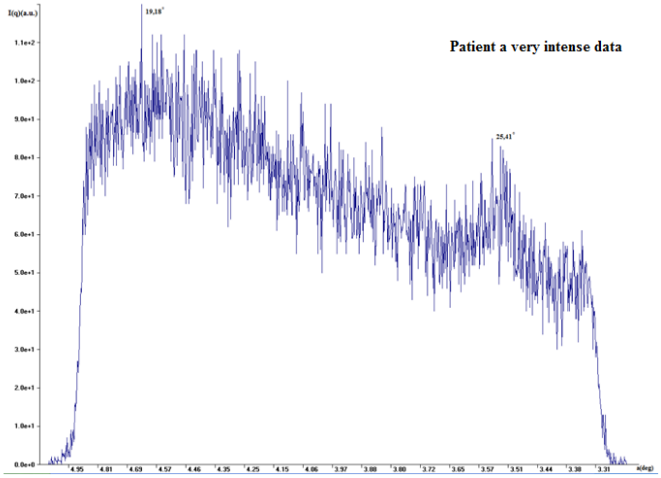


Figure 4: I(q)(a.u)-θ(°) chart for a very dense cartilaginous tissue specimen of a bone-cartilage sample from a female patient.

In Fig. 4, the two peak θ values for female patient with very dense cartilaginous tissue samples were determined as 19,18 ° and 25,35 °, respectively. The differences in the calcification touch in males indicate a very different crystalline structure in the WAXS profiles.

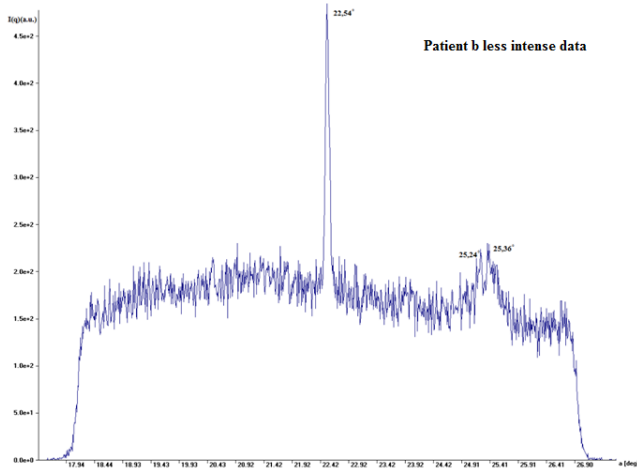


Figure 5: I(q)(a.u)-θ(°) chart for a less dense cartilaginous tissue specimen of a bone-cartilage sample from a male patient.

In Figure 5, the three-point θ values for male patient with less dense cartilaginous tissue samples were determined to be 22.54 °, 25.24 ° and 25.36 °, respectively.

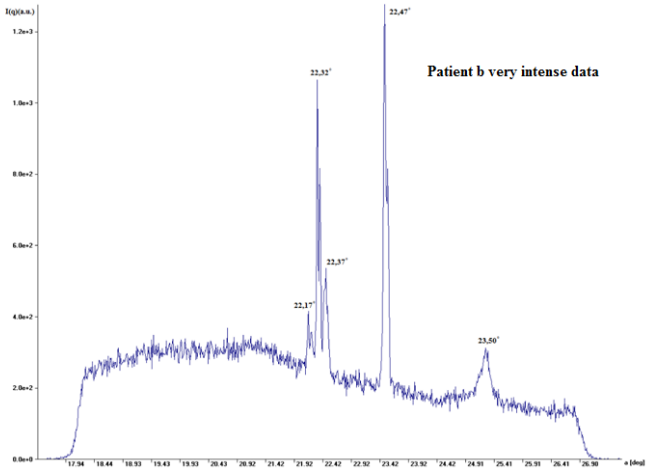


Figure 6: I(q)(a.u)-θ(°) chart for a very dense cartilaginous tissue specimen of a bone-cartilage sample from a male patient.

In Fig. 6, the five peak θ values for the male patient with a very dense cartilaginous tissue sample were determined to be 22,17 °, 22,32 °, 22,37 °, 22,47 ° and 23,50 °, respectively.

IV. CONCLUSION

In this study, bone and cartilage samples of male and female patients were examined using wide-angle X-ray (WAXS). The scattering angles of the WAXS were calculated by taking into account the Bragg diffraction law. Two peaks were obtained in the very dense portion of the female patient while five peak values were obtained in the very dense portion of the male patient. The cartilage area of male and female patients differs greatly depending on the severity of scattering intensity. The scattering angles obtained between the bone and cartilaginous sections due to the less dense and very dense center are almost the same. This has shown us that both bone and cartilage buildup is deteriorated due to osteoarthritis. And if we were to make comparisons for men and women; we can explain that there may be more deformations in the female.

REFERENCES

[1] Blanchet CE and Svergun DI, "Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution." *Annu Rev Phys Chem* vol. 64, 2013.

[2] Pe´rez J and Nishino Y, "Advances in X-ray scattering: from solution SAXS to achievements with coherent beams." *Curr Opin Struct Biol*, vol. 22 pp.670-678, 2012.

[3] Schneidman-Duhovny D and Kim S, Sali A, " Integrative structural modeling with small angle X-ray scattering profiles." *BMC Struct Biol*, vol. 12, pp.17, 2012.

[4] S. R. Kline, "Reduction and analysis of SANS and USANS data using IGOR Pro," *J. Appl. Crystallogr.*, vol. 39, pp. 895–900, 2006.

[5] P. V. Konarev, M. V. Petoukhov, V. V. Volkov, and D. I. Svergun, "ATSAS 2.1, a program package for small-angle scattering data analysis," *J. Appl. Crystallogr.*, vol. 39, pp. 277–286, 2006.

[6] A. V. Semenyuk and D. I. Svergun, "GNOM. A program package for small-angle scattering data processing," *J. Appl. Crystallogr.*, vol. 24, no. pt 5, pp. 537–540, 1991.

[7] G. Fritz and O. Glatter, "Structure and interaction in dense colloidal systems: evaluation of scattering data by the generalized indirect Fourier transformation method," *J. Phys. Condens. Matter*, vol. 18, pp. S2403–S2419, 2006.

[8] D. Franke and D. I. Svergun, "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering," *J. Appl. Crystallogr.*, vol. 42, pp. 342–346, 2009.

Design of A Triple Band Absorber Metamaterial for C-Band Radars with Golden Ratio Approaches

Z. KOCAMAN[1], A. KOCAKUSAK[1], I. B. BASYIGIT[2], S. OZEN[1], S. HELHEL[1]

Department of Electrical and Electronics Engineering, Akdeniz University, Antalya, Turkey[1]
Department of Electrical and Communication Engineering Süleyman Demirel University, Isparta, Turkey[2]namacokpenyez@gmail.com, atalaykocakusak@akdeniz.edu.tr, bahadirbasyigit@sdu.edu.tr, sukruozen@akdeniz.edu.tr, selcukhelhel@akdeniz.edu.tr

***Abstract*— Metamaterials are man-made artificial materials that are not a natural material but rather have different electromagnetic properties. One of the features mostly studied in recent years is that it nearly absorbs all incoming electromagnetic waves. These features make it possible to invisibility against remote sensing systems. Thus, the radar cross-section area (RCA) representing the radar visibility can be reduced by covering the target plane with metamaterial. As a first step in this work, the unit cell of the metamaterial, which has a resonance effect, is designed to achieve high absorption at 6.0 GHz. Then, using the unit cell's scalability feature for this metamaterial, two more-unit cells with resonance effect at 5.0 GHz and 7.0 GHz were designed, respectively. Subsequently, all three were brought next to each other and super-cell structure was introduced. The structure obtained shows 94.62% absorption at 5.08 GHz, 99.95% absorption at 5.98 GHz and 90.42% absorption at 7.02 GHz, respectively, in the C band region (4-8 GHz).**

Keywords*— *radar absorber, radar cross-sectional area , c-band absorber, meta-materials.

I. INTRODUCTION

Artificially obtained metamaterials are artificial composite periodic structures which are not inherent in existing materials but exhibit highly useful electromagnetic properties for other purposes [2]. These materials are used for invisibility, perfect focusing, radar absorbtion and miniaturization of antennas[3-6]. The structures created by adapting the geometric parameters of the absorber-based metamaterials are used as ultra-absorbent coatings from the microwave to the optical frequency region, as well as being very thin. [7].

A fundemantel starting point for detecting and determing the the target is the radar cross-sectional (RCS) area of the target. And it basically expresses the reflection ratio of the electromagnetic wave falling on the object itself [1,8]. That means lowering RCS of any target makes it almost invisible at certain frequency ranges. In this context, the reduction of RCS is important for not be detected. Radar Absorbing Materials (RAM) are used for reducing RCS of target that reduction in RCS is achieved by controlling the absorption, reflection and transmission parameters of the RAM. Low-cost, easily producible and highly efficient metamaterial-based, excellent signal absorbers have given a new look to the concept of RCS [9,10]. Permeability (ε) and magnetic permeability (μ) of an object have negative values, this object is metamaterial for this frequency range. This causes electromagnetic waves to exhibit different properties within the material. Metamaterials are materials that provide the desired properties by combining the materials that exist in nature in different forms, rather than a chemical change. Theoretical studies have shown that metamaterials with negative ε and μ values are negative for refractive indices in contrast to natural materials [13].

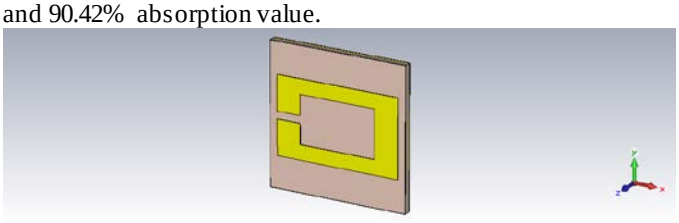
In the second part of this work, the design, optimization and simulation of the triple band radar absorber metamaterial were performed for C-band applications, and the results of the study were discussed in the third and last section.

II. DESIGN AND SIMULATION RESULTS

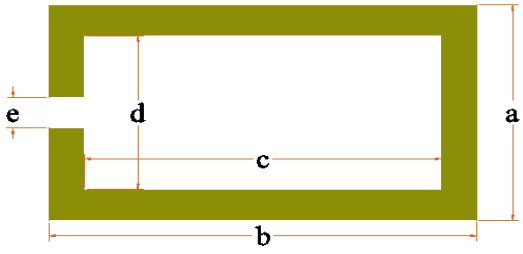
The front view of the unit cell geometry of the proposed structure together with the field views is shown in Fig 1. The structure has a thickness of 1.016 mm, two metallic layers separated by FR-4 with a relative permeability ($\varepsilon_r = 4.3$) and a dielectric tangent loss (tanδ=0.025) . The top layer is made up of three golden rectangular rings with different dimensions. As shown in Figure-2, The length parameters of these rings are given in Table-1. The bottom surface is completely communicated and thus the transmission is zero. Both upper and lower layers are made of copper with a conductivity (σ) of 5.8 × 10 7 S / m and a thickness of 0.035 mm. Since the bottom surface is completely covered with copper, the transmitted power | S_21 (ω) | ^ 2 is zero. Therefore, by properly designing the building parameters, the reflected power | S_11 (ω) | ^ 2 can be reduced most. The absorption (ω) in the structure can therefore be calculated as shown in (1).

$$A(\omega) = 1 - |S_{11}(\omega)|^2 - |S_{21}(\omega)|^2 \quad (1)$$

The structure was simulated using the CST Microwave Studio and showed three different minimum S21 values of -12 dB, -33.1 dB and -10.2 dB at 5.08 GHz, 5.98 GHz and 7.02 GHz, respectively, as shown in Fig. 3, indicating 94.62%, 99.95% and 90.42% absorption value.



(a)



(b)

Figure 1. Unit cell geometric view (a) and ring dimension (b)

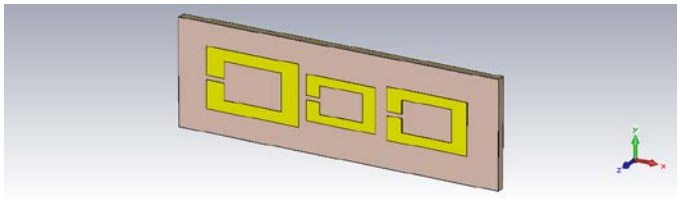


Figure 2. Three golden ratio rings with different dimensions.

Table 1

	a	b	c	d	e
1. Ring	7.1	11.49	7.1	4.39	0.5
2. Ring	6.04	9.77	6.04	3.73	0.5
3. Ring	5.28	8.54	5.28	3.26	0.5

Figure-4 shows the distribution of electric and magnetic fields at all absorption frequencies; 5.08 GHz, 5.98 GHz and 7.02 GHz. Explains the source of electrical and magnetic stimuli in different parts of the structure at the absorption frequencies. At the 5.08 GHz, 5.98 GHz, and 7.02 GHz absorption frequencies, the middle ring, the ring on the right and the ring on the left, respectively, were observed to have a high contribution to high localized electromagnetic fields.

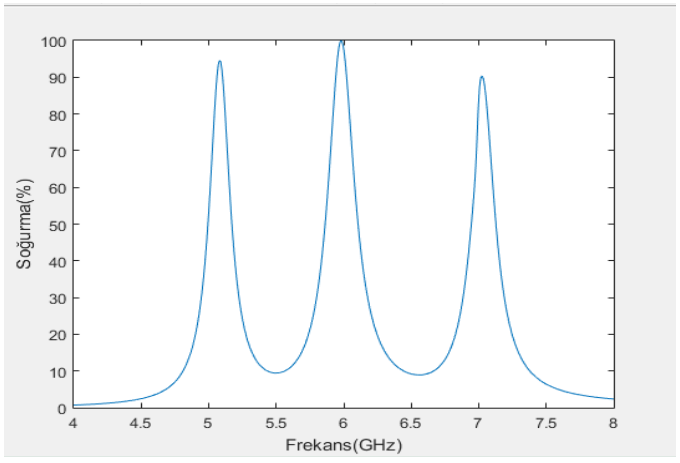


Figure-3 Absortion

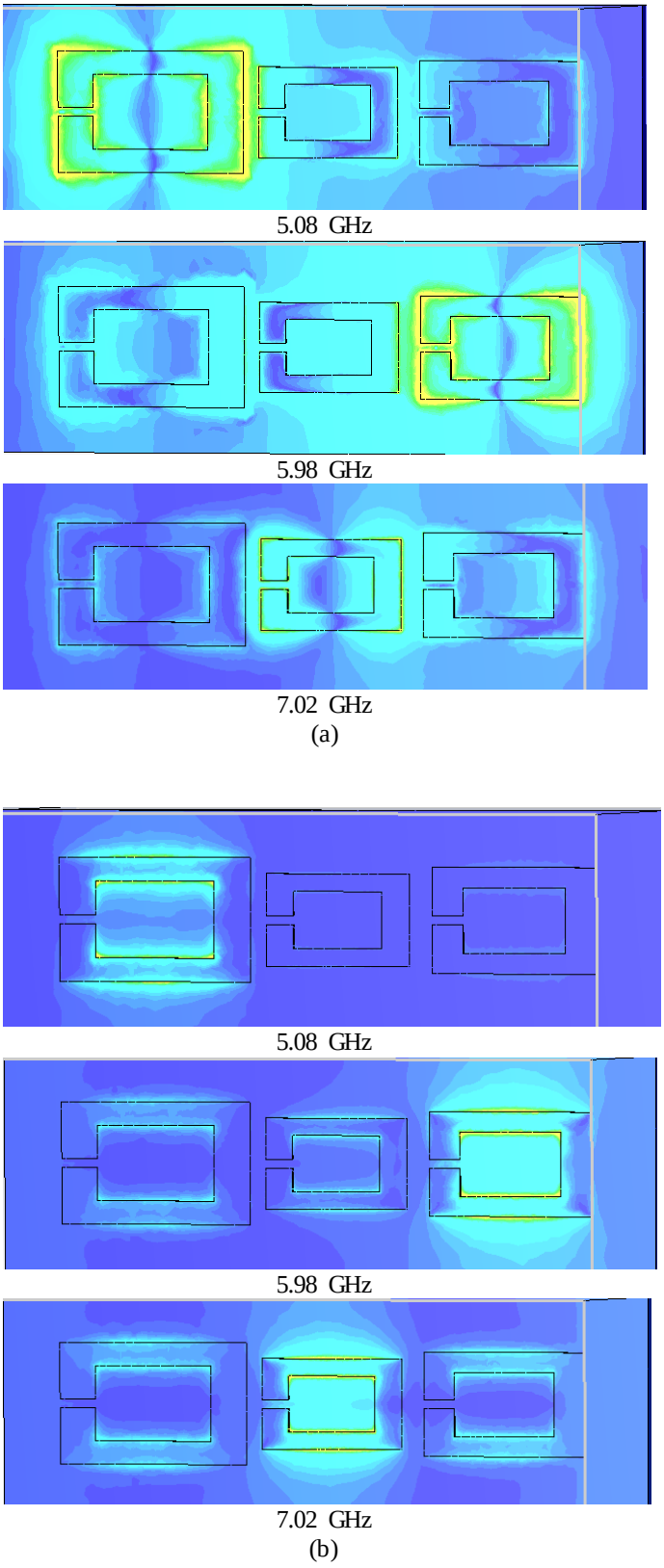
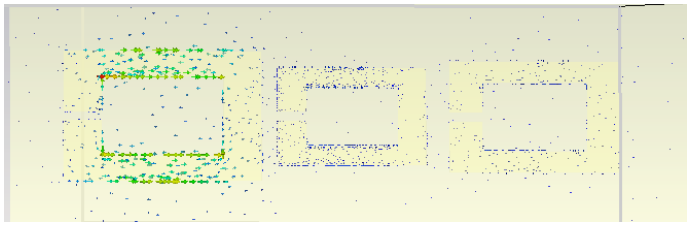
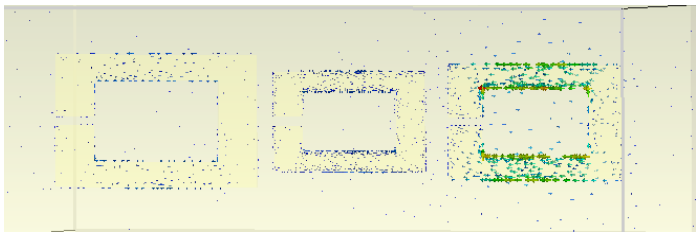


Figure-4 (a) E-Field (b) H-Field for proposed structure

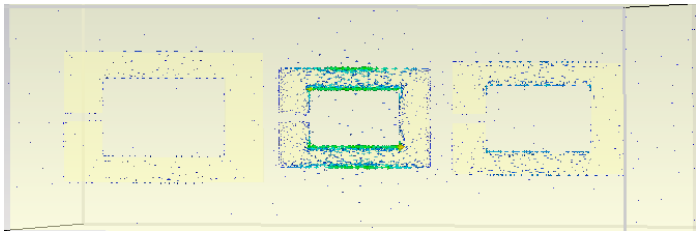
Figure 5 shows the surface current distributions of the top and bottom layers of the structure proposed to understand the absorbtion mechanism, and it is evident that the ring forms separate absorption peaks. At 5.08 GHz the current is often confined to the left, ie, the longest ring, while at 7.02 GHz the current has a high density in the middle ie the smallest ring. The surface current at the center frequency 5.98 GHz is limited on the right, that is, in the medium size ring. It is also seen that the direction of the surface currents in the upper and lower layers is not parallel and therefore constitutes the circulation ring perpendicular to the direction of the magnetic field causing the magnetic excitation. On the contrary, the excited electric field in the structure is elevated along the direction of the electric field in the square, thus creating an electric excitation. For this reason, both electrical and magnetic excitations occur at the absorbing frequencies, which leads to strong electromagnetic absorption.



5.08 GHz



5.98 GHz



7.02 GHz

Şekil-5 Surface currents

III. RESULTS

A thin triple band metamaterial absorber consisting of three golden ratio rectangular rings was presented. The dimensions of the rings are optimized to obtain three absorption peaks in the microwave band (band C). In order to understand the physical mechanism of absorption, electric and magnetic field distributions along with surface current distribution are shown for all peak absorption frequencies. Various parameters have been analyzed parametrically to observe the absorption mechanism of the structure. Finally, simulations of the designed structure were made and the responses obtained were analyzed and their results discussed. The designed radar absorber can be used in many potential applications, such as aerial surveillance radar, thermal spreaders, wireless communications, phase shifter and defensive applications.

ACKNOWLEDGMENT

This study was supported by project number 1919B011700666 under TUBITAK 2209-A. We would also like to thank Akdeniz University Industrial Medical Applications and Microwave Application and Research Center (DPT-2007K120530) for its contributions.

REFERENCES

[1] K. Ozden, A.Ozer, O. M. Yücedag ve H. Kocer, "Metamalzeme tabanlı geniş bant ışıma emici yapılar kullanılarak radar kesit alanının azaltılması", Journal of the Faculty of Engineering and Architecture of Gazi University, Vol.31-4, pp. 1105-1112, 2016.

[2] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, "Composite medium with simultaneously negative permeability and permittivity," Phys. Rev. Lett., vol. 84, pp. 4184-4187, 2000.

[3] K. Saurav, D. Sarkar and K. V. Srivastava, "CRLH Unit-Cell Loaded Multi-Band Printed Dipole Antenna," IEEE Antennas Wireless Propag. Lett., vol. 13, pp. 852-855, 2014.

[4] N. I. Landy, S. Sajuyigbe, J. J. Mock, D. R. Smith, and W. J. Padilla, "Perfect metamaterial absorber," Phys. Rev. Lett., vol. 100, pp. 207402, 2008.

[5] S. Enoch, G. Tayeb, and P. Vincent, "A metamaterial for directive emission," Phys. Rev. Lett., vol. 89, pp. 3901–3904, 2002.

[6] A. Rajput, and K. V. Srivastava, "Design of a two-dimensional metamaterial cloak with minimum scattering using a quadratic transformation function," J. Appl. Phys., vol. 116, no. 12, pp. 124501, 2014.

[7] R. Marques, F. Martin, and M. Sorolla, Metamaterials with Negative Parameters(Wiley, New Jersey, 2008), p. 43–90.

[8] Prof. Dr. Selçuk HELHEL Microwave Techniques Lecture Notes, 2017.

[9] Culhaoglu A.E., Osipov A.V., Russer P., Mono-and Bistatic Scattering Reduction by a Metamaterial Low Reflection Coating, IEEE Transactions on Antennas and Propagation, 61 (1), 462-466, 2013.

[10] Yang H., Cao X.Y., Gao J., Li W., Yuan Z., Shang K., Low RCS Metamaterial Absorber and Extending Bandwidth Based on Electromagnetic Resonances, Progress in Electromagnetics Research M, 33, 31-44, 2013.

[11] Veselago V.G., The Electrodynamics of Substances with Simultaneously Negative Values of ε and μ, Physics-Uspekhi, 10 (4), 509-514, 1968.

[12] Shelby R.A., Smith D.R., Schultz S., Experimental Verification of a Negative Index of Refraction, Science, 292 (5514), 77-79, 2001.

[13] Landy N.I., Sajuyigbe S., Mock J.J., Smith D.R., Padilla W.J., Perfect Metamaterial Absorber, Physical Review Letters, 100 (20), 207402, 2008.

Antenna Selection Techniques for Digital Relaying based Cooperative MIMO Channels

Ö.ÖZDEMİR

Selcuk University, Konya/Turkey, ozgur@selcuk.edu.tr

***Abstract* - In this paper, performance of joint transmit and receive antenna selection in cooperative MIMO channels is investigated over Rayleigh fading environments. A cooperative MIMO communication scenario in which all of the terminals are equipped with multiple antennas is considered where source want to transmit data to the destination with the help of the relay. The techniques developed in literature to combat error propagation effect in digital relaying networks such as cooperative maximal ratio combining (C-MRC), virtual noise (VN) based detection and selective relaying (SR) are exploited in this study to obtain new antenna selection criteria for cooperative MIMO systems. Only one antenna is selected and activated at each of the three terminals in a manner to minimize the end-to-end bit error rate (BER) of the overall system. This approach eliminates the requirement of space-time signaling in system design and reduces the complexity and cost by minimizing the number of the required RF chains in the network. Unlike most of the studies in the literature, we assume that the direct link between source and destination exists. Numerical results have shown that a diversity order of $(n_S + n_R) * n_D$ can be achieved for VN and SR based transmissions where n_S, n_R, and n_D are the total number of the antennas at the source, relay and destination, respectively. Among the considered schemes, the best performance is obtained in case of SR based system model.**

***Keywords* - Cooperative MIMO, antenna selection, error propagation, digital relaying.**

I. INTRODUCTION

THE main problem that arises in wireless communication systems is the time varying character of the channel [1]. Diversity is one of the most important methods used to solve this problem which is known as fading and sometimes decreases the signal-to-noise ratio significantly [1]. The diversity is based on the principle that multiple copies of the message to be transmitted are sent from different channels. This method increases the quality of the transmission since it reduces the possibility that all the copies are simultaneously exposure to a very severe fading. Various diversity techniques exist in literature. Among these approaches, spatial diversity is the most important one because of its superior performance and ease of use with other techniques. Spatial diversity provides very high capacity increments when realized by using antenna arrays at both receiver and transmitter terminals. Such a communication model, referred to as a Multi-Input Multi-Out (MIMO) system, can provide increases in diversity level and performance as well as requirements such as data rate, capacity, spectral efficiency, link quality and low transmit power [2].

Recent research activities yielded a cooperative diversity approach that enables spatial diversity by sharing the resources in a wireless communication system with multiple single antenna terminals [3]. In cooperative diversity, the retransmissions of the neighboring users which are known as relays are utilized to obtain the antenna diversity which is inherent in the multi-user wireless networks. In addition to the total number of terminals collaborating, the signal processing techniques classified as analog and digital have an important role in determining the performance of cooperative networks [4]. In this paper cooperative networks with digital relaying are considered.

The idea of integrating conventional MIMO technology with cooperative relaying systems is an attractive solution to enhance the capacity of the relay networks through additional spatial diversity and multiplexing gains [5]. The multi-antenna based cooperative systems can further increase the end-to-end performance of the wireless relay communication by allowing the use of specific transmission schemes such as space-time coding techniques [6-8] and beamforming [9, 10]. The main drawback of both space-time coded and beamforming based cooperative MIMO systems is that they increase system complexity and hardware cost by increasing the total antennas and radio frequency (RF) chains in the system. An effective method to combat this situation is antenna selection where only a subset of all antennas is used. The performance of antenna selection technique in cooperative MIMO systems is one of the hot research topics in wireless communications literature. In [11], joint relay and antenna selection strategies for dual-hop analog relaying based MIMO relay networks are studied. A new source transmit antenna selection based on both channel state information and transmission scheme for MIMO digital relaying networks is proposed in [12] where full diversity order is achieved. In [13], the performance of multi-antenna two-way relay networks is considered for both analog and digital relaying strategies.

In this paper, we investigate the performance analysis of digital relaying based cooperative MIMO channel, where the source and destination are equipped with multiple antennas and communicate with each other with the help of a multi-antenna relay. This paper is organized as follows: The channel and transmission model is given in Section II. In Section III, the antenna selection criteria are presented. Simulation results and concluding remarks are provided in Section IV and Section V, respectively.

II. SYSTEM MODEL

We focus on the three terminal digital relaying based cooperative MIMO network with n_S–antenna source and n_D–antenna destination which communicate with help of a n_R–antenna relay. All nodes are assumed to be half-duplex meaning that they transmit and receive either at different times, or over different frequency bands. The relay adopts the digital relaying protocol and direct link between S and D is considered in detection protocols. Only single antenna is selected at each of the terminals during the two time slot communication protocol. The best single transmit and receive antennas at the source, relay and destination are determined according to various strategies such as cooperative maximal ratio combining (C-MRC) [14], virtual noise (VN) based detection [15] and selective relaying (SR) [16] that involve minimization of the overall error propagation effect.

Let $\mathbf{H_{SD}}$, $\mathbf{H_{SR}}$ and $\mathbf{H_{RD}}$ be n_S x n_D, n_S x n_R and n_R x n_D channel matrices of the links S → D, S → R and R → D, respectively. The elements of these matrices represented by h_{ik}^{SD} ($i = 1, \ldots, n_S$ and $k = 1, \ldots, n_D$), h_{ij}^{SR} ($i = 1, \ldots, n_S$ and $j = 1, \ldots, n_R$) and h_{jk}^{RD} ($j = 1, \ldots, n_R$ and $k = 1, \ldots, n_D$) are independent complex Gaussian fading coefficients. Note that in all of the three strategies, only single antenna is selected at each of the three terminals and we will refer to this structure as an (n_S, 1; n_R, 1; n_D, 1) system. This situation not only eliminates the need for space-time signaling but also minimizes the number of RF chains required in realization of the network so that a considerable reduction in the system complexity and hardware cost is achieved. It is assumed that the channel matrices are known by the corresponding receivers by exploiting training signaling and binary phase shift keying (BPSK) is utilized as the modulation type. Two time slots are used for communication. In the first time slot, the source sends its data block x and the received signals can be written as

$$y_{SD}(n) = h_{ik}^{SD}\sqrt{E_s}\,x(n) + z_{SD}(n) \tag{1}$$

$$y_{SR}(n) = h_{ij}^{SR}\sqrt{E_s}\,x(n) + z_{SR}(n) \tag{2}$$

In (1-2) $z_{SD}(n)$ and $z_{SR}(n)$ are the additive noise (AWGN) samples modeled as independent zero-mean complex Gaussian random variables with variance $N_0/2$ per dimension. h_{ik}^{SD} is the element at the *i*th row and *k*th column of $\mathbf{H_{SD}}$ and h_{ij}^{SR} is the element at the *i*th row and *j*th column of $\mathbf{H_{SR}}$ respectively. The instantaneous signal to noise ratio (SNR) of these links are defined by $\gamma_{ik}^{SD} = E_s\left|h_{ik}^{SD}\right|^2 / N_0$ and $\gamma_{ij}^{SR} = E_s\left|h_{ij}^{SR}\right|^2 / N_0$.

The last time slot of the transmission scheme depends on the antenna selection criteria and the detection approach. In CMRC [14] and VN [15] schemes, the relay retransmits the detected symbol block $\hat{x}$ to the destination. The received signals in the second phase in these schemes can be given as

$$y_{RD}(n) = h_{jk}^{RD}\sqrt{E_s}\,\hat{x}(n) + z_{RD}(n) \tag{3}$$

while in SR based communication

$$y_{RD}(n) = \begin{cases} h_{jk}^{RD}\sqrt{E_s}\,\hat{x}(n) + z_{RD}(n) & \text{for state } s_1 \\ 0 & \text{for state } s_0 \end{cases} \tag{4}$$

In (3) and (4), h_{jk}^{RD} is the element at the *j*th row and *k*th column of $\mathbf{H_{RD}}$ and the instantaneous SNR of this link is defined by $\gamma_{jk}^{RD} = E_s\left|h_{jk}^{RD}\right|^2 / N_0$. It is seen from Expression (4) that in SR strategy relay may retransmit the received signal from source (state s_1) or remain silent (state s_0). If it remains silent source starts transmitting the next data [16].

III. ANTENNA SELECTION CRITERIA

In this section, antenna selection criteria for C-MRC, VN and SR based transmissions proposed in literature for combating error propagation effect are given. Unlike the existing antenna selection criteria introduced for digital relaying based MIMO relay systems in the literature, the antenna subset selection criteria used in this study are adopted from the methods proposed for combating error propagation effect in a manner to eliminate the requirement of space-time signaling [17-18]. As mentioned in the previous section, in all of C-MRC, VN and SR based schemes; single antenna is activated at each node of the network.

A. Antenna Selection Rule for C-MRC based Cooperative MIMO System

In the case of CMRC technique, the signals are combined as

$$y_C(n) = w_d\,y_{SD}(n) + w_r\,y_{RD}(n) \tag{5}$$

where $w_d = h_{SD}^*$, $w_r = \gamma_e h_{RD}^* / \gamma_{RD}$, γ_e is the equivalent SNR of S→R→D link defined as

$$\gamma_e = 0.5\left\{Q^{-1}\left[P_{eq}^{b}\left(\gamma_{ij}^{SR}, \gamma_{jk}^{RD}\right)\right]\right\}^2 \tag{6}$$

where [14]

$$P_{eq}^{ijk}\left(\gamma_{ij}^{SR}, \gamma_{jk}^{RD}\right) = \left[1 - P_{ij}^{SR}\left(\gamma_{ij}^{SR}\right)\right]P_{jk}^{RD}\left(\gamma_{jk}^{RD}\right) + \left[1 - P_{jk}^{RD}\left(\gamma_{jk}^{RD}\right)\right]P_{ij}^{SR}\left(\gamma_{ij}^{SR}\right) \tag{7}$$

In (7), $P_{jk}^{RD}\left(\gamma_{jk}^{RD}\right)$ and $P_{ij}^{SR}\left(\gamma_{ij}^{SR}\right)$ are the conditional BERs at both hops calculated by $P_{jk}^{RD}\left(\gamma_{jk}^{RD}\right) = Q\left(\sqrt{2\gamma_{jk}^{RD}}\right)$ and $P_{jk}^{SR}\left(\gamma_{ij}^{SR}\right) = Q\left(\sqrt{2\gamma_{ij}^{SR}}\right)$ for BPSK signaling respectively [14].

Then the antenna selection rule for C-MRC based transmissions can be given as

$$\{I,J,K\} = \underset{1\le i\le n_S}{\arg\min} \underset{1\le j\le n_R}{} \underset{1\le k\le n_D}{} \left\{ P_{eq}^{ijk}\left(\gamma_{ij}^{SR},\gamma_{jk}^{RD}\right)\right\} \quad (8)$$

where I, J, and K are the best antenna indices at source, relay and destination respectively, which minimize the end-to-end BER of the MIMO relay system.

B. Antenna Selection Rule for VN based Cooperative MIMO System

The conditional BER for this detection model can be expressed by [15]

$$P_c^{ijk}\left(\gamma_{ij}^{SR},\gamma_{jk}^{RD},\gamma_{ik}^{SD}\right) = \left[1-Q\left(\sqrt{2\gamma_{ij}^{SR}}\right)\right]Q\left(\sqrt{\frac{2\left(\gamma_{ik}^{SD}+\gamma_{EQ}\right)^2}{\gamma_{ij}^{SD}+\gamma'_{EQ}}}\right) + Q\left(\sqrt{2\gamma_{ij}^{SR}}\right)Q\left(\sqrt{\frac{2\left(\gamma_{ik}^{SD}-\gamma_{EQ}\right)^2}{\gamma_{ik}^{SD}+\gamma'_{EQ}}}\right) \quad (9)$$

where [15]

$$\gamma_{EQ} = \frac{\left|h_{ij}^{SR}\right|^2\left|h_{jk}^{RD}\right|^2}{\left(\left|h_{ij}^{SR}\right|^2+\left|h_{jk}^{RD}\right|^2\right)N_0},\quad \gamma'_{EQ} = \frac{\left|h_{ij}^{SR}\right|^4\left|h_{jk}^{RD}\right|^2}{\left(\left|h_{ij}^{SR}\right|^2+\left|h_{jk}^{RD}\right|^2\right)^2 N_0} \quad (10)$$

Then the antenna selection rule for VN based detections can be given as

$$\{I,J,K\} = \underset{1\le i\le n_S}{\arg\min} \underset{1\le j\le n_R}{} \underset{1\le k\le n_D}{} \left\{ P_c^{ijk}\left(\gamma_{ij}^{SR},\gamma_{jk}^{RD}\right)\right\} \quad (11)$$

C. Antenna Selection Rule for SR based Cooperative MIMO System

The end-to-end BER of this model can be written as

$$P_{e2e} = \begin{cases} P_{s_1}\left(\gamma_{ik}^{SD},\gamma_{ij}^{SR},\gamma_{jk}^{RD}\right) & \text{for state } s_1 \\ P_{s_0}\left(\gamma_{ik}^{SD}\right) & \text{for state } s_0 \end{cases} \quad (12)$$

where [16]

$$P_{s_1}\left(\gamma_{ik}^{SD},\gamma_{ij}^{SR},\gamma_{jk}^{RD}\right) = \left[1-Q\left(\sqrt{2\gamma_{ij}^{SR}}\right)\right]Q\left(\sqrt{2\left(\gamma_{ik}^{SD}+\gamma_{jk}^{RD}\right)}\right) + Q\left(\sqrt{2\gamma_{ij}^{SR}}\right)Q\left(\frac{\gamma_{ik}^{SD}-\gamma_{jk}^{RD}}{\sqrt{\left(\gamma_{ik}^{SD}+\gamma_{jk}^{RD}\right)/2}}\right) \quad (13)$$

$$P_{s_0}\left(\gamma_{ik}^{SD}\right) = Q\left(\sqrt{2\gamma_{ik}^{SD}}\right) \quad (14)$$

The antenna selection rule can be designed in manner to minimize the end-to-end BER of the SR system and it can be expressed by

$$\{I,J,K\} = \underset{1\le i\le n_S}{\arg\min} \underset{1\le j\le n_R}{} \underset{1\le k\le n_D}{} \left\{\min\left[P_{s_1}\left(\gamma_{ik}^{SD},\gamma_{ij}^{SR},\gamma_{jk}^{RD}\right),P_{s_0}\left(\gamma_{ik}^{SD}\right)\right]\right\} \quad (15)$$

IV. SIMULATION RESULTS

In this section, we present computer simulation results for the MIMO relay system. We set the data block length to 100 bits. A symmetric scenario where the variances of all the channels are equal to 1, is considered. The performance comparisons are done at a BER of 10^{-7}.

Figure 1 illustrates the performance of the C-MRC, VN and SR based detections when all the terminals have only one antenna (n_S = 1, n_R = 1; n_D = 1) and no antenna selection algorithm is employed in the network. It is seen that all of the detection schemes give full diversity gain and almost the same performance in case of (1, 1; 1, 1; 1, 1) system configuration.

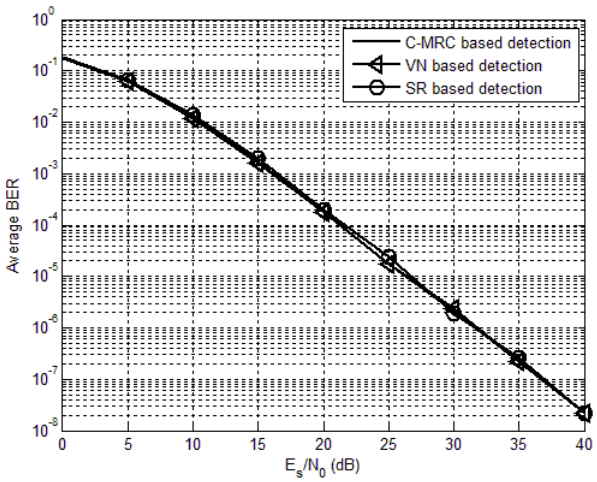


Figure 1: The performance curves of the (1, 1; 1, 1; 1, 1) system

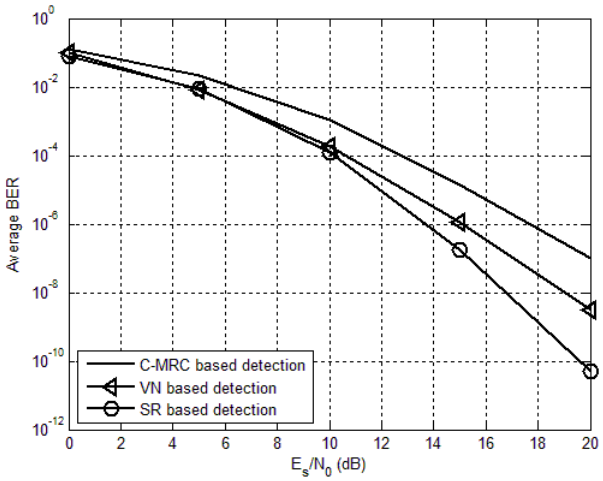


Figure 2: The performance curves of the (2, 1; 2, 1; 2, 1) system

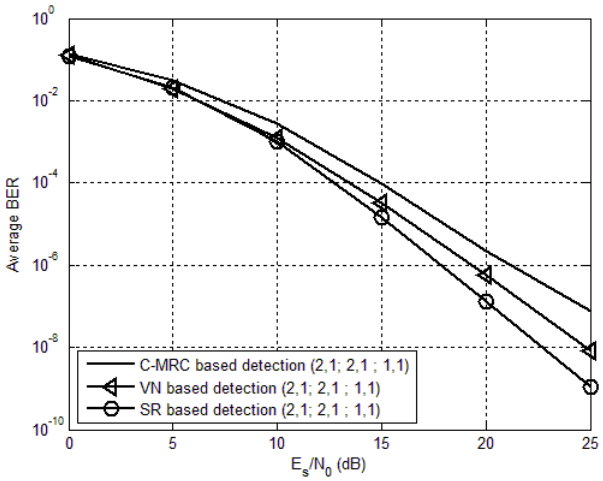


Figure 3: The performance curves of the (2, 1; 2, 1; 1, 1) system

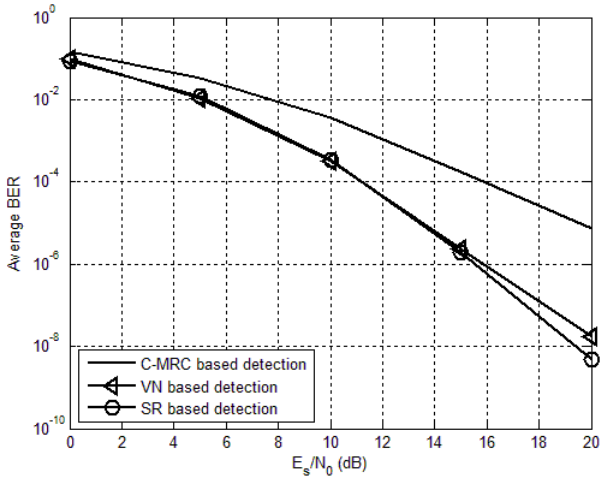


Figure 4: The performance curves of the (2, 1; 1, 1; 2, 1) system

Figure 2 shows the performance curves of the C-MRC, VN and SR based detections when all the terminals have two antennas (n_S = 2, n_R = 2; n_D = 2). Unlike the performance curves given in Figure 1 for (1, 1; 1, 1; 1, 1) system model, there are considerable SNR differences between performances of different the antenna selection schemes. The best performance is obtained for SR based detections. The SNR difference between SR and VN and C-MRC based detections are about 2 and 5 dB respectively. It is seen that as the number of the total antennas at the terminals are increased, not only the performance of the network improves but also the SNR differences between the different schemes are significantly increases.

Figure 3 and Figure 4 demonstrate the performance curves of the C-MRC, VN and SR based detections for (2, 1; 2, 1; 1, 1) and (2, 1; 1, 1; 2, 1) system models respectively. It is obvious that VN and SR based schemes provide better diversity levels with respect to C-MRC based transmissions. The SNR gain given by SR scheme with respect to VN based detections is 2 and 0.5 dB for (2, 1; 2, 1; 1, 1) and (2, 1; 1, 1; 2, 1) system configurations respectively.

V. CONCLUSION

Cooperative MIMO systems with antenna selection are considered where only one antenna at each terminal is selected and activated. Digital relaying protocol is employed and antenna selection criteria are derived by using the techniques in the literature developed for mitigating error propagation. Simulation results have shown that the best performance is obtained for SR technique which provides full diversity gain.

REFERENCES

[1] A. Goldsmith, *Wireless Communications*. New York: Cambridge University Press, 2005.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[3] A. Sendonaris, E. Erkip, and B. Aazhang, "User Cooperation Diversity Part I and Part II," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1927–1948, Nov. 2003.

[4] J. N. Laneman, David N. C. Tse, and Gregory W. Wornell: "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behaviour", *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004.

[5] Y. Fan and J. S. Thompson, "MIMO Configurations for Relay Channels: Theory and Practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–86, May 2007.

[6] Q. Yang, K. S. Kwak, and F. Fu, "Outage performance of cooperative relay using STBC systems," *Wireless Pers. Commun.*, vol. 52, no. 4, pp. 789–797, 2010.

[7] K. Yang, J. Yang, J. Wu, C. Xing, and Y. Zhou, "Performance analysis of DF cooperative diversity system with OSTBC over spatially correlated Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1270–1281, Mar. 2014.

[8] H.-M. Kim, T.-K. Kim, M. Min, and G.-H. Im, "Low-complexity detection scheme for cooperative MIMO systems with decode-and-forward relays," *IEEE Trans. Commun.*, vol. 63, no. 1, pp. 94–106, Jan. 2015.

[9] M. K. Arti and M. R. Bhatnagar, "Maximal ratio transmission in AF MIMO relay systems over Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 64, no. 5, May 2015. DOI: 10.1109/TVT.2014.2334631.

[10] K. Xiong, P. Y. Fan, Z. F. Xu, H. C. Yang, and K. B. Letaief, "Optimal cooperative beamforming design for MIMO decode-and-forward relay

channels," *IEEE Trans. Signal Process.*, vol. 62, no. 6, pp. 1476–1489, Mar. 2014.

[11] G. Amarasuriya, C. Tellambura, and M. Ardakani, "Joint relay and antenna selection for dual-hop amplify-and-forward MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 493–499, Feb. 2012.

[12] X. Jin, J.-S. No, and D.-J. Shin, "Source transmit antenna selection for MIMO decode-and-forward relay networks," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1657–1662, Jul. 2013.

[13] K. Song, B. Ji, Y. Huang, M. Xiao, and L. Yang, Performance Analysis of Antenna Selection in Two-Way Relay Networks, *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2520–2532, May 2015.

[14] T. Wang, A. Cano, G. B. Giannakis and J. N. Laneman, "High-performance cooperative demodulation with decode-and-forward relays," *IEEE Trans. Commun*., vol. 55, no. 7, pp. 1427–1438, July 2007.

[15] D. Kim, H.-M. Kim, and G.-H. Im, "Improved network-coded cooperative transmission with low-complexity adaptation to wireless channels," *IEEE Trans. Commun*., vol. 59, no. 10, pp. 2916–2927, Oct. 2011.

[16] F. A. Onat, A. Adinoyi, Y. Fan, H. Yanikomeroglu, J. S. Thompson, and I. D. Marsland, "Threshold selection for SNR-based selective digital relaying in cooperative wireless networks," *IEEE Trans. Wireless Commun*., vol. 7, pp. 4226–4237, Nov. 2008.

[17] K. Yang, J. Yang, J. Wu, C. Xing, and Y. Zhou, "Performance analysis of DF cooperative diversity system with OSTBC over spatially correlated Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 63, no. 3, pp. 1270–1281, Mar. 2014.

[18] N. Varshney and A. K. Jagannatham, "MIMO-STBC based multi-relay cooperative communication over time-selective Rayleigh fading links with imperfect channel estimates," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6009–6025, Jul. 2017.

Comparative Analysis of the Distributed Energy Resources Connections to Distribution Network

A. ARIFJANOV[1], R. ZAKHIDOV[2] and N. A. ALMAGRAHI[3]

[1] SICICT at Tashkent University of Information Technologies, Tashkent/Uzbekistan arifjanov@yandex.ru
[2] Scientific & Technical Center of JSC "Uzbekenergo" Tashkent /Uzbekenergo, r.zakhidov@mail.ru
[3]Karabuk University, Karabuk/Turkey, nuri_almagrhy@yahoo.com

***Abstract* - In the last decade, technological innovations and a changing economic and regulatory environment have resulted in a renewed interest for distributed generation. In most cases distributed generation systems are represented in the form of autonomous energy centers, since the connection of distributed generation to centralized electrical networks is being limited to the lack or imperfection of the regulatory framework and effective means of controlling technological regimes. In this paper is carried out a comparative analysis of the operation principles and options for connecting to the distribution network of autonomous power plants using different types of distributed energy resources. While connection of distributed energy resources to distribution network, one of the main tasks is distribution of loads between generating capacities. An algorithm for solving this problem is also presented in the paper.**

***Keywords* - Distributed generation, power plants, distributed energy resources, distribution network, circuitry solutions.**

I. INTRODUCTION

Trends in the development of the electric power industry in the world are associated not only with increasing the scale of electricity production in traditional large power plants, but also with increasing the share of distributed generation (DG) based on renewable energy sources.

In the last decade, technological innovations and a changing economic and regulatory environment have resulted in a renewed interest for distributed generation. According to the International Energy Agency (IEA) the following factors contribute to this interest:

- Development of distributed generation technologies
- Constraints on the construction of new transmission lines
- Increased customer demand for highly reliable electricity
- Liberalization of the electricity market
- Environmental pollution.

IEA defines distributed generation as a generating plant serving a customer on-site or providing support to a distribution network, connected to the grid at distribution-level voltages. The technologies of DG include engines, photovoltaic systems, small (and micro) turbines and fuel cells. [1].

The share growth of DG in electric power systems has not only positive aspects, but also creates certain technical problems that are associated with changes in the properties of systems, their control capabilities under normal and emergency conditions. The control of such power systems with DG is more actual nowadays, because for these systems considerable territorial distribution and heterogeneity. Heterogeneity of system can be related to use of distributed energy sources (DER), as wind and solar power stations. The main distinguishing feature of DG is the stochastic character of the parameters of the primary energy source. This applies to helioenergetics and wind power. Therefore, the electrical energy generated from DER creates new problems before the already overloaded distribution network.

The task of modern electrical systems is constant balancing of supply and demand, flexible network control and ensuring optimal levels of energy efficiency. These problems can be solved, but at the same time the dispatching and automatic control of the electric power system becomes more complicated. It is necessary to develop new mathematical models to justify the development of the electric power system and power supply systems, analyze their regimes and control them.

The operation of power systems with DER requires the systematization of existing devices that a make a possible the formation of the most efficient power supply systems. To find the most optimal technological solutions for the combined system, need to analyze each component separately and the system as a whole. Based on the foregoing, the integration and effective control issues of the operation modes for various types of power plants based on DER are considered in paper. In the report is carried out a comparative analysis of the operation principles and options for connecting DER to the distribution network. While connection of DER to distribution network, one of the main tasks is distribution of loads between generating capacities. An algorithm for solving this problem is presented in the paper.

II. ANALYSIS OF THE OPERATION PRINCIPLES AND CONNECTION OPTIONS OF AUTONOMOUS POWER PLANTS TO THE DISTRIBUTION NETWORK

The operation feature of any electric power system is one-stage production and consumption of electricity. This requires an equality of energy, generated by the power plants and consumed by the consumers. Violation of such equality leads to a change of the network parameters by voltage and frequency, and at large deviation - to loss of dynamic stability and the disruption of normal functioning of the system.

There are two ways to maintain a continuous in time equality of generated and consumed electricity in the power supply

systems based on DERs. The first way is on continuously monitoring of the power consumption and on appropriate power regulation of the electricity sources so that to continuously maintain the equality between two above mentioned power. In this option the energy storage systems keep the energy while energy consumption is reduced and after converting energy is supplied to the electric grid when a certain level of consumption is exceeded.

A promising option for creating autonomous power supply systems for electricity consumers in decentralized zones is the use of hybrid power plant such as wind-diesel, photo-diesel or wind-photo-diesel power plants (Figure 1).

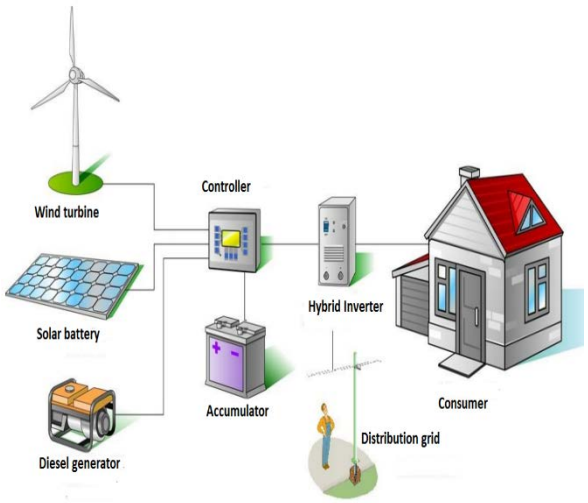


Figure 1: Structural diagram of a hybrid autonomous power supply system.

In general, such a system consists of the following main blocks:

• Diesel generator

• Energy storage device (battery, flywheel, superconducting magnetic energy storage, etc.)

• Distributed energy sources (photovoltaic, wind turbines, biomass, fuel cell, etc.).

The operation principle of an autonomous power plant is as follows (Figure 2) [2]:

1. To ensure high-energy efficiency of the autonomous power supply system, the energy generated by the primary energy source should be directly delivered to the consumer (Path 3).

2. In periods of excessive energy generation, the energy should be stored by the energy storage devices (Path 1).

3. If there is a shortage in the energy generated by a primary source, the accumulated energy from secondary sources should be delivered to the consumer, covering the existing shortage (Path 2).

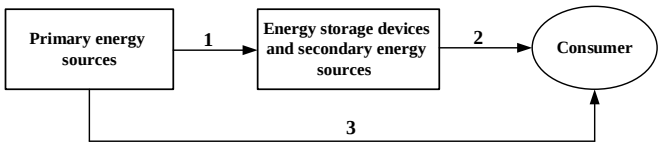


Figure 2: Main components interaction of the autonomous power supply system.

The effective control system of the station should provide not only a controlling strategy of the autonomous power plants, but also synchronization of the start-up of the units and their further synchronous operation.

As a rule, there are three methods of the power flow controlling in the power systems with DER [3]:

1) Dropping of the excess energy
2) Energy storage
3) Load shedding (Figure 3).

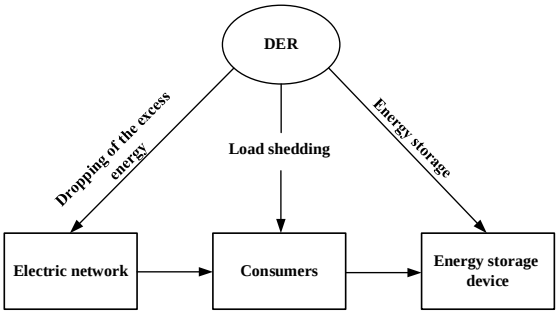


Figure 3: Scheme of power flow control methods from DER to consumers.

Due to the non-stationary nature of the generation process by DER, the practical implementation of these control methods is a serious problem.

Let's consider options for connecting DER based power plants to the distribution network in hybrid energy systems. There are different options of conjugation of diesel generator plants (DGP), wind power plants (WPP) and photovoltaic plants (PVP), working to general consumer, which can significantly differ by the composition of the used electrical equipment and by the technical and economic characteristics.

Nowadays several types of hybrid systems based on DER have been developed:

- AC/AC power supply system;
- DC/DC power supply system;
- Systems with mixed connection.

Combined AC power systems can be divided into two categories: centralized and decentralized (distributed). In the centralized AC power system, all connected units to the main AC bus in front of the consumer (Figure 4, a). Wind turbine and diesel generator generate an alternating current, so they can be directly connected to the main AC bus or AC/AC converters. The photovoltaic module produces DC, hence the inverter must be used before connecting to the main AC bus. For batteries with a DC stream, a bi-directional inverter must be used.

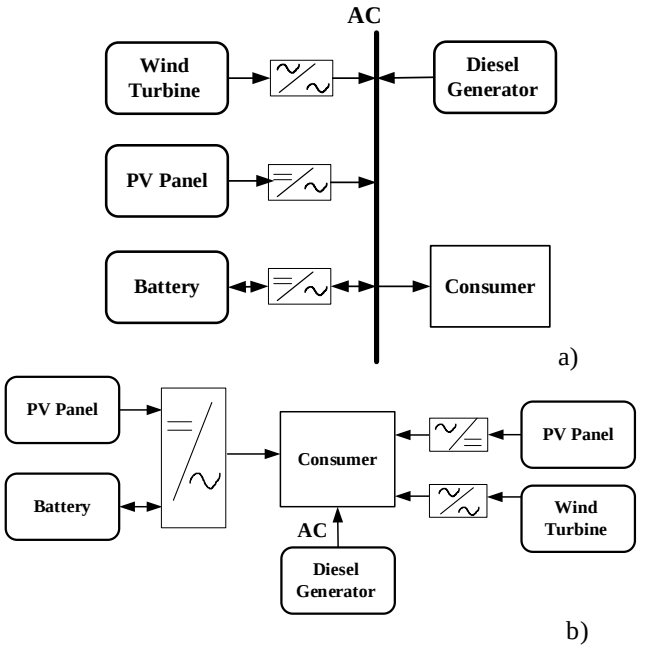


Figure 4: Centralized (a) and decentralized (b) AC power system.

In a decentralized AC power system, all units are distributed or decentralized. The units of the system are not connected to backbone AC bus, some or all of units are individually connected with the consumer (Figure 4, b).

In a decentralized system, energy sources should not be connected to one common bus, as in the previous case. In this configuration, the energy sources should not be located close to each other, since each source of the system is connected to the consumer separately. The DC from PV panel and batteries must be converted to AC by the corresponding inverters in front of the consumer. The advantage of this type system is some various sources can be located remotely from each other, so that more convenient locations can be selected (places with greater illumination for photovoltaic cells or windy for wind turbines). In addition, this system has disadvantage in terms of control difficulties.

The analysis of comparing centralized and decentralized systems show that the centralized power supply systems can be controlled more easily and conveniently then decentralized system [4-7]. In a centralized DC system (Figure 5, a), all units are connected to the main DC bus directly in front of the user. DC bus connection with AC bus is executed through the main inverter.

In Figure 5, b is illustrated a widespread version of a hybrid power plant scheme based on DER, where sources are connected directly to the distribution network without intermediate electricity conversion.

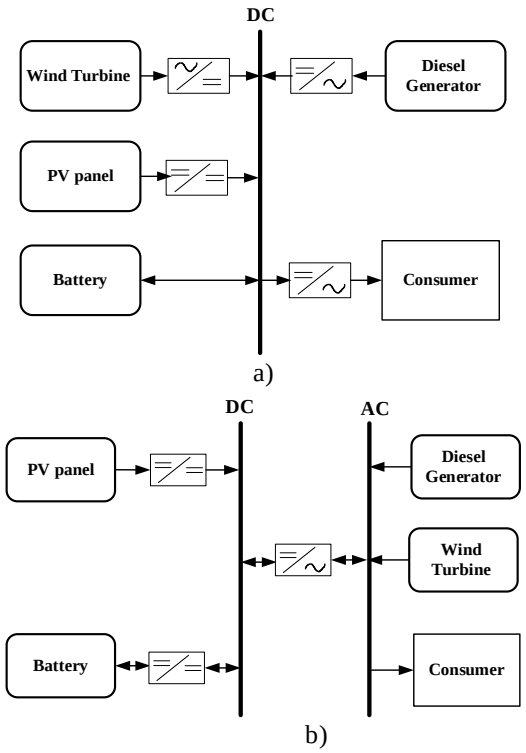


Figure 5: Centralized DC system (a) and hybrid power system (b).

III. Circuitry Solutions for Connection DER-based Power Plants to the Distribution Network

The autonomous power supply system presented in Figure 6 is simple to implement. It can be scale by installing, for example, several wind turbines. Due to the absence of additional converters, it is ensured a high efficiency factor of the whole system. However, it is essential the presence of specified, identical and constant values (output voltage) of electrical generators and frequency of the network. There is necessity using of a wind turbine with sophisticated systems for aerodynamic stabilization of the rotor speed or using of wound rotor induction motors with a proper control from inverter [8, 9]. Such wind turbines are suitable for large wind energy, but they find extremely limited application in small wind energy conversion systems due to their high cost.

These disadvantages are absent in the connection scheme shown in Figure 7.

Despite the more complex structure, this scheme has the following advantages in compare with Figure 6 [9, 10]:

1) There is no essential to coordinate the operation modes of DGP, WPP and PVP, that allows to control these units based on the required optimality criteria.
2) The system is easy to scale and simply enough to solve the problem of electro-magnetic compatibility.
3) High quality of the supplied electricity is provided due to the consumers' power source from an inverter.

This variant of hybrid energy systems has found a widely application at small and medium power plants with capacities 1 – 100 kW [11-14].

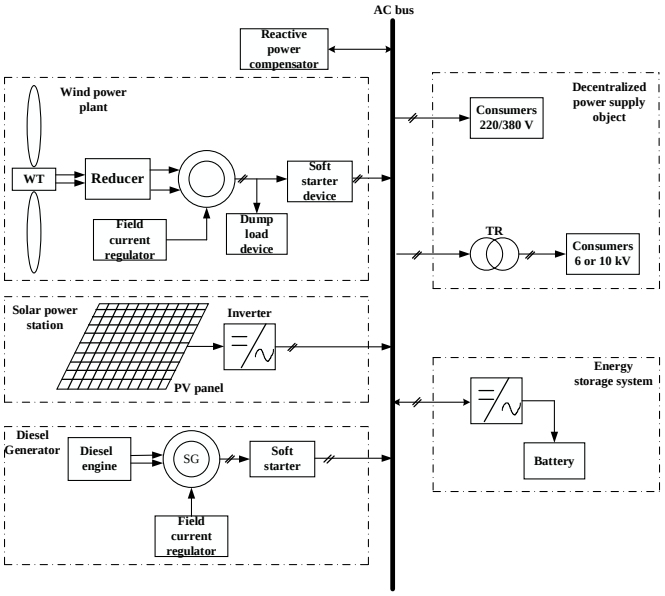


Figure 6: Hybrid power plant scheme with direct connection of DER to distribution network.

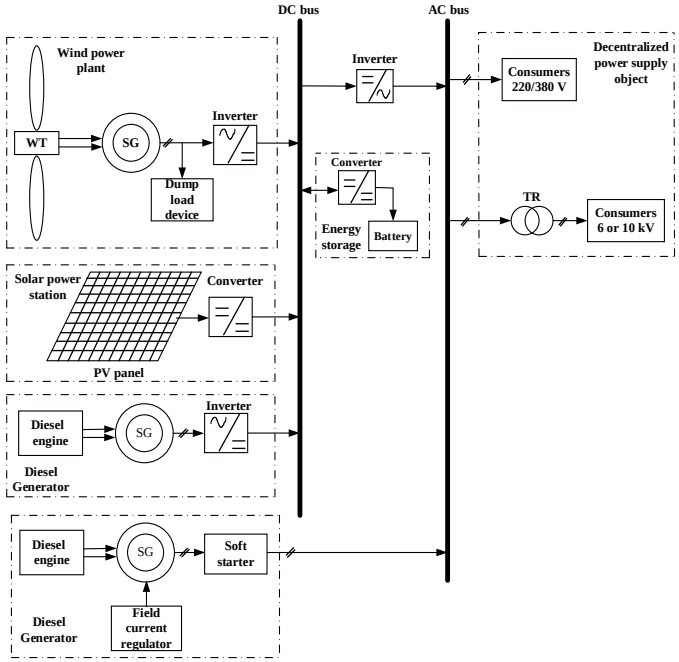


Figure 7: The scheme of a hybrid power plant with connection of DER to an intermediate DC bus (and with a mixed connection).

The hybrid power plant based on conjugation of power plants by means of an auxiliary bus operating at a high frequency is shown in Figure 8.

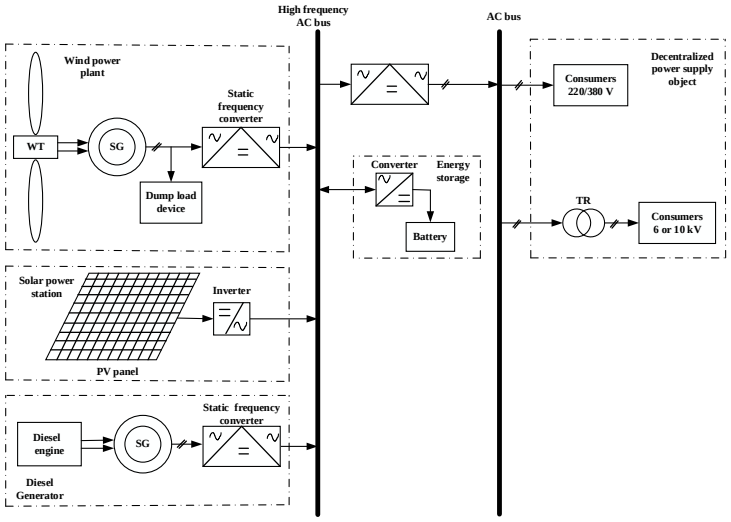


Figure 8: The scheme of a hybrid power plant with the connection of DER through a high-frequency AC bus.

This method is mainly used in creating power supply networks for air and space vehicles. It allows to minimize the number of reactive elements in the system, to reduce the weight, dimensions and the cost of hybrid power systems. However, by the reason of the geometric disconnection of individual aggregates (DGP, WPP, PVP and etc.), while using of this scheme, we have some problems related to the power losses in auxiliary network, electromagnetic compatibility, etc. [5-9, 13].

A comparative analysis of the autonomous power plants schemes has shown that the most promising option of the interfacing of DER within the power system is to utilize of intermediate DC bus. In this case, the designed hybrid power system on an aggregate principle, easily scaled and, if necessary, rearranged. In addition, it is possible to integrate the structure and construction of electronic power converters. The using of DC bus insertion makes it possible to more simply summarize and distribute power flows from generating sources and implement effective control algorithms for this process.

The block diagram of the autonomous power supply system with a DC bus is shown in Figure 9. This system consists of separate *n-number* generating power sources PS_1, PS_2, ..., PS_n. Each power source includes a corresponding power unit PU_1, PU_2,... , PU_n, which designed on one or another physical principle, and a controlled by static converter SC_1, SC_2,..., SC_n . As power plants can be used wind turbines, PV panels and diesel generators. The static converter should be selected according to the type of power plant, for example, for wind turbines and diesel generators are used boosting rectifiers, and for photovoltaic modules – DC/DC converters. An important advantage of the proposed scheme is the possibility of a significant expansion of the energy storage system (ESS) functions, once input signals such as wind speeds (from the wind turbine) and the intensity of solar radiation (from PV) are entered to control system. In this case, the output signals from ESS come to the controlling converters providing the mode of the maximum power selection from power plants. When we use inverter type diesel generator sets in autonomous power system,

the output signal from ESS transmits to the actuator to control position of the fuel pump rod. Thereby, it provides optimization of the diesel generator operation modes by the criterion of minimum fuel consumption.

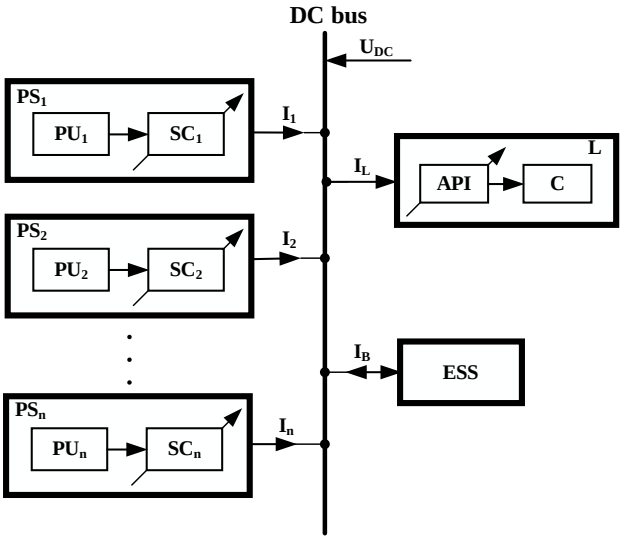


Figure 9: Structural diagram of the autonomous power supply system with DC bus.

I_1,.., I_n - currents of power sources; I_l - load current; I_B - current of energy storage system; U_{DC} – voltage on DC bus.

This set of functions can be implemented by installing the corresponding additional expansion modules into ESS. Due to additional modules, the system receives and processes information about environmental conditions from the meteorological complex; calculates in real time the optimal values of the current loads for each generating power unit, based on the principle of maximum power takeoff and generates control actions for each converter of the generating source.

Thus, the scheme of an autonomous power station has the following advantages:

- inclusion in the system of any autonomous power plant irrespective of the installed power equipment;
- program configuration of the energy storage system for a specific power complex by connecting to a ESS control system of a personal computer when performing this operation;
- control of power flows not only between the system and ESS, but also between all power plants and the load (if the installed equipment allows);
- effective use of the potential of the wind plant by installing additional modules to collect information on environmental conditions and development of control actions for the purpose of managing the power plants of the complex.

The main distinguishing feature of energy-generating complexes based on DER is the stochastic character of the parameters of the primary energy source. This applies to helio- and wind power. In this case, it is necessary maximum use of distributed energy resources in all operation modes of the hybrid power system. In conditions of randomly varying solar activity, wind speed and variable electrical loads, providing a reduction in the load of the supply line.

IV. AN APPROACH TO THE DISTRIBUTION OF LOADS BETWEEN THE GENERATING AGGREGATES

When connecting to the distribution network of autonomous power plants based on DER, one of the main tasks to be performed for each hour of the calculation period is the distribution of loads between the generating aggregates. The following approach is used to solve this problem:

1. It is determined (based on climatic conditions and astronomical parameters of the Sun) the potentially possible generation of power by PVP and WPP (at their specified installed capacity).
2. It is determined losses of active power in the transmission of potentially possible generated energy by PVP and WPP to the consumer.
3. If the summary power of generated by PVP and WPP energy is not less than the load of the consumer, the all it is covered by renewable energy sources.
4. If there is surplus of power, then it is transmitted to charge of the accumulators. In the case when the accumulators have already been fully charged, then on the corresponding amount reduces the power of DER. Initially regulated of wind turbine (as the most wearing), and then PVP.
5. If the total capacity of distributed energy resources is less than the load of consumers, the lack of capacity is covered by the accumulators.
6. If the accumulators are discharged (the charge 30% of the nominal value), then the DGP is switched on. It turns on at full power and turns off when the storage battery reaches a certain threshold value (50% of the nominal value).
7. If the power of the DGP in conjunction with DER and discharged accumulator is less than the load of the consumer, then there is a shortage of electricity.

V. CONCLUSION

A comparative analysis of the schemes of autonomous power plants showed that the most promising option for conjugation different types of power plants in one power system is the use of an intermediate DC insert.

The proposed scheme of the autonomous power supply system with DC insertion provides the following possibilities:

- inclusion in the system of any autonomous power plant irrespective of the installed power equipment;

-software configuration of the energy storage system for a specific energy complex by connecting to the control system of the ESS of a personal computer when performing this operation;

- it is simpler to summarize and distribute energy flows from generating sources and implement effective control algorithms for this process;

-- management of energy flows not only between the system and ESS, but also between all power plants and the load (if the installed equipment allows);

- effective use of the potential of the wind power plant by installing additional modules for collecting information on environmental conditions and developing control actions for the purpose of managing the power plants of the complex.

REFERENCES

[1] A.Sh.Arifjanov, and R.A.Zakhidov, "An approach to the creation of the adaptive control system for integration of non-steady power sources into a common electric power grid", *In Proceedings of the tenth international conference on management science and engineering management. Springer*, pp. 563–574, 2016.

[2] A.Sh.Arifjanov, and R.A.Zakhidov, "Management of technological modes of system distributed generation electric energy on the basis of daily schedules of electric loadings", *In Proceedings of the eleventh international conference on management science and engineering management*. Springer, pp. 1198 – 1209, 2017.

[3] V.I.Chindyaskin, and D.V.Grinko, "The choice of the optimal solution for the application of combined installations based on renewable energy sources", *Proceedings of the Orenburg State Agrarian University*, 1 (45), 2014.

[4] X. Liu, P. Wang, and P. C. Loh, "A hybrid AC/DC microgrid and its coordination control," IEEE Trans. Smart Grid, vol. 2, no. 2, pp. 278–286, Jun. 2011.

[5] L. A. Souza Ribeiro, O. R. Saavedra, S. L. de Lima, and J. Gomes de Matos, "Isolated micro-grids with renewable hybrid generation: The case of lençóis island," IEEE Trans. Sustain. Energy, vol. 2, no. 1, pp. 1–11, Jan. 2011

[6] P. C. Loh, L. Ding, C. Yi Kang, and F. Blaabjerg, "Autonomous Operation of Hybrid Microgrid With AC and DC Subgrids," Power Electronics, IEEE Transactions on, vol. 28, no. 5, pp. 2214–2223, 2013.

[7] C. Jin, P. Wang, J. Xiao, Y. Tang, and F. H. Choo, "Implementation of hierarchical control in DC microgrids," IEEE Transactions on Industrial Electronics, vol. 61, no. 8, pp. 4032–4042, 2014.

[8] A.Sh. Arifjanov and R.A.Zakhidov, "Grid connection management of distributed generators on the basis of renewable energy sources", *Applied Solar Energy*, 2017, Vol.53, No.4, pp. 347-353. Allerton Press. Inc., 2017

[9] Chauhan A., Saini R.P. A review on Integrated Renewable Energy System based power generation for stand-alone applications: Configurations, storage options, sizing methodologies and control // *Renewable and Sustainable Energy Reviews*. Vol. 38. pp. 99–120, 2014.

[10] J. Driesen and F. Katiraei, "Design for Distributed Energy Resources," *IEEE Power and Energy Magazine*, vol. 6, pp. 30–40, May/June 2008.

[11] J. M. Guerrero, J. C. Vasquez, J. Matas, M. Castilla, and L. G. de Vicuna, "Control Strategy for Flexible Microgrid Based on Parallel Line-Interactive UPS Systems", *Industrial Electronics, IEEE Transactions on*, vol. 56, pp. 726-736, 2009.

[12] H. Chaoyong, H. Xuehao, and H. Dong, "Hierarchical control techniques applied in micro-grid", in *Power System Technology (POWERCON), 2010 International Conference on*, 2010, pp. 1-5.

[13] W.Fedak, S.Anweiler, R.Ulbrich, B.Jarosz, "The Concept of Autonomous Power Supply System Fed with Renewable Energy Sources", *Journal of Sustainable Development of Energy, Water and Environment Systems*, Vol. 5, issue 4, pp. 579-589, 2017.

[14] Ackermann, T., Andersson, G., and Söder, L., "Distributed generation: a definition", *Electric power systems research*, vol. 57, pp. 195-204, 2001.

[15] S.G.Obukhov and I.A.Plotnikov, "Comparative analysis of schemes for constructing autonomous power plants using renewable energy installations", *Industrial Energy*, No.7, pp.38-51, 2012.

3D Visualization Thyroid CT Images Using Marching Cubes Algorithms

A.H. YURTTAKAL[1] , H. ERBAY[2] , T. İKİZCELİ[3], S. KARAÇAVUŞ[4], G. ÇINARER[5]

[1] Bozok University, Yozgat/Turkey, ahyurttakal@gmail.com
[2]Kirikkale University, Kirikkale/Turkey, hxe68@yahoo.com
[3]University of Health Sciences, İstanbul/Turkey, turkan.ikizceli@sbu.edu.tr
[4] University of Health Sciences, Kayseri/Turkey, seyhan.karacavus@sbu.edu.tr
[5]Bozok University, Yozgat/Turkey, gokalp.cinarer@bozok.edu.tr

Abstract **- Thyroid cancer is the type of cancer caused by the cells of the thyroid gland. This is less common than other types of cancer. If the correct diagnosis is made and appropriate treatment is given, the disease can be completely removed.**
In this study, 3D models of thyroid cancer were modeled using DICOM images. As is known, dicom is the de-facto file standard in medical imaging, and these files contain many metadata. We used some of these meta-attributes to calculate the Hounsfield Unit. In addition, the pixel values are calculated according to the average attenuation of the tissue corresponding to a scale of -1024 to + 3071 on the Hounsfield scale. The DICOM images used in the study were obtained from real patients under the supervision of specialist doctors. Thyroid cancer tumors were modeled as 3D using the pixel values of Marching Cubes Algorithm.

Keywords **– Computed Tomography (CT), Marching Cubes Algorithms, Hounsfield Unit (HU)**

I. INTRODUCTION

Thyroid cancer is the type of cancer caused by the cells of the thyroid gland.Thyroid cancer is relatively uncommon compared to other cancers. In the United States it is estimated that in 2016 approximately 64,000 new patients will be diagnosed with thyroid cancer, compared to over 240,000 patients with breast cancer and 135,000 patients with colon cancer. If the correct diagnosis is made and appropriate treatment is given, the disease can be completely removed [4].

Computed tomography (CT) is a diagnostic imaging test used to create detailed images of internal organs, bones, soft tissue and blood vessels. The cross-sectional images generated during a CT scan can be reformatted in multiple planes, and can even generate three-dimensional images which can be viewed on a computer monitor, printed on film or transferred to electronic media. CT scanning is often the best method for detecting many different cancers since the images allow your doctor to confirm the presence of a tumor and determine its size and location [5].

Volume visualization has been applied in many problem domains and, as such, has become an important tool for exploring data and discovering knowledge. Commonly, the domain data to be visualized is scalar volumetric data. Scalar volumetric data can be defined as a collection A of 3D points $P_i=(x_i,y_i,z_i)$, each of which has its own scalar value or property [6].

This paper presents a study of the Marching Cubes Algorithm used for generating 3D structures from 2D dicom images. We used Thyroid CT dicom images and thresholded it to values over 350 HU and 0 HU, thus isolating all bone structures and water structures, and then use the marching cubes algorithm to create a 3D mesh of those structures.

Dicom images were taken from a local hospital. Our solution was made in Python programming language and open source scientific libraries like Pydicom, Matplotib, Scikit-Learn.

II. MARCHING CUBES ALGORITHM

The Marching Cubes (MC) algorithm by Lorensen and Cline is most used algorithm for extraction of isosurface out of volumetric data. This algorithm produces a triangle mesh by computing iso-surfaces from discrete data. By drawing all these triangles, we can build a three dimensional representation of the CT-Scans [7].

Marching cubes uses a divide-andconquer approach in which volume data is processed through voxels. For the processing, multi-image slices are arranged as a multidimensional array. And two adjacent slices are taken into consideration at a time as shown in Figure 1 [10].

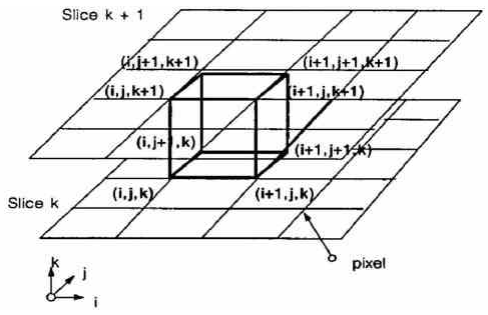
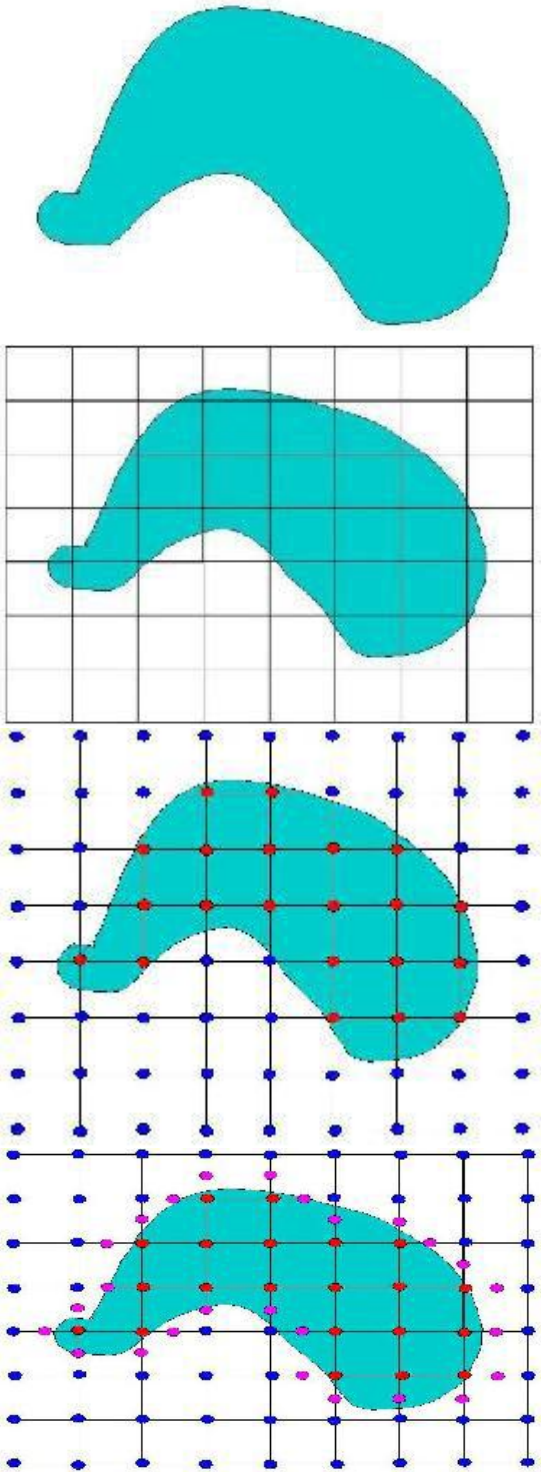


Figure 1: Cube formed on lattice [3]

The marching cube algorithm can be divided into two main parts. The first part is to figure out how to define the sections of surface that cut up an individual cube. There are a total of 256 possible configurations of corner classifications if we classify each corner as either being below or above the isovalue. In these

256 possible configurations, two of them are trivial. They are the cases where all points are inside or outside the cube, which does not contribute to the isosurface. For the rest of the configurations, we have to decide whether the isosurface crosses along each cube edge, and use these edge intersection points to construct one or more triangular patches for the isosurface. When the value of the vertex is greater than or equal to isovalue it is internal, and when it is less than isovalue it is external. We assign 0 to vertex outside the surface and assign 1 to vertex inside the surface in Figure 2 [7].



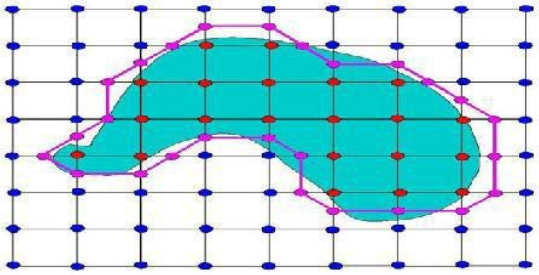


Figure. 2: Pre-Processing
(A) Image (B) Segmented Image (C) Intersected Point (D) Offset Points (E) Join all Offset points [8].

When symmetries cases are considered in Figure 4, there are only 14 unique configurations in the rest of the 254 configurations indeed in Figure 3.

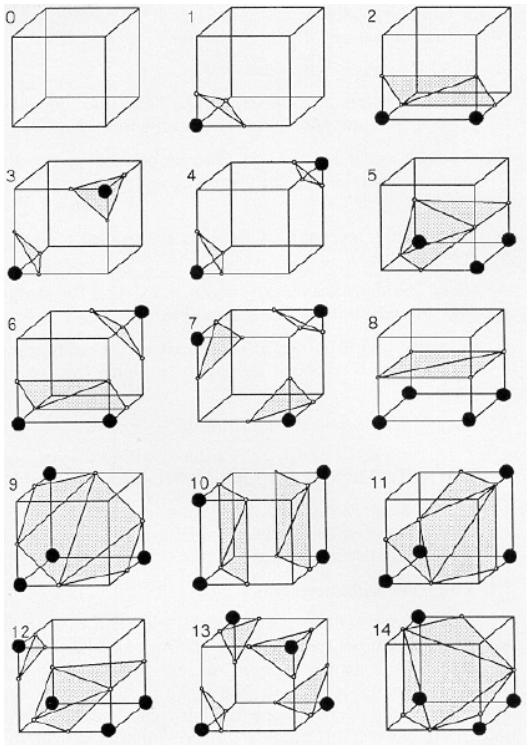


Figure 3: Unique configurations

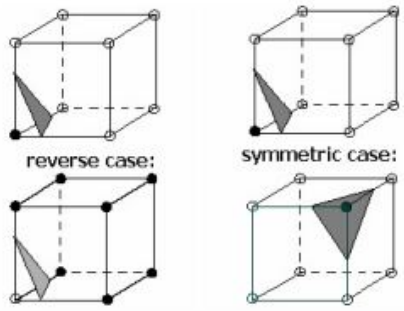


Figure 4:Symmetry cases

In the case that only one corner is less than the isovalue, it forms a single triangle which intersects the edges which meet at this corner, with the patch normal facing away from the corner.

There are 8 similar configurations belongs to this case. We should calculate an index for the cube by comparing the eight density values at the cube vertices with the surface constant.in Figure 5 [7].

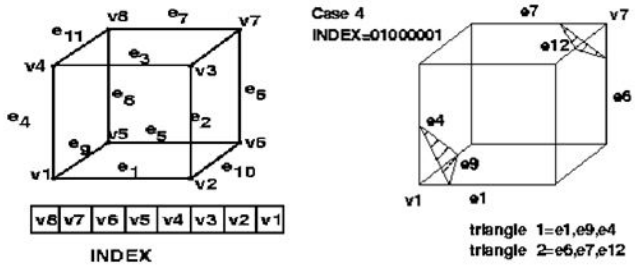


Figure 5: Triangulation

The final step in marching cubes is to calculate a unit normal for each triangle vertex. It is used to create Gouraud-shaded images in the rendering algorithms. Interpolate the normal to each triangle vertex and output the triangle vertices and vertex normals [7].

To calculate surface normal, we need to determine gradient vector g. The gradient at cube vertex (i , j, k), is estimated using central differences along the three coordinate axes by:

$G_x(i,j,k)=D(i+1,j,k)-D(i-1,j,k) \ / \Delta x$
$G_y(i,j,k)=D(i,j+1,k)-D(i,j-1,k) \ / \Delta y$
$G_z(i,j,k)=D(i,j,k+1)-D(i,j,k-1) \ / \Delta z$

D (i, j, k) is the density at pixel (i, j) in slice k. Δx, Δy, Δz are lengths of the cube edges [3]

III. HOUNSFIELD UNIT

The pixel value in a CT image give us a direct way to roughly identify the tissue type each pixel belongs to. Thus, simply by means of thresholding the image to known ranges, we can extract given types of tissue straight from the image without hassle so we calculated Hounsfield Unit [2].

The Hounsfiled scale is a quantitative measure of radiodensity. Pixels in an image obtained by CT scanning are displayed in terms of relative radiodensity. The pixel value is displayed according to the mean attenuation of the tissue that it corresponds to on a scale from –1024 to over 3000 on the Hounsfield scale. Water has an attenuation of 0 Hounsfield units (HU) while air is –1000 HU, bone is typically +400 HU or greater and metallic implants are usually +1000 HU [2].

The algorithm requires a data volume and an isosurface value. For example, in CT imaging Hounsfield units of +700 to +3000 represent bone. So, one potential input would be a reconstructed CT set of data and the value +700, to extract a mesh for regions of bone or bone-like density [9].

We converted raw pixel-data to HU. A typical CT image is not saved directly in HU. In 99% of the cases the 'raw' pixel-data needs to be rescaled. Therefore, DICOM images of CT examinations typically contain two values in their metadata, most commonly called RescaleSlope and RescaleIntercept. These values are simply the a and b in a typical linear transformation. The formula for calculating the HU out of the raw pixel value is [2]:

HU = PixelValue * RescaleSlope + RescaleIntercept

The materials corresponding to the Hounsfeld value are shown in Table 1

Table 1: Hounsfiel Unit-Material [1]

Material	**Hounsfield Unit**
Air	-1000
Lung	-500 to -200
Fat	-200 to -50
Water	0
Blood	25
Muscle	25 to 40
Bone	200 to 1000

IV. EXPERIMENTAL RESULTS

Thyroid dicom images were taken from a local hospital. The properties of the images used in the study are given in Table 2. This means we have 3,27 mm slices, and each voxel represents 1,36 mm. A CT slice is typically reconstructed at 512 x 512 voxels, each slice represents approximately 700 mm of data in length and width.

Table 2:Properties of images

Properties	**Value**
Slice Thickness (mm)	3,27
Pixel Spacing (row, col) (mm)	1.367188, 1.367188
Number of Slices	247
Dimensions	512x512

Using DICOM's metadata, we can find the size of each voxel as the slice thickness. That's why we have resampled each slice into 1x1x1 mm pixels and slices to view the CT in 3D isometric form and to compare different scans. After sampling, new dimensions are 808x700x700.

CT images are shown in Figure 6 by skipping every 10 slices to get an overview of the work.

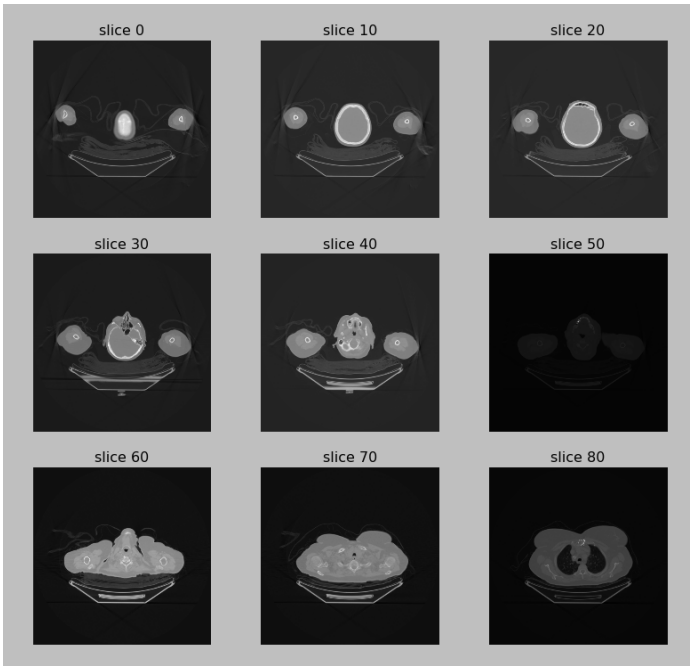


Figure 6:Slices

The histogram of all voxel data is shown in Figure 7. According to the histogram Mostly there is plenty of soft tissue a little fat,

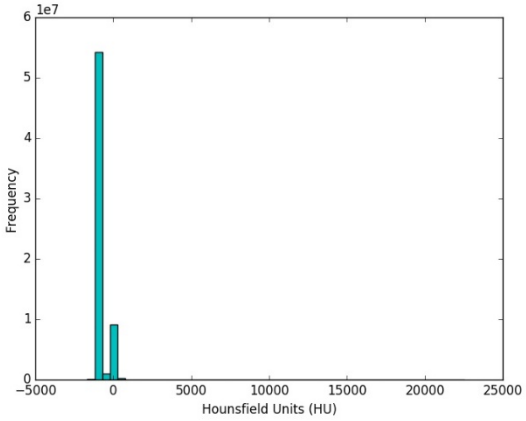


Figure 7: The histogram of all voxel data

Thyroid CT dicom images thresholded it to values over 350 HU, thus isolating all bone structures and then use the marching cubes algorithm to create a 3D mesh of those structures. In Figure 8, bone structures are shown.

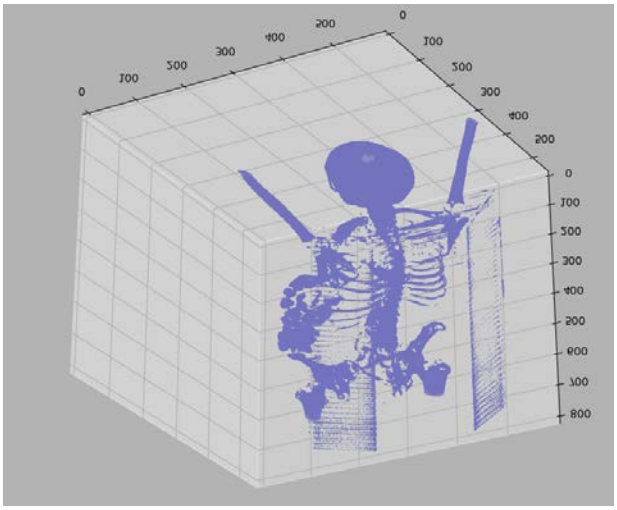


Figure 8: Bone Structures

Dicom images thresholded it to values over 0 HU, thus isolating all water structures and then use the marching cubes algorithm to create a 3D mesh of those structures. In Figure 9, water structures are shown.

A practical application of this is in evaluation of tumors, where, for example, an tumor with a radiodensity of less than 10 HU is rather fatty in composition and almost certainly a benign [11].

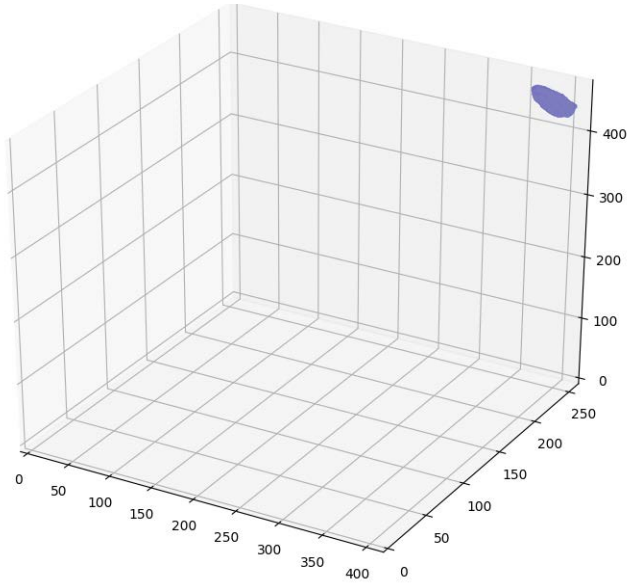


Figure 9: 3D Segmented Tumor

Calculated verts matrix is [[361. 252. 453.][362. 252. 452.][362. 251. 453.],[411. 254. 443.][411. 254. 444.][411. 254. 445.]]
Calculated faces matrix is [[2 1 0][4 3 0][0 3 2],[4494 4508 4495][4495 4508 4509][4495 4509 4491]]
As a result, the tumor volume is calculated. According to the results obtained, the tumor is benign.

V. CONCLUSION

Calculation of tumor volume and 3D modeling are very important in cancer treatment. Rendering of medical data in 3D is an appreciated tool that will help doctors in identifying abnormalities and treating with better sureness.

In this paper, we are presenting a useful tool implemented by ourself for 3D Tumor modeling, which could generate a three dimension models from two dimension slice data. Our approach is primarily based on the marching cube algorithm. The coordinates of the extracted model are compatible with the dicom image

REFERENCES

[1] Hounsfield scale. (2018, March 14). Retrieved March 21, 2018, from https://en.wikipedia.org/wiki/Hounsfield_scale

[2] Surface Extraction: Creating a mesh from pixel-data using Python and VTK. (2014, October 26). Retrieved March 21, 2018, from https://pyscience.wordpress.com/2014/09/11/surface-extraction-creating-a-mesh-from-pixel-data-using-python-and-vtk/

[3] William E. Lorensen, Harvey E. Cline, "Marching Cubes: A high resolution 3D surface construction algorithm" Computer graphics, volume 21, July 1987.

[4] Thyroid Cancer. (n.d.). Retrieved March 21, 2018, from https://www.thyroid.org/thyroid-cancer/

[5] Computed Tomography. (n.d.). Retrieved March 21, 2018, from https://www.radiologyinfo.org/en/submenu.cfm?pg=ctscan

[6] Jones M, Leu A, Satherley R, Treavett S. Glossary. In: Chen M, Kaufman A, Yagel R, editors. Volume graphics. London: Springer; 2000. p. 395–406.

[7] Sio, C. C., Ngan, M., Yi, J., & Chen, X. (2011). Volume Rendering with Marching Cube Algorithm. *University of Southern California, date Feb, 11*.

[8] An Implementation of the Marching Cubes Algorithm. (n.d.). Retrieved March 21, 2018, from http://www.cs.carleton.edu/cs_comps/0405/shape/marching_cubes.html

[9] Marching Cubes. (n.d.). Retrieved March 21, 2018, from http://scikit-image.org/docs/0.13.x/auto_examples/edges/plot_marching_cubes.html

[10] Parmar, B. N., & Bhatt, M. T. (2016). Volume Visualization using Marching Cubes Algorithms: Survey & Analysis.

[11] "Adrenal Adenoma Imaging." *Overview, Radiography, Computed Tomography*, 9 Mar. 2017, emedicine.medscape.com/article/376240-overview.

Design and Static Stress Analysis of Lifting Hook

A. OZDEMIR[1] and Y.A. ONUR[2]

[1]Turkish Hardcoal Enterprise, Zonguldak/Turkey, ozdemiralp@gmail.com
[2]Bulent Ecevit University, Zonguldak/Turkey, aytaconur@beun.edu.tr

Abstract **- Lifting hooks are the components that ensure reliable and economical lifting and transport of loads. By selecting the shapes, dimensions and appropriate materials to be used, a reliable and economical lifting hook is obtained. In this study, stress analyzes of lifting hooks were carried out primarily with two different analytical approaches. Then, modeling of the hook was made in computer environment and the finite element analysis was done in the computer environment. The analysis results and standardized values for different loads are compared under the same conditions. The study shows that computer-aided finite element analysis can be applied reliably and economically in studying the stress states of lifting hooks.**

Keywords **– Lifting hook, static stress analysis, 3D modeling.**

I. INTRODUCTION

Lifting hooks are the most reliable and economical load carrying elements and used to be most commonly for lifting or transporting of loads. It is expected that the lifting hooks will be able to connect and grasp the load in a short period of time to provide work safety. At the same time, it is expected that the cargo handling operations will be carried out with minimum labor. It is basically a hoisting fixture that is designed to engage a ring or link of a lifting chain or the pin of a shackle or cable socket and it must be suitable for health and safety guidelines [1]. Choosing the proper shape of hook and proper material increases the load capacity of the hook and makes the transport operations easier [2-4]. The lifting hooks are complex and curved machine elements in terms of their geometric structure. Due to the hook construction, it is difficult to reliably determine the stress caused by the load effect on the lifting hooks. There are many alternative approaches for calculating the tension in the lifting hooks. Some of the analysis methods in the hooks are the simple beam theory method, the curved beam theory calculation method, the finite element analysis method in the order of historical development. In recent years the development of computer technology has also been used frequently for complex stresses in the computer environment. In this study, the lifting hook with number 05 where numbers are described in DIN 15401 norm is used. Maximum stress values of the lifting hook considering 25 kg, 250 kg, 500 kg and 1000 kg have been calculated and presented. 1000 kg load is the maximum working load of the lifting hook with number 05. The stress analysis of the number 05 hook was made according to the approximate calculation method, the curved beam theory method and the finite element analysis (FEA). As a result of the comparison of the stress analyzes, it has been determined that the design of the lifitng hook is safe and economical with FEA.

II. LIFTING HOOKS AND ANALYTICAL STRESS CALCULATION

The lifting hook geometries have complex curved axes for easy handling of the load and it enables maximum load transfer with small hook sizes. In general, simple lifting hooks and ramshorn hooks are widely used. It is produced from materials specified in DIN 15400 [5]. The simple lifting hooks are standardized in accordance with DIN 15401 norm [6]. There are many kind of hooks for special purposes such as point hooks, ramshorn hooks, eye hooks, cargo hooks and cross hooks.

It is difficult to calculate stresses occurred on lifting hooks due to having curved geometry. There are two basic analytical approaches to calculate maximum stresses in lifting hooks.

The first approach is known as approximate calculation method. In approximate calculation, the simple hook is assumed to be like a beam which is loaded on its axis with a small force and the stress distribution is assumed to be linear. The control calculations of the two critic cross-sections are done separately. Force should be calculated as if the load is hanged on the sling [7].

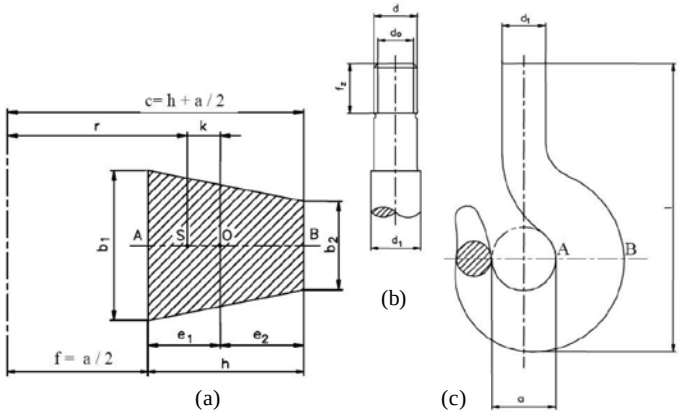


Figure 1: Cross-section and dimensions of a simple hook

Figure 1a, 1b, 1c show the dimensions of the hook section. In Figure 1a and Figure 1b , values of b_1, b_2, h, f, c, a are taken from simple hook standard [5]. Q is the lifted load, A is the cross-sectional area of the trapezoidal section, M is the bending moment. Accordingly, the distances to the center of curvature are; distance of outer fibers is c, distance of the inner fibers is f, distance between S neutral axis and center of curvature is r, the

S-O distance is expressed as k and the distance between O centroidal axis and center of curvature is R. Equations (1) to (6) show required maximum stress calculation equations of lifting hook considering approximate calculation method. Maximum stress (σ_{max}) occurs inner curved point A.

$$I = ((b_{1^2} + 4.b_1.b_2 + b_{2^2}) / (b_1 + b_2)).(h^3 / 36) \quad (1)$$

$$e_1 = ((b_1 + 2b_2) / (b_1 + b_2)).h / 3 \quad (2)$$

$$W = J / e_1 \quad (3)$$

$$M = Q.R \quad (4)$$

$$A = ((b_1 + b_2) / 2).h \quad (5)$$

$$\sigma_{max} = (M / W) + (Q / A) \quad (6)$$

In the second analytical stress approach, the hook curve is treated as curved beam. This method is called as curved beam theory where centroidal axis and neutral axis are not coincident. Analysis is made by assuming that the cross section has an axis of symmetry in a plane along the length of the beam, plane cross sections remain plane after bending, the modulus of elasticity is the same in tension as in compression. Equations (7) to (9) show required maximum stress calculation equations of lifting hook considering curved beam theory. Maximum stress (σ_{max}) occurs inner curved point A.

$$r = \frac{A}{b_2 - b_1 + \left[(b_1 c - b_2 f) / (c - f)\right] \ln(\frac{c}{f})} \quad (7)$$

$$R = r + k \quad (8)$$

$$\sigma_{max} = (M.h_1) / (A.k.f) + (Q / A) \quad (9)$$

$$h_1 = r - f$$

III. Computer Aided Lifting Hook Modeling and Finite Element Analysis

The development of computer technology in recent years has resulted in high processor power and high memory capacities that enable computers to use in engineering calculations. Through finite element analysis these sophisticated simulations provide valuable information for designing and developing new products [8]. In order to perform stress analysis in a computer environment, the lifting hook must first be reliably modeled. In this study, the hook with DIN 15401, 05 number is modeled in Solidworks 2016 software [5].

The finite element method is used reliably and efficiently in the computer environment, especially in the stress analysis of solids with complex geometries. In this study, Ansys Workbench 16.0 finite element analysis software was used to analyze the modeled hook.

IV. Stress Analysis with Different Approaches

Investigated lifting hook with 05 number has 1000 kg lifting capacity according to the 2_m drive group [9] and strength class S [5]. The material of the hook is selected to be 34CrMo4 (Young's Modulus (E) is $2.1x10^5$ N / mm^2, yield strength is 390 N / mm^2 and Poisson's ratio is 2.88 [10].

25 kg, 250 kg, 500 kg and 1000 kg loads were applied to the designed lifting hook respectively. The static analysis of the hook was made by the approximate calculation method, the curved beam method and the finite element method. In the FEA, the tetrahedron finite element with 10 nodes having 7 mm size is used in ANSYS Workbench software. The hook model is divided into 20574 elements with 31218 nodes. Results obtained by using different calculation methods are presented in Table 1.

Table 1: Results obtained by using different approaches

Load	Maximum stress values (MPa)		
	Approximate calculation method	Curved beam theory	FEA
25 kg	2.17	3	3.53
250 kg	21.65	30.01	35.32
500 kg	43.31	60.02	70.65
1000 kg	86.61	120.03	141.22

In this study, safety factor is also calculated and presented in Table 2 where allowable maximum working load of 1000 kg is considered.

Table 2: Safety factors of investigated hook

Load	Safety factor		
	Approximate calculation method	Curved beam theory	FEA

1000 kg	4.50	3.25	2.76

Approximate calculation method and curved beam theory are used to calculate maximum stresses occurred on lifting hook investigated for 4 specific loads. In the same way, the FEA was carried out considering boundary conditions.

For the lowest applied load of 25 kg, the approximate calculation method gives a lower maximum stress value than the other methods (2.17 MPa) and the maximum stress value in the FEA is the highest (3.53 Mpa). The maximum stress (141.22 MPa) is obtained by FEA at 1000 kg, which is the maximum allowable load of the hook according to the operating standards, is still 17.65% higher than curved beam. There is a good harmony in the results between curved beam theory and FEA. Approximate calculation method results give lowest maximum stress values among other methods.

V. CONCLUSION

This study demonstrates the calculation of the stress values of lifting hooks which are used as load holding and lifting duties. The static stress analyzes of the number 05 hook have been done when the hook was exposed to loads of 25 kg, 250 kg, 500 kg and 1000 kg under certain operating conditions. Two different analytical methods are used to determine maximum stresses. The FEA of the designed hook is then carried out. One of the remarkable results that approximate calculation method results are quite different and low stress value compared to other methods. The reason for this is that in the approximate calculation method, the curved structure of the hook is assumed to be straight.

FEA at 1000 kg load results in 141.22 MPa maximum stress and 2.76 safety factor. There is a good harmony in the results between curved beam theory and FEA. Approximate calculation method results give lowest maximum stress values among other methods. With finite element analysis, stress analysis of the lifting hook can be made economical as well as reliable.

REFERENCES

[1] Rashmi Uddanwadiker; "Stress Analysis of Crane Hook and Validation by Photo-Elasticity," *Engineering*, 2011, vol.3, pp. 935-94.

[2] N. N., *DIN Taschenbuh 185 Krane und Hebezeuge 2*, Beuth Verlag, Berlin, 1995.

[3] H. Ernest, Die Hebezeuge Band 1, *Grunderlagen und Bauteile*, Verlag, Braunschweig, Germany, 1973.

[4] H.J. Zebich, Fördertechnik 1 Hebezeuge, Vogel,Verlag, Germany, 1975.

[5] DIN 15400, Lifting hooks; materials, mechanical properties, lifting capacity and stresses, *Deutsches Institut für Normung*, 1990.

[6] DIN 15401, Lifting hooks for lifting appliances; Single hooks; Unmachined parts, *Deutsches Institut für Normung*, 1983.

[7] . Ozer Derya, A. B. Erdil, I. Gerdemeli Ismail, "Finite element approach to 3-D modelling and stress analysis for crane lifting hooks," *Eleventh International Research/Expert Conference Trends in the Development of Machinery and Associated Technology*, Hammamet, Tunisia, pp. 1007-1010, 2007.

[8] Y. A. Onur, C. E. Irmak, "Design and static stress analysis of elevator car suspension during operation," *12th International Research/Expert Conference Trends in the Development of Machinery and Associated Technology*, pp. 821-824, Istanbul, August 2008.

[9] FEM 9.511/86, Rules for the design of lifting equipment series; Classification of mechanisms, *European Materials Handling Federation*, 1986.

[10] DIN 17100, Steels for General Structural Purposes, *Deutsches Institut für Normung*, 1980.

A Voice Encryption Application Based on a Chaotic System with Single Parameter

S. KAÇAR[1], F. YALÇIN[2], B. ARICIOĞLU[3] and A. AKGÜL[4]

[1]Sakarya University, Sakarya/Turkey, skacar@sakarya.edu.tr
[2]Sakarya University, Sakarya/Turkey, farukyalcin@sakarya.edu.tr
[3]Sakarya University, Sakarya/Turkey, baricioglu@sakarya.edu.tr
[4]Sakarya University, Sakarya/Turkey, aakgul@sakarya.edu.tr

Abstract **- In this study, a voice encryption application is realized by using a four dimensional (4D) chaotic system with single parameter. The main purpose of use of a single parameter chaotic system is reduction of computational load in order to run the voice encryption application on low performance hardware. Since the chaotic system is four dimensional, the system has four state variables and for every state variable different values of initial conditions can be used. This makes the key space length sufficiently long. In the application, the state variables are obtained with Runge-Kutta (RK4) method. The voice encryption is realized by XORing the obtained state variables with the voice data. The security performance of the voice encryption is proved by comparing the encrypted and original voice data in both time and frequency domain.**

Keywords **- Data security, Voice encryption, Chaotic systems**

I. INTRODUCTION

DATA security is one of the most popular nowadays subject in the literature. A secure communication between persons is very important subject of data security. Accordingly, a chaos based voice encryption application is realized in this study.

The use of chaotic system in data security has become very common in the literature [1-5]. As it can be seen in these studies, the chaotic systems have become a very effective tool in ensuring of data security. In addition, chaotic system must meet some properties to be used in data security applications, especially in voice encryption application. In this study, a four dimensional (4D) chaotic system with one parameter which have the necessary properties to be used in the voice encryption application [6]. The system state variables obtained with Runge-Kutta 4 (RK-4) algorithm are used in the voice encryption application.

In the second section, the chaotic system used in the encryption is presented. In the third section, the voice encryption based on the chaotic system is presented. In the fourth section, the results of the encryption are presented. In the last section, Conclusion is given.

II. 4D CHAOTIC SYSTEM WITH SINGLE PARAMETER

In this section, 4D chaotic system with one parameter is briefly introduced. The chaotic system is given in Equation 1.

$$\begin{aligned} \dot{x} &= a(y-z)-w \\ \dot{y} &= w-y-zw \\ \dot{z} &= w-z+xy \\ \dot{w} &= x+y+zy \end{aligned} \tag{1}$$

As it can be seen in Equation 1, the behavior of the system is only depend on parameter a. This enables to obtain the results of the system more quickly. Also, the system has three nonlinear terms that the system is not very complex. In other words the computational complexity of the system is quite low. In the previous study [6], it is showed that the system in Equation 1 exhibits chaotic behavior and it is also proved that the system has sufficient randomness for the encryption. Accordingly, the chaotic system has the necessary properties for the voice encryption application such as low computational complexity and sufficient randomness.

III. VOICE ENCRYPTION

In this section, the realized voice encryption application is explained. In Figure 1, flow chart of the encryption process is given. As it can be seen in Figure 1, the values of the system parameter, initial conditions and iteration step interval for RK4 algorithm is entered as a first step. In the study the system parameter is set as a=1.5, the initial conditions are set as x_0=y_0=z_0=w_0=1 and the iteration step interval is set as 0.01 [6]. After this step, the chaotic system's equations are solved with RK4 algorithm and the system state variables are obtained. Then, the obtained state variables are converted from floating to binary format. At the same time, the voice data is converted to binary format. The encryption process is realized by XORing the bit series obtained from original data with the bit series obtained from the chaotic system. Then, the encrypted bit series is converted back to the floating point format to obtain the encrypted voice data. For the decryption process, the original voice is replaced with the encrypted voice in the flow chart.

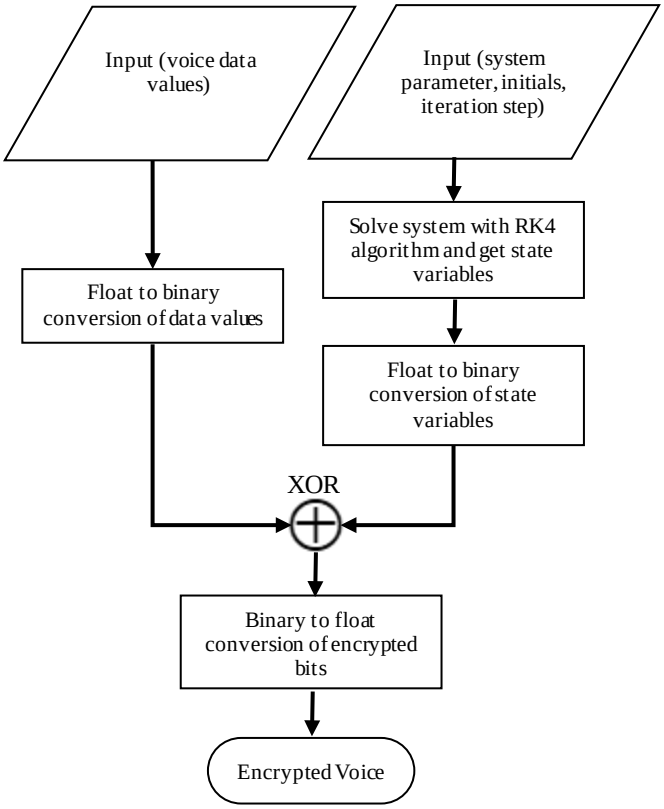


Figure 1: Encryption process.

IV. RESULTS OF APPLICATION

In this section, the results of voice encryption application is presented. The graphs of the original, encrypted and decrypted voices are given as in Figure 2,3 and 4 respectively when the system parameter is set as a=1.5, the initial conditions are set as x_0=y_0=z_0=w_0=1 and the iteration step interval is set as 0.01. If Figure 2 and 3 are examined, there is no similarity between the original and the encrypted voices. If Figure 2 and 4 are examined, after the encryption and decryption processes the original data is obtained without any loss and distortion. As it can be seen in the figures the encryption and the decryption processes are realized successfully. To evaluate the performance of the voice encryption application, the frequency spectrum of the original, encrypted and decrypted voice data are examined as a different method. The frequency spectrum of the original, encrypted and decrypted voice data is given in Figure 5, 6 and 7 respectively. As it can be seen in Figure 5 and 7, the original voice data and the decrypted voice data have the same frequency spectrum. This shows that, there is no data loss or distortion in the both encryption and decryption processes. The frequency spectrum given in Figure 6 is completely different from the spectrums given in Figure 5 and 7 and it is almost distributed homogeneously over the spectrum. This shows that, the encrypted signal is completely structurally different signal than the original signal, and the original signal cannot be obtain with any filtering process. The frequency spectrum analysis shows that the successful performance of the both encryption and decryption processes.

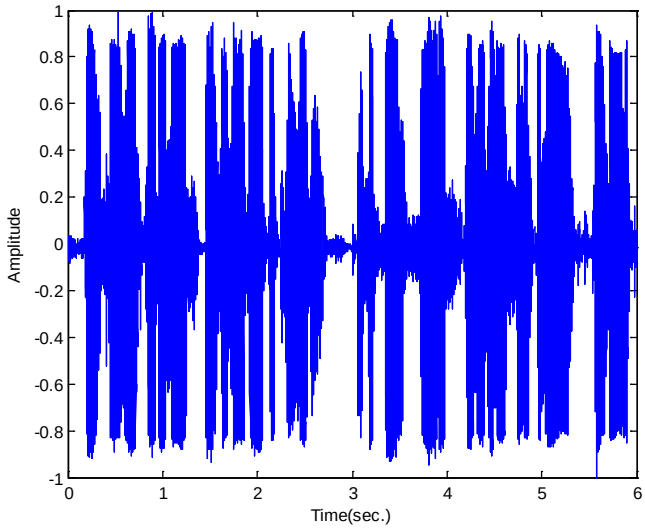


Figure 2: Original voice.

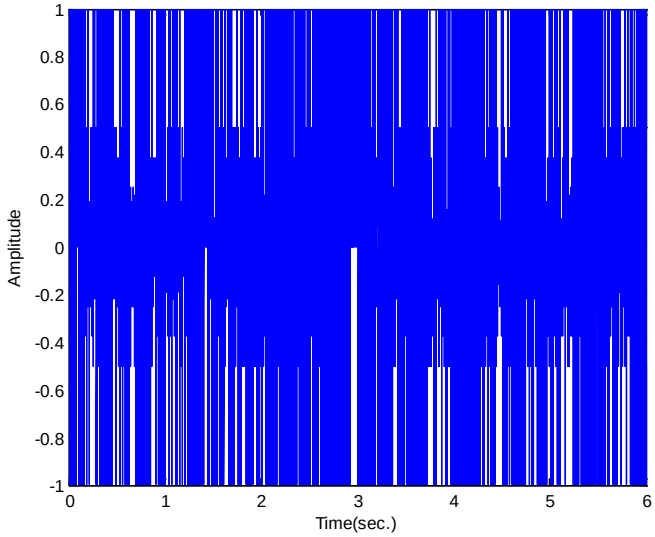


Figure 3: Encrypted voice.

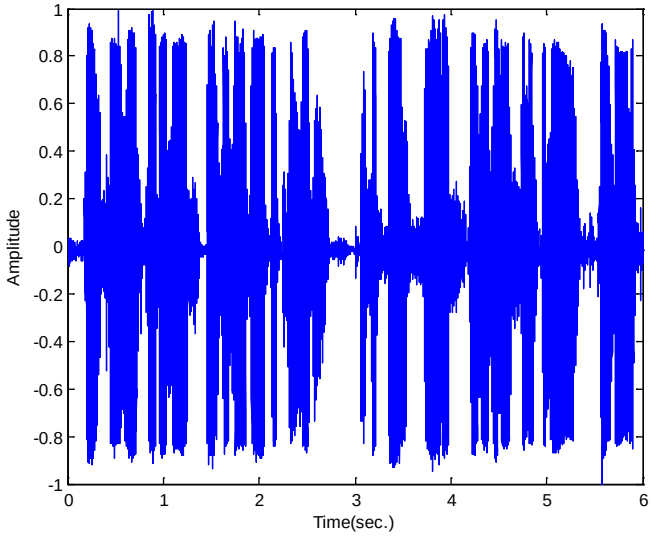


Figure 4: Decrypted voice.

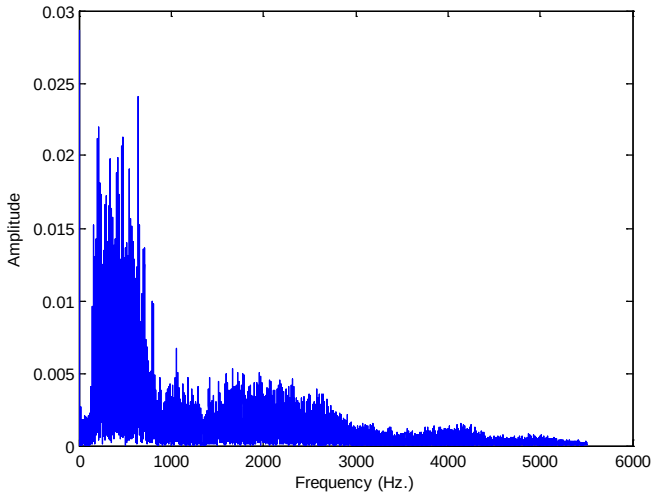


Figure 5: Original voice spectrum.

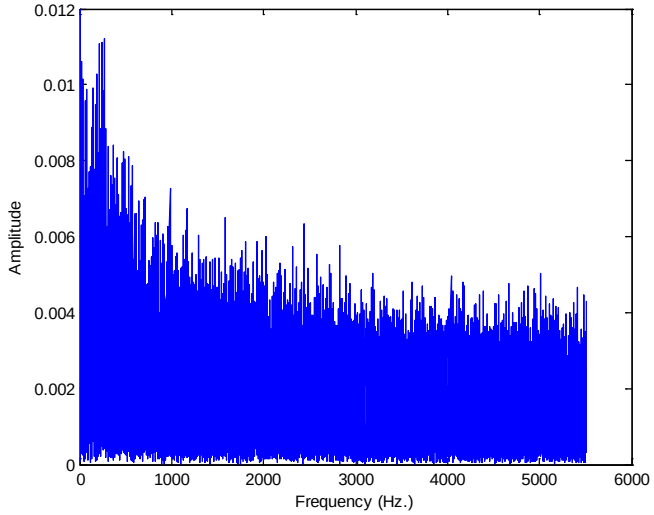


Figure 6: Encrypted voice spectrum.

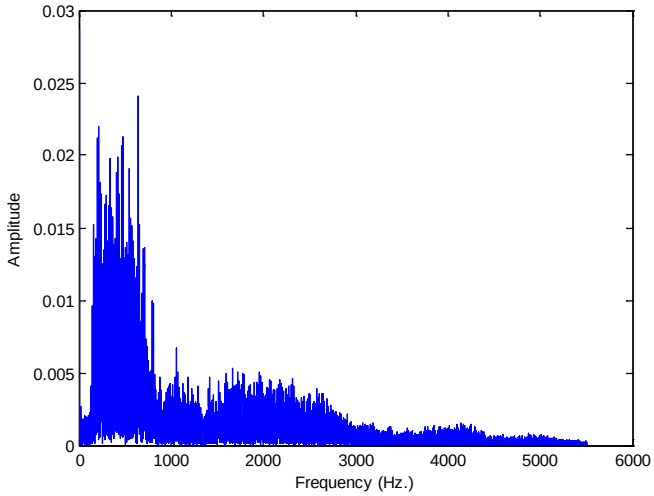


Figure 7: Decrypted voice spectrum.

V. CONCLUSIONS

In this study, a voice encryption application based on a 4D chaotic system with one parameter is realized. The successful performance of both encryption and decryption processes is shown with both time and frequency domain analysis. Hence, this proves that the chaotic system can be used in encryption applications. In the future studies, encryption of different kinds of multimedia data like video and image with the chaotic systemused in this study will be realized.

REFERENCES

[1] S. Vaidyanathan, A. Akgul, S. Kaçar, and U. Çavuşoğlu. "A new 4-D chaotic hyperjerk system, its synchronization, circuit design and applications in RNG, image encryption and chaos-based steganography." *The European Physical Journal Plus*, vol.133, no. 2 pp. 46, 2018.

[2] A. Akgul, S. Kacar, and B. Aricioglu, "A new two-level data hiding algorithm for high security based on a nonlinear system." *Nonlinear Dynamics*, vol .90(2), pp. 1123-1140, 2017.

[3] X. Wang, A. Akgul, S. Kacar, and V. Pham. "Multimedia Security Application of a Ten-Term Chaotic System without Equilibrium." *Complexity* vol.2017, 2017.

[4] Ü. Çavuşoğlu, A. Akgül, S. Kaçar, İ. Pehlivan, and A. Zengin. "A novel chaos-based encryption algorithm over TCP data packet for secure communication." *Security and Communication Networks*, vol.9 no. 11 pp. 1285-1296, 2016.

[5] A. Akgul, I. Moroz, I. Pehlivan, and S. Vaidyanathan. "A new four-scroll chaotic attractor and its engineering applications." *Optik-International Journal for Light and Electron Optics* vol. 127, no. 13 pp. 5491-5499, 2016.

[6] S. Kacar, Ü. Çavuşoğlu, A. Sevin and C. Bayilmis. " A New 4D Chaotic System with Single Parameter and Its RNG Application." in *Conf. Rec. 2017 the 8th INTERNATIONAL ADVANCED TECHNOLOGIES SYMPOSIUM*, pp. 4045-4051

A Pseudo Random Number Generator Design Based on a Four Dimension Chaotic System

S. KAÇAR[1], F. YALÇIN[2], B. ARICIOĞLU[3] and A. AKGÜL[4]

[1]Sakarya University, Sakarya/Turkey, skacar@sakarya.edu.tr
[2]Sakarya University, Sakarya/Turkey, farukyalcin@sakarya.edu.tr
[3]Sakarya University, Sakarya/Turkey, baricioglu@sakarya.edu.tr
[4]Sakarya University, Sakarya/Turkey, aakgul@sakarya.edu.tr

***Abstract* - In this study a novel four dimensional (4D) chaotic system is introduced. To prove chaotic behavior of the system, time series and phase portraits are presented. A novel pseudo random number generator (PRNG) application of the chaotic system is realized to show the system is suitable for engineering applications like encryption and data hiding. In order to realize PRNG, the chaotic system is discretized with numerical methods. The next step is selection of different number of bits from different state variables obtained from the chaotic system. NIST 800-22 statistical tests, the highest international standard, are performed for the generated random numbers to prove the proposed PRNG has sufficient randomness and successful results are obtained.**

***Keywords* - Random Number Generator, NIST Tests, Chaotic systems.**

I. INTRODUCTION

CHAOTIC systems and its applications have become one of the topical subject in the literature in the recent years. Random number generation and encryption applications are the most common applications of the chaotic system [1-5]. The chaotic systems may look like random but, in fact, they can be expressed with mathematical equations. This enables the usage of the chaotic systems in applications such as random number generation and encryption.

In this study, a novel four dimensional (4D) chaotic system and its pseudo random number generator (PRNG) application are presented. NIST 800-22 statistical tests are performed to prove the designed PRNG has sufficient randomness,

In the second section, the novel 4D chaotic system is presented. In the third section, the PRNG application of the 4D chaotic system and its NIST 800-22 statistical tests results is presented. In the last section, Conclusion is given.

II. PROPOSED 4D CHAOTIC SYSTEM

In this section, the novel 4D chaotic system whose mathematical model is given in Equation 1 is presented. When Equation 1 is examined the system has five parameters including one constant parameter and two nonlinear terms. Accordingly, the system model is not very complicated however, it can offer sufficiently large key space for encryption applications due to it has five system parameters. In other words, the system can provide large key space with low computational power.

$$\begin{aligned} \dot{x} &= a(y-x)-z \\ \dot{y} &= b(x-y)-w^2 \\ \dot{z} &= c(z-w)-xz \\ \dot{w} &= d(x+y)+z+e \end{aligned} \tag{1}$$

When the initial condition of the system are set as x_0=1, y_0=z_0=w_0=0 and the parameter values are set as a=2.75, b=1.25, c=0.75, d=0.95, e=0.025, the time series of the system is obtained as in Figure 1 and the phase portraits of the system is obtained as in Figure 2-7. As it can be seen in Figure 1, the time series of the system are varying between certain amplitude range in an aperiodic and random form. Thus, it can be said that the system exhibits chaotic behavior.

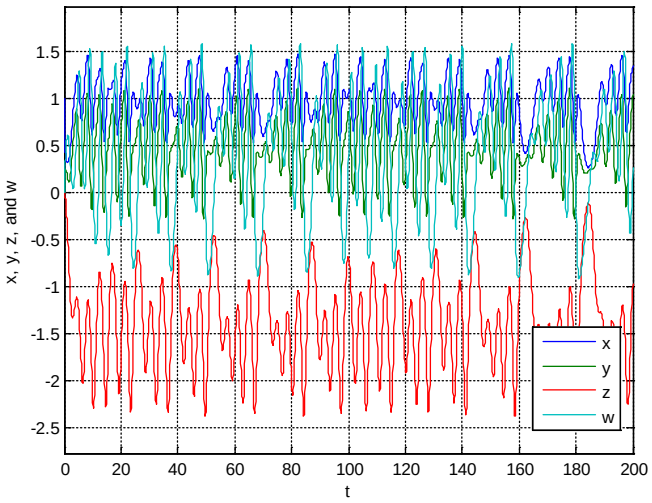


Figure 1: Time series of state variables x, y, z and w.

The phase portraits of the system state variables are given between Figure 2 and 7. When the phase portrait are examined, agian it can be said that the system exhibits chaotic behavior

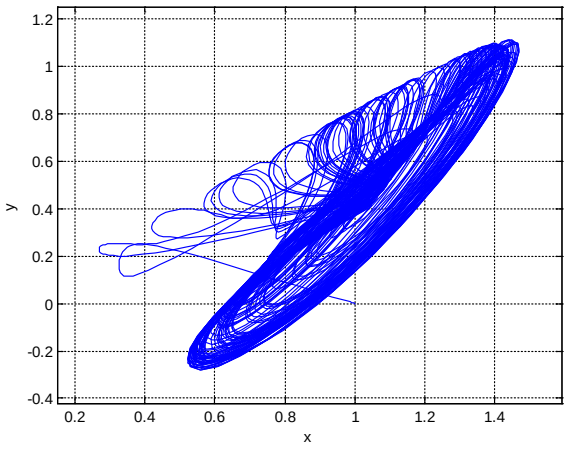


Figure 2: Phase portrait of x-y.

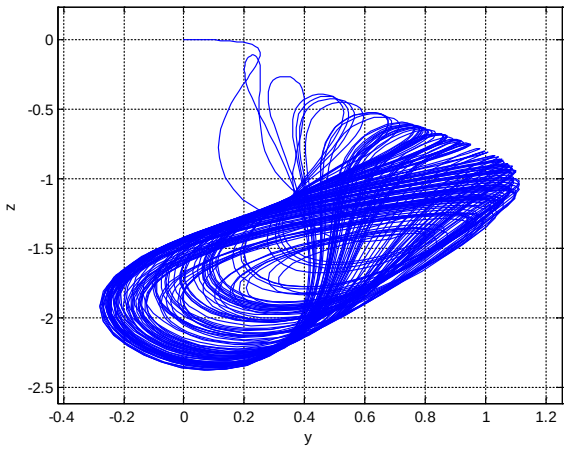


Figure 5: Phase portrait of y-z.

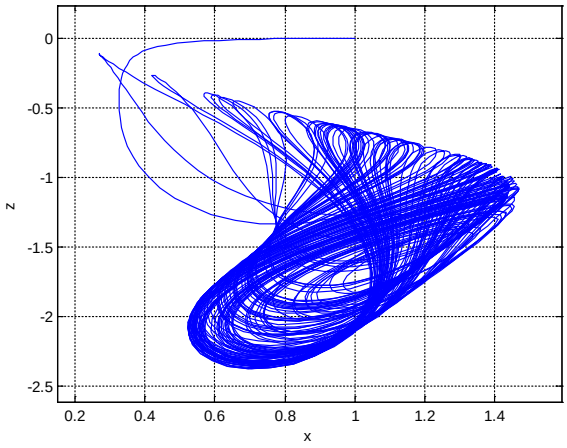


Figure 3: Phase portrait of x-z.

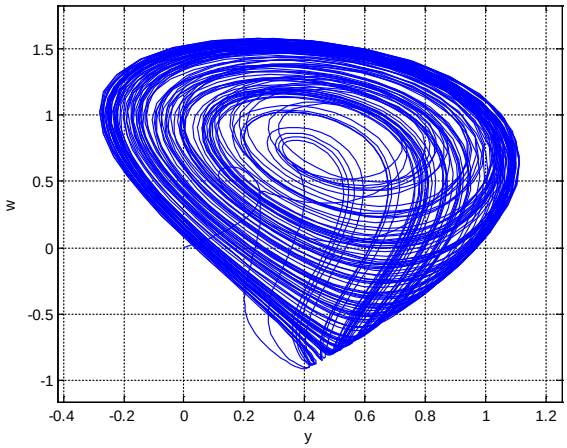


Figure 6: Phase portrait of y-w.

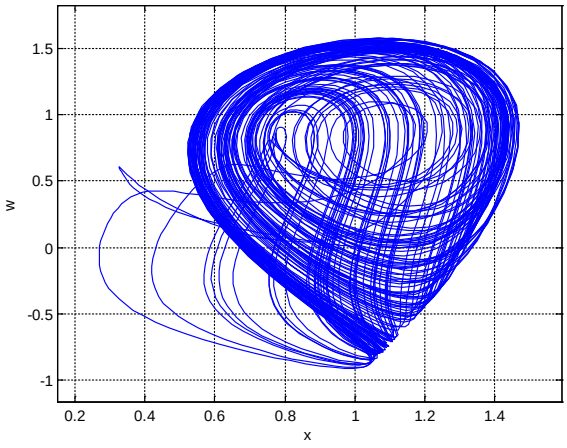


Figure 4: Phase portrait of x-w.

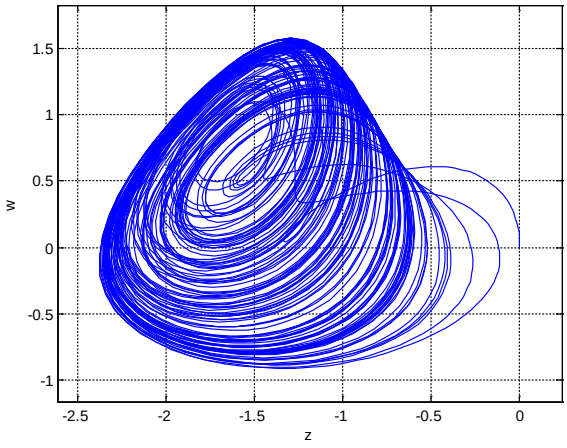


Figure 7: Phase portrait of z-w.

III. PSEUDO RANDOM NUMBER GENERATOR

Random number generators are software (pseudo) or hardware (true) applications that generates random number with certain procedure. In the study, a PRNG application based on the 4D chaotic system given in Equation 1 is realized.

The design steps of the PRNG applications are listed as:

1- Determination of appropriate value for the system parameters and initial conditions for the system to exhibit chaotic behavior and entering the iteration step interval and number of taken bits for the computation.

2- Calculation of the chaotic system state variables using Runge-Kutta 4 algorithm.

3- Conversion of the state variables from float to binary in every iteration.

4- Generation of 1000000 bit series for each state variable by selecting certain number of bits from binary form of state variable in every iteration.

5- Generation of different bit series by XORing the bit series generated in step 4.

6- The evaluation of success of the generated series a with NIST statistical tests.

7- The PRNGs which successfully pass the NIST tests can be used in different applications.

The NIST 800-22 tests which are the international standard for evaluation of random numbers contains 16 different statistical tests [6]. For a PRNG to be considered to have sufficient randomness, it must pass all the 16 statistical tests successfully.

In the application the system parameter values are set as a=2.75, b=1.25, c=0.75, d=0.95, and e=0.025; and the initial values are set x_0=1, y_0=z_0=w_0=0. The iteration step interval is set as 0.025 in the calculation and the maximum number of bits selected from the state variable in each iteration is set as 10. Out of the 1000000 bit series generated under these conditions, the bit series generated from x and z state variable and the bit series generated from XORing of x and y pass the NIST tests successfully and their results are given in Table 1. The results given in Table 1 show that the numbers generated from the realized PRNG have sufficient randomness and the PRNG can be used in different applications. Moreover, the results prove that the chaotic system have sufficient randomness and it is suitable for different applications like encryption.

IV. CONCLUSION

In the study a novel 4D chaotic system is presented and its chaotic behavior is examined. It is shown that the system exhibits chaotic behavior for appropriate parameter and initial values. A PRNG application based on the novel chaotic system is realized to show the system can be used in different applications. Then, to prove the realized PRNG has sufficient randomness the NIST 800-22 statistical tests are conducted.
The tests results prove that random number are generated with three different ways with the proposed chaotic system. To sum up, it can be said that the proposed chaotic system can be used in different areas such as encryption and random number generation. In the future studies, encryption application of multimedia data like audio, video and image based on the proposed chaotic system and PRNG will be realized.

Table 1: NIST-800-22 tests of proposed system based PRNG

Statistical Tests	**P-value (X)**	**P-value (Z)**	**P-value (X⊕Z)**	**Result**
Frequency (Monobit) Test	0,80568223	0,91399569	0,37132189	Successful
Block-Frequency Test	0,92887992	0,97506242	0,24576930	Successful
Cumulative-Sums Test	0,84533821	0,92177712	0,32537270	Successful
Runs Test	0,91087138	0,13002287	0,75260903	Successful
Longest-Run Test	0,94136351	0,01694575	0,75653314	Successful
Binary Matrix Rank Test	0,61315193	0,82581630	0,41220447	Successful
Discrete Fourier Transform Test	0,13470823	0,40367645	0,63323058	Successful
Non-Overlapping Templates Test	0,01056791	0,01190767	0,09676641	Successful
Overlapping Templates Test	0,46344395	0,39153148	0,83965145	Successful
Maurer's Universal Statistical Test	0,41722306	0,90132604	0,67250829	Successful
Approximate Entropy Test	0,37587501	0,33831343	0,95102247	Successful
Random-Excursions Test (x = -4)	0,12233302	0,09421901	0,95271643	Successful
Random-Excursions Variant Test (x = -9)	0,745967435	0,73060771	0,59545416	Successful
Serial Test-1	0,01238893	0,37408265	0,80010688	Successful
Serial Test-2	0,08082051	0,95847372	0,85985942	Successful
Linear-Complexity Test	0,30057023	0,88544250	0,40441237	Successful

REFERENCES

[1] S. Vaidyanathan, A. Akgul, S. Kaçar, and U. Çavuşoğlu. "A new 4-D chaotic hyperjerk system, its synchronization, circuit design and applications in RNG, image encryption and chaos-based steganography." *The European Physical Journal Plus*, vol.133, no. 2 pp. 46, 2018.

[2] A. Akgul, S. Kacar, and B. Aricioglu, "A new two-level data hiding algorithm for high security based on a nonlinear system." *Nonlinear Dynamics*, vol .90(2), pp. 1123-1140, 2017.

[3] X. Wang, A. Akgul, S. Kacar, and V. Pham. "Multimedia Security Application of a Ten-Term Chaotic System without Equilibrium." *Complexity* vol.2017, 2017.

[4] Ü. Çavuşoğlu, A. Akgül, S. Kaçar, İ. Pehlivan, and A. Zengin. "A novel chaos-based encryption algorithm over TCP data packet for secure communication." *Security and Communication Networks*, vol.9 no. 11 pp. 1285-1296, 2016.

[5] A. Akgul, I. Moroz, I. Pehlivan, and S. Vaidyanathan. "A new four-scroll chaotic attractor and its engineering applications." *Optik-International Journal for Light and Electron Optics* vol. 127, no. 13 pp. 5491-5499, 2016.

[6] A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, "A statistical test suite for random and pseudorandom number generators for cryptographic applications." Booz-Allen and Hamilton Inc. Mclean Va 2001

An Automated Vulnerable Website Penetration

A. MURZAEVA[1], S. AKLEYLEK[2]

[1] Ondokuz Mayis University, Samsun/Turkey, azhar.murzaeva@bil.omu.edu.tr
[2] Ondokuz Mayis University, Samsun/Turkey, sedat.akleylek@bil.omu.edu.tr

***Abstract* – SQL Injection vulnerability is one of the most important and prevalent vulnerabilities. It is important to make pen tests to develop secure applications. In this paper, we perform a penetration test and implement SQL Injection attack. The results are tested in the demonstration platform testphp.vulnweb.com. By this work, we aim to emphasize the importance of secure systems and make people aware of that. Otherwise, existence of such vulnerabilities in a system can bring to bad results and bad situations. Then, we make an automated tool for SQL Injection penetration test with the aim to finalize the test quickly and to provide a convenience for pen testers.**

***Keywords* – website vulnerabilities, automated penetration testing tool, w3af_console, sqlmap, SQL Injection.**

I. INTRODUCTION

NOWADAYS, the availability of data sharing is one of the important things. And Internet is the most popular tool used to make the data sharing easier. Nearly all information systems are built as web-based database applications (e-government, e-tickets for transportations, remote education, e-books etc.). So, with the spread of Internet's usage the amount of users is increasing, hence the amount of servers is increasing too and new technologies such as the Cloud Systems are being developed. Thus, the amount of data transmission increases and the security problems are emerging with it. And for safety of system and data on it, those security problems must be avoided. To avoid and solve these security problems in a system, we need in testing of that system.

Penetration test, also known as a pen test, is an attack on a system that is performed intentionally by pen-testers. Various tools for penetration provide an access to the destination system. The purpose of pen test is to check the security level of system and report it to the authorized user. Today, penetration testing is the most frequently and commonly applied of all security best practices [1].

It this paper, we tried to demonstrate the SQL Injection (sqli) attack. SQL Injection (sqli) is one of the most widely spread and dangerous vulnerabilities of web applications. It occurs because of the mistakes made in SQL query statements, which are used in web applications that make use of SQL-based databases.

II. PENETRATION TEST

A. Test Platform

Kali Linux: It is a Debian-based Linux distribution which was developed by Offensive Security. Its purpose is to be used in various information security tasks [2].

Web Application Attack and Audit Framework (w3af): It is an open source web application security scanner. It helps to identify and exploit vulnerabilities in web applications [3].

Sqlmap: It is an open source penetration tool that is used to detect and exploit SQL injection flaws [4]. It automatically comes with a Kali Linux.

As a target website we used **testphp.vulnweb.com** which is used for demonstration of a vulnerable website.

B. Web site analyzing with Web Application Attack and Audit Framework (w3af)

w3af is one of the popular tools used in penetration tests. It checks a target website for various vulnerabilities and can detect the existence of malicious code in that website. Under the "audit" selection in w3af tool we select vulnerabilities we want to check for. For example, we check sqli checkbox and then started audit. In Figure 1, w3af GUI tool is demonstrated.

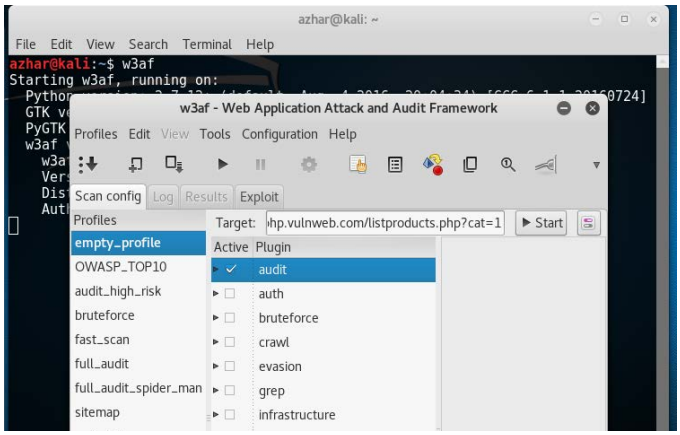


Figure 1: w3af GUI tool.

After clicking on "Start" button, w3af tool begins to give you information about founded errors. In Figure 2, a report about existence of SQL Injection vulnerability in a target system is given.

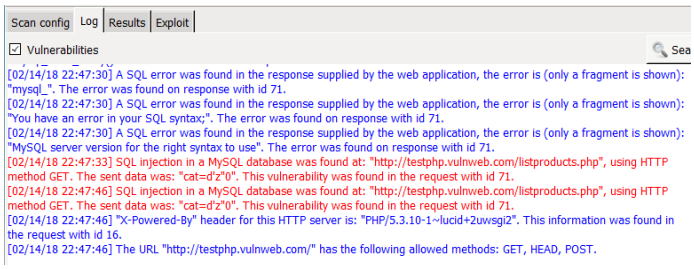


Figure 2: "Log" tab.

Under the "Log" tab we saw reports of all requests sent to the target website by the tool. Then on the "Results" tab vulnerabilities of the given system in appropriate order is given. The results are provided in Figure 3.

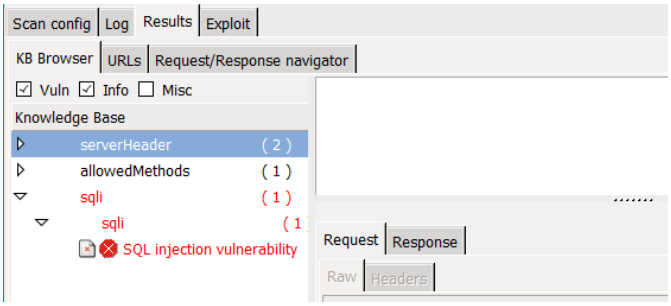


Figure 3: "Results" tab.

C. *Exploiting SQL Injection vulnerability using SQLMap tool*

SQLMap is an advanced tool that is used for detecting and exploiting SQL injection flaws. It is developed using python. Then it can dump databases which is a dangerous situation. This tool comes installed on Kali Linux.

Firstly, we must test a connection to a target website. For that we must type on terminal this command:

sqlmap –u "yourtarget.website.com"

Here, -u parameter defines your destination site.

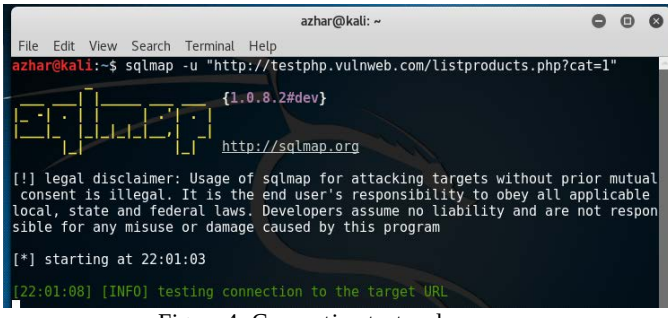


Figure 4: Connection test, sqlmap.

After running a command shown on Figure 4, we get some reports on the screen that is shown on Figure 5. There we can see that system is injectable. After getting this report we can start our attack.

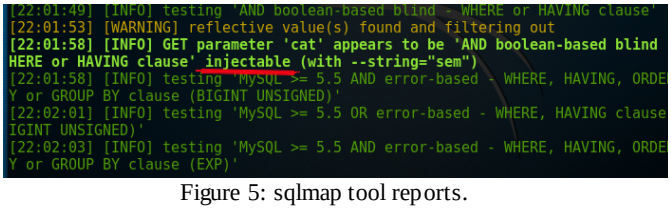


Figure 5: sqlmap tool reports.

Firstly we try to learn about databases on the target system by running this command:

sqlmap –u "yourtarget.website.com" –dbs

Here, --dbs parameter means –databases.

As a result we see a database named acuart. See Figure 6.

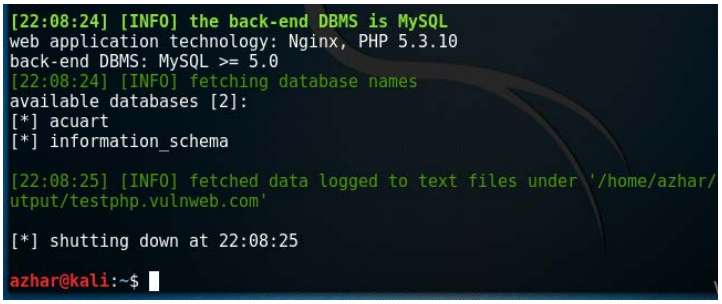


Figure 6: Existing databases, sqlmap.

In the same way we get tables from this database and results are shown on Figure 7. From these tables we choose the table that contains more important data then others do. In most attacks, as in this pen test, the most important table actually stores information about users. So, from the list of tables on the Figure 7, we choose "users" table. (It could be named like "directors", "admins", "students", "teachers" etc. depending on a work place)

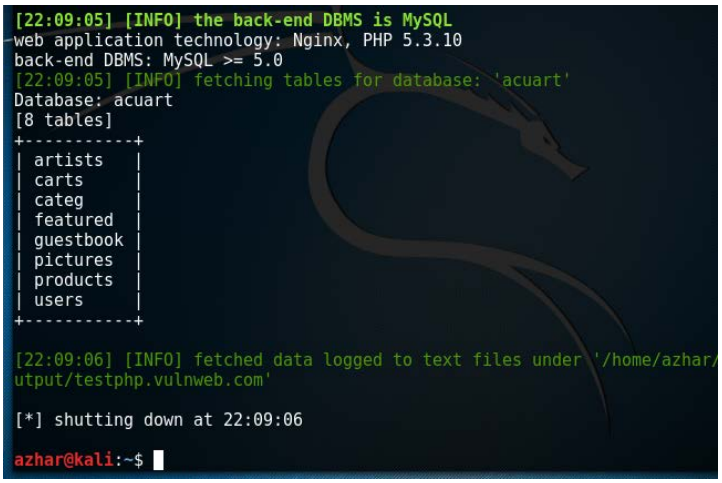


Figure 7: List of tables, sqlmap.

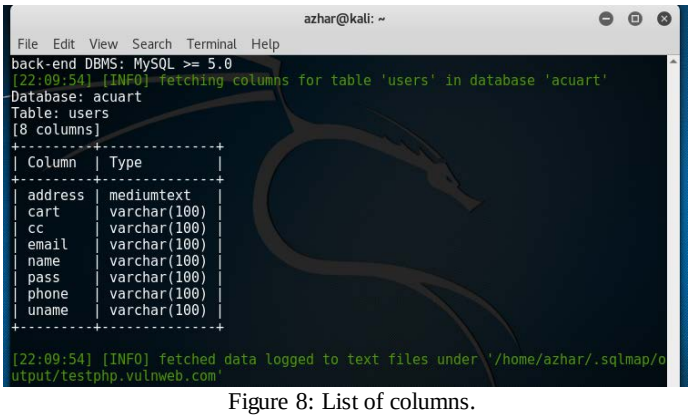


Figure 8: List of columns.

After defining a table we run a command to list columns of that table (To see columns' list, look at Figure 8). Then, we dump the data from columns that are potential for having an important data (Figure 9).

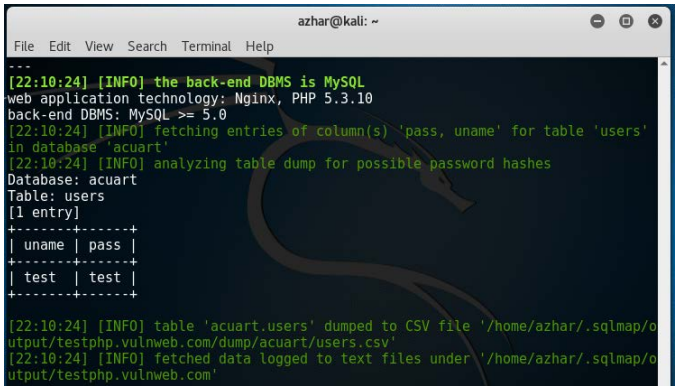


Figure 9: Dumped data, sqlmap.

As we see, we got the usernames and passwords that exist in that database of a vulnerable website. There is just one user with a "test" username and with a "test" password. Using that information, we can log on in the website.

In Figure 10, we enter the **uname** and **pass** data that we have accessed by exploiting SQL Injection vulnerability.

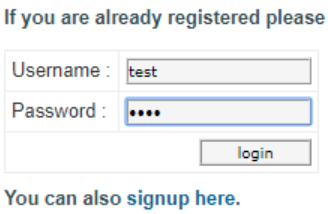


Figure 10: Entering data.

In figure 11, we see that we logged on the website using obtained users' data. After logging in as an authorize user, all information about user (such as credit card number, address, photo, etc.) appears and an attacker can use it as he wishes.

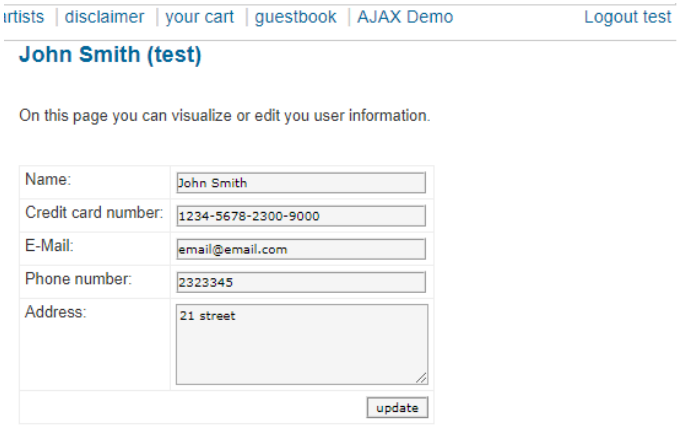


Figure 11: Access to personal information.

III. Related Works

There are many works made on penetration test and many articles are written. For example, a penetration test was performed on web pages that are introduced to academic personal of a university [5]. That pen test was implemented using tools on a BackTrack operating system. There are also other pen testing examples such as smartphone penetration test, traffic sniffing, hacking WPA Protected Wifi, Man-in-the-Middle attack, hacking remote PC, etc. For example, M. Denis and others explained the manual penetration test towards smartphone, Bluetooth etc. [6]. With the emerging of the Internet of Things (IoT), IoT devices have risks in insecure data transmission. Hence, to provide a secure communication penetration test is required for them, too. For example, the Penetration Testing Tool For Internet of Things was developed [7]. The main purpose of all these works is to demonstrate the penetration test and nearly all these works made manually.

IV. Automatization

In this work we tried to automatize the penetration testing process that we performed in section II. PENETRATION TEST. For that, we prepared an automated pen test tool and command line tool named "SQLi PentestEd", which is programmed using Ruby programming language.

Firstly, it checks a target website for SQL Injection vulnerability by running w3af tool mentioned above. In our tool we used a console version of w3af tool instead of w3af gui explained in previous section.

Then, if the existence of SQL Injection is detected in a target website, tool starts SQL Injection by running sqlmap tool in background.

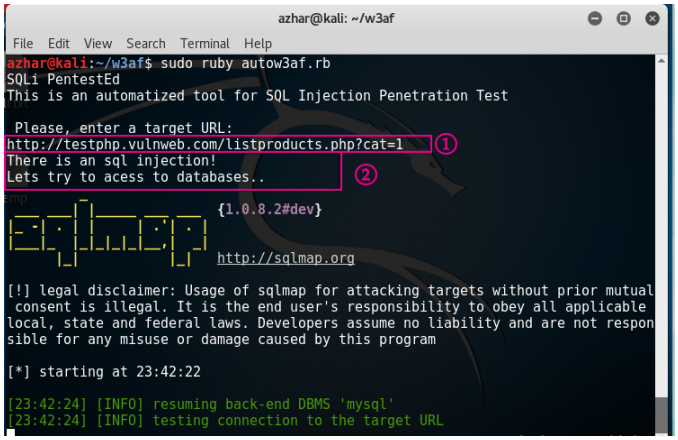


Figure 12: "SQLi PentestEd" tool.

Figure 12 demonstrates an interface and usage of our automatized tool. The mark 1 on the figure indicates on a target website entrance field, and the mark 2 gives us results of the testing for SQL Injection vulnerability. If the system is vulnerable, tool continues for pen testing, otherwise it congratulates for secure system.

In our situation, target website is vulnerable and injectable. So, let us take a look on Figure 13.

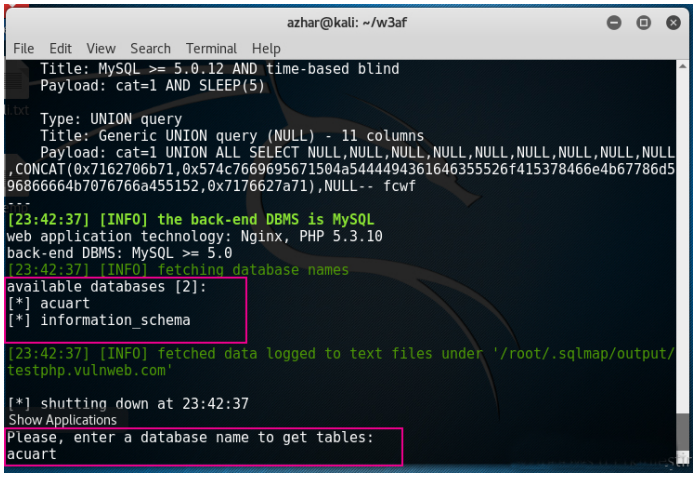


Figure 13: List of databases, automated tool.

Here, we see that automatized tool accessed to the databases of a vulnerable website and asks us to enter the database name from the list it gave us. After entering a database name, it will access to the tables of that database (see Figure 14).

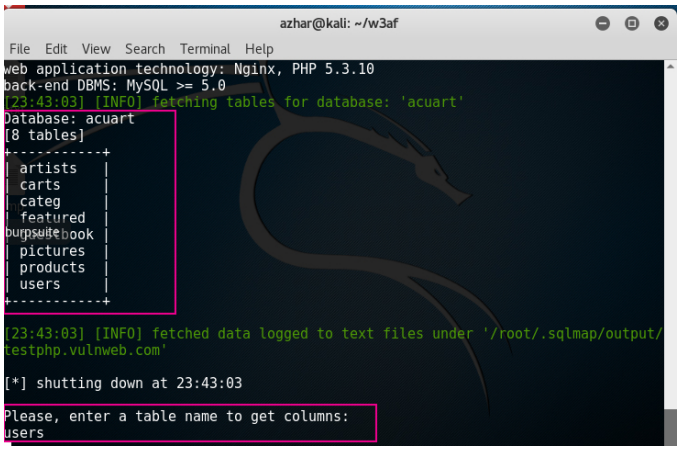


Figure 14: List of tables, automated tool.

After listing tables of the database you have chosen, it will ask you to choose a table name to access to its columns. Then, you are going to choose potential columns.

In all these required parameters you must enter the most potential ones for accessing to an important data. For example, on Figure 15, the potential columns are **uname** and **pass**, because using them we can log on the website and take an authorized access.

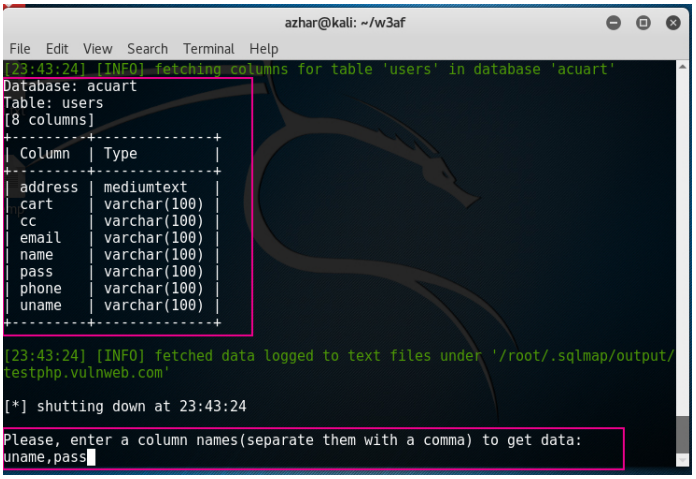


Figure 15: Accessed columns' list.

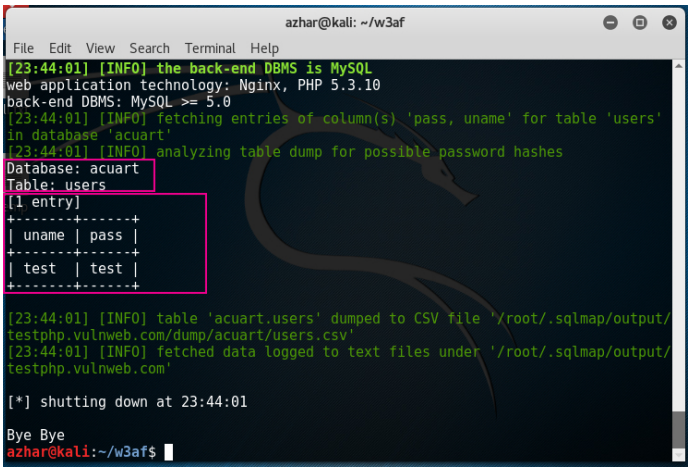


Figure 16: Dumped data.

As we see, we accessed to the usernames and passwords that exist in that website's database. There is just one user with a "test" username and with a "test" password. Using that information, we can log on in the website as demonstrated in Section II.

V. RESULTS

In this work, penetration testing was performed on a test and demonstration website. Here, we obtained the personal data of the user such as username and password that allowed us to log in to the website and access to other information such as credit card number, address etc. If the user was an admin of the system, it would allow attacker to do more things like adding users, changing data (for ex., credit card number) that would create important problems in real life. As we see, with the development of the Internet technology, as well as positive results negative results will be produced, too.

We need to make pen tests like this for producing and developing secure applications. And performing such penetration testing like this takes a time. Hence, we made an automated tool for SQL Injection penetration test with the aim to win the time and to provide a convenience for pen testers.

Also, this work was made with the purpose to make people

aware of such vulnerabilities to take precautions, to prevent such bad results and to prevent bad unexpected situations.

ACKNOWLEDGMENT

This work is partially supported by OMÜ under grant no.PYO.MUH.1906.17.003. The first author would like to thank to Taha Yasir Kiroglu for his support during this work.

REFERENCES

[1] B. Arkın, S. Stender, G. McGraw, *Software Penetration Testing*, IEEE SECURITY&PRIVACY, 84-87 pp.

[2] Kali Linux, "Kali Linux Documentation", "What is Kali Linux?" [Online]. Available: https://docs.kali.org/introduction/what-is-kali-linux

[3] w3af tool, [Online]. Available: https://github.com/andresriancho/w3af

[4] sqlmap tool, [Online]. Available: http://sqlmap.org/

[5] Y. Yilmaz, "Fırat Üniversitesi Personel Bilgi Sistemi Penetrasyon(Sızma) Testi", 1st International Symposium on Digital Forensics and Security (ISDFS'13), 20-21 May 2013, Elazığ, Turkey

[6] M. Denis, C. Zena and T. Hayajneh, "Penetration testing: Concepts, attack methods, and defense strategies," 2016 IEEE Long Island Systems, Applications and Technology Conference (LISAT), Farmingdale, NY, 2016, pp. 1-6. doi: 10.1109/LISAT.2016.7494156

[7] V. Visoottiviseth, P. Akarasiriwong, S. Chaiyasart, S. Chotivatunyu, "PENTOS: Penetration Testing Tool for Internet of Things", IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

RC4 Stream Cipher Based Digital Color Image Encryption Using Chaotic Systems

S.TUNÇER[1], C. KARAKUZU[1] and F. UÇAR[1]

[1]Bilecik Seyh Edebali University, Bilecik/Turkey, sefa.tuncer@bilecik.edu.tr
[1] Bilecik Seyh Edebali University, Bilecik /Turkey, cihan.karakuzu@bilecik.edu.tr
[1] Bilecik Seyh Edebali University, Bilecik /Turkey, firatucar94@gmail.com

Abstract **- RC4 is an algorithm that encrypts a data in the form of a bit string with a specified key. The security of RC4 with high encryption speed depends on the random key. Speed of image encryption is very important parameter due to size of the data. Performance analysis was examined using color image encryption, because RC4 is usually used in speed-critical applications. RC4 showed that the desired values could not be obtained when examining the histogram, correlation coefficient and information entropy analysis results. Because of this, chaotic systems are used to increase the performance criteria of image encryption with RC4. Chaotic systems are very sensitive due to their inherent dependence on the initial conditions and dynamic variables. Not random and non-periodic oscillations these systems are performed in a certain frequency range. The RC4 algorithm are enhanced by using chaotic system-based encryption algorithms because of its key size capability and speed. Performance criteria have been improved using 2D Cat Map, Tent Map and Lorenz chaotic systems. When we examine only the encryption made with RC4 and the encryption made with RC4 supported with chaotic systems; more successful results are obtained from histogram, correlation coefficient and knowledge entropy analysis. In addition, the structure of chaotic systems is increased key sensitivity and key size and thus a more secure algorithm is obtained.**

Keywords **– Image encryption, RC4 cipher, chaotic systems, cryptanalysis methods, data security.**

I. INTRODUCTION

RC4 (Rivest Cipher 4) is a key stream encryption. It is a symmetric encryption algorithm because it uses the same key for encryption and decryption. Information is XORed with the key, encrypted and sent to the receiver. The algorithm is serial, which is why the state inputs need to follow each other depending on the key sequence. Very intensive calculations may be required. The RC4 algorithm usually uses a variable-length key. RC4 encryption is used in Wireless Encryption Protocol.

Kumar used the RC4 Enrichment Algorithm Approach in image encryption. This approach is derived from the standard RC4 algorithm. With cryptanalysis methods, many weaknesses, vulnerable points and attacks have been detected in the RC4 algorithm. Kumar reported that there were deficiencies in the security of RC4. Therefore, RC4 based new enrichment approach is designed to strengthen the RC4 algorithm. This approach was based on the new KSA and PRGA algorithm process, which are two phases of the RC4 algorithm [1]. Ginting et. al. proposed an algorithm for image encryption based on RC4 stream encryption algorithm and chaotic logistics map in [2]. They generated a key flow based on the chaotic logistics map. Processing a permutation and the result is then XOR-ed with bytes stream of digital image. As a result, it was seen that the visual structure of the encrypted image was unclear, there is no correlation between the plain image and encrypted image, and there was no loss in decryption. The same result was achieved by checking hash value (MD5) of original image and the decrypted image [2].

The proposed algorithm by Mondal in [3] destroyed, correlation among pixels of the plain image by the shuffling of pixels in the permutation process, which is then subjected to substitution using RC4 key stream generator. Ibrahim has proposed a chaos-based (Henon Map) encryption algorithm to encrypt digital images. It was seen that the performance analysis of the encrypted color image was very successful [4].

Prusty et al. [5] mixed the image using Arnold Cat Map chaotic system. They generated key with random numbers using Henon Map and then encrypted images. They have achieved successful results by encrypting images in different formats. Li et al. [6] implemented image cryptography using Tent Map and Lorenz chaotic systems. It has successfully passed key widths greater than 256 bits, data randomness, histogram analysis and correlation tests. For this reason, it can be said that this approach is a successful encryption. Liu and Wang [7] performed encryption and decryption by combining red, green and blue (R, G, B) values in three different images having the same size. They used the SHA-256 hash function for key in encryption and Lorenz chaotic system to generate random numbers. All of these show that chaotic systems are quite effective in image encryption.

In this study, RC4 stream cipher was modified with chaotic systems to obtain a more secure digital color image encryption algorithm. Performance analysis of the proposed algorithm has been performed and it has been tested for image encryption.

II. RC4 CIPHER

The RC4 algorithm has a simple and understandable encryption structure. A variable-length key of from 1 to 256 bytes (8 to 2048 bits) is used to initialize a 256-byte state vector S, with elements S[0], S[1], …, S[255]. S contains a permutation of all 8-bit numbers from 0 through 255. For encryption and decryption, a byte k is generated from S by

selecting one of the 255 entries in a systematic fashion. As each value of k is generated, the entries in S are once again permuted.

The next part of RC4 is the key-scheduling algorithm (KSA), listed below:

```
for i=0:255
   S[i] = i
end
j = 0
for i=0:255
   j = (j + S[i] + key[i mod keylength]) mod(256)
   swap(S[i],S[j])  // replace S[i] with S[j]
end
```

KSA creates an S-sequence with 256 entries with numbers between 0 and 255, like the following table:

0	1	…	…	n	n+1	…	…	254	255

Each of the 256 inputs in S is then replaced by the j-th input calculated in (1) for S.

$$j = (j + S(i) + key(i \ mod \ keylength])mod(256) \quad (1)$$

where j is the previous j value (which is initially zero). *S[i]* is the value of the current entry in *S*. *key[i mod keylength]* is either a zero or a one.

The next part of RC4 is the pseudo-random generation algorithm (PRGA). The PRGA is below:

```
i = 0
j = 0
while i < MAX_SIZE
   i = (i + 1) mod 256
   j = (j + S[i]) mod 256
   swap(S[i], S[j])
   output S[(S[i] + S[j]) mod 256]
end
```

RC4 has the following features:

- Variable key length.
- Security depends on the use of random permutations.
- There is no known bad key.
- The encryption speed is approximately Megabytes / sec.

III. CHAOTIC SYSTEMS

Chaos can be defined as a regular state of irregularity that is overly sensitive to initial conditions and has a power spectrum such as noise.

Chaotic systems exhibit an unlimited number of nonlinear behaviors with varying periodic oscillations, whose amplitudes and frequencies are undetectable. However, these dynamics take place in a limited area with chaotic signals. Depending on these signals, the current state of the dynamic system, the past state, and the set of possible states are also known [8]. The characteristics of chaotic systems are given below:

- Moves irregularly in time dimension.
- Sensitive to initial conditions.
- Includes an unlimited number of periodic oscillations. It is not random.
- Amplitude and frequency are uncertain.
- Changing signals in limited area.

Tent Map, Cat Map and Lorenz are chaotic systems used in this study.

A. Cat Map

2B Cat Map provides the ability to change the location of the pixels in the image by producing different coordinate values continuously in a given area.

Cat Map is the most widely used map in the literature. If a $N \times N$ gray level image and the pixel values of this image are determined as coordinates $C = \{(x, y) \mid x, y = 1, 2, \dots, N\}$, then the Cat Map is defined as follows [9]:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = Q \begin{pmatrix} x \\ y \end{pmatrix} mod(N) = \begin{bmatrix} 1 & p \\ q & pq+1 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} mod(N) \quad (2)$$

where q and p are positive integers. (x, y) and (x', y') are the original and new positions, respectively, of the coordinate values. Cat Map pseudo-random generation algorithm is below:

```
N = dimension // in x and y plane
for i=1:N
   for j=1:N
      newi = mod(((i-1)  + p*(j-1)),N)+1
      newj = mod((q*(i-1)  + (p*q+1)*(j-1)),N)+1
      CatMap_Image(newi,newj,:)  = Original_Image(i,j,:)
   end
end
```

B. Tent Map

Tent Map chaotic system allows you to generate non-correlated pixel values when applied to any image. This reduces the relationship between the input image and the output image. The one-dimensional discrete-time chaotic Tent Map system is expressed by (3).

$$f(a, x) = \begin{cases} \left\lceil \frac{M}{a} x \right\rceil, \ 0 \le x \le a \\ \left\lfloor \frac{M}{M-a} (M - x) \right\rfloor + 1, \ a < x \le M \end{cases} \quad (3)$$

where a is a user-specified integer ($a \in [1, M]$) and $\lfloor x \rfloor$ and $\lceil x \rceil$ represent the upper and lower bounds of x, respectively. The M value is usually chosen according to the plain text and $M = 256$ for an 8 bit image. Tent Map pseudo-random generation algorithm is below:

```
S = 256 // constant value
for i = 1:S
   if image(j,i)<=a
      Tent (j,i)  = ceil((S/a)  * image (j,i))
   elseif x(j,i)>a
      Tent (j,i)  = floor((S/(S-a))  * (S - image (j,i)))+1
   end
end
```

where ceil rounds the result to upper integer and floor rounds it to lower integer.

C. Lorenz

Lorenz is a 3-dimensional (3D) chaotic system. a, b and c are considered as constant parameters of the system, the dynamics of the chaotic Lorenz systemare expressed by (4).

$$Lorenz = \begin{cases} x'(t) = a(y(t) - x(t)) \\ y'(t) = -x(t)z(t) + cy(t) \\ z'(t) = x(t)y(t) - bz(t) \end{cases} \quad (4)$$

In this study, Lorenz is used to increase the randomness of the results of the Tent Map. We prefer Lorenz because the 3D image is encrypted.

IV. RC4 Cipher with Chaotic Systems (CRC4)

The RC4 algorithm is a fast algorithm by itself, but it is not completely secure. Chaotic systems can be used to strengthen this algorithm to use in image encryption. Image permutation is done with Cat Map. RC4 is applied and then random values are obtained with Tent Map. RC4 cipher with chaotic systems (CRC4) pseudo-randomgeneration algorithm is below:

CRC4 pseudo-random generation algorithm:
Read(Image)
CatMapImage = **CatMap**(Image)
run RC4 key-scheduling
CipherImage = **RC4_Encryption**(CatMapImage)
Generate random numbers with **Lorenz**
Result = **TentMap**(CipherImage, Lorenz)

The parts in the above algorithm are explained in the previous sections.

V. Performance Analysis

Baboon and Peppers original images (a, b) and encrypted images (c, d) results with CRC4 are shown in Figure 1, respectively. It is not possible visually to make any detection between the encrypted image and the original image.

A. Key Space

RC4 key can be in a different length. However, 128 bit key is used in this study. The key length with chaotic systems reaches up to 168 bit for CRC4. It is long enough for data security.

B. Key Sensitivity

Key sensitivity generally is specific to the chaotic systems. Chaotic systems are sensitive to initial conditions, so a small change in the input of chaotic systems makes the encrypted data completely different. If the input of the Lorenz chaotic system are changed at 1×10^{-14}, the encrypted image will be completely different.

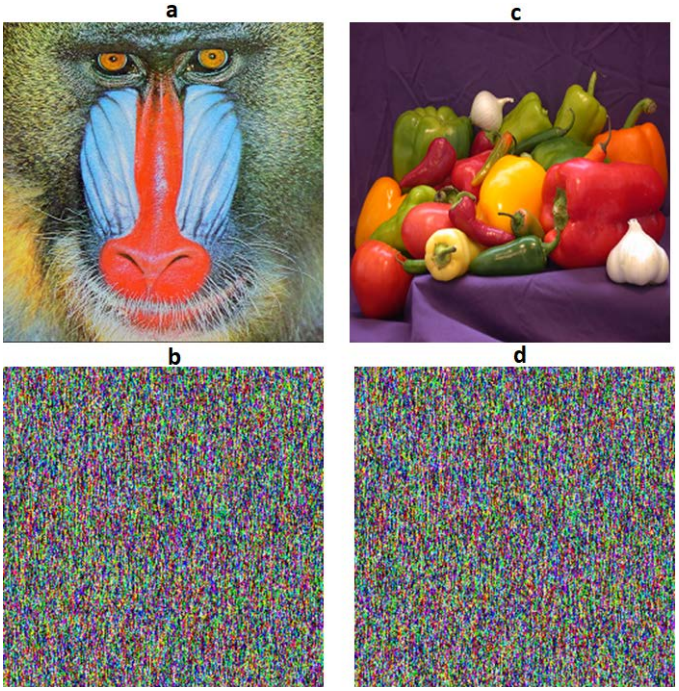


Figure 1: Baboon (a, b) and peppers (c, d) images encryption visuals.

C. Entropy Analysis

Entropy represents the degree of uncertainty in a system. The information entropy $E(m)$ of an m message is calculated as in (5).

$$E(m) = -\sum_{j=0}^{2^n-1} p(m_j) log_2 \frac{1}{p(m_j)} \quad (5)$$

$p(m_j)$ denotes the probability of occurrence of m_j. In a random process each symbol has equal probability. $p(m_j) = 2^{-8}$, when $n = 8$ is taken. Thus distribution of entropy is $E(m) = 8$.

Entropies of red, green, blue (RGB) values are examined separately for baboon and peppers images. Entropy of CRC4 appears to be more than RC4 in digital image encryption in Table 1.

Table 1: RGB entropies of RC4 and CRC4.

Entropy values of images	RC4		CRC4	
	Baboon	Peppers	Baboon	Peppers
Red	7.40	7.39	7.90	7.90
Green	7.39	7.39	7.89	7.90
Blue	7.39	7.40	7.89	7.89

D. Correlation Coefficient Analysis

Correlation explores the relationship between two or more variables in terms of increase and decrease. The correlation between adjacent pixels in the original image is quite high. An effective encryption algorithm should reduce the relationship between adjacent pixels [10]. Correlation coefficient (CC) k_{xy} calculated as in (6).

$$CC = \begin{cases} e(x) = \frac{1}{N_l}\sum_{i=1}^{N_l} x_i \\ d(x) = \frac{1}{N_l}\sum_{i=1}^{N_l}(x_i - e(x))^2 \\ cov(x,y) = \frac{1}{N_l}\sum_{i=1}^{N_l}(x_i - e(x))(y_i - e(y)) \\ k_{xy} = \frac{cov(x,y)}{\sqrt{d(x)}\sqrt{d(y)}} \end{cases} \quad (6)$$

where x and y represent the values of two adjacent pixels in the image.

Table 2 shows the correlations of the RGB values of Baboon and Peppers images for RC4 and CRC4. Based on these results, it can be said that both algorithms are successful.

Table 2: Correlation coefficient results.

Correlation values of images	RC4		CRC4	
	Baboon	Peppers	Baboon	Peppers
Vertical correlation	-0.0057	-0.0064	0.339	0.3632
Horizontal correlation	-0.0208	-0.0009	-0.207	0.0035
Cross correlation	-0.0252	0.0090	-0.010	0.0095

E. PSNR

The peak signal to noise ratio (PSNR) is the difference between the pixel values of the image before and after encoding. A high PSNR value means that the difference in the quality of the image is so small. The PSNR value is calculated using (8) by the mean square error (MSE) in (7).

$$MSE = \frac{\sum_{m,n}[I_1(m,n) - I_2(m,n)]^2}{m \times n} \quad (7)$$

$$PSNR = 10log_{10}(\frac{R^2}{MSE}) \quad (8)$$

According to Table 3, PSNR values are higher in image encryption with CRC4. Therefore, CRC4 can be considered to more successful than that of RC4.

Table 3: PSNR value of images.

PSNR of images	RC4		CRC4	
	Baboon	Peppers	Baboon	Peppers
Red	68.76	69.71	72.05	73.21
Green	68.89	74.02	72.60	78.57
Blue	70.04	74.79	73.65	79.47

F. NPCR and UACI

To test the effect of a single pixel change in the original image, the number of changing pixel rate (NPCR) is calculated by Eq. (9) and the pixel change rate is measured. Then the unified averaged changed intensity (UACI) is measured by Eq. (10). NPCR measures the percentage of different pixel numbers between two images. The UACI measures the average intensity of differences between two images [10]. W and H are width and height values of the scrambled image. Both RC4 and CRC4 have achieved the desired results compared with the original and encrypted images.

$$NPCR = \frac{\sum_{i,j} D(i,j)}{W \times H} \times 100\% \quad (9)$$

$$UACI = \frac{1}{W \times H}\left[\sum_{i,j}\frac{|C_1(i,j) - C_2(i,j)|}{255}\right] \times 100\% \quad (10)$$

Table 4 shows that both RC4 and CRC4 are successful at pixel change. NPCR and UACI values are in the expected range for them.

Table 4. NPCR and UACI values of images.

Test results of images	RC4		CRC4	
	Baboon	Peppers	Baboon	Peppers
NPCR	0.996	0.994	0.996	0.995
UACI	0.329	0.255	0.318	0.330

VI. CONCLUSION

RC4 is not enough to encode image alone. Everyone knows that the RC4 algorithm has incomplete aspects, but it is used by modified because it is so fast. The proposed CRC4 algorithm is strengthened by chaotic systems. The randomness of chaotic systems has reduced the relationship between the original image and the encrypted image pixel values. For security the key size is increased, the image is permutated with Cat Map and encrypted more effectively with RC4 and Tent Map. In addition, PSNR, NPCR and UACI results are sufficient for performance. Based on all performance data, as a result, it can be said that CRC4 is a secure image encryption algorithm.

REFERENCES

[1] P. Kumar and P. K. Pateriya, "RC4 encrichment algorithm approach for selective image encryption," *International Journal of Computer Science & Communication Networks*, vol. 2, pp. 181–189, 2013.

[2] R. U. Ginting and R.Y. Dillak, "Digital color image encryption using RC4 stream cipher and chaotic logistic map," *Information Technology and Electrical Engineering (ICITEE)*, pp. 101–105, 2013.

[3] B. Mondal, N. Sinha and T. Mandal, "Digital color image encryption using RC4 stream cipher and chaotic logistic map," *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics*, pp. 227–237, 2016.

[4] A. S. Ibrahim and M. I. Sameer, "A new approach of color image encryption based on RC4 algorithm and chaotic map," *International Journal of Computer Applications Technology and Research*, vol. 6, pp. 422–429, 2017.

[5] A. K. Prusty, A. Pattanaik and A. Mishra, "An image encryption & decryption approach based on pixel shuffling using Arnold Cat Map & Henon Map," *International Conference on Advanced Computing and Communication*, pp. 1–6, 2013.

[6] J. Li, Y. Xing, C. Qu and J. Zhang, "An image encryption method based on tent and lorenz chaotic systems," *Software Engineering and Service Science (ICSESS*, pp. 582–586, 2015.

[7] H. Liu, X. Wang, "Triple-image encryption scheme based on one-time key stream generated by chaos and plain images," *The Journal of Systems and Software*, vol. 86, pp. 826–834, 2013.

[8] A. Akgul and I. Pehlivan, "A new three-dimensional chaotic system without equilibrium points, its dynamical analyses and electronic circuit application," *Technical Gazette*, vol. 23, pp. 209–214, 2016.

[9] D. Xiao, X. Liao and P. Wei, " Analysis and improvement of a chaos-based image encryption algorithm," *Chaos, Solitons and Fractals*, vol. 40, pp. 2191–99, 2009.

[10] S. Tunçer, C. Karakuzu, "Information Encryption and Hiding into an Image By Steganography Methods to Improve Data Security," *Journal of New Results in Science*, vol. 5, pp. 170–177, 2016.

Utilizing RFM Analysis and Apriori Methods on Turkey State Supply Office Data for Regaining Lost Customers

H.A. TURKOZ[1], I. BABAOGLU[2]

[1] Devlet Malzeme Ofisi, Dept. of Project and Business Development, Ankara, Turkey, ali.turkoz@dmo.gov.tr
[2]Selçuk University, Dept. of Computer Engineering, Konya, Turkey, ibabaoglu@selcuk.edu.tr

***Abstract* - The use, interest and analysis of e-sales systems are growing by the innovations the internet data processing technologies. Institutions, which want to provide better services to the customers, must first handle the demands, needs, and requirements of the customers. State Supply Office is an institution that meets to a large extent of purchasing needs of institutions in Turkey. In this study, a scorecard was created for each customer by segmenting the customers using recency, frequency and monetary (RFM) analysis method on the state supply office dataset. These scorecards were used to determine which the customers are loyal and which customers are being lost. Then, Apriori Algorithm was utilized to conduct a joint analysis on the products purchased by the lost customers for tendering campaigns and discounts. In order to regain the lost customers, the campaign and discount schemes were constructed considering scope of the right products obtained by the Apriori Algorithm.**

***Keywords* - Apriori, RFM, Data Mining, Scorecard, Association Rules.**

I. INTRODUCTION

OVER the last four decades, many organizations have generated a large amount of machine-readable data in the form of files and databases. Today, information retrieval from this data is main concern. Reaching the important and useful information provided by data mining concepts.

Data mining uses database technology to discover additional knowledge or patterns in the data. One of the major technologies in data mining involves the recommendation mechanisms for online sales sites like which products bought together.

The State Supply Office (DMO) is a public enterprise, which is an affiliate of the Ministry of Finance. The customers of DMO are other public enterprises and the main duty of DMO is to meet their needs.

The product groups of procured and delivered by DMO are as below; Computers, hardware & software, Furniture & furnishings, Office machinery, Stationery, Cleaning & Illumination supplies, Medical equipment & hospital furnishings, Office and industrial type white goods, School, dormitory and guesthouse supplies, Transportation vehicles, Construction equipment, Substructure vehicles, Motor vehicles.

DMO offers various procurement methods to its customers like stock, catalogue, protocol and miscellaneous purchase upon buyer demands. The protocol sales can be done with applying to DMO or DMO's e-commerce site (online sales site). Online sales site contains only catalogue and stock products so with this site, only catalogue and stock products can be bought by customers.

Customers can make orders and provide their needs from this online sales site. One of the main concerns of DMO is to direct the customers to this site because giving online order is faster in terms of the sale procedures and more useful for both customers and DMO employees. Therefore, increasing the online sales will be very advantageous for DMO.

To provide convenience to customers and increase the sales, a recommendation mechanism that is integrated to this website can be very useful. It is aimed to develop a recommendation mechanism offers to customers that which product is bought with which product most in this study.

We should analyze customers by segmenting customers for to provide a better service to customers. In this Analysis, we can see if the customers are loyal. Thus, we can organize the campaign intended for losing customers to win them back.

DMO serves as a total of over fifteen thousand products, it is recommended that a specific brand / model is not proper. Categories are used as base instead of products. In other words, the final result of this study is the product groups (categories).

In order to accomplish this study, data-mining methodologies are used. There are different data mining approaches for solving different problems. Product recommendation problem is similar to "Market Basket Analysis" problem and Apriori Algorithm is implemented to solve the problem.

The result of this study is the information of which categories are bought together from online sales site and the results will be presented to executives. To integrate the recommendation mechanism to online sales site can be thought as the future work and it is out of the scope of the study.

II. BACKGROUND

A. RFM Analysis

RFM analysis is based on three concepts; Recency-Frequency-Monetary [1].

This analysis method can be used if a customer of a sales unit is to be better analyzed. While analyzing customers' purchases according to RFM Analysis; data should be considered regarding when the last time the customer shopped

(RECENCY), how often the customer shop (FREQUENCY) and how much cost the customer shopped (MONETARY). The most basic objective of this analysis is to identify the most valuable customers. The most valuable customers are the customers who spends the most money on short periods of time [2].

Customers should be divided into groups according to each parameter of the analysis and points should be given to the them. For example;

- For R (Recency); 5 points if the last shopping was done within the last 30 days, 4 points if it is between 30 days and 2 months, 3 points if it is between 2-6 months, 2 points if it is earlier than 1 year should be given.
- For F (Frequency); 5 points if the frequency of order is 20 or more than 20, 4 points if it is between 11-15, 3 points if it is between 3-5 and 2 points if it is between 1-2 should be given.
- For M (Monetary), 5 points if the total cost of shopping is more than 100.000 or more, 4 points if it is between 75.000-100.000, 3 points if it is between 50.000-75.000, 2 points if it is between 20.000-50.000 and 1 point if it is less than 20.000 should be given.

Three different parameters point of each customer are calculated by multiplying the points given after evaluations and the weight coefficient of the parameter. These scores are summed up to form a Customer-Value table and the customers with the highest score are described as the most valuable customers.

Customers are divided into groups as "Most Valuable", "Valuable", "Medium Valued" and "Being Lost". In this way, each group will receive special campaigns or discounts. By the developing strategies that could raise all customers to "Most Valuable" level, sales figures of the supermarket will be increased.

Campaigns for the "Being Lost" customers should be offered on the products that can attract those customers' interest. For this reason, that customers' purchases should be examined and the most purchased products should be presented as discounted to them with the products they will purchase. The method of Association Analysis composes the basis of the discount campaigns to be made at this point.

B. Data Mining

We use sql statements to get information from database but for some situations, these sql statements are not enough. For example, if we want to get the information of "the number of students who is registered to mathematics class", we can use a sql statement but if we wonder that "which characteristics affect the students whom will probably register the mathematics class", we should think a different solution [3].

The subject of data mining is to mine or extract knowledge in terms of patterns or rules from vast amounts of data. The goals of data mining can be thought as constructed by following four classes: prediction, identification, classification, and optimization [4].

The problem concerning this study is the subject of classification category because classification deals with the different classes or categories can be identified based on combinations of parameters. The aim of the classification is to categorize the data for analyzing afterwards [4].

During the data mining process, some useful information is discovered including association rules, classification hierarchies, sequential patterns, patterns within time series and clustering. For most of the applications, the combination of these information is the desired result. [4]

In this study, it is desired to obtain the knowledge of association rules.

C. Association Rules

One of the major technology in data mining is the discovery of association rules. Association rules correlate the presence of a set of items with another range of values for another set of variables. These rules are for retrieving the useful and important relations between the objects.

X => Y is the notation of association rules. Here the rule can be explained as "If X is valid, Y is valid too". The relation of the left side and right side is not a certain relation but a probabilistic one so it is more appropriate to explain this rule as "If X is valid, Y is quite likely valid too". "The customers who buy celery buy carrot too" or "The students who register to math class not register to the gym class" are the examples of association rules [5].

There are two measurements to evaluate the association rules namely support and confidence values.

- Support: With support value, it is shown that how frequently a specific item or item set occurs in the database.
- Confidence: With confidence value, it is shown that the probability of the realization of the right side with given left side. The confidence of X => Y rule is computed as Support (X U Y)/Support(X) [4].

Market Basket Analysis is a common example of the usage of association rules.

D. Apriori Algorithm

Apriori Algorithm is used for finding the frequent item sets among the data. The algorithm uses "bottomup" approach.

Firstly, Apriori algorithm finds frequent item sets have one item that provide minimum support value and then, frequent subsets are extended one item for every step using previous subsets. Every subset provides the minimum support value.

The input of the Apriori Algorithm is the transactions and the minimum support value, and the algorithm gives the frequent item sets as the output. The pseudo code of the Apriori Algorithm is given in Fig. 1.

Join process: Ck is generated by joining Xk-1 with itself
Prune process: Any(k-1)- item set that is not frequent can't be a subset of a frequent k-item set
Pseudo-code: Ck: Candidate item set of size k

```
Xk: frequent item set of size k
X1= {frequent items};
For (k=1; Lk !={};k++){
Ck+1=candidates generated from Xk;
For each transaction t in database do
Increment the count of all candidates in CK+1
That are contained in t
Xk+1= candidates in Ck+1 with min-support
End
Return UkXk;
```

Fig 1. The pseudo code of Apriori Algorithm [4]

III. RELATED WORK

There is no study utilized to find which product groups are bought together from E-Sales site in DMO before as far as we know. For this study, related works can be thought as solution of the market basket problem, studies about online sales site recommendation and usage of Apriori Algorithm.

There are various application areas for data mining in e-commerce area. These application areas can be listed as customer profiling, recommendation systems, web personalization and customer behavior [6].

The recommendation systems are for providing the related items to the users that they will probably be interested in. The users of the recommended system are the customers in e-commerce activities. Final recommendation systems provide the recommended content to the users. These systems recommend the related items to the customers using the other users' activities and choices. There are certain algorithms implemented for recommendation systems [6].

Previous online sales site recommendation studies contain Apriori Algorithm, exemplification algorithm, frequent-pattern (fp) tree, fp-growth algorithm and partition algorithm techniques [6].

There are many studies in e-commerce recommendation with Apriori Algorithm. Apriori is the most influential algorithm among Association rules algorithms. Apriori Algorithm is a the first thing coming to mind algorithm for finding the association rules in the frequent item sets. Most of the association rules algorithms is the interpretation and improvements of the Apriori Algorithm [7].

IV. PROBLEM DEFINITION

The aim of the study is to find the categories bought together. There are more than fifteen thousand products in DMO's online sales database so recommending a specific brand/model will not be proper for the results that we want to achieve.

DMO use Oracle's ERP system and all data about inventory, stocks, orders, sales, accounting, etc. is stored in ERP database. Online sales site uses its own database apart from ERP database but necessary information for web site is obtained from ERP database.

Data used for web site is retrieved from ERP database to web site database and after customers give orders from web site, related data is send to ERP database from web site database. Web site and ERP databases are separate from each other but use each other's data.

For data selection phase, firstly, it is analyzed that which tables in ERP database contain related information about web orders. After finding and deciding the tables, the details of the information stored in these tables investigated and the general information about these tables and tuples is learned. After observing the ERP tables, corresponding web site tables are investigated and analyzed. After realizing web site database tables are clear, simpler, more comprehensible and unnecessary data rating in web site tables is better than ERP database tables, it was decided using the web site database instead of ERP database.

The format of the tables and data in online sales database is not proper to implement the algorithm easily so the first step must be arranging the tables and data to implement the algorithm easily.

In addition to these, the Apriori Algorithm is utilized for finding the frequents item sets but the results must be the association rules which shows the probability of purchasing other products if a product (or products) is purchased. Based on this need, in addition to Apriori Algorithm, a method to calculate the confidence values and a method to find the association rules must be implemented too.

V. PROGRESS

The data selection and formatting the tables and the data can be thought as the first step for the implementation of the study.

There are various tables in online sales database but the relevant tables to this study are "Urunler", "UrunDetaylari", "Kategoriler", "Musteriler" and "Siparisler" tables.

"Musteriler" table is to store the information of customers (public enterprises) of DMO.

"Urunler" table is used to keep the catalogue and stock products information and detailed information about these products are stored in "UrunDetaylari" table.

"Kategoriler" table contains the information of product categories.

The most important table is "Siparisler" table because all orders and the order items data are stored in this table. Because using these tables as they are would be more difficult for coding, we created new tables which makes easier coding phase.

There are so many product categories and breakdown of the categories in online sales database. Each product is under one category but that category can be the lowermost breakdown. Therefore, we found each product's top level category and replaced the products' categories with that top level category. We created a new table called "Items" contains new category id.

The "Siparisler" table consist of the orders and order details but it's format is not proper for implementing the market basket approach. To overcome this problem, we created a new table called "Orders". The table has transaction id and items columns. For every transaction, all top level categories of products bought in that transaction are stored in items column. If a transaction has only one product, the table does not involve that transaction because our concern is the transactions have more than one product. We used the data which is obtained between the years 2013 and 2014.

The "Categories" table is for demonstrating the category names. Because some names have unnecessary information, these names are updated and unnecessary columns are removed.

The result of the algorithm is both frequent item sets and association rules according the last frequent item set. The application needs minimum support and confidence values to find the frequent item sets and association rules. The Result of Fridges Orders are given in Table 1 and Table 2 with Support:70 & Confidence:70.

Steps	Products	Support
1.Step	Fridges	293
	File Cabinets	95
	Table Sets	108
	Executive Seat	96
2. Step	Fridges-File Cabinets	95
	Fridges-Table Sets	108
	Fridges-Executive Seat	96
	File Cabinets- Table Sets	82
	File Cabinets- Executive Seat	74
	Table Sets- Executive Seat	84
3.Step	Fridges-File Cabinets-Table Sets	82
	Fridges- File Cabinets- Executive Seat	74
	File Cabinets- Table Sets-Executive Seat	70
4.Step	Fridges-File Cabinets-Table Sets-- Executive Seat	70

Table 1. Result of Frequent Item sets; Support:70

Statu	Associations	Confidence
-	Fridges=>File Cabinets-Table Sets-Executive Seat	0,2389
+	Fridges-File Cabinets=>Table Sets-Executive Seat	0,7368
+	Fridges-File Cabinets-Table Sets=>Executive Seat	0,8536
+	File Cabinets=>Table Sets-Executive Seat-Fridge	0,7368
+	File Cabinets-Table Sets=>Executive Seat-Fridge	0,8536
+	File Cabinets-Table Sets-Executive Seat=>Fridge	1
-	Table Sets=> Executive Seat-Fridge-File Cabinets	0,6481
+	Table Sets-Executive Seat-Fridge-File Cabinets	0,8333
+	Table Sets-Executive Seat-Fridge=>File Cabinets	0,8333

Table 2. Result of Association Rules; Confidence:70

VI. CONCLUSION

The result of this study is to find if a category is bought, the probability of the other categories is bought. In order to obtain this result, data mining methodologies are used. This problem looks like Basket Market Analysis and to solve this problem, it is necessary to find the frequent item sets. To create the frequent item sets, Apriori Algorithm is implemented. After finding frequent item sets, to find the association rules and their confidence values, other methods are implemented in addition to implementation of Apriori Algorithm.

The results show us which categories (product groups) are bought together from online sales site. For now, these results are useful for DMO and they will be presented to executives. Thus, we can organize the campaign intended for losing customers to win them back.

If this recommendation mechanism will be integrated to online sales site, it will not be enough to show to the customers only product groups. It can be shown best-selling products of that category, reduced products of that category or the products of that category that is decided by DMO like stock products.

Additionally, to make the recommendation, the time and the location of the orders can be considered individually. Producing results according to both location and time will be very useful in terms of making comparison.

VII. REFERENCES

[1] Wei Jo-Ting, Lin Shih-Yen ve Wu Hsin-Hung (2010), "A review of the application of RFM mode", African Journal of Business Management Vol. 4(19), pp. 4199-4206

[2] Derya Birant (2011). Data Mining Using RFM Analysis, Knowledge-Oriented Applications in Data Mining, Prof. Kimito Funatsu (Ed.), ISBN: 978-953-307-154-1

[3] Ünal Yarımağan (2000), Veritabanı Sistemleri, Ankara, Akademi ve Türkiye Bilişim Vakfı

[4] Wei Jo-Ting, Lin Shih-Yen ve Wu Hsin-Hung (2010), "A review of the application of RFM mode", African Journal of Business Management Vol. 4(19), pp. 4199-4206.

[5] Erdoğan, Gülcan ve Karampaşa (2015), "BİRLİKTELİK KURALLARI VE UYGULAMALARI: LİTERATÜR TARAMASI (2000-2014)", 13.Uluslararası Türk Dünyası

[6] Ming CHEN¸Research on Recommender Technology in E-commerce Recommendation System, 2010 2nd International Conference on Education Technology and Computer (ICETC), Shanghai, pp. 409-412, June 2010.

[7] WANG Pei-ji ,SHI Lin, BAI Jin-niu, ZHAO Yu-lin, Mining Association Rules Based on Apriori Algorithm and Application, IFCSTA '09. International Forum on Computer Science-Technology and Applications, Chongqing, pp. 141-143, December 2009.

Blockchain : A Decentralized Approach to Big Data

Senem Pehlivan Kaplan[1] and Assist. Prof. Dr. Serap Şahin[2]

[1] Sirnak University, Sirnak/Turkey, senemkaplan@sirnak.edu.tr
[2]Izmir Institute of Technology, Izmir/Turkey, serapsahin@iyte.edu.tr

***Abstract* - The blockchain technology is a hot topic and a new technology in recent years. It is not only an underlying technology for many applications like Bitcoin application, but also it is a kind of thinking including cognitive and mental processing and understanding for artificial intelligence and human enhancement. All data or services are digitized. So, this leads to deal with big data. It is a challenge to deal with big data from the perspective of performance, scalability, availability and privacy in centralized systems. Blockchain is applicable to big data and brings different perspective how to process, store, read and write data. Also, the aim of this paper is to show that better solutions are possible in a decentralized way. Even the technology is in its early stages, the blockchain technology will be in future due to its superior features. Therefore, it is better to adopt this technology as soon as possible to place in future. This paper gives a brief about how the blockchain could approach to big data and analyzes existing information regarding the challenges of big data from the side of the blockchain.**

***Keywords* – Blockchain, big data, decentralized systems, scalability, performance, availability, privacy.**

I. INTRODUCTION

Today, data is growing larger rapidly, since everything is digitized. Data could be structured or unstructured. Dealing and processing big data is a big challenge without sacrificing security, performance, scalability and availability. First applications to come to mind are social networks, IoT, mobile apps, healtcare etc. using videos, images, documents, electronic records, medical measurements. Also, these applications based on centralized client-server model has some limitations and concerns.

In recent years, a new technology called blockchain has drawn lots of attention and got much interest with its most popular and known application named Bitcoin. The first appearance of the blockchain is when Satoshi Nakamato comes up with a digital currency called Bitcoin using blockchain as an underlying technology in his white paper [1] in 2008.

Blockchain (BC) is a shared, distributed and decentralized ledger technology. BC uses P2P communications. Moreover, there is no trusted third party in the blockchain. It is an underlying technology for broad range of applications. Moreover, [12] emphasizes that BC is not only underlying technology, but also it is a kind of thinking for machines and humans. In BC thinking, processing occurs in a decentralized way.

The blockchain is categorized in three types such as public – permissionless, private – permissioned and consortium BCs in [5]. Bitcoin and Ethereum are the examples of the public ones. HyperLedger Fabric is an example of the private one. In the public BC, any node can join and leave the network. There is no restriction about it. On the other hand, in the private BC, to access the network, permission is required. If any node requests to join the network, it must be approved from other nodes in the BC private network. The study of [10] claims that this allows for accountability and compliance with laws and regulation and also supports openness and collaboration with participants. In consortium BC, preselected nodes determines the consensus process [5].

In BC, each transaction is stored in a block. After the transaction and block is verified, it is chained with other blocks. So, each node in the blockchain network has a copy of whole transactions, which provides transparency. Each record is stored permanently. In the private blockchain, only verified nodes can join the network. Because any trusted authority doesn't exist in BC, trust concept is based on cryptographic proof. So, this reduces costs and provides easiness. Even it is in early stages, it has already been used in many areas. Banking and finance, government processes, social media, healthcare, education operations, notary, real estate, IoT (Internet of Things) etc. Its approach is also applicable to big data and brings better solutions to handle big data.

Processing big data and keep sensitive information secure is a concern even in centralized systems. If the centralized system crashes, it could lead to data loss. Even GFS (Google File System) is based on the idea that component failures are the norm rather than exception [16]. This prevents data loss and guarantees availability. In the blockchain, all records is stored in all nodes in the network. Because the number of nodes are thousands or millions, and even some nodes crash in the network, data will be available all the time. However, storing any data in all nodes could lead to some storage costs or capability problems, which are still researched and there exists different type of solutions about these issues. Because there is no third party in the blockchain, transactions are done in roughly ten minutes, while they take a few days in centralized systems due to being intermediary in between.

It is possible to store all transactions including big size data like videos, pictures etc. Scalability is an issue in the blockchain while data gets larger. There are different studies and approaches to big data from the blockchain perspective.

In the blockchain network, anything can be exchanged. It could be currency, goods or service such as user post in social

networks, digital currency in banking, service in government agencies etc.

In existing systems, there are double-spending problem in banking and finance, higher costs not to stay anonymous or privacy issues from users' side. To overcome these, BC could be an alternative solution.

We know the blockchain technology is in its early stages. Scalability and security issues are still on table. To specify and draw its usage and range of application is not known yet, for it is the technology could applied to whole network, or mediator, public or private or even hybrid, we can extend or narrow it according the our application. Therefore, we need to consider this technology from the wider perspective due to its flexible feature.

In the next section, the important terms on the blockchain are described. In the third section, the limitations of centralized systems are reviewed. In the fourth section, pros and cons of the blockchain and its solutions for big data and, also, the existing applications of blockchain are analyzed and examined.

II. IMPORTANT TERMS IN BC

This section includes the definitions of blockchain related terms and concepts. These are important to explain the blockchain domain.

Node — It is a computer in the blockchain network.

Miner — A special node is usually rewarded to mine on the BC network.

Block — A block is a structure where each record is stored. The blocks are timestamped and then are chained in the blockchain. Each block consists of the block header, merkle tree and previous block's hash.

Distributed Systems — It consists of collection of independent computers. Whole system behaves like a single coherent system. The idea of BC is based on distributed systems.

Distributed ledgers — All records are shared and replicated among all nodes in the blockchain network.

Fork — A block could reference two or more blocks. Any user of the Bitcoin system may maintain a local copy of the blockchain and resolves conflicts by believing in the longest chain [14].

Consensus Process — In BC, each transaction or record must be reviewed, since after verifying transactions, they are immutable and irreversible. Also, there is no central authority in BC. Therefore, consensus is necessary and there are some consensus algorithms we will review here.

Consensus algorithms :

1. *Proof of Work (PoW)* : A block is hashed with a nonce in order to get a hash with zero bits at the beginning of hash. This consensus algorithm could use SHA-256 as used in [1]. Proof of Work effort is dependent on energy consumption and CPU power.

2. *Proof of Stake (PoS)* : This is an alternative consensus algorithm of Proof of Work. In PoS, mining is dependent on the amount of digital assets each node has.

3. *Proof of Activity* : This is a consensus algorithm of hybrid of Proof of Work and Proof of Stake.

4. *Proof of Intelligence* : It is a consensus algorithm dependent on reputational asset on the blockchain thinking.

In addition to these most known consensus algorithms, there are other consensus algorithms such as Proof of Importance, Proof of Authority etc., which we will not mention in this paper.

Mining — It is a process to make transactions valid at the expense of computational power. In Bitcoin BC, on an average, every 10 minutes, a new block is appended to the blockchain through mining [15].

Incentive — Mining is rewarded. For example, in Bitcoin, it is rewarded with coins. The reward is 12.5 BTC at the time of writing. Also, in [9], MedRec, healthcare management system based on BC, the reward is to access aggregate, anonymized medical data. Another example is that [6] representing a traffic announcement system has a reward mechanism informing other drivers with announcement of the accident to earn some coins.

Smart Contract — It is a small computer code which is executed on the blockchain. Ethereum blockchain nodes are capable enough to execute any kind of smart contracts.

III. LIMITATIONS OF CENTRALIZED SYSTEMS

The existing systems mostly adopt centralized client-server model. However, some businesses change their system to distributed ones. Distributed systems solve some issues, but it is not enough without combining both distributed and decentralized systems. The reason for transferring data to distributed systems is why centralized systems even looks like better to control of data, focus on data and be more consistent, and also it is not flexible and brings some issues to process a large variety and amount of data sets.

For most of applications, it is critical to access data all the time such as banking applications, government processes etc.. Depending on the centralized system has risks at crashing servers affecting whole system unavailable. This could be a disaster for banks and users.

Privacy is another concern from the side of centralized systems. Social networks gather all personal data of users and user preferences. Users has no control of how their data is processed and at whose hands. The privacy of users is violated due to imposing irreversible user preferences. The same problem occurs at in different area called Internet of Things (IoT). IoT objects has direct connections between users and IoT object providers. These providers gather all different kind of personal data including actions and habits of users. In addition to these, mobile applications has privacy issues for users. Once a user accepts a privacy policy of a mobile

application by installing it, it is unchangeable unless the mobile app is uninstalled.

In centralized systems, while large data sets grow exponentially, scalability and performance issues come up. Moreover, servers have permitted capability. Therefore, exceeding this capability leads to degrade performance to respond clients and Denial of Service (DoS). That is, these systems have scalability bottlenecks. For example, we can tackle healthcare systems. Healthcare systems deal with large size of data such as imaging, documents, electronic records, etc. The number of patients and their documents increases and their capability is limited. So, existing healthcare systems, have scalability concerns and, also, are not interoperable and convenience because different hospitals can not see their medical records, which is highly vital to track of patient's medical history and provide fast feedback to patients.

IV. BIG DATA ON BLOCKCHA IN

How BC approaches big data is a key concept to shape both today and future. BC has the distributed and decentralized architecture. It is immutable and irreversible. In BC network, any node has a whole copy of ledger. There will be thousands or millions nodes. Even some of computers crash in BC network, data loss is prevented and data is kept available by this way, since the decentralized nature of the BC is fault-tolerant. Trusted third party models such as banking, social networks are unavailable for security and maintenance purposes. BC enhances the availability of data while comparing it to traditional centralized systems by eliminating third parties. For example, file not found or server unavailable error messages will be mitigated at a significant percentage.

While we tackle social media networks, Internet of Things, mobile applications or data or digital content ownership applications suffer from especially privacy due to centralized systems. The existing applications don't guarantee user privacy. User has a lack of control of their data. The reference [4] proposes a user-centric permissioned blockchain based social media network called Ushare. The consensus algorithm of Ushare is Proof of Stake to validate transactions and blocks. As pointed out in [4], Ushare provides more secure social media network, for content is encrypted and Ushare works on a distributed hash table similar structure Bigtable [17]. Bigtable is established as distributed storage system managing structured data for large size data in order to achieve better scalability and higher performance used by Google as stated in [17]. A Personal Certificate Authority (PCA) is introduced to allow only user's circle members to see the user shared content [4].

It is obvious that the number of IoT objects are increasing every day. This brings privacy, scalability and performance concerns with it. Therefore, first of all, it is significant to provide privacy-aware solutions. In [7], a solution is brought for the privacy of IoT and mobile applications with the blockchain. In this design and implementation, the blockchain is an intermediary between users and IoT application or mobile application providers for management of privacy preferences. In the Ethereum blockchain, privacy policies and device's info are embedded in the smart contracts. Accepting and declining privacy policies are done via blockchain connected gateway. If the user accepts the privacy policy, whenever he connects IoT devices through the blockchain, the blockchain network preserves his privacy preferences. The digital signature scheme is based on ECDLP (Elliptic Curve Discrete Logarithm Problem). All user preferences are encrypted and stored in the blockchain network. The user can connect the gateway and review the privacy policies and information of the device. The computational cost of PDSS (Proposed Digital Signature Scheme) is nearly 283 ms considered practical and reasonable in real-life applications as pointed out in [7]. First implementation of Internet of Things based on the Ethereum blockchain network is Slock.it.

One of the area managing big data is healthcare. MedRec is a healthcare application based on BC. MedRec targets to access faster to medical data, to overcome interoperability; to enhance medical research with patient's data. MedRec's aim is to achieve security, privacy, interoperability and scalability issues by keeping immutable logs related to patients. In private blockchain networks especially, some modifications can be made for higher scalability and performance [9]. For this reason, MedRec is based on private, Ethereum BC which utilizes smart contracts categorized in registrar contract, patient-provider relationship contract and summary contract.

There are scalability and performance bottlenecks in the blockchain. While comparing transaction rate of visa and bitcoin, the most common application of BC, it is seen that Visa can process 4000 transactions per second, on the other hand, bitcoin processes 7 transactions per second. However, as a first step, re-parameterization of block size for 4 megabyte in Bitcoin and block intervals for roughly 10 minutes in Bitcoin make difference to get higher throughputs and lower latencies [19]. In addition to these, adopting different protocols will definitely increase scalability and performance of whole broad range of blockchain applications by maintaining the decentralization architecture of the blockchain. The abstraction layers in [19] are defined as network layer, consensus layer, storage layer, view layer and side layer, respectively from bottom to top. To analyze each layer leads us to find solutions to achieve scalability and performance issues. It is emphasized that broadcasting messages are made in network layer and also after propagating transactions to the BC network, a block is propagated. This means that transactions are stored in the block structure, so transactions transmitted twice. In consensus layer, different protocols such as Byzantine Fault Tolerant protocols or Paxos could be used but sharding protocols seem to be necessary to improve scalability of the BC. After authenticated data in consensus layer, in addition to data, smart contracts and views are stored and available in storage layer, which is vital to be developed for growing data [19].

EduCTX is also a consortium blockchain-based platform and works on delegated proof of stake (DPoS) as a consensus algorithm. Its aim is to transfer credits and grades among higher education institutes globally. This platform is

not restricted to any certain languages thanks to smart contracts. In this consortium blockchain, only higher education institutes can join the network after creating a blockchain wallet and verifying by the other nodes. Each student has a blockchain address and then, each ECTX tokens regarding the students' completed courses are transferred to the students' blockchain addresses.

Hyperledger Fabric is mostly known as a cryptocurrency, but actually it is an permissioned blockchain supported by IBM in order to be utilized as an underlying technology for a broad range of area such as finance, IoT, healthcare etc. The reference [10] emphasizes that in Hyperledger Fabric blockchain network, performance optimizations depends on code generated by compiler, the performance of cryptographic algorithms and enhancements specific. When an asymmetric algorithm is established on elliptic curves increases performance significantly.

V. CONCLUSION

To deal with big data, it is important to achieve certain issues such as privacy, availability, scalability and performance. At this point, understanding BC provides alternative and better solutions than existing traditional systems. Analyzing and summarizing applications built based on the BC will give us insight to evaluate and propose different approaches. Retaining the decentralization feature of BC, it is possible to choose the corresponded type of the blockchain and consensus algorithm for different areas.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system", White Paper, 2008.

[2] N. Bozic, G. Pujolle and S. Secci, "A tutorial on blockchain and applications to secure network control-planes," in Smart Cloud Networks & Systems (SCNS), Dubai, UAE, December 2016.

[3] E. Karafiloski, A. Mishev. "Blockchain solutions for big data challenges: A literature review." In: IEEE EUROCON 2017 -17th International Conference on Smart Technologies. July 2017, pp. 763–768. DOI: 10.1109/EUROCON.2017. 8011213 (cit. on p. 16).

[4] A. Chakravorty, C. Rong, "Ushare: user controlled social media based on blockchain", International Conference on Ubiquitous Information Management and Communication, 2017.

[5] M. Turkanovic, M. Hölbl, K. Kosic, M. Hericko, A. Kamisalic, EduCTX: A blockchain-based higher education credit platform. IEEE Access 2018, doi:10.1109/ACCESS.2018.2789929.

[6] L. Li. , J. Liu, L. Cheng, S. Qiu, W. Wang, X. Zhang., Z. Zhang, "CreditCoin: A Privacy-Preserving Blockchain-Based Incentive Announcement Network for Communications of Smart Vehicles", IEEE Transactions on Intelligent Transportation Systems. PP. 1-17. 10.1109/TITS.2017.2777990, 2018.

[7] S. Cha, J. Chen, C. Su, K. Yeh, "A Blockchain Connected Gateway for BLE-based Devices in the Internet of Things",IEEE Access. PP. 1-1. 10.1109/ACCESS.2018.2799942.

[8] G. Zyskind, O. Nathan and A. '. Pentland, "Decentralizing Privacy: Using Blockchain to Protect Personal Data,"2015 IEEE Security and Privacy Workshops, San Jose, CA, 2015, pp. 180-184. doi: 10.1109/SPW.2015.27.

[9] A. Ekblaw, A. Azaria, J. D. Halamka, MD, A. Lippman,n"A Case Study for Blockchain in Healthcare: "MedRec" prototype for electronic health records and medical research data", White Paper, 2016.

[10] N. Mencias, D. Dillenberger, P. Novotny, F. Toth, T. E. Morris, Jr., V. Paprotski, J. Dayka, T. Visegrady, B. O'Farrell, J. Lang, E. Carbarnes. (2018). An optimized blockchain solution for the IBM z14. IBM Journal of Research and Development. PP. 1-1. 10.1147/JRD.2018.2795889.

[11] A. Dorri, M. Steger, S. S. Kanhere and R. Jurdak , "BlockChain: A distributed solution to automotive security and privacy, [online] Available: https://arxiv.org/abs/1704.00073.

[12] Swan M, "Blockchain thinking: the brain as a decentralized autonomous corporation," IEEE Technology and Society Magazine, vol. 34, no. 4, pp. 41-52, December, 2015.

[13] S. Fujimura, H. Watanabe, A. Nakadaira, T. Yamada, A. Akutsu, J.J. Kishigami, "Bright: A concept for a decentralized rights management system based on blockchain", 2015 IEEE 5th International Conference on Consumer Electronics — Berlin (ICCE-Berlin), pp. 345-346, Sept 2015.

[14] M. Moser, R. Bohme, D. Breuker, "An inquiry into money laundering tools in the Bitcoin ecosystem", IEEE eCrime Researchers Summit (eCRS), 2013.

[15] G. Foroglou, A. L. Tsilidou, "Further applications of the blockchain", 2015.

[16] S. Ghemawat , H. Gobioff , S. Leung, "The Google file system", Proceedings of the nineteenth ACM symposium on Operating systems principles, October 19-22, 2003.

[17] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D.A. Wallach, M. Burrows, T. Chandra, A. Fikes, R.E. Gruber, "Bigtable: a distributed storage system for structured data", 2006.

[18] C. Jentzsch,"Decentralized Autonomous Organization to Automate Governance", White Paper, 2015.

[19] K. Croman, C. Decker, I. Eval, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saaxena, E. Shi, E. G. Sirer, D. Song, R. Wattenhofer, "On Scaling Decentralized Blockchains", (A Position Paper). In 3rd Workshop on Bitcoin and Blockchain Research, 2016.

Application of PageRank Algorithm in Linked Data

Y. GÜLTEPE[1], K. AKYOL[1] and A. KARACI[1]

[1] Kastamonu University, Kastamonu/Turkey, yasemingultepe@kastamonu.edu.tr
[1] Kastamonu University, Kastamonu/Turkey, kakyol@kastamonu.edu.tr
[1] Kastamonu University, Kastamonu/Turkey, akaraci@kastamonu.edu.tr

***Abstract* - The main purpose of the semantic web is to develop standards and technologies that will enable well-defined and linked information and services to be easily computer-readable and computer-understandable in the web environment. Linked data is one of the approaches used to acquire meaningful integrity by gathering data-related data collections by creating semantic links between the web pages that make up the content of the semantic web. Linked data is based on RDF (Resource Description Framework) technology. RDF is a data model that provides space-independent formal semantics with respect to chart resources. In a linked data application, the most important decision point is how to access the linked data. Linked data crawler is a program that explores linked data in web by tracking RDF links. In this work, DBLP (Database Systems and Logic Programming) data set is used as a source of Linked Data. DBLP gradually expanded toward all fields of computer science. An example will be presented related to pageRank sorting of RDF resources in the DBLP dataset. As a result; the search area has shrunk and search results have improved.**

***Keywords* - Linked Data, Semantic Web, PageRank, RDF, DBLP.**

I. INTRODUCTION

The semantic web, the linked data shows its existence today as a winning state of reality [1]. The existing data is spread on the web in the form of triples which can be understood by the structural and machine by complying with Resource Description Framework (RDF[1]) standard. With the spread of these triples, datasets are formed in certain areas. The resultant data sets form the basis of the linked data as a result of being associated with each other at the semantic level thanks to RDF links [2, 3].

Tim Berners-Lee first introduced the basic principles of the bound data concept in 2006 [2]. In 2007, A project called Linking Open Data (LOD), aimed at semantic associating open sets of data in open license status, was launched [2]. Thanks to the LOD project, it is open to anyone who wants to publish their data on the web according to their data guidelines. Because of this, the linked data space is constantly under development.

In the web search domain, text based search engines rank documents and domains by their popularity and relevancy levels [3]. However, in the semantic web and linked data fields, ranking has a more complex structure in the search process due to semantic relationships. Most of the targeted ranking methods for semantic web and link data are generated using common web search and ranking algorithms such as PageRank [3].

According to data sources, linked data ranking operations are classified in different forms. [4]. These classifications are ontology ranking, graph ranking, entity ranking, and RDF document ranking.

DBpediaRanker is hybrid ranking algorithm that uses Dbpedia resources. DBpediaRanker searches the DBpedia graph to calculate the value of resemblance for each couple of resources attained over the search process in graph exploration and queries the external information sources one by one [5].

In this work, DBLP data set is used as a source of Linked Data. An example will be presented related to pageRank sorting of RDF resources in the DBLP dataset.

The DBLP contains meta-data known to exist in computer science biography, consisting of more than a few million magazines by more than 1 million authors, and more than 1.8 million publications in conference proceedings. DBLP has begun to focus on two fundamental areas of computer science, database systems and logical programming (abbreviation). Despite this, however, it is rapidly evolving to cover all disciplines of computer science.

The system proposed in [6] combines the advantages of these two approaches, semantic-based and text-based information retrieval, with the ability to make advanced analysis by taking advantage of the results from the most popular search engines. In addition, ranking algorithms have been developed with textual and link analysis. A relative ranking system is different from PageRank style algorithms. The song is shudder. Every node in the graph has no importance on its own, but it is ranked by its neighboring nodes. Each node has a separate value based on the query being expressed. Instead of a single weight for a single source, a weight representing the similarity relation between the sources is calculated as the PageRank style algorithm is in operational mode [6].

[1] https://www.w3.org/RDF/

II. RANKING LINKED DATA

A. Ranking Methods

Linked data ranking works vary in ranking operations according to the data sources used. For example, an ontology-based ranking method can be used based on the use of semantic web technologies for web services. It is also based on the likelihood of advertising using ontology and the similarity between field service parameters. The TF-IDF[2] scoring is the sequencing result of the RDF sources. The RDF triple provides a "language modeling approach" that allows the resulting RDF table to be sorted. The "Maximal-Marginal Level" approach is used to repeatedly rank the top-ranked results and this approach calculates the relevance level as a result. Another approach is entity ranking approach. Entity ranking approach are two types; It is ranking for entities (such as persons, places and organisations) and entity type ranking (named as resource description ranking and property ranking).

B. PageRank Algorithm

PageRank is a value that indicates the population of a web page [1]. Web browsers; the quality of the web pages, the number of visitors and how often they are entered on the page are calculated and valued separately for each web page. Based on a given algorithm, page ratings are calculated one by one for each web page. These algorithms measure and determine the relationship of web pages to each other. Visits to linked pages; they are increasing the likelihood of reaching the values of the links they have established during the visit and the existing web pages linked.

In the 90's, PageRank is an algorithm that creates the original core of Google's search algorithm was evolved as a result of Larry Page and Sergey Brin's work [7].

$$PR(A) = (1-d) + d * (PR(\mathrm{T}_1)/C(T_1) + ... + PR(T_n)/C(T_n)) \quad (1)$$

PageRank equation showed in Equation (1). PageRank or PR(A) can be estimated using a basic iterative algorithm, and corresponds to the main eigenvector of the standardized web's link matrix.

It should be known that PageRank grades each web page separately and does not see web pages as a whole. It is also stated that the PageRank page of page A is replicated in a recursive manner by the PageRanks of the pages linked to page A. The T_i PageRank of the pages linked to Page A does not affect PageRank A of Page A. In the PageRank algorithm, a page's PageRank is always weighted by the count of out links C(T) on the T page. This means that the more out links a page T has, the less will page A draw on a link to it on page T.

The weights of PageRank's T_i pages are after selected. The result of this is that an extra incoming link for pageA will always raises page A's PageRank. Finally, the total of the weighted PageRanks of all T_i pages is producted by a damping factor (d) to set from 0 to 1. Therefore, the extend of PageRank utilization for a page by another page linking to it is droped.

In study [8], the PageRank algorithm is modified to correspond to the matrix of the graph where $P_x(a)$ is a value of node a in iteration x, d is a damping factor (for web graph orginally set to about 0.85), V is the vertex set whose nodes in the graph, U is a set of nodes with link to node a, D symbolize the set of all dangling nodes and w_{ij} is link weight for each link (i, j).

$$P_{x+1}(a) = \frac{(1-d)}{|V|} + d * ((\sum_{u \in U} \frac{P_x(u) * w_{ua}}{\sum_{v \in V} w_{uv}}) + \frac{\sum_{s \in D} P_x(s)}{|v|} \quad (2)$$

In Equation (2), the PageRank value of a page is the total of the PageRank values of the pages referring to this page. The ranking value is published equally for each page on the related pages.

Moreover, the canonical parameters "damping factor" and "number of iterations", PageRank [9] calculations depend naturally on most input charts. On PageRank calculations, Thalhammer and Rettigner [10] show a significant effect of the relationship between filtering and weighting. For example, it has been stated that the output of PageRank accounts in RDF version obtained from Wikipedia may be less related to page-view-based sequences than the PageRank accounts in wikipedia linkage chart.

III. APPLICATION OF PAGERANK ALGORITHM IN LINKED DATA

The DBLP Computer Science Bibliography contains thousands of journals or conferences, and more than 1.8 million academic studies written by over one million authors in these academic studies. DBLP has begun to initially target database systems and logic-oriented programming and grows to include all subsections in the computing domain.

In this study, DBLP rdf extension file is used. Because rdf files are used by various applications. It contains these tags and is a format that uses attributes as well as object definitions. Figure 1 shows a cross-sectional view of the DBLP data set (dblp-2018-04-01.rdf) using the Protégé[3] ontology development editor.

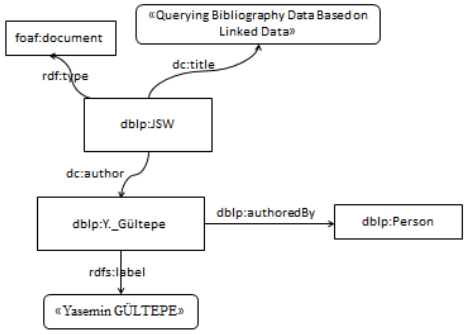


Figure 1: Data records in DBLP.

Any web browser needs a URL database to find all the URLs in the web. When the web browser performs a crawl operation

[2] http://nlp.stanford.edu/IR-book/html/htmledition/tf-idf1-weighting-1.html

[3] http://protege.stanford.edu/

for the PageRank operation, it must create a connection queue. Although this sequence is simple, it is an important process for large volume data sets.

The successive steps of the practice in this study are given below.

- The PageRank algorithm, briefly described in Chapter 2, is used to calculate PageRank scores on RDF graphs.
- Two calculation methods can be used in the PageRank algortihm. The first is the format used to code commonly used RDF graphs such as N-Triples or Turtle.
- A software was developed that runs the PageRank algorithm. The Eclipse IDE was used for this software. The Output file containing the PageRank results consists of a total of 23412131 lines.

In Figure 2, a simple 4-Triples document is taken as input. An N-Triples is a set of RDF terms that represent the subject, predicate, and object of an RDF triples.

```
<PageA> <link> <PageB> .
<PageA> <link> <PageC> .
<PageB> <link> <PageC> .
<PageC> <link> <PageA> .
<PageD> <link> <PageC> .
```

Figure 2: A simple RDF graph.

Figure 3 shows an example of PageRank dissemination. Page A with rank value 1,48 assigns a PageRank of 0,74 to pages B and C. Similary, page B assigns a PageRank of 0,78 to C.

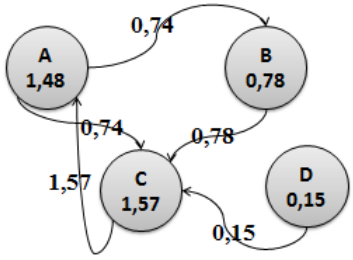


Figure 3: PageRank calculation on a simple graph.

Table 1 presents the total number of records available in the DBLP dataset for the year 2018 in the context of the latest updated data.

Table 1: Records in DBLP[4].

Title of Records	Number
Journal Articles	1669085
Book and Thesis	79461
Data and Artifacts	66
Editorship	37927
Parts in Books or Collections	36272
Informal_Publications	146719
Conference and Workshop Papers	2156740
Reference Works	9991
Withdrawn Items	14
Total	4136275

As stated in Table 1, the DBLP data size is too large. For this reason, PageRank values are calculated according to the specific nodes of the DBLP data set in Table 2.

Table 2: PageRank results belongs to some nodes in the DBLP dataset.

Node	PageRank	
<http://dblp.13s.de/d2r/resource/journals/soca>	0,53	
< http://dblp.13s.de/d2r/resource/journals/twc>	15,59	→ Highest ranking
http://dblp.13s.de/d2r/resource/publications/conf/lrec/2010	2,83	
http://dblp.13s.de/d2r/resource/publications/phd/us/Kim2008	0,1	→ Lowest ranking
< http://dblp.13s.de/d2r/resource/authors/Akemi_Izumiyama>	0,154	
<http://dblp.13s.de/Authors/Akemi+De+Castro>	0,17	

IV. CONCLUSION

In this study, the pageRank algorithm is applied on DBLP, one of the semantically associated data sets on the connected data cloud.

Ranking can be applied to almost every layer of the semantic web. At the lowest levels, triples or XML-related indexes are indexed and ranking, RDF descriptions and document files are ranked, ontology definitions and properties are ranked; and the results are ranked at a higher level.

Linked data browser is a program that searches for linked data on the web by following RDF links. In this study, DBLP (Database Systems and Logical Programming) data set is used as Linked Data source. DBLP has gradually expanded to all areas of computer science. An example of pageRank ranking of RDF resources in the DBLP dataset is presented. As a result; The search field has shrunk and the search results have improved.

REFERENCES

[1] T. Berners-Lee, J. Hendler, and O. Lassila, The Semantic Web, Scientific American, Vol. 284, No. 5, pp. 34-43, 2001.

[2] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data-the story so far". *International Journal on Semantic Web Information Systems*, Vol. 5, No .3, pp.1-22, 2009.

[3] S. Brin, and L. Page, Anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems. Vol. 30, No. 1, pp.107-117, 1998.

[4] S. Yumusak, E. Dogdu, and H. Kodaz, "A short survey of linked data ranking", ACM SE '14, 2014.

[5] R. Mirizzi, A. Ragone, T. D. Noia, and E.D. Sciascio, "Ranking the Linked Data: The Case of Dbpedia", *ICWE 2010*, pp.337-354, 2010.

[6] R. Miriziz, A. Ragone, T. D. Noia and E.D. Sciascio, "Semantic tag cloud generation via Dbpedia", *International Conference on Electronic Commerce and Web Technologies*, 2010, pp.36-48.

[4] https://dblp.uni-trier.de/statistics/recordsindblp

[7] L. Page, S. Brin, R. Motwani and T. Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", 1998, Technical Report, Standford InfoLab.

[8] M. Nykl, K. Jezek, M. Dostal, and D. Fiala, "Linked data and PageRank-based Classification", *IADIS International Conference Theory and Practice in Modern Computing 2013*, pp.61-64.

[9] S. Brin, and L. Page, "The Anatomy of a Large-scale Hypertextual Web Search Engine", *Proceedings of the Seventh International Conference on World Wide Web 7*, pp. 107–117, 1998.

[10] A. Thalhammer, and A. Rettinger, "PageRank on Wikipedia: Towards General Importance Scores for Entities", *ESWC 2016*.

An Analysis of Rumor Spreading Fundamentals with a Case Study on Facebook

Burcu SAYİN[1] and Serap ŞAHİN[1]

[1]İzmir Institute of Technology, İzmir/Turkey, burcusayin@iyte.edu.tr
[1]İzmir Institute of Technology, İzmir/Turkey, serapsahin@iyte.edu.tr

***Abstract* - Nowadays, the effect of social networks on people's lives is quite high. This situation gives rise to the density of information exist over social networks. That is why, analyzing the spreading pattern of information on social networks is an important issue today. The aim of this study is to technically review the background of information spreading, especially the fundamentals of rumor spreading and analyze the well-known methods on SNs. As a result, this article provides an important background for those, who works on the information spreading over SNs.**

***Keywords* – rumor spreading, epidemics, social networks**

I. INTRODUCTION

TODAY'S most popular virtual environment is Social Networks (SNs), with a 2.7 billion user [1]. SNs are mostly represented by graphs. Graphs are generally used to represent relationship among SN users. This relationship dynamically changes by (i) adding or removing new friendships and (ii) changing the profile settings. Some questions on this graph representation are listed below:

- How does a post or personal information spread on SN graph?
- Which model does better explain the information spreading?
- How the speed of information spreading is defined and measured?
- Which methods are more efficient to spread any information on SNs?

When answering these questions, an analogy is used between rumor and information spreading because information spreads like a rumor from one friend to another in SNs. That is why, the objective of this study is to comprehend the fundamentals of rumor spreading, including its mathematical background and state-of-art methods. Furthermore, we analyzed the efficiency of these methods in SNs. As a case study, we selected Facebook as our research domain.

The rest of this paper is organized as follows. In Section 2, we present the fundamentals of rumor spreading. Section 3 covers our experimental work on Facebook. We conclude the paper in Section 4.

II. FUNDAMENTALS OF RUMOR SPREADING

A. Epidemic Models

Previously, rumors and epidemics were likened to each other. Epidemics spread for a time and then lose its effect; rumor also has the same behavior but works with a threshold theorem. The theorem models the population by three types; susceptible (*s*), infected (*i*), and removed (*r*) which is called as SIR model. "Susceptible" ones are ignorant. After a susceptible one gets a rumor, it becomes "Infected". "Removed" means it stopped spreading.

In epidemics, time evolution of a disease according to this theorem is governed by a threshold that is defined based on the number of susceptible, infected and removed rates respectively. For example; in one hand, the probability of infecting the disease in an area that consists of a small number of people is low. On the other hand, if there are many people in the environment, this probability becomes bigger. Thus, population size is an important effect in epidemics [2].

Mainly, epidemics exist with two models: simple epidemics and complex epidemics [3]. Simple epidemics infect the whole population proportionally to the log of the population size. Let the population size be n, then epidemics spread with $\log n$. Additionally, individuals are always either susceptible or infected in this model. On the other hand, in complex epidemics, individuals are either susceptible, infected, or removed. The only complication is to decide when to be removed. In this instance, rumor spreading is deterministically modelled according to equations (1) and (2), given that $s + r + i = 1$.

$$\frac{ds}{dt} = -si \tag{1}$$

$$\frac{di}{dt} = +si - \frac{1}{ctr}(1-s)i \tag{2}$$

The equation (1) shows that susceptible ones will be infected according to the product of number of susceptible and infected populations denoted as "si". The equation (2) shows an interest lost for any infected one in time, which means it has recovered from disease or the individual lost interest to spread this rumor. To simulate these equations, an additional term counter value (ctr) is added, where $\frac{1}{ctr}$ shows the probability of interest loss. In equation (3), the infection function $i(s)$ denotes i as a

function of s with the elimination of time and defined the number of infected population:

$$i(s) = \frac{ctr+1}{ctr}(1-s) + \frac{1}{ctr}\log s \quad (3)$$

Here the value of $i(s)$ goes to zero when the value of s decreases exponentially with the value of ctr according to $s = e^{-(ctr+1)(1-s)}$ [3,4]. This means that if ctr is increased, rumor reaches a bigger portion of network, but it requires more rounds to succeed this. Hence, ctr provides control over the termination and the size of rumor spreading area in a network.
Furthermore, complex epidemics models are divided into two sub-models: complex epidemics with static rumor and complex epidemics with dynamic rumor. During the whole spreading process, if rumor does not have any revision, then it belongs to static rumor sub-model. Otherwise, it becomes a dynamic rumor which is revised by nodes during the process [5].
Overall, rumor spreading algorithms can focus on one of these model structures. Because of the relation between rumors and epidemics, rumor spreading algorithms are also established on these two models. In the following section, the implemented model for our experiments on Facebook SNAP dataset will be explained.

B. Rumor Spreading Models

Rumor spreading can be utilized in a network context to keep all the nodes up-to-date. Here, each node represents an individual. Demers et al. [3] define three common methods for performing this propagation update. These methods are:

- The *direct mail method*, where each update is mailed to all other nodes. Although it is efficient, it is not considered reliable because mails can be lost.
- The *anti-entropy method* is a simple epidemic technique where each node randomly chooses another adjacent node to resolve their differences and, then, exchange a rumor. All nodes perform this action at each round and therefore the communication cost increases. This method is less efficient from the direct mail due to the extra processing cost of difference resolving. Investigations on the cost of anti-entropy gave rise to three update distribution mechanisms: *push*, *pull*, and *push-pull* methods.
    - According to *push method*, in each round one node randomly chooses another one, if the node which made the selection has more updated rumor than the selected node, it pushes the update.
    - *Pull method* does the reverse; having the more updated rumor, the selected node pulls the update.
    - As for *push-pull method*, both push and pull methods are applied in each round.
- The rumor mongering method for update propagation adopts complex epidemics mechanisms to distinguish it from anti-entropy techniques. This method also uses the same update distribution mechanisms (push, pull and push-pull). The scheme considers all nodes as susceptible at the beginning. When a node takes a rumor, it becomes infected and only these nodes start to spread it by doing random choices till it loses interest in spreading the rumor. Due that reason, the communication cost decreases. The main issues of rumor mongering are:
    i. Decide when to stop spreading (the number of susceptible nodes should be close to 0, which is measured by the count of uninformed or residue nodes (rsd))
    ii. The spreading speed of rumor to all population should be minimized. The system should converge to an inactive state (a state that there is no infection, which means spreading is terminated) by the least number of rounds which is defined by S_n. S_n is also related to the speed of rumour spreading.
    iii. If ctr value is increased (as mentioned in equation 3), S_n increases. In this case, rsd value decreases because number of uninformed nodes becomes smaller when the number of rounds increase.

According to the study of Demers et al. [3]; randomized algorithms are used to propagate updates for database maintenance in a distributed environment. The algorithms in this study propose some methods which are listed below. They manage how the system converges according to different ways of interest loss in spreading process.

- *feedback:* a sender loses interest only if the recipient already knows the rumor, this feedback depends on the probability of neighbor node's state,
- *blind:* sender loses interest with probability $\frac{1}{ctr}$ regardless of the recipient state,
- *counter:*
    - counter with feedback: sender loses interest only after ctr unnecessary contacts,
    - counter with blind: sender loses interest after any ctr contacts.

In the proposed rumor spreading model:

i. The main performance variables are rsd (the number of residue uninformed nodes) and S_n (the number of rounds to converge whole network); rsd should be close to 0 and S_n should be optimized.
ii. The input variables are defined by ctr (counter) as mentioned above and the communication traffic m. Communication traffic is measured by the number of update messages sent in each round without regarding to the topology of the network.
    a. The deterministic solutions prove that increasing ctr value with feedback is an

effective way of minimizing the values of rsd and S_n.

b. The average value of traffic for an infected node in each round is formulated as $m = \frac{Total\ rumor\ update\ count}{number\ of\ node}$

iii. S_n is proportional to the value of m. Increasing the value of m, on the other hand, decreases the residue according to $rsd = e^{-m}$.

A relation among the performance variables (ctr, rsd, m and S_n) can, also, be seen in the original article [3].

Anti-entropy and rumor mongering methods both use the same push, pull and push-pull algorithms as mentioned above. But depending on the nature of the network, the advantages, and disadvantages between pull and push algorithms vary. For instance; if a network has very frequent multiple updates simultaneously, then the pull algorithm has advantages in spreading the rumor very fast because there is a high probability that the selected node is already infected. But if a network has very rare updates, then the pull algorithm creates an unnecessary traffic and in that case push algorithm is preferable. Karp et al. [6] compared the push and pull algorithms under the same assumptions such as similar update rate, under uniform distribution and a perfect interconnection without failures. The result of their performance is presented in Figure 1.a and 1.b.

- Push algorithm forwards the rumor to the nodes and the set of infected nodes grows exponentially until reaching to $n/2$ of the population. After this point as shown in Figure 1.a, the set of susceptible populations shrinks with a constant factor at each round. This factor is about $(1 - 1/e)$ since the fraction of players that do not receive a call in a round is approximately $1/e$. Thus, this shrinking phase takes $\theta(lnn)$ rounds and the push algorithm sends $\theta(n)$ messages.

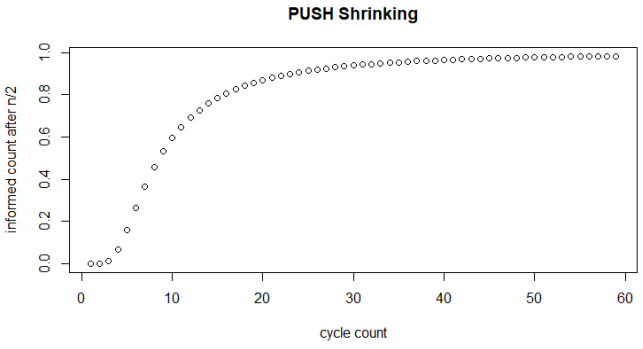


Figure 1.a. PUSH Algorithm Shrinking Phase

- In the implementation of pull algorithm; the infected node must wait for a connection request to start spreading the rumor. Therefore, propagation time can be unpredictable for the first round. After the set of infective node count reaches to $n/2$ of the population as shown in Figure 1.b, pull algorithm has an advantage against the push algorithm due to the fraction of susceptible nodes roughly squares from round to round. This is because in a round starting with ϵ.n susceptible nodes, each node has probability $1 - \epsilon$ to receive the rumor, so that the probability of staying susceptible is ϵ and $\epsilon > 0$. At the end of the round $\epsilon.n(1 - \epsilon) \approx \epsilon^2.n$ susceptible nodes will exist. Thus, shrinking phase takes $\theta(\ln\ln n)$ rounds and $\theta(n\ \ln\ln n)$ messages. "n" factor in the message count comes from the total node number because according to the assumption behind these methods is that each node transmits a message in each round.

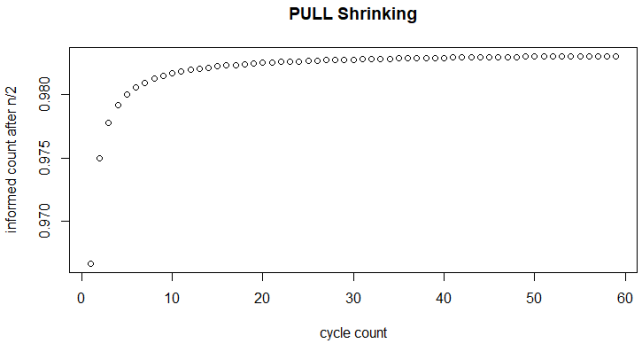


Figure 1.b. PULL Algorithm Shrinking Phase

The goal of R. Karp et al. [6] is to spread the rumor among all nodes with minimized number of rounds and update transmission. They impress that aforementioned algorithms are commonly used for lazy transmission of updates. They investigate that this large communication overhead is coming from the nature of epidemic algorithms and it can be reduced significantly when the rumor is sent in both directions. Combining the benefits of push and pull algorithm provides a more efficient algorithm. Push algorithm works efficiently until $n/2$ of the population becomes informed and pull algorithms works efficiently only after $n/2$ becomes informed. Thus, by using those two algorithms at the same time, we can get a more efficient result. With a simple push-pull algorithm, spreading a rumor to all nodes requires $S_n = O(\ln n)$ rounds and $O(n \ln\ln n)$ transmissions. For instance; the random phone call model uses uniform distribution model and push-pull algorithm [6]. This randomized communication of n players in parallel rounds increase robustness of the rumor spreading model. At each round, a node u randomly picks another node v. Then, u makes a phone call to v and they exchange rumors. At this model, they state that in a round t, rumor can be exchanged in both directions (push-pull). Thus, they claim that the number of transmissions can be reduced significantly by using a simple push-pull algorithm. This algorithm informs all nodes in time as the maximum number of rounds $S_{n_{max}} = log_3 n + O(\ln\ln n)$ using $O(\text{n}\ln\ln n)$ messages.

A problem of this algorithm is determining the optimal termination time. Furthermore, it is very sensitive to any error among nodes which affects the rumor spreading. In order to improve their solution, they devise a distribution termination scheme which is called as median-counter algorithm [6]. According to this algorithm, rumor is defined as r and there are four types of states for each node; A, B, C and D. Nodes that has not taken any rumor yet are in state A. For nodes, which are

in state B a counter is held, which is shown as ctr (v,r) where v denotes the node. Each time a node in state B takes a new rumor r, it increases its counter. If $ctr(v,r) = x$, then node v is in B-x state. If x is equal to the maximum counter value, state is changed to C. If a node v in state A receives r only from nodes in state B, then it switches to state B-1. If a node in state A receives r from a node in state C, then it switches to state C. For nodes in state C; every node stays in this phase for at most $ctr_{max} = O(\ln \ln n)$ rounds, and then switches to state D which terminates the rumor spreading. Hence, the median-counter algorithm uses only $S_n = O(\ln n)$ rounds with $O(\ln \ln n)$ message transmission. How the rsd value is affected and why median-counter algorithm is used are explained in the article thoroughly.

Furthermore, Karp et al. [6] give a general lower bound for the random phone call model using uniform distribution model that any address-oblivious algorithm requires $S_n = \Omega(\ln n)$ rounds in order to inform all nodes and requires $\Omega(n \ln \ln n)$ transmission independent from the number of rounds.

III. EXPERIMENTS

We performed our experiments on SNAP Facebook dataset [7]. This dataset contains real data from Facebook such as users, profile features, friendship relations etc. However, all these data are kept as anonymized to protect user's privacy. Our dataset contains 4039 real Facebook users and 88234 relations between them. We used NetworkX library [8] and created a graph from this dataset.

Before describing our experiments, we first provide some information about Facebook. Facebook is a social network which allow users to add other users as friend (friendship relation), follow the friends' updates, share photo/status (post), like/comment on shared posts, join specific groups etc. Moreover, when a user creates a post, s/he can select a privacy setting for it. There are some specific terms like Friends, Friends of Friends, Public etc. These terms define the visibility area of the posts on social graph. For example, if someone selects "Friends" as privacy setting for his/her post, then only his/her friends can see or access it. Thus, we can observe some subsets i.e. friendship graph over the whole social graph based on specific users. Friendship graph of a user only contains nodes and the corresponding edges which are directly connected with this user via friendship relation.

In Facebook, user's privacy preferences for the posts on their profile directly affect the visibility of those posts in SNs. The user has, also, the rights to put some additional restrictions on the visibility of specific posts, such as direct addressing to a specific user or group to view this post. Apart from the case of restricting the privacy setting to only a specific person or group, the post can be visible by all users in the friendship graph. However, the visibility of the post can be beyond of this friendship graph if the members of it also share this post with or without comments on their friendship graphs. According to these specifications of Facebook we assume that:

i. Each post can be generated randomly by any user and this post spreads on whole Facebook graph through connected users.

ii. Spreading process stops when all connected users in network receive the post and there is no requirement for a counter (ctr) value to define removal times.

In our experiments, we would like to observe theoretical outcomes by the measurement of S_n value for the spreading of a specific rumor on whole network. Hence, we used "complex epidemic model with static rumor, without a counter value" as presented in previous section. The following sections present the observed, analyzed and evaluated experiments.

We refer to Karp et. al. [6] and perform an experiment to observe push, pull and push-pull methods on Snap Facebook dataset.

We propose a detailed explanation of our experiment together with the algorithms in the following part:

Push method: Algorithm 1 shows the steps of this method. We kept and continuously monitored two lists. One for susceptible nodes and other one is for infected (informed) nodes. Initially, susceptible list contains all nodes in the graph and infected list is empty. First round starts after we randomly select a node from susceptible list and give the rumor to that node. The selected node becomes informed, so it is removed from susceptible list and added to our infected node list. One round is completed when all nodes in infected list randomly select a node among its neighbor nodes and pushes the rumor. Each new informed node is removed from susceptible list and added to infected list so list sizes changes dynamically. Push method runs until the whole nodes in the network become informed.

Algorithm 1

Input: *val* [int] – any information/rumor, *network* – social graph

1. SET *infectedNodeList* to an empty list
2. SET *nodeList* to all nodes in network
3. SET *randomNodeID* to a randomly selected node
4. Find the node whose ID is equal to *randomNodeID* and inform it about *val*
5. SET *numberofRounds* to 0
6. SET *infectedNodeCount* to 1
7. WHILE all nodes in network are not informed:
  a. FOR all nodes in *infectedNodeList*:
    i. Select a *neighborNode*
    ii. If the *neighborNode* does not have the rumor, inform it and add *neighborNode* to *infectedNodeList*. Increment *InfectedNodeCount* by 1.
  b. Increment *numberOfRounds* by 1

Output: *numberOfRounds*

Pull method: Algorithm 2 explains this method. We have susceptible and infected lists again but this time susceptible nodes randomly selects a neighbor and try to pull the rumor. All nodes in the network are susceptible at the beginning so susceptible list holds the whole network initially and the infected list is empty. After we randomly select a node and give the rumor, that node becomes informed and we remove it from susceptible list so add to infected list. Then, first round starts.

One round is completed when all nodes in susceptible list randomly selects a neighbor node in network and pulls the rumor. Each new informed node is removed from the susceptible list. Pull method runs until the whole nodes in the network become informed.

Algorithm 2

Input: *val* [int] – any information/rumor, *network* – social graph

1. SET *susceptibleNodeList* to all nodes in network
2. SET *nodeList* to all nodes in network
3. SET *randomNodeID* to a randomly selected node
4. Find the node whose ID is equal to *randomNodeID*, inform it about *val* and remove it from *susceptibleNodeList*
5. SET *numberofRounds* to 0
6. SET *infectedNodeCount* to 1
7. WHILE all nodes in *network* are not informed:
    a. SET *currentInfectedNodes* to empty list
    b. FOR all nodes in *susceptibleNodeList*:
        i. Select a *neighborNode*
        ii. If the *neighborNode* have the *val*, pull it and add this node to *infectedNodeList*. Increment *InfectedNodeCount* by 1.
    c. FOR all nodes in *currentInfectedNodes*:
        i. Remove from *susceptibleList*
    d. Increment *numberOfRounds* by 1

Output: *numberOfRounds*

Push-pull method: Algorithm 3 depicts the steps of this method. Whole network is kept in a list. First round starts after we randomly select a node and give the rumor to that node. One round is completed when all nodes in the list randomly choose a neighbor node and either pushes/pulls the rumor or does not do any operation. If the current node does not have the rumor but randomly selected neighbor node has the rumor, current node pulls the rumor from it. Otherwise, it pushes the rumor to randomly selected neighbor node. If both the current and randomly selected nodes have the rumor or neither of them have it, no operation is needed. Push-pull method runs until the whole nodes in the network become informed.

Algorithm 3

Input: *val* [int] – any information/rumor, *network* – social graph

1. SET *nodeList* to all nodes in network
2. SET *randomNodeID* to a randomly selected node
3. Find the node whose ID is equal to *randomNodeID* and inform it about *val*
4. SET *numberofRounds* to 0
5. SET *infectedNodeCount* to 1
6. WHILE all nodes in *network* are not informed:
    a. FOR all nodes in *network*:
        i. Select a *neighborNode*
        ii. IF the node has the *val* but *neighborNode* does not have; inform it and then increment *InfectedNodeCount* by 1.
        iii. ELSE IF the node does not have the *val* but *neighborNode* has; pull the rumor and then increment *InfectedNodeCount* by 1.
    b. Increment *numberOfRounds* by 1

Output: *numberOfRounds*

Our results support the study of Karp. et. al. [6]. Let the population size be n. After $n/2$ of the population being informed, pull method spreads faster than push method. Below two figures shows the differences.

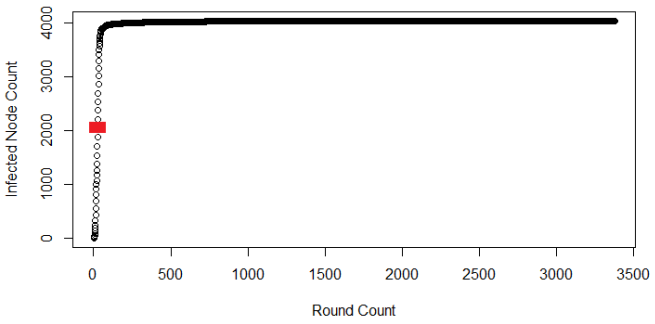


Figure 3.a PUSH Algorithm Shrinking Phase
Population Size: 4039

For the push algorithm shrinking phase, 2042 nodes become informed after 26 rounds (red point in Figure 3.a).After that, 3354 more rounds are performed to inform whole population (4039 nodes). Figure 3.a shows this situation, after $n/2$ of the population becomes informed, it takes too much round to inform whole population because the probability that a randomly chosen neighbor node being uninformed (susceptible) is too low.

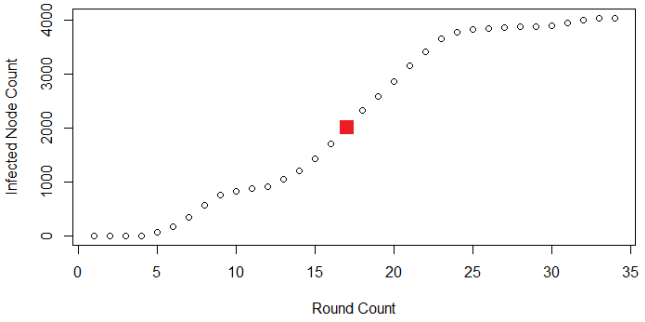


Figure 3.b PULL Algorithm Shrinking Phase
Population Size: 4039

For the pull algorithm shrinking phase, 2017 nodes become informed after 17 rounds (red point in Figure 3.b).After that, 17 more rounds are performed to inform whole population. Thus,

pull algorithm informs whole population in only 34 rounds. It can also be seen in Figure 3.b that after $n/2$ of the population becomes informed, whole population becomes informed in a fast way because the probability that a randomly chosen neighbor node being uninformed (susceptible) is very high.

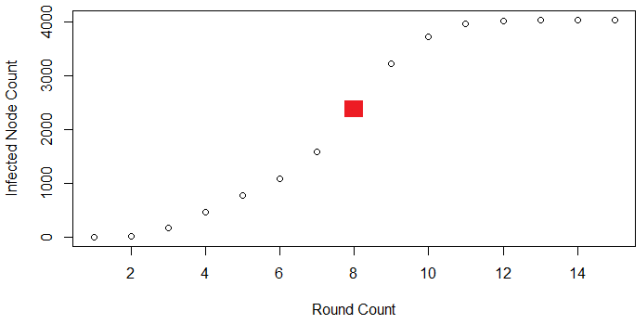


Figure 3.c PUSH-PULL Algorithm Shrinking Phase Population Size: 4039

For the pull algorithm shrinking phase, 2396 nodes become informed after 8 rounds (red point in Figure 3.b).Then, only 7 more rounds are performed to inform whole population. It can be seen in Figure 3.c that push-pull algorithm over performs both push and pull algorithms and spreads the rumor through all nodes in a fast way. It takes 15 rounds to inform 4039 nodes. This experiment supports the claim of the study [9] which says that "if an n-node connected graph has conductance ϕ then rumor spreading successfully broadcasts a message within $S_n = O\left(\frac{log^4 n}{\phi(G)^6}\right)$ steps, with high probability, using the push-pull strategy."

IV. Current Approaches to Rumor Spreading

In this section, we will firstly introduce the studies, which focus on the modified version of SIR model as an adaptation to information spreading, and then demonstrate an information spreading model based on cascades, that allow us to predict how well the information will be spread.

Bao et al. [10] revises SIR model and proposes SPNR model by dividing the Infected state into two: (i) Positive Infected (P: nodes that have been infected, and they support the information) and (ii) Negative Infected (N: nodes that have been infected, but they oppose the information).

Serrano et al. [11] propose an agent-based information spreading model, considering four states: (i) Neutral (initial state), (ii) Infected (believe the information), (iii) Vaccinated (believe the anti-information before being infected) and (iv) Cured (believe the anti-information after being infected).

Cordasco et al. [12] propose that a user may not immediately start spreading when he/she becomes an Infected; so they only become an "Aware". This model consists only these three states: (i) Ignorant, (ii) Aware and (iii) Spreading. Instead of defining a Removed state, they propose a termination rule in the original paper.

Sumith et. al. [13] claims that the assumptions made in SIR model fail in real world. They propose that the whole population do not mix with each other equally, so the distribution is not homogeneous. Moreover, all individuals are not Susceptible initially, and most of them will restrain themselves from interactions. With this approach, authors proposed R_n SIR model, as an extension to SIR model. Hence, they added a new state to SIR model, which is called as R_n. R_n state represents nodes, who restrain themselves from any interaction with other nodes.

Tong et al. [14] describes an information cascade model in SNs. First, they provide an extensive study on cascade scales, the scope of the cascade subgraphs, and topological attribute of spread tree. Then, based on the evaluation results, they analyze the spread of the user's decisions for city-wide activities. Decisions include "want to take part in the activity" and "be interested in the activity". This study introduces three mechanisms to use for making a decision: (i) equal probability, (ii) similarity of nodes, and (iii) popularity of nodes.

Overall, the main aspect of the current approaches for modeling the information spreading is to propose a realistic model that matches with the complex and dynamic mechanism of human behavior. Hence, researchers try to adopt their models with new parameters, such as popularity of nodes, similarity between them, etc.

V. Conclusion

The aim of this study is to propose an enhanced technical review on rumor spreading models and analyze the success of core update distribution mechanisms (push, pull, push-pull) on a popular social network, Facebook.

In this paper, we highlighted how an epidemic model and rumor spreading model resembles each other. Then, we analyzed related mathematical models, rumor spreading algorithms and their success on Facebook data. Results show that complex epidemic model can be adapted to social networks and push-pull algorithm is a realistic approach to observe the actual spreading of information.

Many other implementation and research domains exist. Most of them requires efficient data aggregation algorithms and modelling for dynamic SNs [15,16]. Another application domain could be source detection of any information exist on SNs, which is a valuable evidence to verify any received information and to protect SN user from malware attacks.

References

[1] https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/ "Number of social media users worldwide from 2010 to 2020 (in billions)"

[2] D. J. Daley and D. G. Kendall (1965), Stochastic Rumours, IMA J Appl Math (1965) 1 (1): 42-55. doi: 10.1093/imamat/1.1.42

[3] A. Demers, D. Greene, C. Hauser, W. Irish, J. Larson, S. Shenker, H. Sturgis, D. Swinehart, and D. Terry. Epidemic algorithms for replicated database maintenance. In Proceedings of the sixth annual ACM Symposium on Principles of distributed computing (PODC '87), Fred B. Schneider (Ed.). ACM, New York, NY, USA, 1-12. 1987. DOI=10.1145/41840.41841 http://doi.acm.org/10.1145/41840.41841.

[4] N. T. J. Bailey. The Mathematical Theory of Infectious Diseases and its Applications. Hafner Press, Second Edition, 1975.

[5] Y. Zhang, S. Zhou, Z. Zhang, J. Guan and S. Zhou. Rumor Evolution in Social Networks. Physical Review E, vol. 87, no. 3, Article ID 032133, 2013.

[6] R. Karp, C. Schindelhauer, S. Shenker, and B. Vocking. Randomized rumor spreading. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS'00). IEEE Computer Society, Washington, DC, USA, 565-574, 2000. DOI: 10.1109/SFCS.2000.892324.

[7] https://snap.stanford.edu/data/egonets-Facebook.html

[8] https://networkx.github.io/

[9] F. Chierichetti, S. Lattanzi, and A. Panconesi. Rumour spreading and graph conductance. In Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms (SODA '10). Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1657-1663.9. 2010.

[10] Y. Bao, C. Yi, Y. Xue, and Y. Dong. A new rumor propagation model and control strategy on social networks. In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'13). ACM, New York, NY, USA, 1472-1473, 2013. DOI=http://dx.doi.org/10.1145/2492517.2492599.

[11] E. Serrano, C. Ángel Iglesias, and M. Garijo. A Novel Agent-Based Rumor Spreading Model in Twitter. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). ACM, New York, NY, USA, 811-814, 2015. DOI: http://dx.doi.org/10.1145/2740908.2742466.

[12] G. Cordasco, L. Gargano, A. A. Rescigno, and U. Vaccaro. Brief Announcement: Active Information Spread in Networks. In Proceedings of the 2016 ACM Symposium on Principles of Distributed Computing (PODC '16). ACM, New York, NY, USA, 435-437, 2016. DOI: http://dx.doi.org/10.1145/2933057.2933069.

[13] N. Sumith, B. Annappa, and S. Bhattacharya. Rnsir: A new model of information spread in online social networks. In 2016 IEEE Region 10 Conference (TENCON), pages 2224–2227, Nov 2016.

[14] C. Tong, W. He, J. Niu, and Z. Xie. A novel information cascade model in online social networks. Physica A: Statistical Mechanics and its Applications, 444(Supplement C):297 – 310, 2016.

[15] D. Kempe, A. Dobra, and J. Gehrke. 2003. Gossip-Based Computation of Aggregate Information. In Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS '03). IEEE Computer Society, Washington, DC, USA, 482-491, 2003.

[16] C. Liu, Z.-K. Zhang, Information spreading on dynamic social networks, Communications in Nonlinear Science and Numerical Simulation, Volume 19, Issue 4, April 2014, Pages 896-904, ISSN 1007-5704, http://dx.doi.org/10.1016/j.cnsns.2013.08.028.

Design and Fabrication of High Gain Ultra-Wideband Antenna for Microwave Imaging and Radar Applications

A.R.CELIK[1], M.B.KURT[1] and S.HELHEL[2]
[1] Dicle University, Diyarbakir/Turkey, eeealicelik@gmail.com
[1]Dicle University, Diyarbakir/Turkey, bkurt@dicle.edu.tr
[2]Akdeniz University, Antalya/Turkey, selcukhelhel@akdeniz.edu.tr

***Abstract* – This paper describes an antenna design that is suitable for microwave imaging and radar applications. Microwave imaging technology has attracted many interests nowadays and it has been used in a variety of applications such as: non-destructive testing and evaluation, through-the-wall imaging, concealed weapon detection at security check points, structural health monitoring and medical imaging. Similarly, radar-based applications have been popular in many areas. The basic idea of using microwave imaging and radar systems is to transmit electromagnetic waves from a transmitting antenna to the target material and receive the scattered waves at a receiving antenna. For this reason, the choice of the antenna plays an important role for the system. There is a need for compact sized, low cost and high efficiency antennas which can radiate ultra-wideband signal to transmit short pulses. Furthermore, these antennas should have similarly end-fire radiation pattern to obtain good resolution of the produced images for using both in imaging systems and radar applications. In this study, firstly a conventional compact-sized rectangular patch antenna is designed. Then, various optimizations are performed on that antenna by using High Frequency Structural Simulator (HFSS) software. After that, this antenna is fabricated and tested with Vector Network Analyzer. The fabricated antenna has a simulated and measured bandwidth from 4 GHz to 9 GHz for |S11|<10 dB, respectively. The return loss results show that the good impedance matching is obtained through the working frequency band. The proposed antenna has nearly stable end-fire radiation patterns throughout the frequency range. All of the results exhibit that the designed antenna can be used in high range radar applications and is a good candidate for microwave imaging applications.**

***Keywords* - UWB, High gain, Antenna design, Microwave imaging, Radar applications**

I. INTRODUCTION

Microwave Imaging (MI) methods have shown excellent capabilities in various fields such as civil engineering, nondestructive testing, industrial and security applications, and have in recent decades experienced strong growth as a research topic in biomedical diagnostics [1-3]. Similarly, radar applications such as detecting objects, determining the structure of the target, and estimating the direction and speed have become widespread in recent years [4,5].

Since the authorization of the unlicensed using of the ultra-wideband (UWB) technology in the range of 3.1 GHz to 10.6 GHz by the Federal Communication Commission (FCC) in the USA [6], using the UWB signals in the MI and radar-based applications has become popular. A key component of these systems is the antenna that is used to radiate and receive the UWB pulses. Good impedance matching, high directivity, narrow beamwidth and small size are desirable properties for the antenna to create a successful MI and radar system [7].

Microstrip patch antennas consist of a dielectric substrate, with a ground plane on the other side. They are compact and have capability to integrate with microwave integrated circuits technology due to their low-profile structures. On the other side, the greatest disadvantage of the microstrip patch antennas is their narrow bandwidth. Also, these antennas generally exhibit omni-directional radiation patterns. However, MI systems and radar-based applications need the antenna having broad band signal to transmit short pulses and end-fire radiation pattern to obtain sufficient range resolution, respectively [8]. In order to remove these disadvantages and to improve the performance of the antenna, there are many techniques such as adjusting feed point, adding parasitic elements to the patch side, etching notches and slots on the ground plane etc. A number of UWB antenna designs have been proposed in [9-13] by using these modification techniques. Each has its own merits and drawbacks.

The modified microstrip antennas which has partial ground plane is called planar monopole antenna and they are the most chosen antennas in the UWB applications because of their broad bandwidth, compact and easy assembled structure. They are named according to their radiation patch shapes such as elliptical, square, circular, rectangular or any session of these shapes. In this paper, firstly a square planar monopole antenna is designed. Then, some modifications are made to improve the performance of the antenna by using High Frequency Structural Simulator (HFSS) software.

After the optimum design parameters are obtained, the proposed antenna is fabricated and tested. Simulation and measurement results show that fractional bandwidth and gain values are suitable for the aim of this study.

II. DEVELOPMENT PROCESS OF THE PROPOSED ANTENNA

A. Operation Mechanism of the Ultra-Wideband Planar Monopole Antenna

Conventional microstrip square patch antenna has a complete ground plane as shown in Figure 1. It is capable of supporting only one resonant mode and has narrow bandwidth as mentioned before. The square planar monopole antenna that has a partial ground plane is given in the Figure 2 (a) as a model to be used in applications of UWB technology. It can support the multiple resonant modes [14]. Figure 2 (b) shows the overlapping of closely spaced multiple resonance modes (f_1,f_2,f_3,…,f_N), which all together form a wide bandwidth for planar monopole antennas [15].

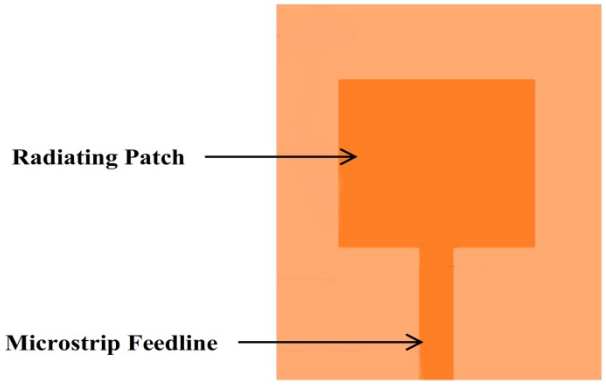


Figure 1: Conventional microstrip square patch antenna.

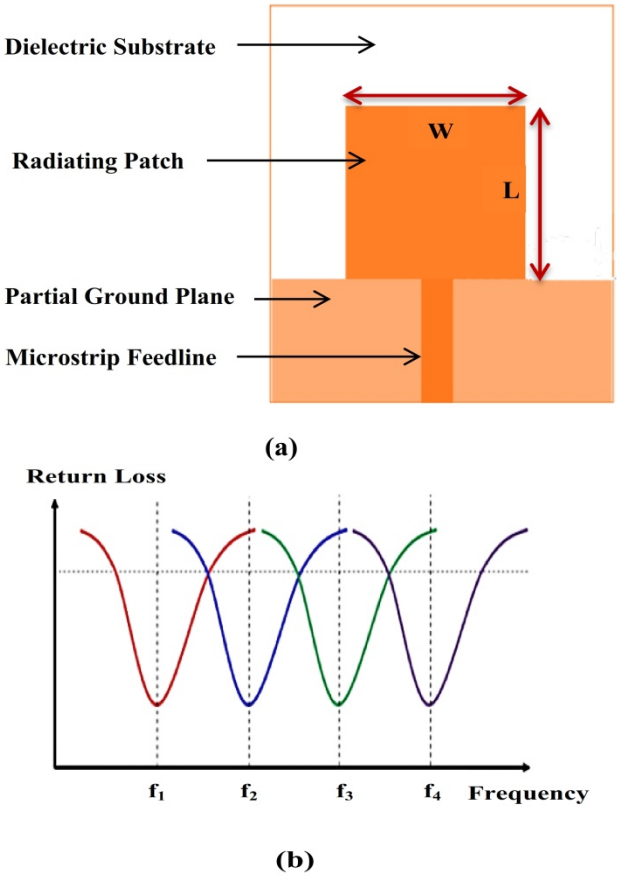


Figure 2:
a) Square planar monopole antenna, b) Multiple resonance modes

B. Determination of the Lower Frequency for Planar Monopole Antennas

In the design process of an antenna, it is required to keep several parameters such as frequency bandwidth (BW), voltage standing wave ratio (VSWR), half power beamwidth (HPBW), gain and directivity in the specific range and values. For this aim, firstly a frequency range is determined, then size and geometry of the radiating patch are calculated according to this bandwidth range. After that, some modifications such as adjusting feed point, adding parasitic elements, etching notches and slots are made to improve performance of the antenna [16].

In the development process of the UWB planar monopole antenna, the size of the patch is used to help for specifying the frequency of the first resonant mode. In generally, the lower working frequency of the antenna corresponding to VSWR = 2 can be approximately calculated by the formula, as given below [17]:

$$\lambda = \frac{L+r}{0.24} \qquad (1)$$

and

$$f_L = \frac{c}{\lambda} = \frac{30 \times 0.24}{L+r}\ \text{GHz} = \frac{7.2}{L+r}\ \text{GHz} \qquad (2)$$

where **L** means the length of the patch, **W** means the width of the patch, and **r** is equal to **W/2π**. L and W are in centimeters. For the square patch **L=W**, so the formula is simpler. Since the FCC permitted the unlicensed use of UWB frequency band covering from 3.1 to 10.6 GHz, we specify the lower frequency of the proposed antenna as 3.1 GHz in this study. In order to obtain this frequency, L is calculated as 20 mm from the Equation (2). A compact size (4x5 mm^2) FR4 Epoxy material is used as substrate. We found optimum value for width of the partial ground plane as 15 mm. Length and width of the feedline are taken as 11 mm and 3 mm, respectively to achieve the 50Ω impedance matching. These design values are shown in Figure 3.

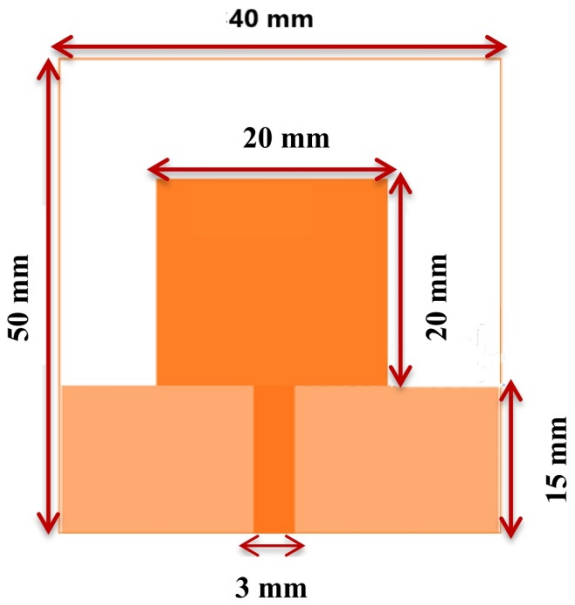


Figure 3: Design parameters of planar monopole antenna

C. Bandwidth Enhancement of the Planar Monopole Antenna with Defected Ground Structure

The return loss (RL) results for the conventional microstrip patch antenna (given in Fig. 1) and planar monopole antenna (given in Fig. 3) are shown at the same graphic in Figure 4. It is clearly seen that the working frequency band (RL<-10dB, VSWR<2) of the monopole antenna is larger than the conventional one. Also, the lower frequency value for the band of monopole antenna is equal to approximately 3.1 GHz as planned in the previously Section. However, this result is not quite sufficient for the MI system, because it uses an UWB pulse which includes low to high frequencies. The lower frequency band ensures adequate depth of penetration while the higher frequency band ensures the enough resolution of the resulting images [18]. For this reason, it is desirable to obtain as large bandwidth as possible.

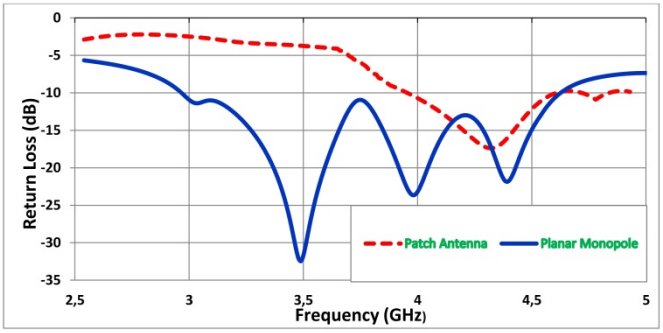


Figure 4: Comparison of return loss for patch and monopole antennas

The bandwidth of the antenna given in Figure 3 can be enhanced by using Defected Ground Structure (DGS) approach more [19,20]. The method of etching slot(s) or notch(es) on the ground plane is one of the most used methods for the bandwidth improvement of a compact planar monopole antennas [21,22]. In this study, two identical rectangular slots are used in order to mitigate the reflection of the surface current for adjusting the antenna impedance and reducing the return loss. Also, a small notch is etched between the ground plane and radiating patch in order to improve return loss level and bandwidth [23]. The optimum sizes of the slots and notch are specified parametrically as 6x3 mm^2 and 1x5 mm^2, respectively. The new antenna is shown in Figure 5.

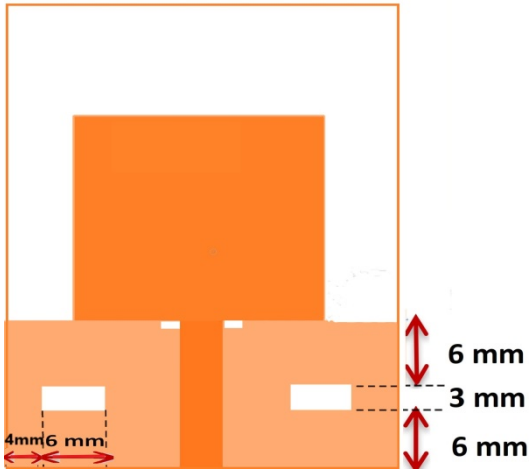


Figure 5: Modified square planar monopole antenna

After these modifications, the new return loss characteristic of the antenna covers the band from 3 GHz to 8 GHz as it is seen from Figure 6.

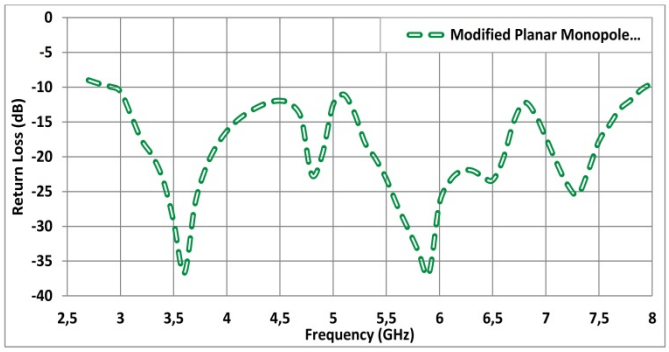


Figure 6: Improved return loss characteristic after some modifications

D. Improving the Directionality and Gain by Using Modifications on the Square Patch

While the conventional microstrip patch antenna exhibits omni-directional radiation pattern, the designed planar monopole antenna given in Figure 5 has two radiation lobes which is shown in Figure 7. However, it is not sufficient for the antenna to be used in the MI systems and radar-based applications. Because, it should have also a low HPBW which increases the resolution capabilities of the antenna to distinguish between two adjacent radiating sources or radar targets through worked frequency band [24].

There are a lot of modification techniques to obtain narrow HPBW and increase the directionality. In this study, we prefer cutting two triangular slots from the bottom corners of the patch. Furthermore, two triangular parts with the same size of the slots are added to top corners of the patch. After these operations, the radiation patterns at 4 GHz, 6 GHz and 8 GHz are obtained as shown in Figure 8. These patterns are stable through frequency band. Hence, it can be said that the proposed antenna is suitable for the aim of our studies.

The development process of the antenna has come to end with the achievement of the desired results. The proposed antenna is given in Figure 9.

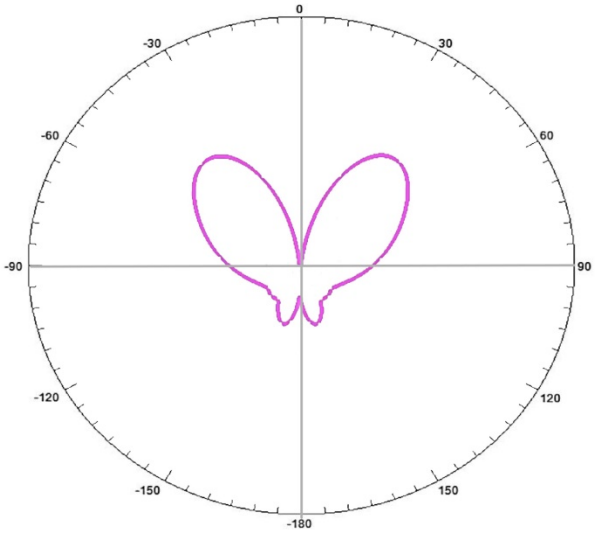


Figure 7: Radiaton pattern of the antenna given in Fig.5

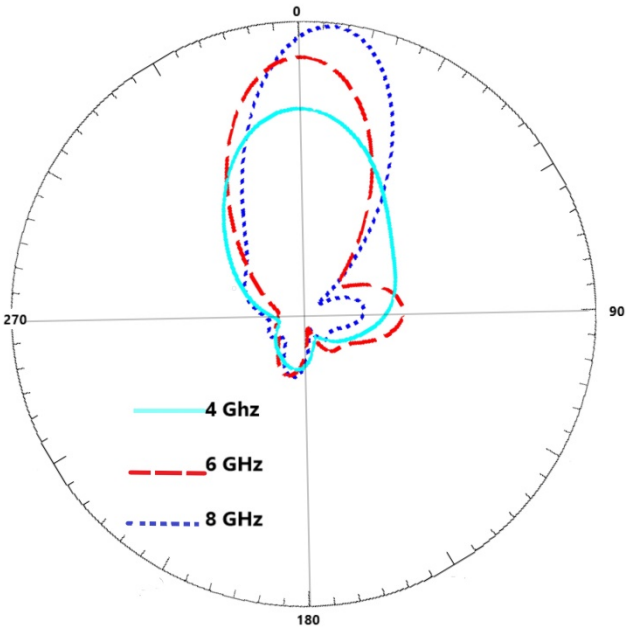


Figure 8: Improved radiation patterns after some modifications

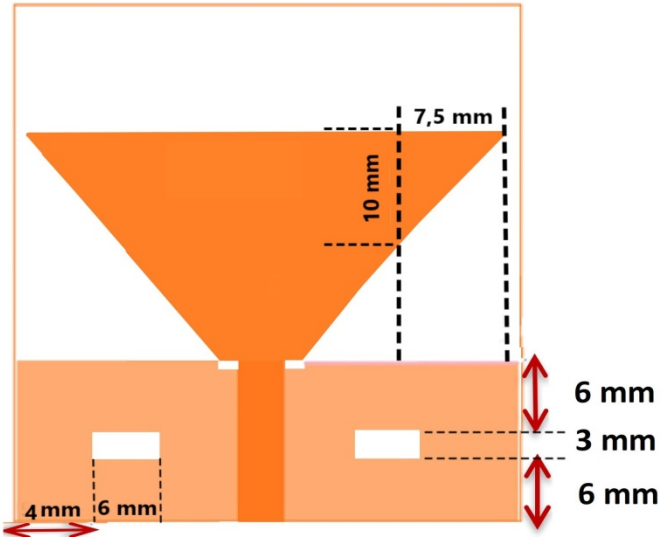


Figure 9: The developed and proposed antenna

III. Experimental Measurement and Results

At the end of the development process, the first shape of the antenna is changed significantly. Hence, the lower frequency and return loss values are altered. According to the final simulation results, the proposed antenna covers the broad frequency range of 4 GHz to 9 GHz (77%) for RL<-10 dB. After the optimum parameters and dimensions are obtained, the proposed antenna is fabricated as seen in Figure 10, and tested by using Anritsu MS2028C Vector Network Analyzer.

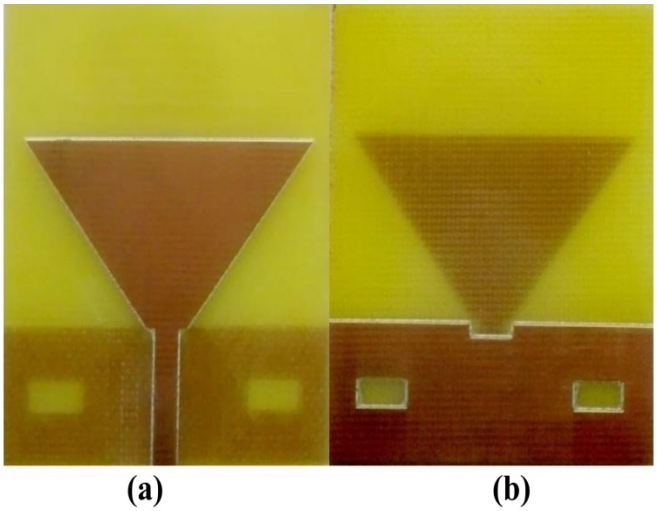


Figure 10: Fabricated antenna (before soldering SMA connector) a) Front view, b) Back view

A. Return Loss Results

Both simulated and measured return loss results of the proposed antenna are given at the same graphic in Figure 11. Although there are some disagreements around a small number of frequencies, two results show the similar broad frequency range. The disagreements in the measured and simulated results may be due to the uncertainties in dielectric material, fabrication constrains, soldering of SMA connector effect, and measurement losses.

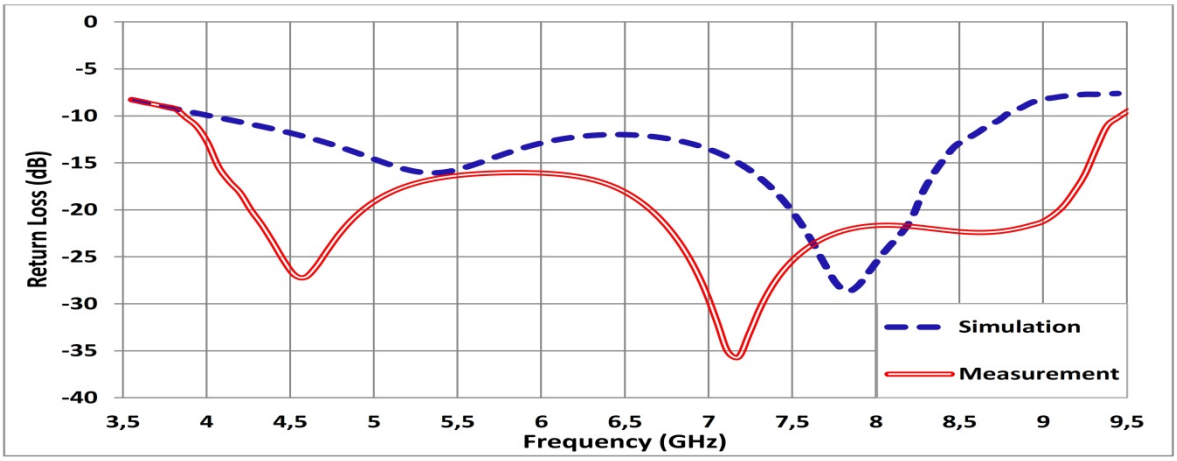


Figure 11: Simulated and measured return loss results of the proposed antenna

B. 3D Polar Plot of the Radiation Pattern

3D radiation pattern of the proposed antenna for the frequency of 6 GHz which corresponds to φ=0° is shown in Figure 12.

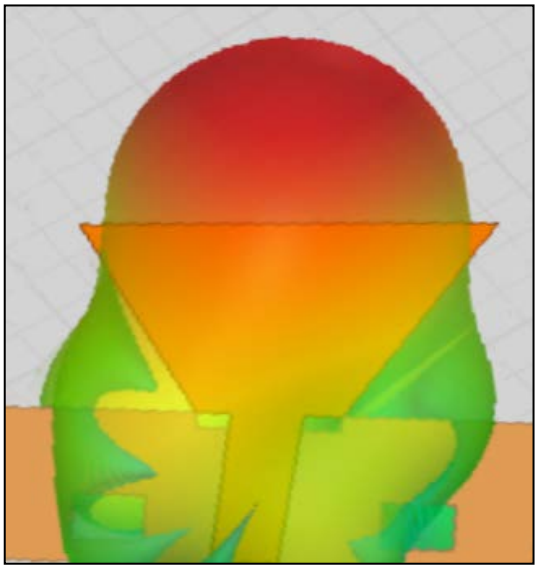


Figure 12: 3D Radiation pattern

C. Gain

Gain result is given in Figure 13 according to θ=90° and φ=10°, where θ is measured from z-axis to x-y plane and φ is measured from x axis to y-z plane.

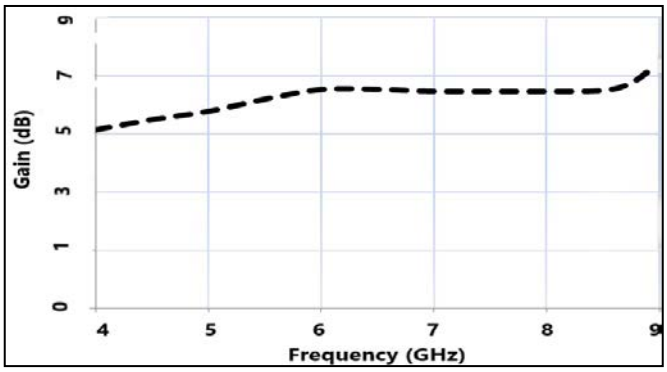


Figure 13: Gain of the antenna for frequency of interest

D. Depth Resolution

In the radar systems, getting a sufficient range resolution is one of the most important requirements. In order to obtain it, the system needs high bandwidth. In a radar-based MI system, depth resolution is as given in Equation [25]:

$$\delta_d = \frac{c}{2B\sqrt{\varepsilon_r}}$$

where **c** is the speed of the light, ε_r is the permittivity and B is the bandwidth of the system. As an example; for the radar-based microwave breast cancer imaging system we take the relative permittivity as 9 [26], since the permittivity of a breast fat is equal to this value. Therefore, the frequency band that we have obtained (4-9 GHz) gives a depth resolution of 10 milimeter corresponding to B=5 GHz and ε_r=9. This result is suitable for the system.

IV. CONCLUSION

In this paper, firstly the importance of using radar-based applications, especially MI has been mentioned briefly. Then, the properties of the radiating and receiving antenna that is the key component of the system has been explained as follows: Good impedance matching, high directivity, narrow beamwidth and small size are desirable to achieve a good MI and radar system. In order to provide antenna, first of all a conventional compact-sized rectangular patch antenna is designed, then various optimizations are performed on that antenna by using HFSS software. One of these modifications is creating partial ground plane to obtain planar monopole antenna instead of conventional microstrip antenna. Then, etching slots and notch on the ground plane is used to improve return loss. Another technique used in the process is modifying the square radiating patch to increase directivity. After all modifications completed, the proposed antenna has been fabricated and tested with VNA device. According to the return loss results, the antenna has bandwidth from 4 GHz to 9 GHz for |S11|<10 dB. Also, the proposed antenna shows stable and directional radiation patterns throughout the frequency range. 3D-polar plot of pattern and gain of the antenna have been shown clearly. Finally, depth resolution has been calculated and explained briefly. All of the results show that the designed antenna can be used in high range radar applications and is a good candidate for MI applications. In the future works, one can use this antenna in the MI and radar systems.

ACKNOWLEDGMENT

This research was supported by the Researching Projects Committee of the University of Dicle (DUBAPK) with the project number 14–MF–71. We are grateful to DUBAPK for financial assistance.

REFERENCES

[1] E. A. Jiya, N. S. N. Anwar and M. Z. Abdullah, "Detection of cracks in concrete structure using microwave imaging technique," *International Journal of Microwave Science and Technology*, vol. 2016, 2016. doi:10.1155/2016/3195716

[2] S. Almazroui, W. Wang and G. Zhang, *Microwave imaging for security applications*. Lambert Academic Publishing, 2016.

[3] E. A. Aydin and M. K. Keles, "Breast cancer detection using K-nearest neighbors data mining method obtained from the bow-tie antenna dataset," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 27(6), 2017. doi:10.1002/mmce.21096

[4] C. Ozdemir, S. Lim and H. Ling, "A synthetic-aperture algorithm for ground-penetrating radar imaging," *Microwave and Optical Technology Letters*, vol. 42(5), pp.412–414, 2004.

[5] M. Klemm, I. Craddock, J. Leendertz, A. Preece and R. Benjamin, "Radar-based breast cancer detection using a hemispherical antenna array - experimental results," *IEEE Transactions on Antennas and Propagation*, vol. 57, pp. 1692–1704, 2009.

[6] The Federal Communications Commission. *Revision of Part 15 of the Commission's Rules Regarding Ultra Wideband Transmission Systems; First Report and Order*, FCC 02–48, 22 April 2002, Washington, DC, USA.

[7] M. Abbak, "Antenna and measurement system for microwave imaging of breast tumors," Ph.D. thesis, Informatics Institute, Istanbul Technical University, 2015.

[8] A. R. Celik, M. B. Kurt, "Development of an ultra-wideband, stable and high-directive monopole disc antenna for radar-based microwave imaging of breast cancer," *Journal of Microwave Power and Electromagnetic Energy*, 2018. Doi: 10.1080/08327823.2018.1458692

[9] I. Unal, B. Türetken, and C. Canbay, "Spherical conformal bowtie antenna for ultrawide band microwave imaging of breast cancer tumor," *Applied Computational Electromagnetics Society Journal, 29*, pp. 124–133, 2008.

[10] C. Uyanik, M. Ö. Yildiz, S. Dogu, A. O. Ertay, I. Akduman, "Log-periodic dipole array antenna design for microwave imaging applications," in *Conf. Rec.* 2016, *Electrical, Electronics and Biomedical Engineering (ELECO)*.

[11] F. Altuntas, I. Develi, M. Turkmen, "Design of microstrip patch antenna with diamond shaped tuning slots for UWB communications," in *Conf. Rec.* 2016, *24th Signal Processing and Communication Application Conference*, doi: 10.1109/SIU.2016.7495990

[12] J. J. Golezani, M. Abbak and I. Akduman, "A novel compact wideband directional monopole antenna for use in radar applications," In *Conf. Rec. 2012 International Symposium on Antenna Technology and Applied Electromagnetics*, pp. 25–28.

[13] R. Kumar and G. Surushe, "Design of microstrip-fed printed UWB diversity antenna with tee crossed shaped structure," *Engineering Science and Technology, an International Journal*, 19(2), pp. 946–955, 2016. Doi: 10.1016/j.jestch.2015.10.006

[14] C. A. Balanis, Circular Patch Antenna Resonant Mode. In Antenna Theory: Analysis and Design, 3rd ed.; Danielle Lacourciere, Eds; John Wiley & Sons: New Jersey, USA, 2015.

[15] O. Haraz and A. R. Sebak, "UWB Antennas for wireless applications" *Advancement in Microstrip Antennas with Recent Applications;* IntechOpen: Rijeka, Croatia, European Union, 2013; 10.5772/51403

[16] A. R. Celik and M. B. Kurt, "Development of a novel ultra-wideband, stable and high directive monopole disc antenna for radar-based microwave imaging of breast cancer", *Journal of Microwave Power and Electromagnetic Energy*, doi: 10.1080/08327823.2018.1458692

[17] G. Kumar and K. P. Ray, "Broadband Microstrip Antennas", Artech House: Norwood MA, USA, 2003,

[18] Y. Zhang, "Microwave imaging for ultra-wideband antenna based cancer detection. PhD thesis, The University of Edinburgh, South Bridge, Edinburgh, Scotland, July 2014.

[19] D. Guha, M. Biswas and Y. M. M. Antar, "Microstrip patch antenna with defected ground structure for cross polarization suppression," *IEEE Antennas and Wireless Propagation Letters* , 4, 455–458, 2005. doi:10.1109/LAWP.2005.860211.

[20] J. Dong, Q. Li and L. Deng,, "Compact planar ultrawideband antennas with 3.5/5.2/5.8 ghz triple band-notched characteristics for internet of things applications. *Sensors*, *17*, 2017. doi: 10.3390/s17020349

[21] M. H. Bah, J. Hong, D. A. Jamro, "Ground Slotted Monopole Antenna Design for Microwave Breast Cancer Detection Based on Time Reversal MUSIC *Progress In Electromagnetics Research C, 59*, 117–126, 2015, . 10.2528/PIERC1508290.

[22] Z. Wani and D. Kumar, "Dual-band-notched Antenna for UWB MIMO applications," *International Journal of Microwave and Wireless Technologies*, *35*, 1-6, 2015. doi: 10.1017/S175907871500152X

[23] L. Liu, S. W. Cheung, and T. I. Yuk, "Bandwidth improvements using ground slots for compact uwb microstrip-fed antennas" in *Conf. Rec* 2011, Progress In Electromagnetics Research Symposium Proceedings (PIERS), , pp. 1420–1423, Suzhou, China.

[24] J. J. Golezani, "Directional wide band printed monopole antenna for use in microwave breast cancer imaging. M.Sc. thesis, Istanbul Technical University, Maslak, Istanbul, June 2012.

[25] X. Liu, X. Xiao, Z. Fan, and J. Yu, "Study on the imaging resolution of ultrawideband microwave imaging for breast cancer detection in *Conf. Rec.* 2009, 3rd International Conference on Bioinformatics and Biomedical Engineering, pp. 8–11, Beijing, China.

[26] Y. Xie, B. Guo, J. Li, P. Stoica, "Novel multistatic adaptive microwave imaging methods for early breast cancer detection," detection in *Conf. Rec.* 2006 Journal on Advances in Signal Processing, 1–13, doi: 10.1155/ASP/2006/91961

VisionDetector: A Low-Cost Mobile Visual Impairment Detection System

Y. ORTAKCI[1], A.E. TOKSOZ[2], B. KESKİN[3]

[1] Karabuk University, Karabuk/Turkey, yasinortakci@karabuk.edu.tr
[2]Karabuk University, Karabuk/Turkey, alperen.toksoz@hotmail.com
[3]Karabuk University, Karabuk/Turkey, burakkeskin2961@gmail.com

***Abstract* – Nowadays, many mobile applications have proposed to provide healthcare service. One of the most important healthcare problem all over the world is failure of the early detection and treatment of visual impairment. In this study, we have developed VisualDetector, a low-cost, user friendly and mobile visual impairment detection system which has two components. First component is an integrated circuit attached to the smartphone that measures the distance between user and smartphone, then, sends the measured distance to the smartphone via Bluetooth. Second component of VisualDetector is android-based mobile application that interacts with the user by executing vision tests. VisionDetector can detect visual acuity and colorblindness problems of the user without requiring to go to an eye clinic. Thus, early detection of visual impairment can be possible, and the expense of treatment will be decreased. The results indicate that our proposed system has a potential to use as an early detection tool for visual impairment.**

***Keywords* – Visual acuity, colorblindness, mobile, vision test, visual impairment.**

I. INTRODUCTION

IN the past, it was quite difficult to get a healthcare services without going to a medical unit. However, today, there exist many mobile applications providing the users healthcare services such as offering a diet [1], breathing exercises [2], estimating blood glucose [3], detecting Parkinson disease [4], navigating the visually impaired people [5] etc. These applications facilitate the detection and control of illnesses and gives feedback to the users or directs them to a medical unit. Thus, healthcare workload of medical units will be decreased.

One of the most important healthcare problem all over the world is deficient number of ophthalmologist and expensiveness of the devices that detect the visual impairment. For instance, in Kenya, only 86 ophthalmologists must take care of approximately 40 million visually impaired patients [6]. In addition, USA spent 7 billion dollars to detect the visual impairment of individuals under 40 years age in 2012 [7]. Detection of visual impairment and decreasing its cost are still important issues of all the countries. According to the result of a research held in 2010, World Health Organization estimates that there exist 285 million people who are visually impaired and %80 of these people can be cured [8].

Visual acuity (VA) is an important parameter in ophthalmologic examination and it should be controlled periodically since it sometimes decreases in time with different reasons such as elderliness [9], diabetes [10], etc. But nevertheless, existing analog methods for measuring VA are difficult to implement or quite expensive and require an ophthalmologic intervention. Two devices are common to measure VA; former is light-box hung on a wall and latter is projector which is very expensive. Both devices only exist in hospital and people do not use them in their daily life.

In this study, we have developed VisionDetector, an Android based mobile systemto detect VA and colorblindness problems of the user without no need to go to a hospital. Users do not need any ophthalmologist or professional support while they are running VisionDetector in their smartphones. The application displays random alphabetic characters on the screen and asks user to vocalize them. It dynamically sets the character size in Snellen's chart according to the distance between smartphone and user's head by using a distance sensor integrated to the smartphone. At the end of the test, the application calculates a rate for the user and if the rate is smaller than the threshold, the user is advised to consult an ophthalmologist. In this way, people can perform basic eye-sight test themselves easily on the smartphone without going to an eye clinic and check periodically whether they have a visual impairment. Thus, workload of ophthalmologists and eye clinics will be decreased and moreover, eye-sight problems can be detected and treated early.

II. RELATED WORK

Vision test are generally done between 40cm and 6m distance as near vision and far vision, respectively. Different vision tests are required to apply to the patient according to the size of test room [11].While Zhang et al. apply a vision test in 3m [12], Lin et al. apply another vision test in 2.5m [13]. Both test can be done in a fixed distance and changing the distance causes errors in vision test. Changing distance is no longer a problem, if the size of Snellen's chart that is used in vision test to measure VA can be changed dynamically. Thus, different size text room will not prevent to apply a vision test.

Huang et al. studied on a method about detection of visual impairment by projecting the characters on the wall [14]. The tester voices the character on the wall and the system checks whether the characters are correct respectively, then finally,

evaluates the eyesight of the tester. The portability of the system is low since it has a detailed devices group.

Black et al. investigated the applicability of the vision tests on iPad devices and inferred that a steady vision test could be performed on iPad devices as long as the brightness problem was solved [15]. Zhang et al. developed a mobile application to detect VA without using projector or lightbox for iPad devices [12]. They performed Tumbling E eye charts to detect VA so that user does not summarize the character order in the test. However, the users cannot perform the test by themselves and they can succeed it in the control of an ophthalmologist. Lin et al. [13] also developed a digital VA detection system that can be controlled overseas and running in client/server model.

Unlike the other application, VisionDetector cannot only detect VA problems but also colorblindness. The distance between smartphone and user is no longer problem for vision test, since we have integrated a distance sensor to the smartphone measuring the distance between user and smartphone to set the size of Snellen's chart dynamically. Also, the characters in the test are created randomly to prevent user to summarize a default pattern. Besides, VisionDetector does not require a detailed and expensive equipment, and even an internet connection to measure VA.

III. VISIONDETECTOR

In this study, an Android based mobile system, including a mobile application and an integrated circuit attached to the smartphone, has been proposed in order to detect the visually impairment of the user. The integrated circuit consists of an Arduino Uno card that has an ultrasonic sensor and a Bluetooth module. The sensor measures the distance between user's head and the smartphone which s/he holds, and Bluetooth module on Arduino Uno sent the measured distance to the mobile application on the smartphone. According to the distance, the mobile application resizes the character in the Snellen's chart on the screen dynamically. In a vision test, the mobile application displays random alphabetic characters, then, user vocalizes the characters which s/he has seen on the screen. The application also recognizes the sound of user and directs him/her according to his/her answer as seen in Figure 1.

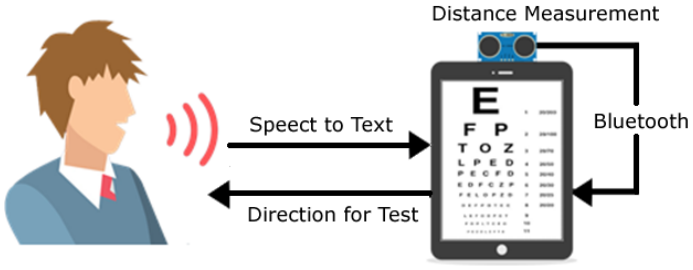


Figure 1. General overview of VisionDetector

A. Distance Measurement

HC-SR04 sensor which emits ultrasonic waves is used in the system to measure the distances (Figure2). When the user runs the system, the sensor emits the wave, then, the emitted wave hits the user's head and returns to the sensor. According to the elapsed time, the distance can be calculated. The range of HC-SR04 sensor is between 2cm and 4m which is quite adequate to measure the distance between user and the smartphone which s/he holds. HC-SR04 sensor can emit waves in every 500 milliseconds periodically, thus, provides a real-time distance measurement.

Figure 2. HC-SR04 ultrasonic sensor [16]

B. Sensor-Smartphone Communication

HC-05 Bluetooth module which was integrated to Arduino Uno has been used to provide the communication between HC-SR04 sensor and user's smartphone. HC-05 Bluetooth module supports Bluetooth 2.0 with 2.4 GHz communication frequency and has approximately 10-meter communication range. The baud rate of both Bluetooth module and the sensor should have the same value for synchronization and have been set as 9600 Bd in VisionDetector. The android application receives the measured distance data via Bluetooth adapter class.

C. Measuring Vision Acuity

We have used Snellen's chart VisionDetector which displays the letters on the screen as pyramidal to detect VA (Figure 3).

Figure 3: Snellen's chart in VisionDetector

When the application is started, first, the letters are created randomly and displayed on the screen in successive nine lines as pyramidal shape [17]. Then, the application calculates the letter size according to the distance measured by the sensor as shown in Equation 1 [18].

$$\text{Letter Size(meter)} = (\text{Tanjant } 5° \text{ x Distance in meter}) \text{ x } 0{,}8\overline{3} \quad (1)$$

Calculated letter size implies the size of the largest letter which is displayed in the uppermost line of Snellen's chart. The letter sizes are being decreased from top to bottom periodically in each successive lines of Snellen's chart. Each line's letter size is calculated by dividing its upper line's letter size to 1.26 [17]. The letter size of the undermost line which can be successfully read by the user is considered as reference letter size in the calculation of VA. Accordingly, VA can be calculated by dividing the test distance to the corresponding reference letter size as shown in Equation 2 [11].

$$Vision\ Acuity(VA) = \frac{Test\ Distance\ (in\ meter)}{Reference\ Letter\ Size\ (in\ M-Unit)} \quad (2)$$

$$where\ 1\ M-Unit = 1.454\ milimeter$$

Calculated VA can be evaluated as follow [11]:

IF ***VA ≥ 1.0****, THEN user is expected to have a healthy eye*
IF ***0.5≤VA≤ 1.0****, THEN user is expected to have a small scale visual impairment*
IF ***VA≤ 0.5****, THEN user is expected to have a serious visual impairment*

During the VA test, the application informs and guides the user with the messages as shown in Figure 4 and VA examination should be done for both eyes separately. On the other hand, letters are created randomly in each run to prevent users to summarize the order of the letters.

Figure 4. Information and guidance messages

D. Color Blindness Detection

Color blindness is a kind of visual impairment which may be congenital or occurs over time based upon the lack of the pigments in the photoreceptor cells of retina. Our system can detect protanopia (defectiveness of red cone cells), deuteranopia (defectiveness of green cone cells), tritanopia (defectiveness of blue cone cells) color blindness disease [19]. Figure 5 compares normal color spectrum with color perception spectrum of the people with color blindness.

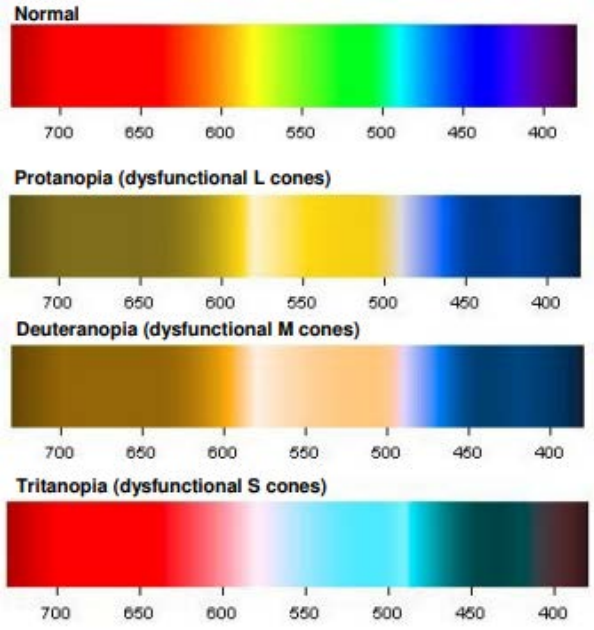


Figure 5: Normal and color blind color spectrums [20].

We have applied Ishihara test [21], which is a color perception test displaying a group of colored dotted plates on the screen, each one includes a colored number as shown in Figure 6, to detect color blindness in our study. Ishihara test includes a group steps at which user is expected to recognize the number on the plates. User should proceed to the next step by vocalizing the number that s/he have seen in the plate on the smartphone. Our application can detect whether the user has a color blindness and the type of color blindness if there exists.

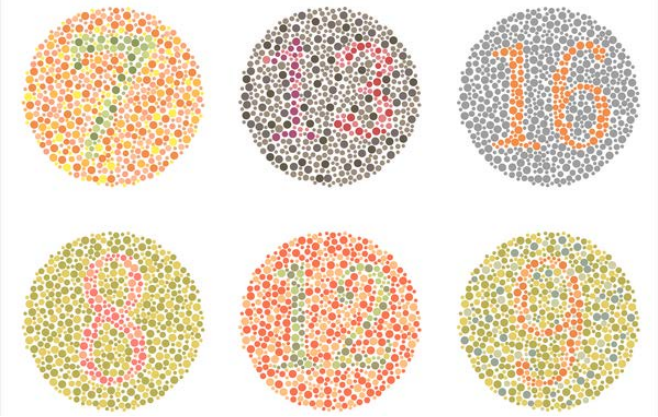


Figure 6: Ishihira plates [22]

IV. Experimental Results And Discussion

Our application has been tested on 20 users (7 women and 13 men) whose ages are changing in the range of 18 and 73 and the obtained result are shown in Table 1.The eye numbers and astigmatism existences of all test users which had been measured by an ophthalmologist before were known in advance.

Table 1. The results of our vision test

User	Age	Gender	VA of right eye	Right eye number	VA of left eye	Left eye number	Astigmatism	Color Blindness
User-1	18	F	0.50	1.75	0.63	1.0	-	-
User-2	20	M	0.2	4.0	0.25	3.0	-	-
User-3	21	M	0.63	0.5	**1.00**	**0.78**	+	-
User-4	22	M	0.10	7.0	0.125	5.0	+	-
User-5	23	F	0.78	0.75	0.63	1.25	-	-
User-6	23	M	**1.0**	**0.5**	0.78	1.0	-	-
User-7	24	M	1.00	0.0	**1.00**	**0.5**	-	-
User-8	25	F	0.50	1.5	0.40	2.25	-	-
User-9	29	F	0.63	1.0	0.40	2.0	-	Protanopia
User-10	30	M	1.00	0.0	**1.00**	**0.25**	-	-
User-11	31	M	0.63	0.75	0.78	0.25	-	-
User-12	35	M	0.20	3.0	0.50	1.5	+	-
User-13	38	M	0.25	0.75	0.32	0.5	+	-
User-14	41	F	0.63	0.25	0.50	0.75	-	-
User-15	44	M	0.52	1.0	0.49	1.5	-	-
User-16	45	F	0.10	-3.5	0.10	-2.5	+	-
User-17	55	M	0.47	1.5	0.54	1.25	-	-
User-18	60	M	0.32	3.5	0.125	5.0	+	-
User-19	62	F	0.40	1.5	0.78	0.5	+	-
User-20	73	M	0.32	2.25	0.20	3.0	-	-

The results indicate that our application can detect a visual impairment by considering VA parameter with %90 accuracy rate. %10 error rate is because of the fact that even if VA values for left eye of user-1, right eye of user-6, left eye of user-7, and left eye of user-10 are calculated as 1.0, they have a VA deficiency. However, our application succeeded to detect visual impairments of 36 eyes for 20 users and a Protanopia color blindness.

The obtained results also show that while VA values are independent of gender parameter, there exists direct relation between age, astigmatism and VA value of a user. Namely, VA value of a user, who has not an astigmatism problem, is higher than a user, who has astigmatism and same eye number with former user. Besides, VA value of a user under 35 years old is higher than a user over 35 years old with same eye number.

Overall, if the eye number of a user is lower than 1.5, VA value is calculated in the range of (0.5-1.0) that implies the user has a small scale visual impairment. On the other hand, if the eye number of a user is larger than 1.5, VA value is calculated under 0.5 that implies the user has a serious visual impairment.

V. CONCLUSIONS

In this study, we have developed a low-cost mobile system which can detect VA and color blindness of the user on his/her smartphone with the support of a small size integrated circuit measuring distance. Thus, visual impairment can be early detected by the user himself without going to of an ophthalmologist. Thereby, the workload of the eye clinics and the ophthalmic expense of a society will be decreased.

As a future work, the vision test results of the users can be collected in remote server for data mining process. Even, the mobile application can be extended to acquire the data of the user's other health problems. In addition, it is also possible to diagnose the other diseases by extending VisionDetecotor. Thus, revealing the ties between diseases and production of new findings about them can be done in a society.

REFERENCES

[1] Escárcega-Centeno, D., et al., *Augmented-Sugar Intake: A Mobile Application to Teach Population about Sugar Sweetened Beverages*. Procedia Computer Science, 2015. **75**: p. 275-280.

[2] Chittaro, L. and R. Sioni, *Evaluating mobile apps for breathing training: The effectiveness of visualization*. Computers in Human Behavior, 2014. **40**: p. 56-63.

[3] Li, J. and C. Fernando, *Smartphone-based personalized blood glucose prediction*. ICT Express, 2016. **2**(4): p. 150-154.

[4] Lan, K.-C. and W.-Y. Shih, *Early Diagnosis of Parkinson's Disease Using a Smartphone*. Procedia Computer Science, 2014. **34**: p. 305-312.

[5] Cecílio, J., K. Duarte, and P. Furtado, *BlindeDroid: An information tracking system for real-time guiding of blind people*. Procedia Computer Science, 2015. **52**: p. 113-120.

[6] Giardini, M.E., *The Portable Eye Examination Kit: Mobile phones can screen for eye disease in low-resource settings*. IEEE pulse, 2015. **6**(6): p. 15-17.

[7] Wittenborn, J.S., et al., *The economic burden of vision loss and eye disorders among the United States population younger than 40 years*. Ophthalmology, 2013. **120**(9): p. 1728-1735.

[8] Organization, W.H., *Universal eye health: a global action plan 2014-2019*. 2013.

[9] Machan, C.M., et al., *Eye examinations improve visual acuity across ages*. Canadian Journal of Ophthalmology, 2013. **48**(4): p. 286-291.

[10] Moss, S.E., R. Klein, and B.E. Klein, *The incidence of vision loss in a diabetic population*. Ophthalmology, 1988. **95**(10): p. 1340-1348.

[11] Universale, C.O., *Visual acuity measurement standard*. Visual Functions Committee, 1984.

[12] Zhang, Z.-t., et al., *A pilot trial of the iPad tablet computer as a portable device for visual acuity testing*. Journal of telemedicine and telecare, 2013. **19**(1): p. 55-59.

[13] Lin, Y.L., N.-F. Law, and C.-w. Do. *Portable vision screenings system*. in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017*. 2017. IEEE.

[14] Huang, Y.-Y., S. Ropelato, and M. Menozzi. *Fully automatic and computerized self-vision-screening system: Vision at own—An E-health service of self vision examination and screening*. in *Digital Information Processing and Communications (ICDIPC), 2015 Fifth International Conference on*. 2015. IEEE.

[15] Black, J., et al., *An assessment of the iPad as a testing platform for distance visual acuity in adults*. BMJ open, 2013. **3**(6): p. e002730.

[16] Alexander Bernard, Timothy Duggan, and C. Welling. *Autonomous Quadcopter for Target Tracking*. 2017 [cited 2018 March, 23]; Available from: https://courses.cit.cornell.edu.

[17] Eğrilmez, S., et al., *Uluslararası standartlara uygun bir Türkçe yakın okuma eşeli.* T Oft Gaz, 2004. **34**(6): p. 404-412.

[18] Öztürk, B., E. Şener, and A. Sanaç, *Görme Keskinliğinin Klinik Değerlendirilmesi.* T Oft Gaz, 2001. **31**: p. 166-172.

[19] Semary, N.A. and H.M. Marey. *An evaluation of computer based color vision deficiency test: Egypt as a study case*. in *Engineering and Technology (ICET), 2014 International Conference on*. 2014. IEEE.

[20] Wolfmaier, T.G., *Designing for the color-challenged: A challenge*. Human-computer Interaction Resources Network, 1999.

[21] Ishihara, S., *Tests for Color Blindness*. American Journal of Ophthalmology, 1918. **1**(5): p. 376.

[22] National Library of Medicine (US), Color Vision Deficiency [Internet]. Bethesda (MD): The Library; March 20, 2018 [cited 2018 March 27]. Available from: https://ghr.nlm.nih.gov/condition/color-vision-deficiency.

Simulation Measurements of Electromagnetic Field Values for Detection of the Breast Cancer

A.R.ÇELİK[1], M.B.KURT[1] and S.HELHEL[2]
[1] Dicle University, Diyarbakir/Turkey, eeealicelik@gmail.com
[1]Dicle University, Diyarbakir/Turkey, bkurt@dicle.edu.tr
[2]Akdeniz University, Antalya/Turkey, helhel@akdeniz.edu.tr

***Abstract* - Detection of the breast cancer at the early stage has gained much attention over last decades. In this paper, a simulation study of a radar-based ultra-wideband microwave system is presented to detect breast cancer tumors. The main principle of this technique is based on the significant difference in the dielectric properties of malignant breast tumors and normal breast tissue in the microwave frequencies. The tumor's electrical properties, in particular conduction and specific absorption rate (SAR), change significantly from those of healthy biological tissue when exposed to microwave radiation. In the measurements, a simple planar breast phantom that consisted low dielectric constant material to represent the fat tissue and high dielectric constant material to represent the tumor is used. An ultra-wideband and high gain antenna is used to measure electromagnetic field data for tumorous and non-tumorous breast tissue. According to the obtained results, the used antenna and microwave system are successful for detecting the breast cancer tumor. Measurement system is developed by using High Frequency Structural Simulator (HFSS) software. Antenna design parameters, properties of the breast phantom, analysis and measurement results are demonstrated and explained clearly in the paper.**

***Keywords* – Breast cancer, Microwave techniques, HFSS**

I. INTRODUCTION

Statistics clearly show the importance of breast cancer disease and its impact on the health. The early detection of the disease could provide to decrease the higher percentages of death cases [1]. It is known that the malignant cells can multiply uncontrolled and spread to other part of the body. So, detecting the malignant tumor at the early stage has a critical role for the treatment of disease. The requirements for an ideal breast cancer detection system can be summarized as low health hazard, sensitivity to malignant tumor, determining the disease at a medicable stage, screening as fast as possible and involving minimal discomfort [2]. X-ray mammography, ultrasound technique, digital tomosynthesis and magnetic resonance imaging are some of the breast cancer detecting and imaging techniques. However, these screening methods don't meet the ideal requirements [3]. For this reason, the alternative methods are actively searching by researchers. One of the emerging electromagnetic techniques is using microwaves in the breast cancer detection systems.

The main principle of this technique is based on the significant difference in electrical properties of malignant tumors and breast tissue in the microwave frequencies [4]. Microwave signals are the non-ionizing form of the electromagnetic waves and there is no risk to health hazard when used at low levels. Microwave detection is less expensive and safer than other modalities of detection. Microwave radiometry, microwave tomography, microwave microscopy, microwave-induced thermal acoustic imaging, ultrawideband (UWB) radar imaging are some of the microwave imaging methods. Among these systems, radar-based imaging has been leading and successful technique in recent years [5-7].

II. ULTRAWIDE BAND RADAR-BASED MICROWAVE IMAGING

Hagness et al. proposed the radar-based UWB microwave imaging [8] that uses an UWB pulse which involves low to high frequencies. The lower frequency band ensures enough penetration depth, and the higher band provides sufficient resolution for the created images. Thus, deeply buried and small size tumor can be detected based on the lower and higher frequency of the UWB bands [9].

There are several measurement configurations such as monostatic, bi-static and multi-static for using in this system. If the same antenna sends and receives the microwave signals, this measurement is named monostatic. If the system has two antennas and the target material is positioned between the receiving and transmitting antennas, this is the bi-static system. Finally, in multi-static configuration, antenna arrays are used. Examples of these configurations are given in Figure 1-3.

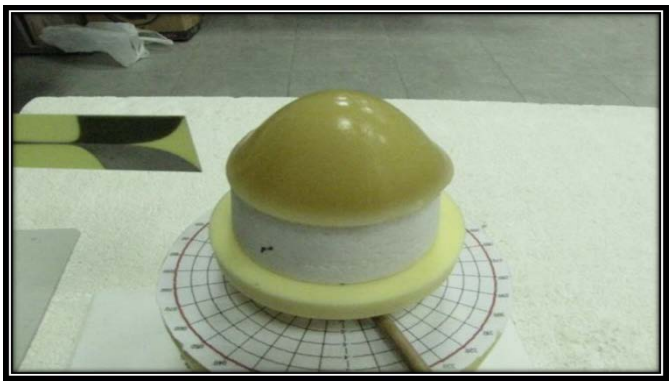


Figure 1: Example of a monostatic measurement [10]

Figure 2: Example of a bistatic measurement [10]

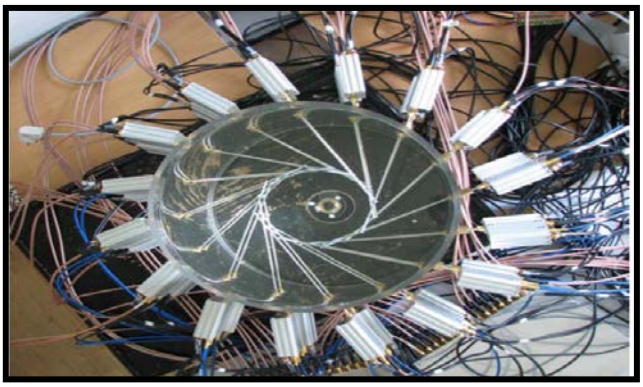


Figure 3: Example of a multistatic measurement [11]

In this study, monostatic radar-based measurement system is preferred and it is developed by using High Frequency Structural Simulator (HFSS) software. In the simulation measurement, a planar monopole antenna which was proposed in [12] is used to perform the task of both sending and receiving the microwave signals.

III. Simulation Measurement

A. HFSS Simulator Software

For simulation findings, HFSS program [13] based on full-wave finite elements method and widely used in the analysis of electromagnetic structures is utilized. HFSS offers state-of the-art solver technologies based on finite element, integral equation, asymptotic and advanced hybrid methods to solve a wide range of microwave, RF and high-speed digital applications. For the measurements; firstly the antenna and breast phantom with tumor are created then the results are obtained in HFSS.

B. UWB Directional Antenna

The key component of the radar-based UWB microwave imaging system is the antenna that is used to radiate and receive the UWB pulses. This system requires UWB, compact, stable and directive antennas as their radiating and receiving sensors. In the study of [12], a modified square planar monopole antenna having broad frequency range and high gain properties has been proposed for using in microwave imaging. The design parameters and front view of the fabricated antenna are shown in Figure 4. The return loss and radiation pattern results of the antenna are given in Figure 5.

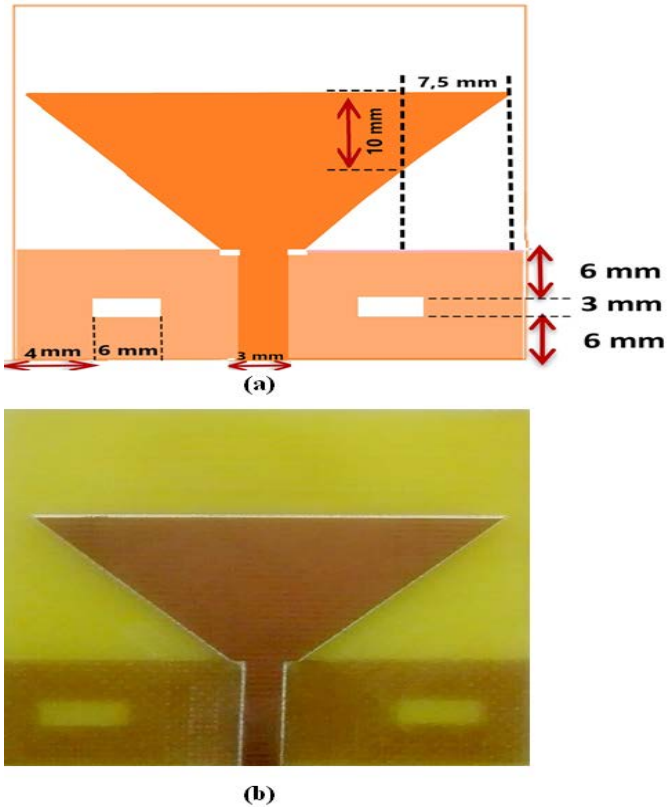


Figure 4: a) Design parameters of the used antenna, b)Fabricated antenna [12]

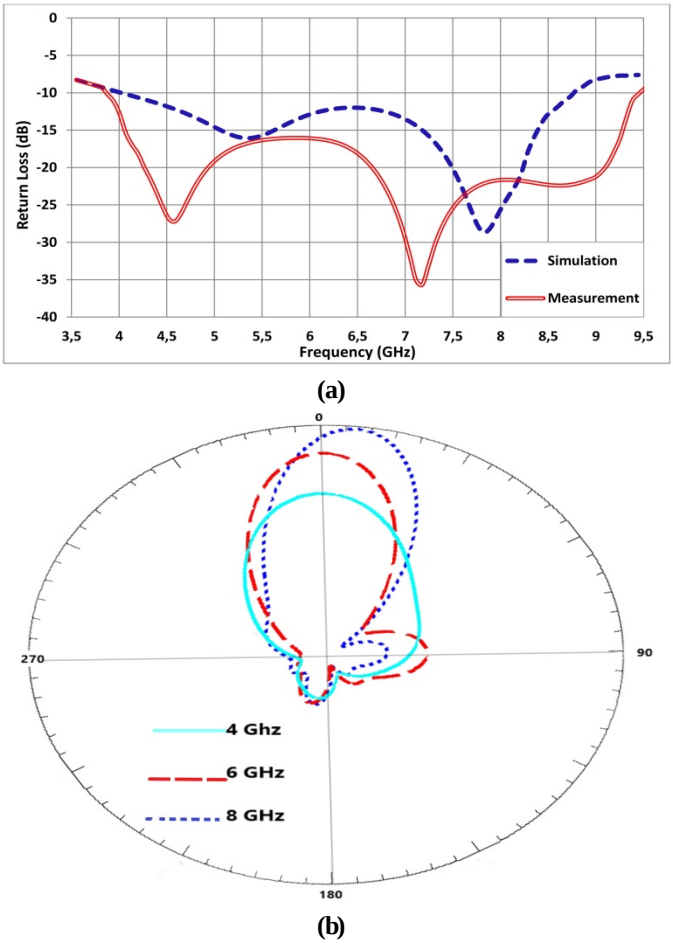


Figure 5: a) Return loss b) Radiation pattern results of the antenna

C. Monostatic Measurement System

For the measurements, a simple planar breast phantom structure with a 150 mm x 50 mm x 70 mm is design to simulate breast tissue layer. Dielectric constants of the skin, breast and tumor issues are 50, 36 and 9, respectively [14]. The skin layer of breast is taken as 5 mm. The tumor is a spherical of 4-mm radius that is inserted inside the breast tissue at different positions. The view of the system when tumor is at the deep point is given in Figure 6 as an example. It is seen from this figure that the phantom is placed according to the direction of main radiation lobe. In the experiment, three different conditions are examined. These cases are: non-tumorous breast tissue, tumorous tissue when tumor is close to the antenna, and tumorous tissue when tumor is away from the antenna.

IV. RESULTS AND DISCUSSIONS

According the simulation results, return loss (S11) vs. frequency with and without tumor in the phantom are given in Fig. 7. Since S11 and the reflection coefficient (Γ) are related to each other according to the (1),

$$S_{11}(dB) = 20\log_{10}|\Gamma| \quad (1)$$

the smaller the magnitude of S11, the larger the reflection becomes. Furthermore, the values of the electric field, magnetic field and flux density of the used antenna for different cases are given in Table 1 showing the values for without tumor, tumor is close to the antenna and tumor is away from the antenna. As a result, these difference values show that the antenna successfully detects the tumor and it can be detected easy when the tumor is close to the antenna.

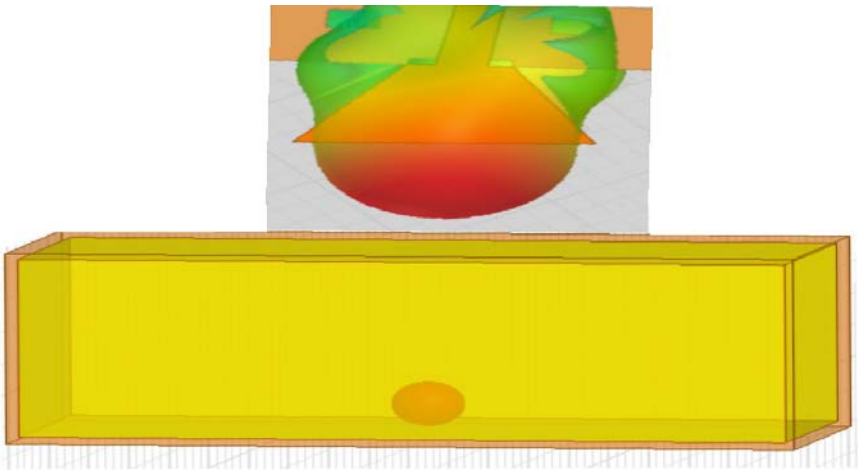


Figure 6: The view of the system when tumor is at the deep point

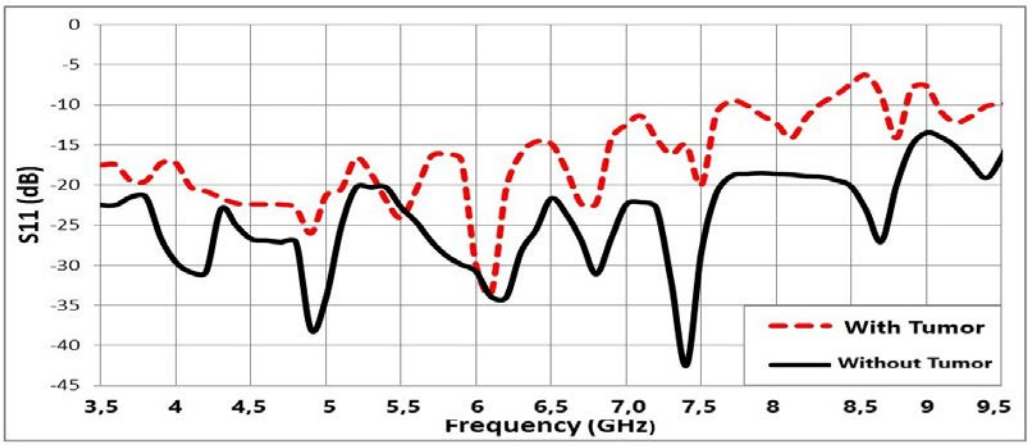


Figure 7: The return loss results when there is tumor and there is no tumor in the phantom

Table 1. The values of electric field, magnetic field and flux density of the proposed antenna

CASE	Max E Field (V/m)	Max H Field (mA/m)	Max J (A/m2)
Without Tumor	192.3	1.1392	43.075
The tumor is away from the antenna	223.4	1.0368	43.768
The tumor is close to the antenna	298.9	1.3845	46.512

V. CONCLUSION

This paper has reported the capability of the radar-based microwave detection system to detect breast tumor by using HFSS simulator software. In order to achieve this purpose, a simple simulation system including a breast phantom which has the similar electrical properties with breast skin, fat and tumor. The simulation measurement has been made using compact and directional UWB planar monopole antenna. Based on the return loss results, it is concluded that the reflection coefficent increases when the antenna gets close to the tumor; it decreases when the antenna is away from the tumor. Also, the obtained electromagnetic field data is evaluated. The maximum electric field value of the non-tumorous breast tissue is 1924.3 mV/m, while it is 2983.9 mV/m in the near position and 2236.4 mV/m in the far position of the tumorous breast tissue. All of these results show that the radar-based microwave techniques for detecting the tumorous breast tissue is a good method, and the compact, UWB, directive antennas are capable of the detecting tumor.

REFERENCES

[1] World Health Organization. Available online: http://www.who.int/cancer/breast_cancer_awareness/en

[2] J. P. Stang, W. T. Joines, Q.H. Liu, G. A. Ybarra, R. T. George, M.Yuan, I. Leonhardt, "3D antenna array design measurement results and image chamber modeling," *in Conf. Rec*. Antennas and Propagation Society International Symposium, Embassy Suites Convention Ctr. North Charleston, SC, USA, 01-05 June 2009, pp. 3–6.

[3] L. Wang, "Early Diagnosis of Breast Cancer," *Sensors* 2017, 17, 1572,

[4] E. C. Fear, X. Li, S. C. Hagness, M. A. Stuchly, Confocal microwave imaging for breast cancer detection: localization of tumors in three dimensions , *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 812-822, 2002.

[5] M. Klemm, I. Craddock, J. Leendertz, A. Preece, and R. Benjamin, "Radar-based breast cancer detection using a hemispherical antenna array | Experimental results," *IEEE Transactions on Antennas and Propagation*, Vol. 57, 1692-1704, 2009.

[6] R.C. Conceicao, M. O. Halloran, E. Jones and M. Glavin, Investigation of classifers for early-stage breast cancer based on radar target signatures," *Progress In Electromagnetic Research*, Vol. 105, 295-311, 2010.

[7] J. M. Sill and E. C. Fear, "Tissue sensing adaptive radar for breast cancer detection - experimental investigation of simple tumor models," *IEEE Trans. Microw. Theory Tech.*, vol. 53, pp. 3312-3319, 2005.

[8] S.C. Hagness, A. Taflove, J. E.Bridges, "Three-dimensional FDTD analysis of a pulsed microwave confocal system for breast cancer detection: design of an antenna-array element," *IEEE Transactions on Antennas and Propagation* 1999, *47*, 783–791, doi: 10.1109/8.774131.

[9] X. Xiao, T. Kikkawa, "Influence of the organism interface on the breast cancer detection by UWB," *Applied Surface Science* 2008, *255*, 597–599, doi: 10.1016/j.apsusc.2008.06.167.

[10] I. Unal, B. Türetken, and C. Canbay, "Spherical conformal bowtie antenna for ultrawide band microwave ımaging of breast cancer tumor," *Applied Computational Electromagnetics Society Journal, 29*, pp. 124–133, 2008.

[11] I. Craddock, M. Klemm, J. Leendertz, A. Preece, and R. Benjamin, "An improved hemispeherical antenna array design for breast imaging," *in Conf. Rec*. The Second European Conference EuCAP, pp. 1-5, 2007.

[12] A. R. Celik, M. B. Kurt and S. Helhel, "Design and fabrication of high gain ultra-wideband antenna for microwave imaging and radar applications" Submitted for publication.

[13] Ansys HFSS. 2014. Ansys Corporation, Canonsburg, USA.

[14] N. Tavassolian, S. Nikolaou and M. M. Tentzeris, "A flexible UWB elliptical slot antenna with a tuning uneven U-shape stub on Lcp for microwave tumor detection", *Asia Pasific Microwave Conference, 2007.*

A Study on BCI Speller Design and Analysis of Signal Window Length

E. ERKAN[1] and M. AKBABA[2]

[1] Karabuk University, Karabuk/Turkey, erdem__erkan@hotmail.com
[2] Karabuk University, Karabuk/Turkey, makbaba@karabuk.edu.tr

Abstract **- Researchers have been studying to understand and classify biological signals for better diagnose diseases and developing assistive technologies. These technologies are sometimes making it possible to communicate in ALS (Amyotrophic lateral sclerosis) patients, sometimes possible to use our computer, faster and more efficient without using our muscular systems. The steady state visual evoked potential (SSVEP) approach currently provides the high performance and reliable communication for the implementation of these technologies. Performance is usually measured by Information Transfer Rate (ITR) and the most important factor affecting ITR is signal window length.**

In the presented paper a SSVEP based BCI (Brain Computer Interface) speller application is introduced and system performance is analyzed for different signal window lengths in experiments. The BCI speller has six box which has six letters in each box on the screen. The six letters in the selected box are distributed as one letter each box after the first selection by application. With the second selection, the letter which desired is displayed on the screen. The application contains Latin letters as well as Turkish letters. Experiments are performed on 3 healthy subjects. Subjects try to choose letter by focusing boxes which has flickering different frequencies. The minimum energy combination (MEC) method is applied to EEG segments that are different length in order to detect SSVEPs. The highest ITR value of 77.55 bit/min is obtained for subject 1 with 2 s signal window length. High accuracy and more useful a BCI system observed when system signal window length set 3 s.

Keywords **- Brain computer interface (BCI), Steady state visual evoked potential (SSVEP), Minimum energy combination (MEC)**

I. INTRODUCTION

Communications without use vocal pathways have been one of the main motivations behind brain computer interfaces. Some methods have shown positive results. SSVEP approach currently provides the high performance and reliable communication for the implementation of a non-invasive BCI systems but still the accuracy and robustness are big concerns. SSVEP based BCI system must reflect flickering lights stimulus at different frequencies to the user. The best response for these stimulus are obtained for stimulation frequencies between 5 and 20Hz [4, 5]. The success of the system is measured by ITR. ITR depends on three factors; speed, accuracy and the number of targets [3, 4, 5]. Reported average accuracy and ITR for SSVEP based BCI systems which has six target are 95.3 % and 58 ± 9.6 bit/min respectively [9]. Another study is reported that average accuracy and ITR are 96.79 ± 7.881 % and 61.70 ± 32.676 bit/min [5]. Another study is reported that ITR is 16.10 bit/min for elderly people, 27.36 bit/min for young group [7]. In another study [8] ITR is reported between 25.69 ± 21.37 bit/min and 39.87 ± 8.37 bit/min for SSVEP based BCI systems by multivariate synchronization index method (MSI).

In this study a non-invasive SSVEP based BCI system is designed, which has six targets. User can type words by choose letters with the system on computer screen. Three subjects participated in the study. Performance of the system are analyzed with different signal window lengths by using MEC methods. Optimum signal window length is detected for designed system. The general structure of the BCI system is given in Fig. 1.

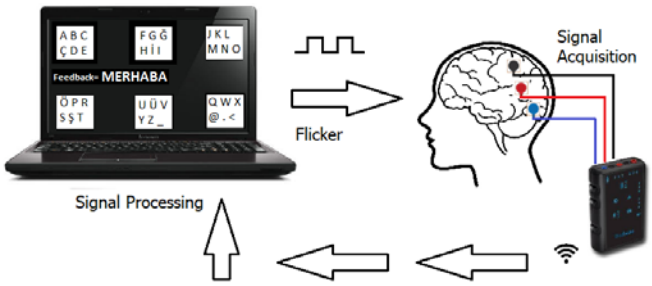


Figure 1: BCI system.

II. MATERIALS AND METHODS

A. Data Acquisition

The data are collected from three healthy subjects at multiple sessions including several trials each. The subjects mean age is 36 years. During the experiment, a subject tried to perform typing the words by focusing on six different square (100 x 100 pixel) which are stimulate by 6.66 Hz, 7.50 Hz, 8.57 Hz, 10 Hz, 12 Hz and 5.45 Hz on the LCD screen(1366 x 768 resolution, 60 Hz refresh rate). In the experiments a computer which has i7 2.8 GHz processors is used for Graphical User Interface (GUI) and signal processing. The computer is run with Windows 10 operating system and 4GB of RAM. Subjects are seated in front of the GUI screen at a distance of about 50 cm. The time series of the electrical brain activity is picked up during these trials

using 8 gold EEG electrodes which is placed on scalp. Electrodes are placed at predefined locations PZ, PO3, PO4, O1, O2, OZ, O9 and FP2. EEG conductive paste is applied between the electrodes and the scalp. Six electrodes are placed near the occipital region in accordance to international 10-20 system. Electrodes placements are shown in Fig. 2. EEG signals are acquired using a biosignal amplifier (Bioradio, Great Lakes NeuroTechnologies, USA). The sampling frequency is set to 500 Hz. The graphical user interfaces (GUI) are designed in the MATLAB 2015a by using Psychtoolbox. Position figures and tables at the top and bottom of each column if possible.

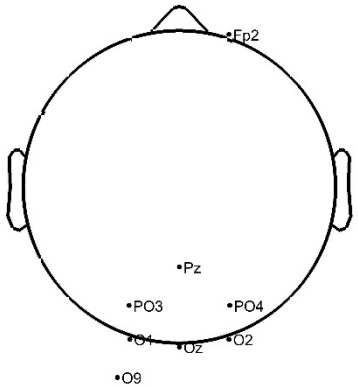


Figure 2: Electrodes placements.

B. Signal Preprocessing

Classification process is applied to raw EEG signals. Signal preprocessing steps are given in Fig. 3.

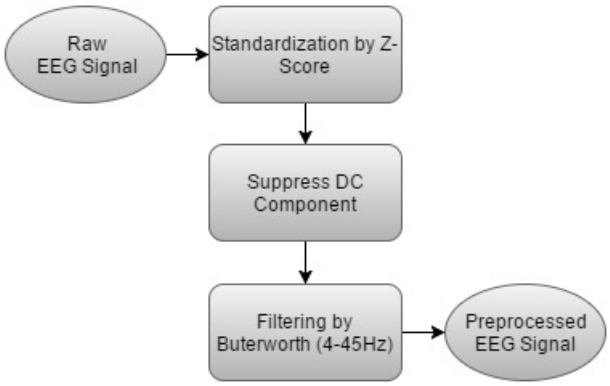


Figure 3: Signal Preprocessing.

Amplitude difference may arise from the result of physiological changes of subject and environmental changes. This difference negatively affects the classification. Therefore, EEG signals are scaled by z-score method, train by train for every channel. This method aims to improve the signal to noise ratio of the EEG signals. Z-score generally represents the distance from the average data. Z-score is described in Equation 1 where $\bar{x}$ represents the signal mean and σ standard deviation value [9].

$$Z = \frac{x - \bar{x}}{\sigma} \quad (1)$$

Standardization of data is obtained by calculating the z-scores of each signals with the help of Equation 1. The signals are filtered by 4-45 Hz Butterworth filter. Before the filtering signal DC component is suppressed.

C. SSVEP Detection

Minimum energy combination method (MEC) [5, 10] is used in SSVEP detection. The method generally consists three steps.

Three steps as follows:

- All information about the frequencies of interest should be extracted from recorded signals. Obtained signals include only information that is uninteresting. Therefore it can be considered as noise components belonging to the original signals.
- A linear combination is achieved, suppressing the noisy signals obtained in the first stage.
- Finally, we apply this linear combination to the original signals to produce low noise signals.

The voltage $y_i(t)$ between the electrode i and a reference electrode at time t, $y_i(t)$ is described in Equation 2 Where f represents stimulus frequency and N_h the number of considered harmonic.

$$y_i(t) = \sum_{k=1}^{N_h} \left(a_{i,k} Sin2\pi kft + b_{i,k} Cos2\pi kft \right) + E_{i,t} \quad (2)$$

The signal consists of two parts. The first part relevant to the SSVEP signal, which is composed of a number of sine and cosine functions with the stimulus frequency and a number of N_h harmonics. $a_{i,k}$ and $b_{i,k}$ represents sine and cosine terms amplitudes. The second part of the signal $E_{i,t}$ is the noise and all the information that are not relevant to the SSVEP signal. EEG signal for the electrode i in the time segment is given in Equation 3 and Equation 4 respectively where N_t represents samples of the signal i, X represents information matrix (size: N_t X $2N_h$) which contains sine and cosine components associated with the N_h harmonics and g_i (size: $2N_h$ X 1) represents amplitudes $a_{i,k}$ and $b_{i,k}$.

$$y_i = X g_i + E_i \quad (3)$$

$$\hat{Y} = Y - (X^T X)^{-1} X^T Y \quad (4)$$

Equation 3 can be generalized for N_y electrodes,

$$Y = XG + E \quad (5)$$

$Y = [y_1, \dots y_{N_y}]$ contains the sampled EEG signals from all the channels. Channel vector s is given in Equation 6 which is a linear combination of the electrode signals y_i where W is a vector of weights associated with the electrode signals. Equation 5 can be generalized for N_s channels in as Equation 6 where channel vector $S = [S_1, \dots S_{N_s}]$ and weight matrix $W =$

$[W_1, \dots W_{N_s}]$.

$$S = YW \tag{6}$$

At first stage, all information about the frequencies of interest should be extracted from recorded signals. Obtained signals include only information that is uninteresting. For this an orthogonal projection is used. This calculation is given in Equation 7.

$$\hat{Y} = Y - (X^T X)^{-1} X^T Y \tag{7}$$

Weight vector W which minimizes the energy of the signal $\hat{Y}$, can be calculate by the minimal eigenvalue λ_i and eigenvector v_i of the matrix $\hat{Y}^T\hat{Y}$. Weight vector matrix W is given in Equation 8 where eigenvalues in ascending order$(\lambda_1, \dots, \lambda_{N_s})$, the corresponding eigenvectors $(v_1, \dots, v_{N_s})$ and N_s is selected channel count which is used for classification.

$$W = [\frac{v_1}{\sqrt{\lambda_1}} \dots \frac{v_{N_s}}{\sqrt{\lambda_{N_s}}}] \tag{8}$$

A frequency power can be calculated by Equation 9.

$$\hat{P} = \frac{1}{N_s N_h} \sum_{l=1}^{N_s} \sum_{k=1}^{N_h} ||X_k^T S_l||^2 \tag{9}$$

Normalized power value is given in Equation 10 where $\hat{P}_i$ is the power of i th signal.

$$P_i = \frac{\hat{P}_i}{\sum_{J=1}^{N_f} \hat{P}_j} \tag{10}$$

Revised power value which is calculated with softmax fuction is given in Equation 11. This revision is necessary for overcoming the problems caused by large N_f values.

$$\acute{P}_i = \frac{e^{P_i}}{\sum_{J=1}^{N_f} e^{P_j}} \tag{11}$$

Finally, the frequency value f_{result} which has max $\acute{P}_i$ value between 1 and N_f is the actual result. This value is given by Equation 12.

$$f_{result} = argmax(\acute{P}_i) \tag{12}$$

D. Calculating System Performance

B_t which is expressed in bits per trial also can be calculated from Equation 13, where p is the classification accuracy and N is the number of targets. The number of targets mean is stimulated squares [11]. In this case N becomes 5.

$$B_t = log_2 N + p log_2 p + (1-p) log_2 \left[\frac{1-p}{N-1}\right] \tag{13}$$

B_m which is ITR and in bits per minute, can be calculate by Equation 14 where C_N is the number of classifications and T is the process time in seconds [11, 5].

$$B_m = \frac{60}{T} C_n B_t \tag{14}$$

III. EXPERIMENTAL RESULTS

Three subjects which has native language Turkish are participated in the experiments. Subjects are tried to type words "anne" (which means mother in Turkish), "baba" (which means father in Turkish), "hello". In the signal process part MEC method is applied to EEG segments in order to detect SSVEPs. Success of subjects by different signal window lengths are given in Table 1, Table 2, Table 3 and Table 4.

Table 1: Success values with 5 s signal length.

	Anne			Baba			Hello		
Subject	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)
S1	8	0	31.02	8	0	31.02	10	0	31.02
S2	8	0	31.02	8	0	31.02	10	0	31.02
S3	11	1	24.02	8	0	31.02	10	0	31.02

Table 2: Success values with 4 s signal length.

	Anne			Baba			Hello		
Subject	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)
S1	8	0	38.77	8	0	38.77	10	0	38.77
S2	8	0	38.77	8	0	38.77	10	0	38.77
S3	10	2	23.19	11	1	30.02	11	1	30.02

Table 3: Success values with 3 s signal length.

	Anne			Baba			Hello		
Subject	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)
S1	8	0	51.70	8	0	51.70	10	0	51.70
S2	8	0	51.70	8	0	51.70	10	0	51.70
S3	11	1	40.03	8	0	51.70	11	1	40.03

Table 4: Success values with 2 s signal length.

	Anne			Baba			Hello		
Subject	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)	Correct Command	Wrong Command	ITR (bit/min)
S1	11	1	60.04	10	2	46.38	10	0	77.55
S2	8	2	41.32	8	2	41.32	11	1	60.04
S3	11	1	60.04	10	2	46.38	12	2	50.12

IV. CONCLUSION

In this study a SSVEP based BCI speller application is introduced. System ITR values are calculated for each subject and signal window lengths. The results of analysis show that while signal window lengths are reduced, ITR values are increased but also wrong command counts are increased. The highest ITR value of 77.55 bit/min is obtained for subject 1 with 2 s signal window length. Subject 1 and subject 2 are successfully completed experiments which has 5, 4 and 3 s signal lengths. System performance is affected negatively from stress of subjects and better signal detection is obtained by naturally calm subjects. Further research might also consider human factors that can influence performance.

REFERENCES

[1] M. A. Pastor, J. Artieda, J. Arbizu, M. Valencia and J. Masdeu, "Human cerebral activation during steady-state visual-evoked responses," *Neurosci*, vol. 23(37) pp. 11621-11627, December 2003.

[2] H. Cecotti, I. Volosyak and A. Graser, "Reliable visual stimuli on lcd screens for ssvep based bci," *European Signal Processing Conference (EUSIPCO)*, vol. 18, pp. 919-923, August 2010.

[3] E. C. Lalor, S. P. Kelly, C. Finucane, R. Burke, R. Smith, R. B. Reilly and G. McDarby, "Steady state vep-based brain-computer interface control in an immersive 3d gaming environment," *EURASIP*, vol. 19, pp. 3156-3164, January 2005.

[4] X. Gao, X. Xu, M. Cheng and S. Gao, "A bci-based environmental controller for the motion-disabled," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11(2), pp. 137-140, June 2003.

[5] I. Volosyak, "Ssvep based bremen-bci boosting information transfer rates," *Neural Eng.*, vol.8(3), pp. 62-71, June 2011.

[6] G. Bin, X. Gao, Z. Yan, B. Hong and S. Gao, "An online multi-channel ssvep-based brain-computer interface using a canonical correlation analysis," *Neural Eng.*, vol.6(4) pp. 046002, August 2009.

[7] I. Volosyak, F. Gembler and P. Stawicki, "Age-related differences in ssvep- based bci performance," *Neurocomputing*, vol. 250 pp. 57-64, February 2017.

[8] Y. Zhang, D. Guo, D. Yao and P. Xu, "The extension of multivariate synchronization index method for ssvep-based bci," *Neurocomputing*, vol. 269, pp. 226-231, June 2017.

[9] E. Kreyszig, Advanced Engineering Mathematics, John Wiley, 9 th edition, 2006

[10] O. Friman, I. Volosyak and A. Graäser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54(4), pp. 742-750, March 2007.

[11] J. R. Wolpaw, H. Ramoser, D. J. McFarland and G. Pfurtscheller, "Eeg-based communication:improved accuracy by response verification," *IEEE Trans. Rehabil. Eng.*, vol. 6(3), pp. 326-333, September 1998.

Dynamic Programming Approach to Selling and Buying of Stocks in XU030 Index of BIST by Forecasting Stock Prices for Next Five Years

O.A.ARIK[1]

[1] Nuh Naci Yazgan University, Kayseri/Turkey, oaarik@nny.edu.tr

***Abstract* - In this paper, historical data of stocks in the XU030 index of BIST are examined by considering their trends and seasonal behaviors in order to forecast next five year's stock prices by using ARIMA or seasonal ARIMA technique. Then, a mixed integer programming model is executed for generating a portfolio management for next five years. The proposed mixed integer programming model is based on knapsack problem. The knapsack problem is one of the most applicable portfolio management models.**

***Keywords* - Forecasting, ARIMA, mixed integer programming, dynamic programming.**

I. INTRODUCTION

THIS paper investigates how to manage a potential future portfolio using mixed integer linear programming technique. It is very important to manage the portfolio in the stock exchange market which is an organized system in which the commercial goods and documents offered to the public are bought and sold within the framework of certain standards and rules. An efficient portfolio management is a significant part of long-term profitability and stability of investments all over the world. This study discusses how to make a profitable portfolio management in the next 5 years by using the mixed integer mathematical model. The monthly stock prices for next five years are forecasted by Autoregressive Moving Average (ARIMA) method. The seasonality and trends of stocks are also considered in forecasting. As a real-life problem, stocks in the XU030 index of Borsa Istanbul (BIST), the first 30 companies that have the most trade volume in the market, are examined and stock prices in view of their seasonality and trends.

The historical data and seasonality of stocks of the XU030 index are taken into consideration. In order to maximize investor's initial investment, a dynamic portfolio management model is proposed by using forecasted stock's prices for next five years. Monthly closing data of XU030 companies were examined from the date of their existence in the stock market. In examining this data, the trends and seasonality of equities were taken into account in order to be able to create a more accurate portfolio management for the future. Then ARIMA method was used to create appropriate ARIMA models considering the autocorrelation and partial autocorrelation values of the stocks. After creating ARIMA models, forecasts for the next five years were made by using them. A mixed integer mathematical model has been proposed for the problem of how to manage portfolios with forecasted data for the next 5 years. The proposed mixed integer mathematical model is presented at the end of the study that will increase the profitability of the investor over the next five years under certain assumptions.

II. LITERATURE REVIEW

In this section, a short literature review is presented for the readers. The portfolio management has been investigated by researchers for more than sixty years. Mansini and Speranza [1] investigated a problem for choosing a portfolio and their problem has been faced with great difficulty in terms of reducing the risk most when considering the return. Their article focused on portfolio management with minimum transaction volumes. They calculated independently of the risk function in order to find a suitable solution.

Beaujon et. al. [2] proposed a mathematical formulation of an optimization model designed on an R&D portfolio based on various constraints is presented. The model was a mixed integer programming problem, previously known in the literature and known as the knapsack problem. Complete solutions to such problems are often difficult but can be achieved using the specialized algorithms.

Calafiore [3] investigated multi-turn sequential decision problems for financial asset location and Calafiore [3] proposed a model by making periodic optimal portfolio adjustments. Calafiore [3] stated that the risk measure must be minimized throughout the investment when estimating the portfolio diversity.

Konno et. al. [4] investigated a mean-variance model that is widely used in fund management. Asset distribution was used to determine the basic principle of a book. Risk management and performance measurement for the portfolio were also taken into consideration and the popularity of the mean-variance model between passive index management and various other asset management models was also used to determine the fund rate to be achieved.

Krink and Paterlini [5] proposed a new multi-objective evolutionary algorithm called as DEMPO (Differential Evolution for Multi objective Portfolio Optimization) for portfolio optimization. They compared DEMPO with quadratic programming and another well-known evolutionary algorithm for multi-objective optimization called NSGA-II.

They stated that the main advantage of their proposed method is its ability to tackle a portfolio optimization task without simplifications, while obtaining very satisfying results in reasonable runtime.

Guastaroba et. al. [6] surveyed different techniques to generate scenarios for the rates of return. Furthermore they compared the techniques by providing in-sample and out-of-sample analysis of the portfolios obtained by using these techniques to generate the rates of return. In their paper, as reference model they used the Worst Conditional Expectation model with transaction costs.

Quaglia et. al. [7] presented a systematic framework for synthesis and design of processing networks under uncertainty. In their paper, through the framework, an enterprise-wide optimization problem was formulated and solved under uncertain conditions, to identify the network (composed of raw materials, process technologies and product portfolio) which is feasible and have optimal performances over the entire uncertainty domain.

Ji et. al. [8] proposed a stochastic linear goal programming model for multistage portfolio management. Their model takes into account both the investment goal and risk control at each stage. They also proposed a scenario generation method that acts as the basis of the portfolio management model. They generated scenarios for multistage portfolio management by incorporating this single-stage method with the time-series model for the asset returns. They solved a real case via the goal programming model and the scenario generation approach which demonstrates the effectiveness of the model.

Rocha and Kuhn's [9] dealt with the financial risks that the privatized electric power supply companies face. They specified a few portfolio methods to protect the company from financial risks that want to best meet the client's claim. In the study of Polak et. al. [10], they investigated how to motivate companies on how to get rid of the worst possible scenarios in the economy. The name of their method was minimax linear programming optimization.

Korhonen [11] tried to predict next three years for the situation of domestic and foreign financial markets under some uncertainties to be experienced in the future, as well as the economic conditions. Korhonen [11] proposed a practical application of goal programming model to the management of the domestic and foreign currency denominated assets and liabilities of a large bank in Finland.

Kawas and Thieley [12] presented a robust optimization approach to portfolio management under uncertainty that builds upon insights gained from the well-known Lognormal model for stock prices, while addressing the model's limitations, in particular, the issue of fat tails being underestimated in the Gaussian framework and the active debate on the correct distribution to use.

In the article of Rabbani et. al. [13], they formulated essential factors in R&D project portfolio selection by a mathematical model, which considers a multi-objective function for maximizing corporate benefit through quantitative and qualitative criteria as well as insight and preferences of decision makers and human resource allocation which does not exceed organization's constraints such as planning horizon, available resources, and interrelationships between projects. In the article of Ostermark [14], a heuristic method called as the Super Genetics Hybrid Algorithm was proposed for portfolio management with very cyclical transaction costs.

April et. al. [15] studied a problem about a loss prevention and performance possibility from a new portfolio optimization and the financial gains of portfolio software. In their article, they used simulation techniques of a new software system to make investment scenarios more strategic and that it is appropriate.

III. Methodology

In this section, the necessary techniques are introduced for the study. Firstly, ARIMA techniques are introduced in order to forecast stock prices. Then, a mixed integer programming model is introduced to create dynamically portfolio management by selling and buying stocks. In these parts, there are some assumptions as follows:

- stock prices are equal for buying and selling prices for
-
- there is no limitation for buying and selling of XU030 stocks,
- there is no transaction cost while selling and buying stocks,
- when an investor wants to sell or buy any stock, he/she can buy immediately.

A. Autoregressive integrated moving average (ARIMA) methods

Autoregressive integrated moving average (ARIMA) forecasting methods were popularized by Box and Jenkins [16]. These techniques, often called the Box-Jenkins forecasting methodology, have the following steps [17]:

- Model identification and selection
- Estimation of autoregressive (AR), integration or differencing (I), and moving average (MA) parameters
- Model checking

The ARIMA technique analyzes and forecasts equally spaced univariate time series data. An ARIMA model predicts a value in a response time series as a linear combination of its own past values, past errors, and current and past values of other time series. An ARIMA model needs three components: AR, I and MA respectively. ARIMA models need stationary time series data. In order to make the time series stationary, the trend in the series is removed with help of differencing. I component of ARIMA model (also known as d) denotes numbers of differencing to create stationary time series. AR component (symbolized as p) denotes the order of autoregressive part of the ARIMA model. MA component (symbolized as q) denotes the order of moving average terms of the ARIMA model. All parameters (p,d, and q) in ARIMA model are integers and they are equal to or greater than 0. Using p, d and q values, ARIMA(p,d,q) model predicts next data of nonseasonal time series. If time series including seasonality, then the ARIMA model needs the second group of parameters for SARIMA(P,D,Q)(t) where t is number of time

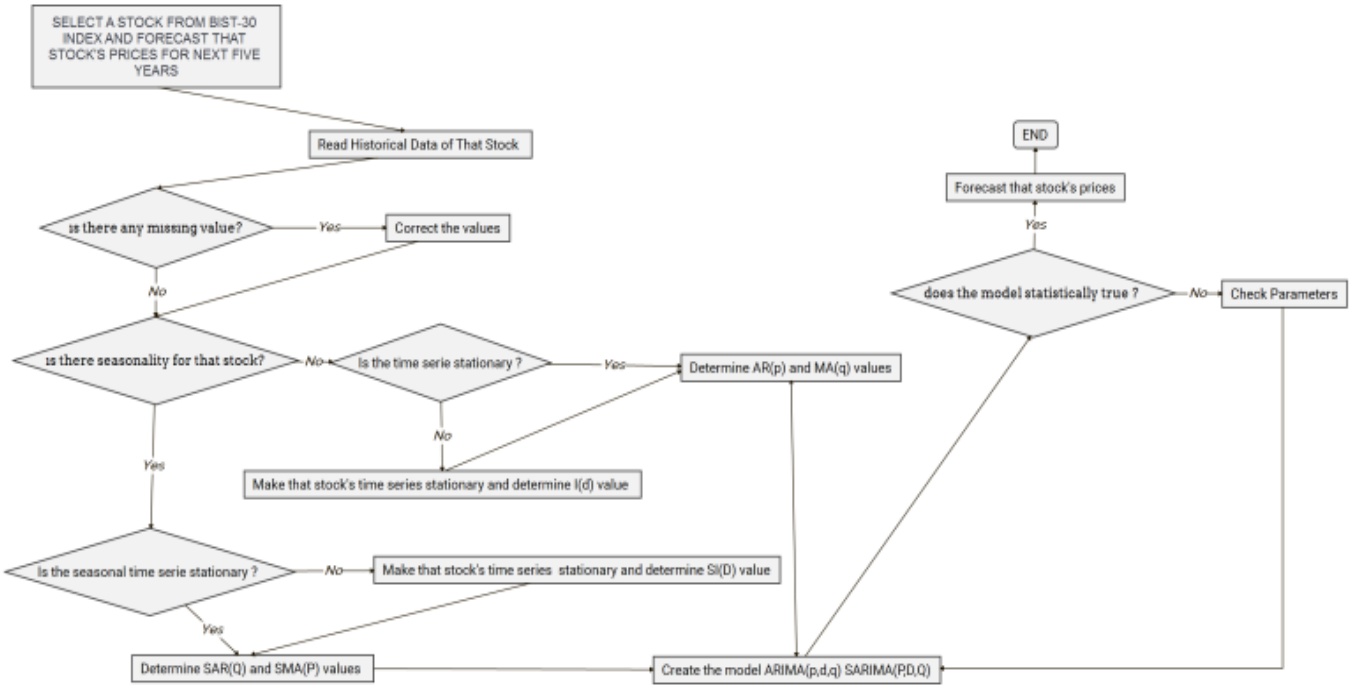


Figure 1: Forecasting Steps for ARIMA and SARIMA

periods in a cycle and t is equal to 12 for a year. Steps for forecasting method in this study are seen in Figure 1. As understood from Figure 1, if the time series have seasonality, then these series must be considered with SARIMA. Whether there is seasonality or not, time series must be stationary for forecasting. After differencing time series to make them stationary, necessary parameters are determined to forecast next values of the time series.

B. *Mixed integer programming model*

After forecasting stocks' prices for next five years, we proposed a mixed integer linear programming model (MILP) as follows:

Indices:
i: index for months
t: index for stocks
Parameters:
$B_{i,t}$: Buying price of stock t on month i
$S_{i,t}$: Selling price of stock t on month i
K: initial investment for stocks in BIST − 30 index
Decision variables:
$Qx_{i,t}$: number of lots for buying of stock t on month i
$Qy_{i,t}$: number of lots for selling of stock t on month i
C_i: cash amount in hand on month i
Cx_i: buying amount in hand on month i
Cy_i: selling amount in hand on month i
Model:

$$Max\ z = C_{60} \quad (1)$$

S.t.:

$$C_1 = K - \sum_{t}^{60} Qx_{t,1} * B_{t,1} \quad (2)$$

$$C_i = C_{i-1} + \sum_{t}^{60} Qy_{i,t} * S_{i,t} - \sum_{t}^{60} Qx_{i,t} * B_{i,t} \ \forall i \ \& \ i > 1 \quad (3)$$

$$\sum_{l}^{i} Qx_{l,t} \geq \sum_{l}^{i} Qy_{l,t} \ \forall i \ \& \ \forall t \quad (4)$$

$$Qy_{1,t} = 0 \ \forall t \quad (5)$$

$$Qx_{60,t} = 0 \ \forall t \quad (6)$$

$$Cx_i = \sum_{t}^{60} Qx_{i,t} * B_{i,t} \quad (7)$$

$$Cy_i = \sum_{t}^{60} Qy_{i,t} * S_{i,t} \quad (8)$$

$$Cx_i, Cy_i, C_i \geq 0 \ \forall i \quad (9)$$

$$Qy_{i,t}, Qx_{i,t} \geq 0 \ Qy_{i,t}, Qx_{i,t} \in N^+ \ \forall i \ \& \ \forall t \quad (10)$$

The objective function (1) is to maximize total money in hand at the end of the last month. Constraint (1) ensures that total money at the end of the first month is equal to the amount that remains after buying different stocks by using initial investment. Constraint (3) shows that buying and selling transactions on each month must be done by considering previous month's total money in hand. This situation is illustrated in Figure 2. Constraint (4) shows that each month' selling transactions for all stocks cannot be greater than the lots of stocks on hand until that month. Constraint (5) ensures that there is no allowance for selling stocks on the first month. Constraint (6) ensures that there is no allowance for buying stocks in the last month. Constraints (7) and (8) show the total buying and selling transactions money quantities on each month. Equations (9) and (10) define the domains of decision variables.

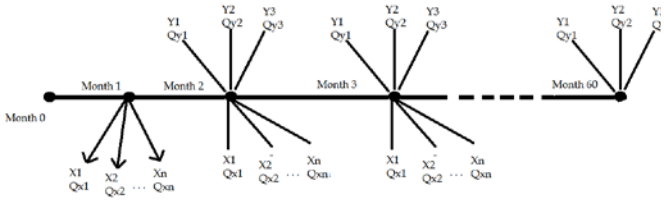


Figure 2: Illustration of constraint (3)

IV. APPLICATION

In this section, an application consisting of stocks in XU030 index dated October 2017 is presented. The monthly closing data of stocks since their existence in BIST were examined for their seasonality and trends. Then, appropriate ARIMA models are found by following the steps in Figure 1 using SPSS software. The models are given in Table 1.

Table 1: ARIMA models for selected stocks

No:	Stock Name	Model	No:	Stock Name	Model
1	AKBNK	(0,2,1)(0,0,0)	16	PETKIM	(1,1,1)(2,1,1)
2	ARCLK	(2,1,2)(0,0,0)	17	SISE	(0,1,0)(0,1,1)
3	ASELSAN	(1,1,1)(0,0,0)	18	SODA	(0,1,0)(0,1,1)
4	BIMAS	(2,1,2)(0,0,0)	19	TAVHL	(0,1,0)(1,1,1)
5	DOHOL	(1,0,0)(0,0,0)	20	TEKFN	(1,1,1)(0,1,1)
6	EKGYO	(1,1,1)(0,0,0)	21	THYAO	(1,1,1)(1,1,1)
7	ENKAI	(2,1,2)(0,0,0)	22	TAOSA	(0,1,1)(0,1,1)
8	EREGL	(2,1,2)(0,0,0)	23	TUPRS	(1,1,1)(0,1,1)
9	FROTO	(2,1,2)(0,0,0)	24	TTKOM	(0,1,0)(0,1,1)
10	GARANTI	(0,2,1)(0,0,0)	25	TCELL	(1,1,1)(0,1,1)
11	SAHOL	(1,1,1)(0,0,0)	26	HBANK	(0,1,0)(1,1,1)
12	ZOREN	(2,1,2)(0,1,1)	27	ISCTR	(0,1,0)(0,1,1)
13	KCHOL	(1,1,1)(0,0,0)	28	ULKER	(0,1,0)(1,1,1)
14	KOZAL	(1,0,0)(2,0,0)	29	VKFBN	(1,1,1)(1,1,1)
15	OTKAR	(1,1,1)(0,1,1)	30	YKBNK	(0,1,0)(0,1,1)

After forecasting monthly closing prices of stocks in XU030 index for next five years, the proposed mixed integer programming model is executed in CPLEX 12.6 software. As initial investment parameter, different K values are used and the results of model for K values are given in Table 2.

Table 2: ARIMA models for selected stocks

Initial investment amount (TL)	Cash amount (TL) after 60 months	Elapsed Times (hh:mm:ss:ms)
10,000.00 TL	76,201.71 TL	00:55:18:67
100,000.00 TL	762,359.05 TL	00:09:52:43
1,000,000.00 TL	7,624,487.30 TL	00:00:03:38
10,000,000.00 TL	76,245,785.48 TL	00:08:25:74

V. CONCLUSION

In this paper, a mixed integer programming model is proposed in order to manage dynamically an investor's stock portfolio using forecasted stock prices. In forecasting phase of this study, ARIMA models are used to investigate seasonality and trends of stock prices. The stocks in the XU030 index of BIST were examined and forecasted for next five years using ARIMA model. The proposed model is executed with Cplex 12.6 software using different initial investment amounts. The results show that theoretical return of investment (ROI) is more than 660%. The proposed model is a kind of multi-period or dynamic knapsack model. In future research, heuristic algorithms for knapsack problems or metaheuristic methods can be applied to the proposed problem. Different forecasting techniques can be applied to this problem in order to invest a stock or stocks.

REFERENCES

[1] Mansini, R., Speranza, M.G. (1999) Heuristic algorithms for the portfolio selection problem with minimum transaction lots, European Journal of Operational Research, 114(2) 213-239.

[2] Beaujon, G.J., Marin, S.P., McDonald, G.C. (2001) Balancing and optimizing a portfolio of R&D projects, Naval Research Logistics, 48(1) 18-40.

[3] Calafiore, G.C. (2008) Multi-period portfolio optimization with linear control policies, Automatica, 44(10) 2463-2473.

[4] Konno, H., Waki, H., Yuuki, A. (2002) Portfolio optimization under lower partial risk measures, Asia-Pacific Financial Markets, 9(2), 127-140.

[5] Krink, T., Paterlini, S. Multi objective optimization using differential evolution for real-world portfolio optimization, Computational Management Science, 8 (1-2), pp. 157-179.

[6] Guastaroba, G., Mansini, R., Speranza, M.G. (2009) On the effectiveness of scenario generation techniques in single-period portfolio optimization, European Journal of Operational Research, 192 (2), pp. 500-511.

[7] Quaglia, A., Sarup, B., Sin, G., Gani, R. (2013) A systematic framework for enterprise-wide optimization: Synthesis and design of processing networks under uncertainty, Computers and Chemical Engineering, 59, pp. 47-62

[8] Ji, X., Zhu, S., Wang, S., Zhang, S. (2005) A stochastic linear goal programming approach to multistage portfolio management based on scenario generation via linear programming, IIE Transactions (Institute of Industrial Engineers), 37 (10), pp. 957-969

[9] Rocha, P., Kuhn, D. (2012) Multistage stochastic portfolio optimisation in deregulated electricity markets using linear decision rules, European Journal of Operational Research, 216 (2), pp. 397-408.

[10] Polak, G.G., Rogers, D.F., Sweeney, D.J. (2010) Risk management strategies via minimax portfolio optimization, European Journal of Operational Research, 207 (1), pp. 409-419.

[11] Korhonen, A. (1987) A dynamic bank portfolio planning model with multiple scenarios, multiple goals and changing priorities, European Journal of Operational Research, 30 (1), pp. 13-23.

[12] Kawas, B., Thiele, A. (2011) A log-robust optimization approach to portfolio management, OR Spectrum, 33 (1), pp. 207-233.

[13] Rabbani, M., Moghaddam, R.T., Jolai, F., Ghorbani, H.R. (2006) A comprehensive model for R and D project portfolio selection with zero-one linear goal-programming, International Journal of Engineering, Transactions A: Basics, 19 (1), pp. 55-66.

[14] Östermark, R. (2005), Dynamic portfolio management under competing representations, Kybernetes, 34 (9-10), pp. 1517-1550.

[15] April, J., Glover, F., Kelly, J., Laguna, M. (2003) A New Optimization Methodology for Portfolio Management, Proceedings - SPE Annual Technical Conference and Exhibition, pp. 2441-2445.

[16] Box, G.E.P., Jenkins, G.M., MacGregor, J.F. (1974) Some Recent Advances In Forecasting And Control-2, Journal of Applied Statistics, 23(2), pp. 158-179.

[17] https://docs.oracle.com/cd/E57185_01/CBPPU/PRHist_ARIMA_intro.htm#CBPPU-pp_user_book_217.

Software Development for Robotic Competition Eliminations Using Expert Systems

A. GUL[1] and E. SOYLU[2]

[1] Karabuk University, Karabuk/Turkey, abdulkadir.gul.95@gmail.com
[2]Karabuk University, Karabuk/Turkey, ekocak@karabuk.edu.tr

Abstract **- Many robot competitions are organized in the national and international arenas. Many different robots compete in many different categories. The difficulties encountered during the competition are tried to be solved later in the competitions. This contributes to the development of automation technology. In this kind of competition, there are robot categories such as sumo robot, mini sumo robot, and line follower robot. Competitions can be done using tournament, elimination and ranking procedures. In this study, automation suitable for all three types of procedure has been developed and expert systems were used. Significant features of the automation include; tournament automation with dynamic elements completely dependent on the field of display, QR code creation and reading, bidirectional and multiple communication, SQL database connection and configuration.**

Keywords **- Automation, robot competition, communication, QR Code, SQL Database.**

I. INTRODUCTION

A a robot is a machine which is programmed to move and perform certain tasks automatically. For certain tasks, robots can be superior to humans in terms of quality of the work that is produced. Basically, parts of robots are sensors, effectors, and control systems [1].

Every year, hundreds of robot competitions are held all over the world. The aims of these competitions are many and varied and include providing forums for enthusiasts, promoting scientific and technical education, marketing a particular make of robot, setting manufacturing challenges, etc. [2]. According to research, robot competitions increase the motivation of engineers [3]. Robot contest participants gain achievements such as critical and creative thinking, communicating effectively, demonstrating a scientific field expertise, gaining work discipline, working individually and in groups, engaging in lifelong learning [4]. Since early 2000, educational robotics competitions for school-age children have increasingly gained the popularity around the world [5].

The biggest problem in these competitions is the determination of the results of the competition of the robot categories and the follow up of the results. Some of the difficulties in robot competitions are the need to clearly determine the rules for comparing results and determining approaches. These rules should be robust, general and applicable [6]. We have developed this automation system to get over these problems. The software is developed for the Karabuk University International Robotic Tournament 2018. Thanks to the flexibility of the system it is possible to use the software for other robotic competitions.

Our software has three parts that are Server, Clients, and Registration.

II. ROBOTIC COMPETITIONS

Many robot competitions are held every year in Turkey. Karabuk University International Robot Tournament, International MEB Robot Competition, International ODTU Robot Days, Okan University Robokan Robot Competition, International Marmara Robot Olympics, Yıldız Robocon, Gebze Technical University Kelebek Robot Olympics, Rotek RoboTeam Anatolian Robot Competition, Istanbul Technical University Robot Olympics are some of them. Contestants from many age groups participate in these competitions.

In Turkey, robot competitions are held in such categories; line follower robot, sumo robot, mini sumo robot, a robot capable of climbing stairs, fire-fighting robot, garbage collecting robot and open projects etc. Generally, the purpose of the robot is to fulfill the task as soon as possible or to eliminate the opponent. A picture of a robot competition in which many categories are shown is given in Figure 1.

Figure 1: Multiple categories robotic competition [7].

We can classify robot categories into two part: elimination and ranking. Robots racing in the elimination category try to win the champions by eliminating their opponents. Robots competing in the ranking category try to do the task in minimum time as soon as possible.

Information about most known general Robot Tournament categories are given below:

A. Sumo Robot

Sumo robots have been inspired by sumo wrestling of Japanese. People interested in robotics wanted to have sumo wrestling done by robots. Sumo robots are robots that have autonomous motion capabilities or can be controlled from the remote control device are programmed for intended movements, designed to combat electronic circuits. They are produced in different standards and categories. Sumo robots face each other on a round ring with certain standards and characteristics called "Dohyo". During the match, sumo robots try to push each other out of the line around the ring. The robots use contrast sensors to detect the white line around the dohyo, try not to get out of the ring, stay in the ring. Various sensors (IR, ultrasonic, laser, etc.) are added to the robots to detect the surroundings and the rival robot in a short time, and the improved tactics algorithms are also installed. The goodness of the mechanical structure, electronic design and software are the main parameters that are effective in winning the sumo robot competition [8]. Sumo robots are racing in the elimination procedure.

B. Mini Sumo

Robot sumo is one of the most popular robotics events based on real Japanese Sumo matches. Robots try to push each other out of dohyo [9]. Mini-sumo category has similar rules with sumo robot category. It is generally desirable that the robot base does not exceed 10 x 10 cm. There is no height limit. It is requested that the weight should not exceed 500g. dohyo's dimensions are rather small compared to the sumo robot, the dohyo's dimensions can change from race to race. Mini-sumo robots are racing in the elimination procedure.

C. Line Follower Robot

A line follower robot is a mobile machine that can detect and follow the line drawn on the floor. Generally, the path is predefined and can be either visible like a black line on a white surface [10]. In the line follower robot competitions bridge, obstacle, ramp, the corner can be added to the race track. Line follower robots are racing in the ranking procedure.

III. SOFTWARE DESIGN

In this study, the software has been developed for the robots that race the elimination and ranking procedure. The software was implemented with the C# language in the Visual Studio environment. SQL database is used. The expert systems are the computer applications developed to solve complex problems in a particular domain, at the level of extra-ordinary human intelligence and expertise. The expert system mimics the expert's reasoning process through the use of IF___THEN___ rules. We used the expert system in this study.

The basic algorithm of the system is given in Figure 2. The user and robot registrations are done via web page before robot tournament. Robots that register via the web page are taken from the database and the appropriate QR codes are created using the Messaging Toolkit QR Code Library [11]. The generated QR codes are read after being placed in the robots and the read robot is marked as "came" in the database.

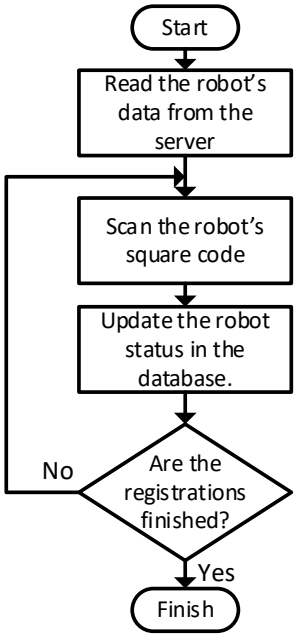


Figure 2: Flow chart of the main algorithm.

Block diagram of the system is given in Figure 3. The server is shown in the middle. Referees connect to the server and enter the results of competitions. Other clients connect to the server to see the results on display units.

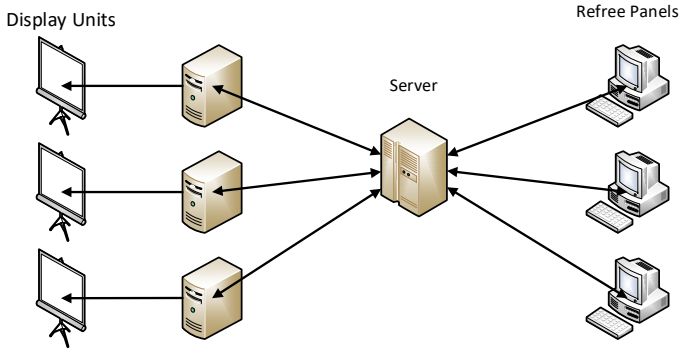


Figure 3: Block diagram of the system.

Console applications have been developed on the server side of the application because of the high memory consumption and long cycle times of visual applications. The server part is coded by taking C # Chat Server as an example [12]. A view of server window is given in Figure 4.

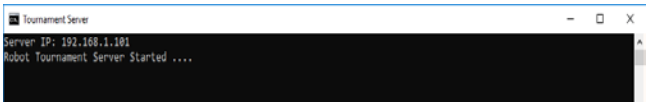


Figure 4: Server window.

Server software reads robot data from the database at the beginning, categorizes all robots, and puts the robots in the race according to the category rules. The client sends robot categorizations on request. In Figure 5 the window to connect to the server is given.

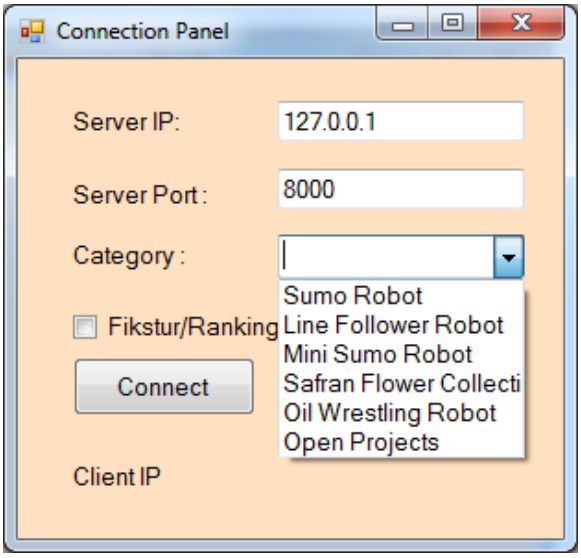


Figure 5: Connection panel.

The basic function of the tournament server here is to listen to the incoming request from clients. The tournament server here is a C# console based application and is listening to the identified PORT on a stat for the connection request from clients. When the server got a connection request, it adds the name of the client into related category clients list and creates a new thread for communicate with the server, then evaluates the requests from the clients and directs the relevant client. All server operations are saved in the log file and can be resumed from where they were restored in case of any disruption. The TCPListener library is used to listen to the port [13], [14].

Tournament category clients are collected in one software. Matching in qualifying competitions and determining the order of competition in ranking competitions is determined by this part. The name of robot or robot is shown on referee panel which will compete. If it is an elimination competition referee enters the winner of the match if it is a ranking competition referee enters the difference time between the start and finish time. Client communication is based on the C # Chat Client example [15].

The Microsoft .NET framework provides two namespaces, System.Net, and System.Net.Sockets for managed implementation of Internet protocols that applications can use to send or receive data over the Internet. The C# Tournament Client here is a Windows based Application and its main function is to send a message to the Tournament Server.

The categories use similar fixture and referee panel. Fixture connects to the server and wants related category. Fixture panels give a response if a request comes from referee panels or when the winner robot information from referee panel comes.

A sample view of referee panel is given in Figure 6. Here the names and QR codes of robots that going to match is shown. To start the match firstly the QR codes of robots are read to validate that robots are coming to match. If two of them came and do the match the winner robot's QR code is read again and this winner can go to the next round. The server responses new robot matches plan. If the rival is disqualified, the winning robot's QR code is read three times.

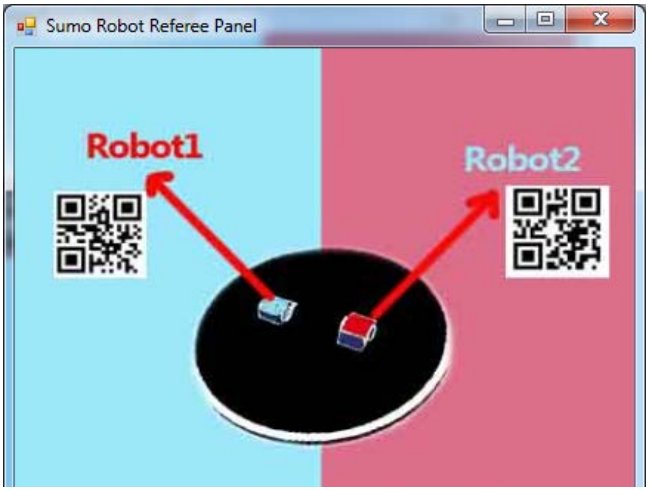


Figure 6: A sample window for elimination category.

The logs of referee and fixture panels have been saved so in case of a mishap, this data is accessed again. In Figure 7 a sample fixture is given for elimination category. This window is dynamic and it depends on the number of robots. The path of the winning robot is shown in red in this schema. The winner robot's name is also written into next stage.

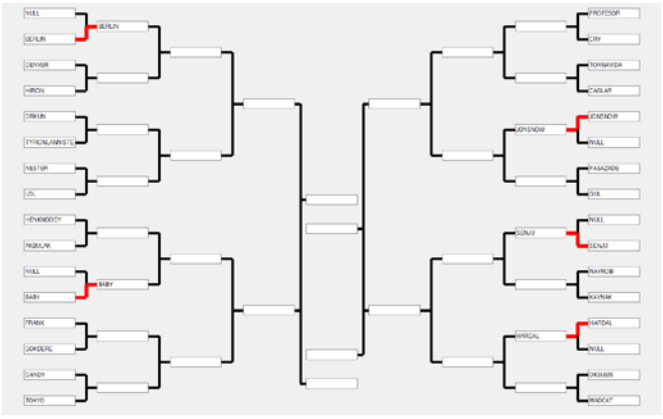


Figure 7: Fixture window.

In Figure 7 a sample window of ranking category competition control panel is given. The referee reads the square code of robot, enters the race time and penalty time. The total time of the race is calculated and the robots are listed according to the time they finished the contest. The log of these data is saved in order to prevent data loss.

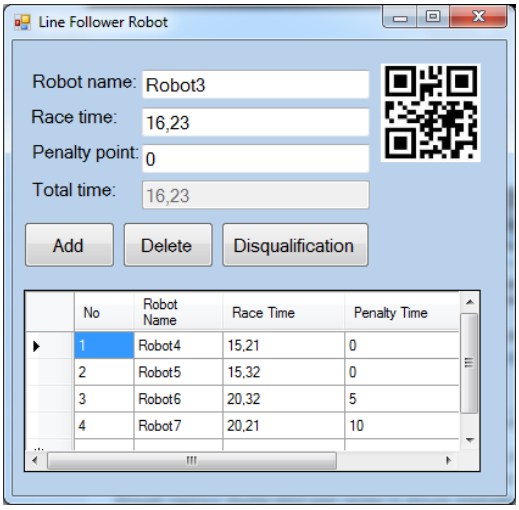


Figure 7: A sample window for the ranking category.

IV. CONCLUSION

In this study, software for robot competition was implemented. There are many robotic competitions including a different type of categories such as sumo, mini sumo, line follower, wrestling, garbage collection etc. Basically, we can classify robot categories into two parts: elimination and ranking. It is necessary to race the robots in many categories at the same time.

To test this study client are connected to the server at the same time and a virtual robot competition organization was realized.

Finally, the program is suitable for real robotic competitions that include ranking and/or elimination categories. The results of the competition can be monitored in real time.

ACKNOWLEDGMENT

Thanks to the "International Karabük University Robot Competition", this has led to the development of this software.

REFERENCES

[1] "Parts of Robots." [Online]. Available: http://www.mind.ilstu.edu/curriculum/medical_robotics/parts_of_robots.php. [Accessed: 22-Mar-2018].

[2] V. Mazzari, "Robotics competitions," 2016. [Online]. Available: https://www.generationrobots.com/blog/en/2016/04/robotics-competitions/. [Accessed: 22-Mar-2018].

[3] P. Xu and D. G. Bailey, "Robotic Competitions : Motivation for Engineering Robotic Competitions : Motivation for Engineering Programmes," no. JULY, 2003.

[4] J. Grimes and J. Seng, "Robotics competition: Providing structure, flexibility, and an extensive learning experience," *Proc. - Front. Educ. Conf. FIE*, pp. 9–13, 2008.

[5] A. Eguchi, "RoboCupJunior for promoting STEM education, 21st century skills, and technological advancement through robotics competition," *Rob. Auton. Syst.*, vol. 75, pp. 692–699, 2016.

[6] J. Anderson, J. Baltes, and K.-Y. Tu, "Improving Robotics Competitions for Real-World Evaluation of AI," *Proc. AAAI Spring Symp. Exp. Des. Real-World Syst.*, pp. 1–8, 2009.

[7] "Robot Competititon," *2017*. [Online]. Available: http://www.eba.gov.tr/video/izle/43882 1b53dc095cd941bd87470c91b296ecbb8869a001. [Accessed: 27-Mar-2018].

[8] P. Benavidez, C. Gleinser, A. Jaimes, J. Labrado, C. Riojas, M. Jamshidi, and L. B. Endowed, "Design of semi-autonomous robots for competitive robotics," pp. 1–6, 2012.

[9] R. Balogh, "A survey of robotic competitions," *Int. J. Adv. Robot. Syst.*, vol. 2, no. 2, pp. 144–160, 2005.

[10] M. Pakdaman and M. M. Sanaatiyan, "Design and implementation of line follower robot," *2009 Int. Conf. Comput. Electr. Eng. ICCEE 2009*, vol. 2, pp. 585–590, 2009.

[11] "MessagingToolkit.QRCode." [Online]. Available: https://www.nuget.org/packages/MessagingToolkit.QRCode/. [Accessed: 29-Mar-2018].

[12] "How to C# Chat Server." [Online]. Available: http://csharp.net-informations.com/communications/csharp-chat-server.htm. [Accessed: 29-Mar-2018].

[13] "TcpListener Sınıfı," 2016. [Online]. Available: https://msdn.microsoft.com/tr-tr/library/system.net.sockets.tcplistener(v=vs.110).aspx. [Accessed: 29-Mar-2018].

[14] "Thread Sınıfı," 2016. [Online]. Available: https://msdn.microsoft.com/tr-tr/library/system.threading.thread(v=vs.110).aspx. [Accessed: 29-Mar-2018].

[15] "How to C# Chat Client." [Online]. Available: http://csharp.net-informations.com/communications/csharp-chat-client.htm. [Accessed: 29-Mar-2018].

Technology Evaluation with HUP-Growth Mining Algorithm

S. ALTUNTAS[1] and M. SEZER[1]

[1] Yildiz Technical University, Istanbul/Turkey, serkan@yildiz.edu.tr, Corresponding author
[1] Yildiz Technical University, Istanbul/Turkey, sezermehmet7@gmail.com

Abstract **- Technology evaluation is quite important field of technology management. Potential technology investments can be assessed thanks to technology evaluation methods in practice. Patent documents are extensively used for technology evaluation in the literature. The use of patent documents provides an objective evaluation of technology in real life engineering applications. In this study, patent documents are analyzed based on mining high utility itemsets. Mining high utility itemsets is a data mining approach. HUP-growth mining algorithm is one of the algorithm proposed for mining high utility itemsets in the literature. HUP-growth mining algorithm is utilized for the analysis of patent documents in this study. To show the application of the algorithm in the field of technology evaluation, all patents related to geothermal energy are retrieved from the United States Patent and Trademark Office (USPTO). The results obtained from this study show that HUP-growth mining algorithm can be easily and effectively used for technology evaluation based on patent documents.**

Keywords **- Technological evaluation, Patent analysis, data mining.**

I. INTRODUCTION

Technology intelligence is the systematic collection, explication and transmission of information about technological behaviors aimed at providing early warning of current opportunities and threats [1]. Technology evaluation can be considered as an integral part of technology intelligence process. A good technology evaluation process helps firms to increase their competitive power in the marketplace. Data mining based approaches are generally performed for a good technology evaluation process in the literature. Data mining is the process of selecting, discovering, and classifying a large amount of data to find and sharpen unknown patterns [2]. Data mining algorithms, which are closely related to knowledge discovery, are structure for creating data mining patterns. In order to create a pattern, the algorithm first analyzes the large data and searches for specific patterns [3].

Patent data is extensively used for various purposes in different periods such as forecasting technology trends, analyzing technology innovation patterns or developing technology strategies [4]. In this study, patent data is used for technology evaluation. HUP-growth mining algorithm, which is a data mining algorithm, is utilized based on patent data for technology evaluation. A case study on geothermal energy technology is conducted to show how to use HUP-growth mining algorithm for technology evaluation in practice.

Technology Intelligence (TI), which is a core process within the discipline of technology management, involves the process of capturing technology related data, converting this data into information by determining relational connections and refining information to produce knowledge that can guide strategic decision makers during strategic planning [5]. Technological change and development are important factors for companies both in practice and in theory in their business area [6]. In this context many studies have shown that companies are inadequate in catching technological changes, or in other words, they cannot achieve the speed of technology change [7], [8], [9], [10], [11]. Since the 1970s, many researchers have offered systematic approaches to capture technological developments in time [12], [13], [14]. The common goal in these approaches is to develop methods to speed up decision-making processes in companies by effectively identifying trends in technology. Patents represent the features of a new technology as a significant output of a large data set and provide a very rich source of data for the analysis of other technologies [15]. Patent analysis is also a valuable approach to obtain information in detail about a particular industry or technology used in forecasting [16]. Therefore, patent analysis based data mining algorithm is used for technology evaluation, which is an integral part of technology intelligence. HUP-growth mining algorithm is applied as a data mining approach in this study. HUP-growth mining algorithm is a type of association rules in the field of data mining.

Association rules are a data mining technique that is created to obtain information through structured databases [17]. Association rules have been widely used for detecting relations between attribute-value pairs of categorical datasets [18]. It extracts relationships between various objects (here CPC codes). Frequent pattern mining (FPM) [19], [20] and association rule mining (ARM) [21], [22] are useful to identify embedded useful knowledge [23]. In addition to these types of association rules, the term of "mining high utility itemsets" first appeared in [24]. The term of "mining high utility itemsets" is related to utility mining, which is a new type of association rules in the literature. In this study, HUP-growth mining algorithm is applied, which is a utility mining algorithm is applied to patent documents.

II. APPLICATION OF HUP-GROWTH MINING ALGORITHM FOR TECHNOLOGY EVALUATION

The proposed method is illustrated in a small sample of data with general outlines in this section. In this extended abstract, the details on HUP-growth mining algorithm is not provided due to the space limitation. Readers can look at [26] to go into the details on HUP-growth mining algorithm.

At the beginning of the application of HUP-growth mining algorithm, the technology to be analyzed has been identified. Although the proposed approach can be applicable on any desired technology, the geothermal energy which was proceed in a previous study [27] has been selected in order to ensure consistency with respect to the number of patent gathered from patent database. The following query is used to gather sufficient number of patent data base on [27].

(((ABST: geothermal AND (power OR energy)) AND ISD:([1991-01-01 TO 2017-08-01])))

It should be noted that this query researches all database and found all relevant patent documents if a patent document has "geothermal AND (power OR energy)" keywords in its abstract and published between 1991-01-01 and 2017-08-01. A special commercial software, namely the AcclaimIP software (http://www.acclaimip.com) which is a patent analysis system, is used for patent analysis.

While the number of patents obtained in the previous study [27] was 241 in 2012, 499 patents obtained in August 2017 with the same key words for this study. Corporate Patent Classification (CPC) codes are examined through the use of HUP-growth mining algorithm. There are two main advantage of this algorithm. First, it considers the repetition number of CPC codes assigned to a patent. Second, the importance of a patent related to a technology under concern can be easily take into account during the calculation process. It should be noted that Patent ID (PID) is considered as transaction ID and the CPC codes are considered as items in a transaction for the application of HUP-growth mining algorithm.

Table 1: A quantitative database in the example

		1	2	3	4	5			
PID	Document No.	C02F	E21B	F16J	F25B	Y10S	TOTAL	Forward Cites	FC/P
1	US4986511	0	2	2	0	2	6	28.00	4.67
2	US4993483	0	0	0	2	0	2	176.00	88.00
3	US5032284	6	0	0	0	3	9	9.00	1.00
4	US5054556	0	2	0	0	0	2	5.00	2.50
5	US5167834	7	0	0	0	3	10	13.00	1.30
	Profit 1	44.8	13.8	6.9	6.9	27.6	100.0		
	Profit 2	6.5	6.2	4.0	76.2	7.0	100.0		

P=total number of CPC codes included in a patent. For example, first patent has 6 codes (E21B=2, F16J=2 and Y10S=2)

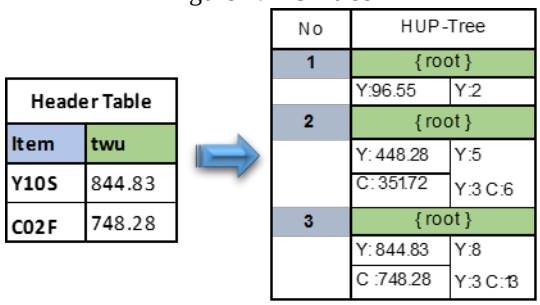
All steps needed to apply HUP-growth mining algorithm, which were identified in [26] are utilized based on Table 1 to show how the algorithm works in practice. In previous studies, the profit value was directly given by user for calculations. Unlike previous studies, two optional formulas are developed in this study to calculate the profit value of each CPC class in patent data. In the first formula, we have calculated a profit value by dividing the total number of CPC class to the number of CPC class in each transactions (Profit 1 in Table 1). In the second formula, we calculated a profit value by considering the forward citation numbers of each patent and distributing it to the CPC class based on the number of CPC lass in the corresponding transaction (Profit 2 in Table 1). In this extended abstract, some of the results obtained from geothermal energy related patents are given for readers. It should be also noted that profit 1 is used to determine the importance of the CPC code. The utilities and the occurrence frequencies of all items in each transaction with minimum high utility (mhu) 35% for Table 1 is given in Table 2.

Table 2: The utilities and the occurrence frequencies of all items in each transaction

		1	2	3	4	5	
PID	Document No.	C02F	E21B	F16J	F25B	Y10S	TU
1	US9660285	0.00	27.59	13.79	0.00	55.17	96.55
2	US9312546	0.00	0.00	0.00	13.79	0.00	13.79
3	US9419292	268.97	0.00	0.00	0.00	82.76	351.72
4	US8701468	0.00	27.59	0.00	0.00	0.00	27.59
5	US9029030	313.79	0.00	0.00	0.00	82.76	396.55
	au	582.76	55.17	13.79	13.79	220.69	886.21
	twu(If twu>au)	748.28	124.14	96.55	13.79	844.83	
						310.17	<-- mhu

HUP tree is constructed based on Table 2. Fig.1 show HIP tree.

Figure 1: HUP tree



The High Utility Itemset(s) for Table 1 is computed and the results are given in Table 3 which shows the first two most important rules based on minimum high utility value, 310.17. As can be seen from Table 2, C02F class as individually and Y10S with C02F classes are high utility opportunity areas. Technology areas related to these CPC codes should be highly focused by investors and decision makers.

Table 3: High Utility CPC classes

Itemset(s)	Utility value
C05F	582.76
C02F + Y10S	748.28

Besides this illustration example on very small data which just includes 5 patents, the result of our big patent data (included 499 patents) related to Geothermal Energy is also provided in Table 4.

Table 4: High Utility CPC classes for Geothermal Energy Technology

Itemset(s)	Utility value
Y02E	6271.96
Y02E + F24J	7422.27

III. CONCLUSION

Technology evaluation is very critical task for companies to increase competitive power of companies. A good technology evaluation can help decision makers and investors to adapt contemporary methods and technologies in their company. Patent data is extensively used for technology evaluation in the literature. Data mining based algorithms are also used for technology evaluation. Integration of data mining methods with patent data provides an objective evaluation of technologies. In this study, HUP-growth mining algorithm as a data mining method, is applied to patent documents related to geothermal energy. In addition to patent data, experts' opinion can be also used and a detailed research based on the several well-known data mining algorithms, such as classification, will be conducted to go into the details of geothermal energy in the future.

IV. REFERENCES

[1] E. Lichtenthaler, Organisation der Technology Intelligence: eine empirische Untersuchung in technologieintensiven, international taetigen Grossunternehmen, Zürich: Dissertation ETH Zurich, 2000.

[2] K. Hian Chye and L. Chan Kee, "Going concern prediction using data mining techniques," *Managerial Auditing Journal*, vol. 19, no. 13, pp. 462-476, 2004.

[3] J. Zhanga, S. O. Williams and H. Wang, "Intelligent computing system based on pattern recognition and data mining algorithms," *Sustainable Computing: Informatics and Systems*, vol. In Press, October 2017.

[4] K. Jeeeun and L. Sungjoo, "Patent databases for innovation studies: A comparative analysis of USPTO, EPO, JPO and KIPO," *Technological Forecasting and Social Change*, vol. 92, pp. 332-345, 2015.

[5] E. Lichtenthaler, "Technological change and the technology intelligence process: a case study," *Journal of Engineering and Technology Management*, vol. 21, no. 4, pp. 331-348, 2004.

[6] H. Tschirky, "The role of technology forecasting and assessment in technology management.," *R&D Management*, vol. 24, no. 2, pp. 121-129, 1994.

[7] W. J. Abernathy and K. B. Clark, "Innovation: Mapping the winds of creative destruction," *Research Policy*, vol. 14, no. 1, pp. 3-22, 1985.

[8] D. Sahal, "Technological guideposts and innovation avenues," *Research Policy*, vol. 14, no. 2, pp. 61-82, 1985.

[9] M. L. Tushman and P. Anderson, "Technological Discontinuities and Organizational Environments," *Administrative Science Quarterly*, vol. 31, no. 3, pp. 439-465, 1986.

[10] R. M. Henderson and K. B. Clark, "Architectural innovation: the reconfiguration of existing product technologies," *Administrative Science Quarterly*, vol. 35, no. 1, pp. 9-30, 1994.

[11] C. M. Christensen, The Innovator's Dilemma: When New Technologies Cause Great Firms To Fail, Harvard Business School Press, 1997.

[12] A. Francis J., Scanning the business environment, New York: Macmillan Co., 1967.

[13] B. James R., "Evaluating Signals of Technological Change," *Harvard Business Review*, vol. 3, no. 4, p. 459, 1988.

[14] C. A.C. and S. D., "Strategic responses to technological threats," *Business Horizons*, vol. 19, no. 1, pp. 61-69, 1976.

[15] C. Changwoo and P. Yongtae, "Monitoring the organic structure of technology based on the patent development paths," *Technological Forecasting & Social Change*, vol. 76, no. 6, pp. 754-768, 2009.

[16] D. Tugrul U., R. Guillermo, M. Hilary and P. Gerdsri, "Forecasting emerging technologies: Use of bibliometrics and patent analysis," *Technological Forecasting and Social Change*, vol. 73, no. 8, pp. 981-1012, 2006.

[17] R. M.D., G.-R. J., M.-S. M., R. M. and M.-B. M.J., "Information fusion from multiple databases using meta-association rules," *International Journal of Approximate Reasoning*, vol. 80, pp. 185-198, 2017.

[19] J. C.-W. L. Wensheng Gan a, F.-V. Philippe, C. Han-Chieh and Z. Justin, "Mining of frequent patterns with multiple minimum supports," *Engineering Applications of Artificial Intelligence*, vol. 60, pp. 83-96, 2017.

[20] H. Jiawei, P. Jian, Y. Yiwen and M. Runying, "Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach," *Data Mining and Knowledge Discovery*, vol. 8, no. 1, pp. 53-87, 2004.

[21] A. Rakesh, I. Tomasz and A. Swami, "Database mining: a performance perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 5, no. 6, pp. 914-925, 1993.

[22] R. v. S. R. Agrawal, "Fast Algorithms for Mining Association Rules in Large Databases," in *VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases*, San Francisco, 1994.

[23] G. Wensheng, L. Jerry Chun-Wei, C. Han-Chieh and Z. Justin, "Data mining in distributed environment: a survey," *Data Mining and Knowledge Discovery*, vol. 7, no. 6, p. e1216, 2017.

[24] C. Raymond, Y. Qiang and S. Yi-Dong, "Mining high utility itemsets," in *Data Mining, 2003. ICDM 2003. Third IEEE International Conference on*, 2003.

[26] L. Chun-Wei, H. Tzung-Pei and L. Wen-Hsiang, "An effective tree structure for mining high utility itemsets," *Expert Systems with Applications*, vol. 38, p. 7419–7424, 2011.

[27] L. Kyungpyo and L. Sungjoo, "Patterns of technological innovation and evolution in the energy," *Energy Policy*, vol. 59, pp. 415-432, 2013.

A Real-time and Secure Patient Monitoring System (RSPMS)

H. KAYA[1], F.V. ÇELEBİ[1], T. YILMAZ[1], M.B. MURATOĞLU[1] and M. S. ERARSLAN[1]

[1] Ankara Yildirim Beyazit University, Ankara/Turkey, hilalkaya@ybu.edu.tr
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, fvcelebi@ybu.edu.tr
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, tugbaylmz001@gmail.com
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, mbetul8@gmail.com
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, erarslanmelike95@gmail.com

***Abstract* - The main objective of this research is design and realization of a wireless, secure, remote monitoring, control and feedback system using GSM technology for the patient health. In this study, CRUD (Create, Read, Update and Delete) operations can be performed on a Windows Communication Foundation (WCF) server using the developed mobile application and MySQL database. This system also includes a ripped QR scanner that prevents the browsers from entering QR code information because a crypto QR definition of the patient's identity number was developed. All the patients' assay results and their radiological views are recorded along their reports. Patients who need regular follow-up can send their daily, weekly and monthly results to the physician using this application. In an instant panic situation, this system can communicate to the 112-emergency services and by the help of a push-to-talk button, system can send the patient's instant location to the phone number indicated by short message service (SMS). If 112-emergency departments support the connection, system can also send this location information automatically to 112-emergency systems. Transmitted data is archived and visualized both on a mobile phone and on a central server. The experiments on the proposed system gave promising results that is accurate in scanning, clear in monitoring, intelligent in decision making and reliable in communication.**

***Keywords* - Patient monitoring system, QR code, secure, emergency**

I. INTRODUCTION

Health is one of the global challenges for humanity [1]. According to the constitutions of World Health Organization (WHO), the highest attainable standard of health is a fundamental right for an individual [2]. In the last years the healthcare monitoring systems have attracted growing interest of the researchers and in these studies, the main goal was to develop a reliable patient monitoring system (PMS) so that the healthcare professionals can monitor their patients, who are either hospitalized or executing their normal daily life activities [3]. PMS is a very critical monitoring system that it can monitor physiological signals including electrocardiograph (ECG), respiration rate, invasive and non-invasive blood pressure, oxygen saturation in human blood (SpO2), body temperature and other gases etc. during the treatment process. It is a quite hard job for medical personnel to monitor each patient for 24 hours [4]. The passive data gathering process may also permit clinicians to focus their efforts on diagnosing, educating, and treating patients, theoretically improving productivity and efficiency of the provided care [5].

The literature on remote patient monitoring (RPM) reveals enthusiasm over its promises to improve patient outcomes, reduce healthcare utilization, decrease costs, provide abundant data for research and increase physician satisfaction [5-7]. One of the main purposes of the RPM is to monitor the status of the patient in a longer period to check a longer-term stability of the status and determine the effect of prescribed medications. The other one is to react immediately in case that some of measured parameters are getting worse than expected or changing in a non-standard way [8]. Advancements in electronic data management and communication have led to remarkably capable systems of remote patient monitoring and the development of telemedicine. The potential for improving Intensive Care Unit (ICU) care by 24-hour remote monitoring (and interaction with the bedside team) was demonstrated almost 10 years ago and has been suggested to deal with the shortage of ICU physicians [9].

Developments in high bandwidth public wireless networks and new generation personal mobile devices give rise to new mobile healthcare services. Due to this emergence, the MobiHealth system provides customizable vital signs tele-monitoring and tele-treatment system based on a body area network (BAN) and a mobile health care (m-health) service platform utilizing next generation public wireless networks. The developed system allows the incorporation of medical sensors via wireless connections, and the live transmission of the measured vital signs to healthcare experts as well as real-time feedback to the patient [10].

E-Nabız is a personal healthcare system provided by Ministry of Health in Turkey. Anybody who has an account can access the personal health information from a single location, regardless of where the examinations and treatments are conducted. It is the world's largest and most comprehensive health information subdivision that can be accessed safely on the internet or by mobile apps, which allows to assess the personal health records in the context of patient's own, time-bound, and limited jurisdictions so that they can be evaluated

by physicians, thereby increasing the quality and speed of the diagnosis and treatment process and establishing a strong communication network between the patient and the physician [11].

Another remote healthcare system Doktorderki [12] is a consultation model that works with professional doctors from different branches. Patients can consult some experts for free, share their labs by the help of the messaging system and they can get a quick return from the doctors. A commercial application iCare [13] provides personal health programs (diet, sports) and health risk assessment for users by collecting and analyzing the health information with wearable device that makes phone measurement.

In this study, a wireless, secure, remote patient monitoring, control and feedback system was designed and developed using GSM technology for patient health. Doctors can access their regularly monitored patients' assay results and their radiological views that are recorded along their reports by scanning their QR code. This QR code includes the patient's personal information in an encrypted form. Patients who need daily follow-up (such as blood pressure, blood sugar, heart rate) can send their daily, weekly and monthly results to the physician using this application. If there is an extreme situation in their results out of the reference interval, doctor will be notified automatically. Also, a step counter application is available for the users.

II. PROPOSED METHOD (RSPMS)

The main objective of the proposed system is design and realization of a wireless, secure, remote patient monitoring, control and feedback system using feedback and GSM technology for patient health. System can be assessed in two main parts, for patient's and the doctor's usages.

a. Patient's Interfaces

Patient will be registered in this system by the regularly following doctor and after the registration process, a specific QR code will be created for the patient that has not been applied before. By the help of the QR code in the patient's file and on the patient's application screen, doctor can quickly access to the patient's all recorded information.

For chronic diseases, patient can send the necessary data to the doctor in daily, weekly and monthly periods. System tends to send a notification to the doctor in case that the tracking records are out of the normal range. The application will automatically generate graphs for the tracked health records of the patient to let the follow-up values of the patient visualized and monitored by the doctor more easily.

On the patient's application screen, also a 112-emergency button is available which can be accessed in case of emergency and the emergency phone numbers screen lets to add and edit the desired numbers. Using the Appointment section, patient can create a suitable appointment from the list of the date and time offers in a user-friendly environment. Doctor and the patient can send messages to each other within the application in an interactive communication environment.

b. Doctor's Interfaces

With the QR code created for the patient, the doctor will scan the QR code instead of reviewing it from the file and quickly access the patient's file and all assay information. Application automatically reports the data regularly entered by the patient and automatically creates the graph at specific intervals, so doctors can make their follow-up more effectively. If the patient's tracking values are out of the normal value ranges, the doctor will be automatically alerted. At this point the doctor will be able to make the necessary medical intervention on time.

Patient can access the desired numbers and 112 services in an emergency. In this case, doctor is also informed by the system via SMS. With in-app messaging and appointments, doctor-patient interaction is improved, and it provides a more convenient and secure environment for patients.

Use-case diagram of RSPMS can be seen in Fig. 1.

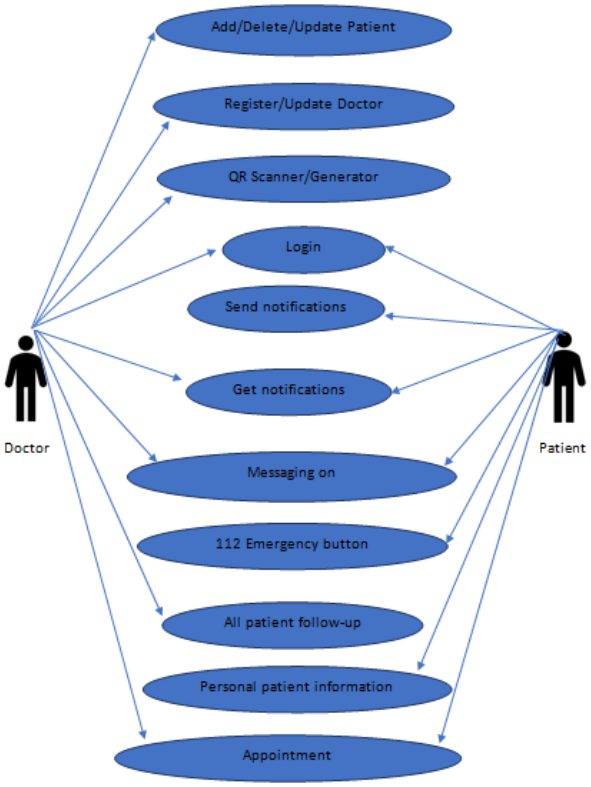


Figure 1. Use-Case diagram of RSPMS.

CRUD (Create, Read, Update and Delete) operations can be performed on a Windows Communication Foundation (WCF) server using the developed mobile application on MySQL database. WCF is one of the latest technologies of Microsoft

that is used to build service-oriented applications. Based on the concept of message-based communication, in which an HTTP request is represented uniformly, WCF makes it possible to have a unified API irrespective of diverse transport mechanisms [14]. The MySQL Database gives power to Web, E-commerce and Online Transaction Processing (OLTP) applications. It is a fully integrated transaction-safe, ACID compliant database with full commit, rollback, crash recovery and row level locking capabilities. It delivers the ease of use, scalability, and performance and it's a popular open source database. Some of the world's most trafficked websites like Facebook, Google, Twitter, Uber, and Booking.com rely on MySQL for their business-critical applications [15].

Database diagram of RSPMS can be seen in Fig. 2.

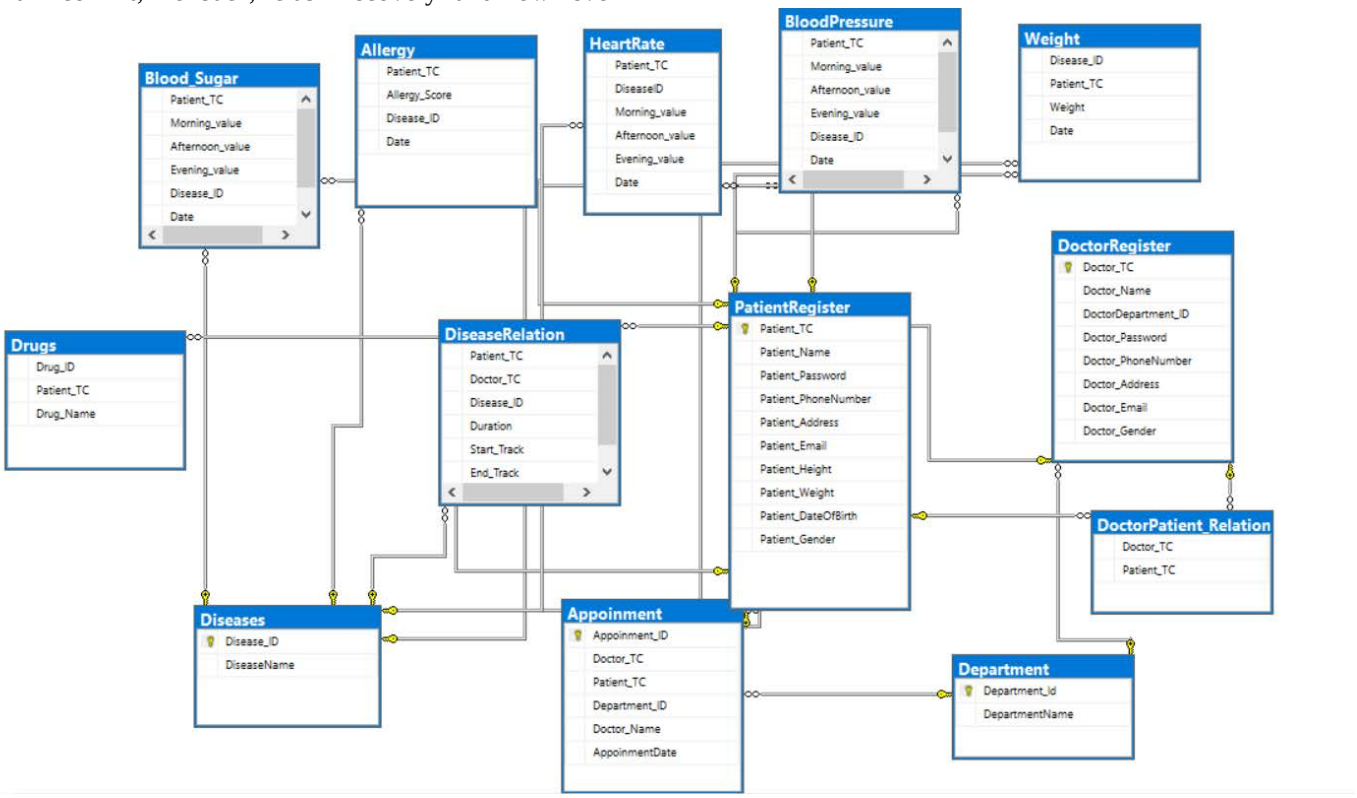


Figure 2. Database diagram of RSPMS.

QR Code or Quick Response Code is a two-dimensional barcode that can be read by modern smartphones and special QR Code scanner devices. A QR code consists of black squares arranged in a square grid on a white background. There are several variants of QR codes depending on their symbol size, layout, encoding and structure. QR generator and scanner is developed by using Zxing library on Java language. Also, the generated QR codes are encrypted before storing on the database.

Functional requirement specifications of QR generator and QR scanner can be seen in Table 1 and Table 2, respectively.

Table 1. Functional requirement specifications for QR generator

Use Case Name	QR Generator
Primary Actor	Doctor
Trigger	Doctor generated QR code for new patient.
Precondition	The doctor accesses the patient tracking systemusing "Add new patient". There is at least one doctor in the system.
Basic Path	1. Patient goes to doctor's clinic and requests to enroll in the system. 2. Doctor accesses to patient tracking systemby clicking "add new patient" button. 3. The system presents a blank grid to enter the patient information. 4. The doctor enters the information of patient and records to registration form. 5. The system checks if the required fields are not blank. 6. Doctor clicks QR generator button. 7. QR code is generated according to each patient's specific identification number.

	8. The system displays the message of "New patient is added".
Alternative Path	*System Fails 1. The system displays error message "Please, try again later". 4.a. If in step 4, there is any blank field, the system displays error message "Blank fields must be filled" Go to step 3 4.b. If in step 4, there is already a patient that has this identity number in the system, system displays warning message of "This identity number is already recorded in the system." Go to step 3 4.c. If in step 4, system does not validate the patient information, system displays error message of "Invalid information". Go to step 3.
Postcondition	The system registers the patient and generates the QR code.
Exception Path	The doctor may abandon the operation at any time.
Other	The patient information includes patient TC, name, password, date of birth, gender, address, phone number, email, chronic diseases and drugs.
Frequency of Occurrence	High (High, Medium, Low)

Table 2. Functional requirements specification for QR scanner

Use Case Name	QR Scanner
Primary Actor	Doctor
Trigger	Doctor's scanner QR code for registered patient.
Precondition	The doctor accesses the patient tracking system using "QR scanner". There is at least one doctor in the system.
Basic Path	1. Patient goes to doctor's clinic and requests to be treated. 2. Doctor accesses to patient tracking system who clicks QR scanner button. 3. Doctor scans the patient's QR code to access the patient information.
Alternative Path	*System Fails 1. The system displays error message of "Please, Try again later". 3.a. If patient's QR code is not available, doctor can access patient information using my patient screen. 3.b. If scanner cannot find patient QR code correctly, doctor can access patient information using my patient screen.
Postcondition	The doctor scans QR code successfully and reach the patient information.
Exception Path	The doctor may abandon the operation at any time.
Frequency of Occurrence	High (High, Medium, Low)

RSPMS includes one general user interface for signing up the system that can be seen in Fig. 3.

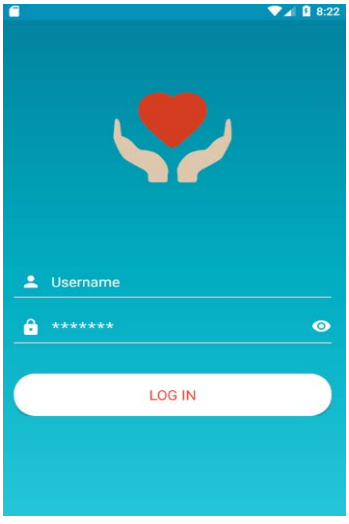


Figure 3. Sign up interface of RSPMS

Patient interfaces and doctor interfaces can also be seen in Fig. 4 and Fig. 5, respectively.

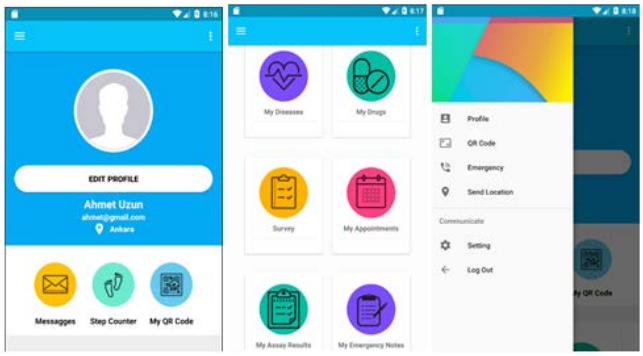


Figure 4. Patient interfaces of RSPMS

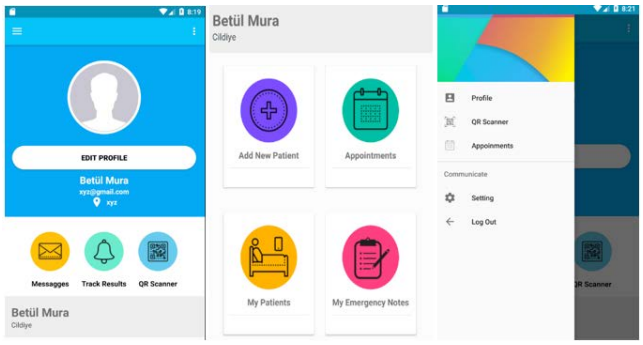


Figure 5. Doctor interfaces of RSPMS

III. CONCLUSION

In this study, a wireless, secure, remote patient monitoring, control and feedback system was developed using GSM technology for patient health. RSPMS aims to facilitate the follow-up procedure of regularly monitored patients by the doctor. This system can be used by private health institutions, private clinics etc. User-friendly design, securely storing information structure and the ability to receive all kinds of summarized reports make the system preferable. Conducted user tests on the system gave promising results. As a future work, we would like to incorporate the medical sensors via wireless connections to this systemto automate the information sending-out process.

REFERENCES

[1] Global Challenges for Humanity available at http://www.millenniumproject.org/millennium/challenges.html
[2] A Right to Health available at http://www.who.int/mediacentre/factsheets
[3] A. Abdullah, A. Ismael, A. Rashid, A. Abou-Elnour, M. Tarique (2015). Real Time Wireless Health Monitoring Application Using Mobile Devices. International Journal of Computer Networks & Communications. 7. 13-30. 10.5121/ijcnc.2015.7302.
[4] S.M.Y. Jafri (2010). IP Based Patient Monitoring Systems, Final Master's Degree Thesis, Electrical and Biomedical Engineering, University College of Borås.
[5] H. Oh, C. Rizo, M. Enkin, A. Jadad (2005). What is eHealth (3): a systematic review of published definitions. J. Med. Internet Res. 7, e1.
[6] S. Ajami, F. Teimouri (2015). Features and application of wearable biosensors in medical care. J. Res. Med. Sci. 20, 1208–1215.
[7] S.R. Steinhubl, E.D. Muse, E.J. Topol (2015). The emerging field of mobile health. Sci. Transl. Med. 7, 283rv283.
[8] G. András, C. Kft, ReHealth System, FI-PPP Workshop, Budapest, June 2013.
[9] S. Dang, A.G. Golden, H.S. Cheung, B.A. Roos (2010). Telemedicine Applications in Geriatrics, Elsevier Brocklehurst's Textbook of Geriatric Medicine and Gerontology (Seventh Edition), Chp 128, 1064-1069. https://doi.org/10.1016/B978-1-4160-6231-8.10128-X
[10] K. Wac, R. Bults, B.V. Beijnum, I.Widya, V.Jones, D.Konstans, M.H.Vollenbroek, H.Hermens (2009). Mobile Patient Monitoring: the MobiHealth System, Conf Proc IEEE Eng Med Biol Soc. 1238-41. doi: 10.1109/IEMBS.2009.5333477.
[11] Ministry of Health, E-Nabız Personal Healthcare System available at https://enabiz.gov.tr/
[12] Doktorderki Consultancy Model available at http://doktorderki.com
[13] iCare State Monitoring System available at https://play.google.com/store/apps/details?id=comm.cchong.BloodAssistant&hl=tr
[14] Tutorials on Windows Communication Foundation (WCF) Server available at https://www.tutorialspoint.com/wcf/wcf_overview.htm
[15] MySQL Database System available at https://www.mysql.com/it/trials/
[16] Zxing QR Code Generating by Java available at https://www.callicoder.com/generate-qr-code-in-java-using-zxing/

Decreasing Failure and Turnover Rates In Double-Glazing Production With Failure Mode And Effect Analysis and An Application

A. AKTEPE[1], B. B. SAKAR[1], U. AYDIN[1], A. N. HAYYAOGLU[1], S. ERSOZ[1]

[1]Kırıkkale University, Faculty of Engineering Department of Industrial Engineering, Kırıkkale, aaktepe@gmail.com
[1]Kırıkkale University, Faculty of Engineering Department of Industrial Engineering, Kırıkkale, b.bestesakarr@gmail.com [1]Kırıkkale University, Faculty of Engineering Department of Industrial Engineering, Kırıkkale, muutaydinn@gmail.com
[1]Kırıkkale University, Faculty of Engineering Department of Industrial Engineering, Kırıkkale, hyy.ayse@gmail.com
[1]Kırıkkale University, Faculty of Engineering Department of Industrial Engineering, Kırıkkale, sersoz40@hotmail.com

Abstract **- The technology to calculate the fire and failure rate of the glass from the glass line in the glass factories that are manufactured is facilitated by the technology. Waste and mistakes must be identified and improved for product and service quality. One of the methods used to calculate these rates is Failure Mode Effect Analysis (FMEA). The FMEA technique is a very powerful numerical analysis technique for preventing errors before they occur. In this study, a sample application of the Process FMEA study, one of the FMEA variants, was carried out for the glass sector. The problem to be addressed is defined as determining the error and waste rates of the glass from the thermopane line in a quality oriented business with the Perfect Scanner machine to reduce these ratios and increase the productivity. Types of errors obtained from Perfect Scanner data; the questionnaire survey conducted in the production area and the feedback from the customers were collected in 5 issues. These; scratches, stains (fingerprints), coating failure, stains between laminating glass, air bubbles. We determined and implemented the changes that will improve the process to eliminate or reduce the reasons for the types of errors encountered. Thus, by measuring the performance of the thermopane line of this company which has an important place in the glass industry, the negativity of the performance in production is minimized and customer satisfaction is increased.**

Keywords **- Glass Industry, Failure Mode Effect Analysis, Thermopane.**

1. INTRODUCTION

Glass, one of the oldest and mature materials of the world, shows that it will serve humanity in future times with its different characteristics. It is estimated that the world annual glass production capacity is approximately 180 million tons in quantity and 130-140 billion dollars in value. The world glass industry is growing at an annual average rate of 2-4%. According to 2017 values; 53% of the glass sector belongs to glass packaging, 29% to flat glass, 5% to glassware, 2% to glass fiber and 11% to other products. The production capacity of our country's glass industry is about 3.5 million tons. [1]

In processes from raw material to final product in production facilities, raw material losses which can't be used for the product of total raw material can be defined as waste. Every management function wants to have maximum productivity from the raw materials and auxiliary materials it receives, and this is also inherent in the glass industry.

Productivity refers to the physical relationship between the amount of output (output) and the amount of production factors (input), as well as being an abstract concept that measures the sensitivity and effectiveness of vehicles in achieving economic goals. This application is designed to reduce the error and waste rates of products in glass production or to find out where the faults originate. [2]

Production errors can create adverse conditions, such as reduced quality image, especially on the basis of customer satisfaction after sales, high risk of legal risk or low performance in manufacturing. Manufacturers of glass products have to choose intelligent and flexible automation systems to control product quality and production process 100%.

Failure Mode and Effect Analysis is intended to prevent existing errors, but also to eliminate the effects that would occur if these errors were generated by destroying them at the source that is not present but likely to occur. The ease of implementation and availability for all sectors makes FMEA technology more advantageous than other techniques. In this study FMEA technique was used in the detergent sector. Firstly, Introduction to Quality Function Dynamics, Experimental Design, Simultaneous Engineering Techniques, Failure Mode and Effect Analysis Technique are explained in detail. In a factory that produces detergents afterwards, Process FMEA was applied for the production process and the results were evaluated. [3]

Determination of the most important foot risks of the management system. There are many methods of risk analysis. In Failure Mode and Effect Analysis (FMEA), systematic faults are a method that enables the detection of hazards without causing accidents and starting from the top priority. In this study, the risk assessment of the health and safety of workers, risk management, types of error and analysis of the risks were taken into consideration and risk assessments were made with FMEA (Failure Mode and Effect Analysis), 197 risk factors were determined and 166 remedies were proposed. (Risk Priority Number) average was reduced to 51.72. [4]

A theoretical framework for Failure Mode and Effect Analysis has been established and an implementation study has been carried out within this framework. In the first phase of the study, information on Failure Mode and Effect Analysis was given. Firstly, it was emphasized when and how this method emerged and when it should be started. In addition, the purpose of the study and the benefits to be provided to the company are

stated. In the second phase, the implementation process of Failure Mode and Effect Analysis is explained. The steps to be followed for the work to be carried out have been specified and as a result of these steps the transfer to the Failure Mode and Effect Analysis Form has been shown. In the third phase, the types of Failure Mode and Effect Analysis, including system, design, process and service, the relationships between them and the application phases are mentioned. In the fourth stage, the application work done at a firm is explained. The application work was done in two parts as design and process type of error and analysis of the effects, and the results obtained were given in forms. [5]

In the risk appraisal, Failure Mode and Effect Analysis (FMEA) method, which is used in various sectors, was applied to a textile factory including pre-finishing, dyeing, printing, finishing processes. In addition, the similarities and differences have been investigated by considering Hazard and Operability (HAZOP) method which is similar to FMEA method and frequently used in chemical processes. The results of the process FMEA risk assessment were numerically calculated and risk priority numbers were created and ranked according to their importance. Biological and radiation factors have not been studied in the operation, and the most common risk factors are fire and explosion, mechanical and chemical agents. [6]

The aim of the risk assessment laboratory is to establish and analyze the causes of physical risk factors that may affect the health of employees and to establish a healthy laboratory work environment for all staff in all processes related to laboratory services provided in the field of biochemistry in hospitals. [7]

Fuzzy Failure Mode and Effect Analysis (Fuzzy FMEA) method; iron and steel forming (rolling) operations in order to prevent the risk of loss of goods and time, we worked on a sample rolling mill. Risk Assessment Methods Inputs obtained by the FMEA method are different from the exact values in the classical FMEA method, and the Fuzzy Logic method has been introduced to detect the existing risk risks. [8]

With the help of Fuzzy Analytic Hierarchy Process (Fuzzy AHP) and Gray Relational Analysis (GIA), he planned to eliminate these deficiencies. Fuzzy AHP is used to give importance to decision makers and to weight decision factors of first risk priority number (ROS1) and second risk priority number (ROS2). Put forward to prove the effectiveness of this methodology in a case study was carried out yarn manufacturing company in Turkey. [9]

The aims of this study;

- Performing remedial remedies,
- Reduction of waste and faults in the lines that significantly affect the profit / loss situation with Failure Mode and Effect Analysis (FMEA) metod,
- Increase productivity and reduce waste and errors.

In the next section, the FMEA method is discussed. Then, we discuss application of the study and finally we interpert the results of study.

2. METHOD: FAILURE MODE AND EFFECT ANALYSIS (FMEA)

There are studies in the literature to reduce the mistakes in glass production. In this study, the performance of thermopane line is measured and the loss of performance in production is minimized. In this case, classification, zoning etc. In error analysis. analyzes have been done to make it easier to work with.

The problem to be addressed in this study is to determine the error and waste rates of the glasses coming out of the thermopane line in an enterprise oriented towards quality assurance, to reduce these rates and increase the productivity.

The technology to calculate the fire and failure rate of the glass from the thermopane line in the glass factories that are manufactured is facilitated by the technology. Waste and mistakes must be identified and improved for product and service quality. One of the methods used to calculate these rates is Failure Mode and Effect Analysis. The FMEA technique is a very powerful numerical analysis technique for preventing errors before they occur.

The FMEA method is listed below. [10]

3. APPICATION

A. Creation of Working Team

In figure 1, the workflow diagram we created for our solution approach is shown.

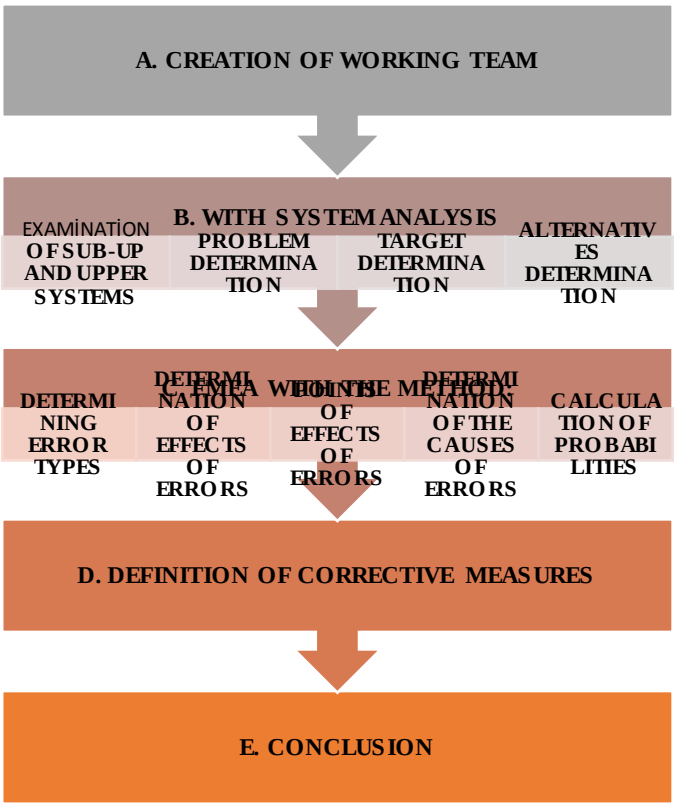


Figure 1. Work Flow Chart

For the "Process FMEA" study, which is one of the FMEA variants, a team consisting of the production manager, the

school lecturers and the students carrying out the project has been established from the quality manager in the business related to this process. This team has decided on the next steps in the form of a brain storming and has created a plan for what to do.

B. System Analysis

A flow chart of the process is drawn for the FMEA.

Table 1. Flowchart of Thermopane Glass Production

NO	O	▭	⇨	D	▽	Definition of Activity
1			*			Arrival of raw material
2					*	Raw material storage
3	*					Cutting glass to desired size
4			*			Carrying of broken glass
5	*					View of glass roving
6			*			Tempering of the glass which is subjected to the
7	*					Glass tempering process
8			*			Moving tempered glass to laminating line
9	*					Glass laminating process
10			*			Moving the laminating glass to the glass line
11	*					Thermopane process
12			*			Moving glass that has become an inscription to
13				*		Waiting for shipment of ready-made glass

The production started with the arrival of raw material factory is shaped according to the order received from the customer. In order to see the tempering process of the windows interrupted by the customer's desired scale, it is necessary to move and rode to the line. Laminated glasses are processed in a tempering machine. Depending on the customer's wishes, the laminating glass is moved to the thermopane line. The glasses that are made of the thermopane are packed and ready for shipment.

C. Application of the FMEA Method

i.Specifying Error Types

Table 2. Types of Errors

Error No	Type of Error
1	Object inside thermopane
2	Hair, stain between two glass (laminated)
3	Grid fault
4	Incorrect bail use
5	Air bubble formation between glass (laminated)
6	Crack in the glass
7	Coating defect, coating waste
8	Plane-appearance-direction fault in glass
9	Scratch in glass
10	Stain on the glass

Possible types of errors in the process have been determined. Possible error types have been defined since the system has been separated into its components. The type of error is the type of error observed. When the fault is connected to a function of the occurrence type, the type of fault is the inability or inability to perform the functions of a system. More generally, the type of error is the inability to perform an unsuccessful function or function. The type of error is defined by the physical properties. Causes from faulty system design or human-related problems. The type of possible error usually includes the type of error and the definition of the effect of the systemon its operation.

These are shown in Table 2.

Types of errors obtained from Perfect Scanner data; the questionnaire survey conducted in the production area, and the return from the customer. These are scratches, stains, (fingerprints) and coating failure.

ii.Identifying Potential Impacts of Errors

Table 3. Potential Effects of Errors

Error Type	Potential Effects of Error
Scratch	* Customer feedback. * Requires additional cleaning. * Processing a stained glass.
Stain	* Remanufacturing cost. * Time lost due to remanufacturing. * Customer feedback.
Coating Error, Coating Break	* Remanufacturing cost. * Time lost due to remanufacturing. * Customer feedback.

The potential effects of the types of errors are identified after determining the types of errors. It describes the reaction of the customer who encountered the error. The customer there may be the next section or the person or end user to process. In practice, it is often seen that the customer is chosen as the end user. The reason for this is that the amount of product buying is related to the user's level of satisfaction. In addition to this, the group, system, product, intermediate customer, supplier industry, compliance with the law, the consequences on the user safety, that is, the effects can be determined. [10]

iii.Scores of Effects of Errors

Table 4. Scoring Criteria

Error Occurrence Frequency	Scoring
Very High	10
	9
High	8
	7
Medium	6
	5
	4
Low	3
	2
Very Low	1

Table 5. Error Types and Scores

Error Type	Error Scores
Scratch	10
Stain	9
Coating Error, Coating Break	7

Here we grade the occurence frequency of errors. In this scope, the most common method is scoring from 1 to 10 according to the severity of the scale effect. The level of possible error results reflected by the severity and the customer is assessed. As the level effect of error increases, the weight increases. A Severity (S) score must be specified and recorded to identify a specific violence.

iv.Identifying Causes of Errors

The reasons for the error types are defined after the effects are identified. The factors that will be effective in the creation of the possible error path are defined as the cause. The causes of the faults indicate the reasons for the occurrence of problems during design. In order to reveal the cause of the error, the answer was searched in the question "What are the causes of the possible process variables in the type of error?".

Table 6. Error Types and Causes

Error No	Error Type	Causes of the Error
1	Scratch	* Machine-welded scratches, welded scratches, operator-originated scratches.
2	Stain	* Machine-welded stains, transport-welded stains, operator-welded stains.
3	Coating Error, Coating Break	* Machine welded glass welded, non-welded contact.

v.Calculation of

The Occurence (O) score should be determined if the likelihood of errors occurring in the error causes is scored. This scoring is also chosen according to the likelihood of occurrence of a fault from a scale ranging from 0 to 1.

After the detection of the causes, it is necessary to identify the types of error that are currently being detected in the system or in the process. It is important to note here that error control methods are not what is desired, but rather the system or process controls.

Control methods of faults are scored according to their ability to detect faults. The Detection (D) score is again determined according to the ability of the scale control method to detect an error from 0 to 1.

Once the above three factors (severity, occurrence, detection) are identified, the Risk Priority Count value can be calculated. The Risk Value is the number of criticality indicators and is calculated by multiplying these three factors by their scores. The Risk Value is the numerical value determined for each type of error or its cause based on three risk factors such as "severity", "occurrence" and "detection". In calculating the The Risk Value, the values assigned to a certain number of risk factors, which are verbally or probabilistically defined, are taken. With the The Risk Value, the risks for each type of fault are defined and the corrective measures to be taken to reduce the shortest period in the short term, from the longest period starting with the largest The Risk Value, are determined. [10]

Table 7. Rating Table

Error Type	Occurrence (Probability)	Severity (Effects)	Detection (Ease of Detection)
Scratch	0.7	8	9
Stain	0.6	6	9
Coating Error, Coating Break	0.3	4	8

Table 8. Risk Value

Error Type	Occurrence * Severity * Detection	Risk Value
Scratch	0.7*8*9	50.4
Stain	0.6*6*9	32.4
Coating Error, Coating Break	0.3*4*8	9,6

D. Corrective Measures

Verification that the proposed corrective measures have been taken to be implemented adequately, new results should be examined and assessed. It is of great importance in terms of the introduction of corrective measures. At this stage, the solutions are examined and evaluated until the critical risk values are lifted. [10]

The scratch as the largest type of error is the highest value among the risk values shown in table 8 and is the first most important fault type to be improved. Weekly maintenance of the washing machine is done regularly, reducing the number of errors caused by scratches.

To improve the error type coating error, which is smaller than the risk value, it was informed about the use of gloves for workers working on the current line in order to minimize the bare hand contact of the glass.

4. CONCLUSION

Within a few days a week in the factory during working hours, the Perfect Scanner machine and production errors were investigated. By observing how these mistakes are detected by these examinations, a discussion on how this might be done, if there is a way to prevent mistakes, has been obtained. 10 error types were determined based on the results of the analyzes. The types of errors mentioned are shown in Table 2.

Types of errors obtained from data; the production workshop survey and the return from the customer were collected in 3 chapters and identified as scratches, stains (fingerprints) and coating faults.

Specified types of faults were inspected at the beginning of the machine during production and tracked through glass passages along the line. As of January and February 2018, observations were made at the beginning of the machines during working hours every day on weekdays at the factory and data collection continued. It was not only limited to the thermopane line but also observed in other lines.

Inspection of defective lenses from customers during the factory time period was also very useful in terms of project flow by continuing to review the collection of Perfect Scanner and observational data. The plant has been actively working within the working hours, making observations, exploring the literature, exploring different solution methods and examining them and extending the perspective of the project.

- Stains, scratches, coating defects deeply located on the glass after the tempering process are a significant number of observations during observations and examinations.
- Lane glass is an important finding of stain and air bubble retention.
- As a solution to the use of old gloves that cause stains, corrective actions have been developed to draw attention to the maintenance frequency of the cleaning of the water in the washing machines, using new gloves.
- As a result of these activities, negative customer feedbacks have been reduced, contributing to the factory.
- In future work, the machines can be examined in more detail and applied in different sectors by further reducing the error and waste rates and making improvements.

REFERENCES

[1] Ministry of Science, Industry and Technology of Turkey, Industry General, Directorate, Sectoral Reports and Analyzes Series, Glass Sector Report, 2017.
[2] TÜBİTAK, Glass Industry Report, 1998.
[3] Yasemin Çeber, Application of Failure Mode and Effect Analysis (FMEA) in Manufacturing Sector, 2010.
[4] Ömer Kahraman, Risk Analysis with Failure Mode and Effect Analysis (FMEA) Method in Occupational Health and Safety in an Automobile Plant, 2009.
[5] Demet Gönen, Failure Mode and Effect Analysis and an Application Study, 2004.
[6] Filiz Aydin, Risk Assessment Method Application of FMEA to a Textile Factory, 2016.
[7] Mustafa Kemal Uslu, Assessment of The Physical Risks Faced By Laboratory Employees with Failure Mode and Effect Analysis (FMEA): A University Hospital Sample, 2016.
[8] Muhammet Çakmak, Application of Risk Assessment İn The İron and Steel Sector Together With Fuzzy and Risk Analysis (FUZZY FMEA) Method, 2015.
[9] Ali Kaan Pastırmacı, A Method For Risk Management: A Hybrid Method İn FMEA, 2014.
[10] Anonymous, (2012), Failure Mode Effect Analysis (FMEA), (Access Link: http://endustrimuhendisligi.blogspot.com.tr/2012/12/fmea-hata-turleri-ve-etkileri-analizi.html?m=1 , Date of Access: 10.03.2018).

A Literature Review on RFID Applications in Advanced Manufacturing Systems

Süleyman ERSÖZ[1], Ali Fırat İNAL[2], Adnan AKTEPE[3], Ahmet Kürşad TÜRKER[4]

[1] Kirikkale University, Kirikkale/Turkey, sersoz40@hotmail.com
[2] Kirikkale University, Kirikkale/Turkey, afinal@windowslive.com
[3] Kirikkale University, Kirikkale/Turkey, aaktepe@gmail.com
[4] Kirikkale University, Kirikkale/Turkey, kturker@yahoo.com

Abstract **- Radio Frequency Identification (RFID) is not a new technology, but it is a new tool for improving performance of the manufacturing systems. There are many references to RFID implementation in the literature. Increase of the efficiency and speed of processes and improvement of information accuracy are some of the common benefits of RFID implementation. But there are many benefits not seen at first glance. Also there is a cost of RFID implementation as well as benefits. In this study, the objective is to carry out a detailed literature review in this field and to compare the studies in the literature for discussing similarities and differences. Thus, the study leads to discover new opportunities to extend research in this field.**

Keywords **- RFID, RFID Implementation, Literature Review**

I. INTRODUCTION

Radio Frequency Identification (RFID) technology is an automatic recognition system that uses radio waves to identify products and objects. RFID is becoming increasingly important in the mobile and wireless communication technologies, and it is known that it will affect many sectors with different usage areas.

RFID technology basically comes from a tag and reader. RFID tags can perform tasks such as retrieving, storing and sending object information such as code, route and coordinates of a part.

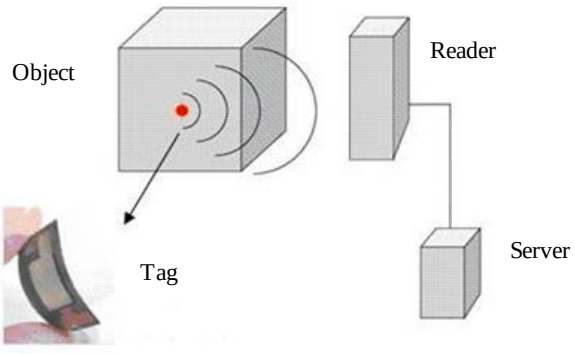


Figure 1: The working principle of RFID.

By reading the tag placed on the product or objects, the information can be automatically saved or changed. With this information, it is possible to easily determine which product belongs to which order, instantaneous changes can be made in the production schedule, and inventory tracking can be done more easily.

In the literature, much work has been done with RFID technology. While some of these studies are examining the benefits provided by RFID technology, a large part is about RFID integration in various sectors.

RFID technology is ideal for solving problems caused by insufficient control when operating. In order to read tags with RFID, you do not need direct view like Barcode system and more than one tag can be read at the same time. In addition, RFID can store more information than the Barcode system and can change them instantaneously.

RFID technology will have many benefits, but this technology will also have costs. Firms that want to take advantage of RFID technology want to know what they will do with this investment. At this stage, cost-benefit analysis is mentioned in the literature.

Cost-benefit analysis is a method of analysis used to compare alternative investment projects. Both private enterprises and, if necessary, various public institutions are often opposed to the use of limited resources such as labor, capital, machinery-equipment and raw materials in their hands in alternative investment areas. Selecting and sorting investment projects is a necessity to use scarce resources in the most efficient way, as the use of resources in a particular area will mean giving up on investments to be made in other areas. Cost-benefit analysis as a method of evaluating a project involves comparing the expected income incurred during the economic lifetime of the investment to the expenditure required by the project. More specifically, the expected revenues during the economic lifetime of the investment are reduced to the initial value of the investment (present value) with the aid of a certain discount rate, which is compared with the new product expenditure made.

The purpose of the study is to discuss RFID applications in the literature. Also in this study, in order to observe costs and benefits of different RFID systems, several studies on RFID implementation process are surveyed and listed in the next section. And finally, in the Conclusion Section, we discuss different aspects of RFID applications in the literature.

II. Literature Review

In this section, the studies in the literature are examined under four main topics.

A. Studies Based on Productivity in Manufacturing

Firstly, Vlachos [1] has shown that RFID applications provide a noticeable improvement in distribution systems and inventory control. Demiralp et al. [2] demonstrated cost savings opportunities between supply chain partners using RFID in the construction industry with a simulation study. Poon et al. [3] designed an RFID-based factory management system in which resources were tracked in real time to increase productivity and savings. Lee et al. [4] developed an RFID-based resource monitoring system for the apparel industry that needs to meet rapidly changing customer demands by companies. This shows that the fuzzy logic based resource monitoring system driven by the simultaneous production monitoring information will play a major role in controlling the plant. Moon et al. [5] used the RFID system to monitor the working efficiency of suppliers. They have also validated the benefits of the RFID system for a supply chain through performance testing and efficiency measurement. Han and Ko [6] have provided a useful example of the literature by integrating RFID into a tire company. Oner et al. [7] proposed an installation plan for RFID design, configuration and deployment, and they have created an RFID project that uses RFID tags in the denim industry. They also shown in the economic analysis, they can reduce the cost of human power and lost products. Ding et al. [8] designed an RFID-based social production system platform to track and manage production and transport processes. They also validated the effectiveness of the proposed platform by showing examples of simultaneous data collection, processing, and process monitoring.

B. Studies Based on Production Control and Scheduling

RFID technology is also very important for production control. Chongwatpol and Sharda [9] proposed different dynamic scheduling rules for the retail outlets with an RFID-based system that can be applied to store floors. Proposals have been proven through simulation tests based on actual data that the rules are better than traditional methods such as first-in-first-out (FIFO) or earliest-due-date (EDD). Zhong et al. [10] developed an RFID-based real-time production execution system, integrating it into the backdrop of a store specializing in mass-customization manufacturing. In addition, they have developed a scheduling rule using real-time scheduling pools and rules and have proven that the proposed algorithm eases process control. Barenji et al. [11] proposed an RFID-based, intelligent agent architecture to design a control and monitoring system at a manufacturing workshop. Lee et al. [12] developed an RFID-based process mining system for quality assurance. They used fuzzy relationship rules to show the link between process parameters and product quality. They have proven that by working in the apparel industry, improvements in the number of imperfections, in the production run and in reprocessing costs. Guo et al. [13] proposed an RFID-based decision support system to ease process monitoring and scheduling in a production environment requiring intensive labor.

C. Studies in Automotive Sector

A lot of research has been done in the literature in order to look at RFID technology in the automotive sector. Khan et al. [14] used RFID technology in the supply chain of the automotive industry in a 'closed loop' style due to the high RFID labes price and lack of infrastructure. Schmitt et al. [15] conducted a case study on the adaptation of RFID to the automotive sector. Foster et al. [16] have designed an architecture in which RFID tags are actively used for an automotive company that manufactures OEM parts so that parts procurement can continue uninterrupted. Manik et al. [17] showed the importance of Auto-ID technology in combination with RFID to solve data integrity problems arising from the manual entry of data into the system in supply chain operations in the automotive sector. In addition, they did a cost-benefit analysis of RFID integration to obtain simultaneous information on semi-finished products and stocks.

As mentioned in previous studies, the supply chain in the automotive sector is composed of many participants. Therefore; Close cooperation between participants for cost savings, cycle time reduction and efficiency enhancement is a critical factor for success and RFID based systems are ideal for this. As you can see, RFID technology is already used for various reasons in the automotive sector and very important investments are made every year in the automotive industry for the development of RFID and transportation systems. By using RFID technology in the automotive industry, firms are trying to eliminate the disadvantages of old technologies like Barcode. As a result, object identification can provide benefits such as acceleration, decrease in human error, increase in production and more healthy information flow. RFID technology will certainly benefit a lot in operating, but integrating such technology into the business is a difficult and costly process. Cost-benefit analysis is an ideal method to compare this benefit with cost.

D. Studies Based on Cost-Benefit Analysis

When we look at studies on the cost-benefit analysis of RFID integration, we found many studies done in various sectors. Kang et al. [18] have made a comprehensive literature review and designed a system for scheduling with RFID in a factory that manufactures automobiles. Also they developed an RFID-based sequence-error proofing system to avoid accidental line stops due to incorrect part sequencing and they used cost-benefit analysis to decide that this system should be installed or not. Hammer et al. [19] conducted a cost-benefit analysis of the integration of the UHF-RFID system to identify and instantaneously track animals in a farm operation. Kim [20] used RFID technology to improve productivity in a cargo company in Korea. In addition, RFID has made a cost-benefit analysis study by determining the investment cost and the benefits it will provide after installation. Roper et al. [21]

proposed a model for using RFID to design an automated monitoring system at a hospital and tested this investment by cost-benefit analysis. Erminio and Pilloni [22] conducted a cost-benefit analysis in a blood bank, suggesting that the RFID system should be used as an alternative to the labeling system so that the blood bags are transported more safely and that the blood in different specifications does not mix with each other. Tege and Verma [23] did a cost-benefit analysis of the idea of integrating the RFID system into the shelves of a store. With this system, empty shelves can be detected immediately or the amount of products stocked on the shelves can be seen. Uckelmann [24] conducted a comprehensive performance appraisal and cost-benefit analysis of the integration of RFID and the Internet of Things in the field of logistics. Khoo and Cheng [25] worked in a company that manufactures electronic parts, to set up a track monitoring system with RFID in order to prevent the stealing of some precious parts continuously. They have demonstrated that this system will be profitable for the company in the long run by making cost-benefit analysis of this system. Kumar et al. [26] highlighted the major challenges to be encountered in this integration process by examining a number of studies of cost-benefit analysis of RFID integration in the production supply chain. Üstündağ and Satoğlu [27] used simulation modeling to evaluate the effect of RFID on a pallet pooling system. They used the simulation output of this RFID application on a pallet supplier and a supply chain consisting of three customers in cost-benefit analysis. Lee and Lee [28] have worked on the investment process of RFID integration in the supply chain and have created a model that helps investors to make decisions using cost-benefit analysis. Wu et al. [29] focused on the shortcomings of the cost-benefit analysis method by examining many studies on RFID integration. They have devised a new methodology based on the idea that the cost-benefit analysis method is not flexible enough because future uncertainties do not add up. They have presented with proof that this method they have designed works reasonably.

III. CONCLUSION

i. As can be seen, many studies have been made in the literature on the cost-benefit analysis of RFID implementation. The common thought that these studies embrace is; Although the initial investment cost of RFID integration is high, the profit to be provided over time after the implementation process is completed will be high.

ii. As seen in the literature, RFID has been found to have a wide variety of uses. These are mainly the automotive industry, supply chain management and logistics sector.

iii. In the majority of the studies, RFID integration was found with the cost-benefit analysis method. This demonstrates the importance of companies and the academic community for RFID integration.

iv. In order to be able to classify the usage areas of RFID more clearly, a pie chart is shown in Figure 2 below.

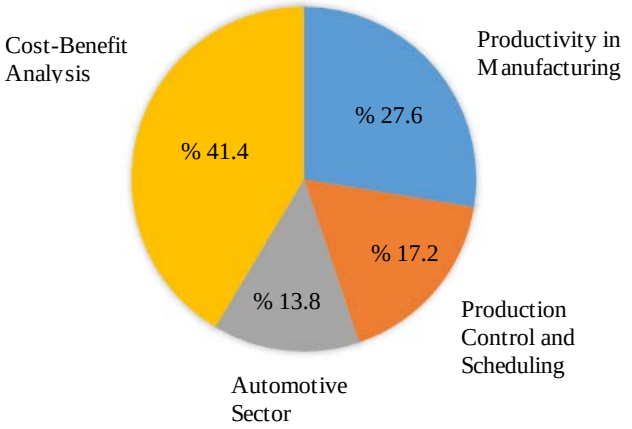


Figure 2: Pie chart of the studies carried out between 2007-2018.

REFERENCES

[1] Vlachos I., A Hierarchical Model Of The Impact Of RFID Practices On Retail Supply Chain Performance, Expert Syst. Appl., 41, 5-15, 2014.

[2] Demiralp G., Guven G., Ergen E., Analyzing The Benefits Of RFID Technology For Cost Sharing In Construction Supply Chains: A Case Study On Prefabricated Precast Components, Autom. Constr., 24, 120-129, 2012.

[3] Poon T., Choy K., Chow H., Lau H., Chan F., Ho K., A RFID Case-Based Logistics Resource Management System For Managing Order-Picking Operations In Warehouses, Expert Syst. Appl., 36, 8277-8301, 2009.

[4] Lee C., Choy K., Ho G., Lay K., A RFID-Based Resource Allocation System For Garment Manufacturing, Expert Syst. Appl., 40, 784-799, 2013.

[5] Moon S., Xu S., Hou L., Wu C., Wang X., Tam V.W., RFID-Aided Tracking System To Improve Work Efficiency Of Scaffold Supplier: Stock Management In Australasian Supply Chain, J. Constr. Eng. Manag, 144, 04017115, 2017.

[6] Han J., Ko Y., Knowledge Exploitation And Entrepreneurial Activity In A Regional Innovation System: First Adaption Of RFID At Kumho Tire In Gwangju, Korea, Eur. Plan. Stud., 25, 867-885, 2017.

[7] Oner M., Üstündağ A., Budak A., An RFID-Based Tracking System For Denim Production Processes, Int. J. Adv. Manuf. Technol., 90, 591-604, 2017.

[8] Ding K., Jiang P., Su S., RFID-Enabled Social Manufacturing System For Inter-Enterprise Monitoring And Dispatching Of Integrated Production And Transportation Tasks, Robot. Comput.-Integr. Manuf., 49, 120-133, 2018.

[9] Chongwatpol J., Sharda R., RFID-Enabled Track And Traceability In Job-Shop Scheduling Environment, Eur. J. Oper. Res., 227, 453-463, 2013.

[10] Zhong R.Y., Dai Q.Y., Qu T., Hu G.J., Huang G.Q., RFID-Enabled Real-Time Manufacturing Execution System For Mass-Customization Production, Robot. Computer-Integr. Manuf., 29, 283-292, 2013.

[11] Barenji R.V., Barenji A.V., Hashemipour M., A Multi-Agent RFID-Enabled Distributed Control System For A Flexible Manufacturing Shop, Int. J. Adv. Manuf. Technol., 71, 1773-1791, 2014.

[12] Lee C.K.H., Ho G.T.S., Choy K.L., Pang G.K.H., A RFID-Based Recursive Process Mining System For Quality Assurance In The Garment Industry, Int. J. Prod. Res., 52, 4216-4238, 2014.

[13] Guo Z.X., Ngai E.W.T., Yang C., Liang X., An RFID-Based İntelligent Decision Support System Architecture For Production Monitoring And Scheduling In A Distributed Manufacturing Environment, Int. J. Prod. Econ., 159, 16-28, 2015.

[14] Khan O., Scotti A., Leverano A., Bonino F., Ruggiero G., Dörsch C., RFID In Automotive: A Closed-Loop Approach, In Proceedings of the 12th International Conference on Concurrent Enterprising, Milano, Italy, 26-28, Haziran, 2006.

[15] Schmitt P., Thiesse F., Fleisch E., Adoption And Diffusion Of RFID Technology In The Automotive Industry, In Proceedings of the 15th European Conference on Information Systems (ECIS), St. Gallen, Switzerland, 7-9 Haziran, 2007.

[16] Foster P., Sindhu A., Blundell D., A Case Study To Track High Value Stillages Using RFID For An Automobile OEM And Its Supply Chain In The Manufacturing Industry, In Proceedings of the 2006 IEEE International Conference on Industrial Informatics, Singapore, 16-18 Ağustos, pp.56-60, 2006.

[17] Manik D., Toth L., Dobrossy P., Analysis Of RFID Application Through An Automotive Supplier's Production Processes, In Proceedings of the International Symposium on Computational Intelligence and Intelligent Informatics (ISCIII 2007), Agadir, Morocco, 28-30 Mart, pp.177-181, 2007.

[18] Kang Y.S., Kim H., Lee Y.H., Implementation Of An RFID-Based Sequencing-Error-Proofing System For Automotive Manufacturing Logistics, Applied Sciences, 8, 109, 2018.

[19] Hammer N., Pfeifer M., Staiger M., Adrion F., Gallmann E., Jungbluth T., Cost-Benefit Analysis Of An UHF-RFID System For Animal Identification, Simultaneous Detection And Hotspot Monitoring Of Fattening Pigs And Dairy Cows, Landtechnik, 72, 130-155, 2017.

[20] Kim M.S., A Cost-Benefit Analysis Of Case-Level Radio Frequency Identification Tagging-Based System On Logistics Service, Advanced Science Letters, 21, 3324-3328, 2015.

[21] Roper K., Sedehi A., Ashuri B., A Cost-Benefit Case For RFID Implementation In Hospitals: Adapting To Industry Reform, Facilities, 33, 367-388, 2015.

[22] Erminio P., Pilloni M.T., RFID Systems For Risk Reduction In Blood Bags: A Cost-Benefit Analysis, International Journal of Mechanics and Control, 16, 39-58, 2015.

[23] Tege S., Verma D.S., Cost Benefit Analysis Of Shelf Replenishment Of RFID In Retail Outlet, 2012.

[24] Uckelmann D., Performance Measurement And Cost Benefit Analysis For RFID And Internet Of Things Implementations In Logistics, Quantifying the Value of RFID and the EPCglobal Architecture Framework in Logistics, 71-100, 2012.

[25] Khoo B., Cheng Y., Using RFID For Anti-Theft In A Chinese Electrical Supply Company: A Cost-Benefit Analysis, IEEE Proceedings of the Wireless Telecommunications Symposium, 1-6, 2011.

[26] Kumar S., B. Kadow B., K. Lamkin M., Challenges With The Introduction Of Radio-Frequency Identification Systems Into A Manufacturer's Supply Chain-A Pilot Study, Enterprise IS, 5, 235-253, 2011.

[27] Üstündağ A., Satoğlu Ş., Cost-Benefit Analysis For RFID Based Pallet Pooling Systems, 2010.

[28] Lee I., Lee B.C., An Investment Evaluation Of Supply Chain RFID Technologies: A Normative Modeling Approach, International Journal of Production Economics, 125, 313-323, 2010.

[29] Wu X., Yue D., Bai J., Chu C., A Real Options Approach To Strategic RFID Investment Decision, IEEE International Conference on RFID, 314-321, 2009.

A Simulation Approach of Revenue Maximization Strategies for Turkish Domestic Airlines

A.YUCEKAYA [1]

[1] Kadir Has University, Istanbul/Turkey, ahmety@khas.edu.tr

Abstract **- The air transport has become the widely used option for the travel. The time saving and comfortable conditions encourage people to choose the airlines. The fierce competition and attractive market forced companies develop different pricing strategies to gain a competitive advantage. The ticket price, the charges for the luggage weight, special good carrying and seat selection are just many of the strategies developed by airline companies. The reaction of the customer is usually to choose the most economic option. Airline companies need a decision support system that will simulate the customer behavior against their strategies considering the strategies of their competitors. In this research we develop a simulation methodology for the domestic market in which we include different prices imposed by the airline companies and the buying preferences of the customers with the assumption that customer always choose the ticket with the minimum price. The airline companies then can analyze the consequences of their price strategies and expected strategies of their competitors.**

Keywords **- Air line transportation, simulation, decision support systems, Linear programming.**

I. INTRODUCTION

There is a fierce competition in the airline market and companies aim to maximize their revenues in this competitive environment. Customers may be confused to buy travel tickets given that companies provide multiple ticket options. The usual behavior of customers is to search offers and find out the best one that fits their expectations. Both good service and the cheapest price are expected by the customers.

Generally, an airline ticket consists of two parts: main and additional prices. Main ticket price is defined by the airline company without tax and service cost. This fare is offered into the market first and then prices rise as the departure draws closer and the seats are sold. The general policy would seem to be to sell a number of seats (perhaps 10) at the lowest fare and then increase the price in linear segments. A similar number of seats are then made available at the next price interval. The sales policy is then resumed until the price cap is reached or all the seats are sold. The dynamic pricing is often applied in other industries as well.

There are many services that are offered by airline companies. Some services require extra cost and therefore customers pay more ticket price. The services include many components. Tax, service price, travel insurance fee, extra baggage weight fee, special baggage fee, seat selection fee, fee for carrying pet animals, fee for taking SMS validation, meal cost and reservation fee are some of them.

Companies try to sell their products or services to specific customer community. According to their customer segmentation, they define their pricing strategies like low-cost strategy. It is possible to observe companies that apply different pricing strategies in airline industry. However, each strategy might have a different reaction from the customers and airline companies need to analyze the effect.

Some researches present simulation models for the market management in airline industry [1-2]. The simulations are aimed to help decision makers about their market operations. There are also researches about airline operations and revenue management considering cancellations and overbooking [2-5]. The revenue maximization for a single company is analyzed in the researches and the price dynamics are not discussed. The capacity management is also studied for an airline [6]. The capacity is the most valuable resource and it needs to be well managed.

We propose a linear model for the air companies to analyze the customer preferences for company selection. We develop a spreadsheet based decision support system that will automatically get the related airline data and simulate the different price offers and return the best strategy for the customer. The decision support system will be able to include customer preferences, the airline data and pricing strategies of airline companies. Then the system will simulate the passenger assignment process for all the customers and all the airline companies considering their capacity and price information.

II. PROBLEM DEFINITION AND SOLUTION APPROACH

Customers usually has more than one option to select among airline companies. The companies implement different pricing strategies to gain competitive advantages and to maximize their revenue in this competitive environment. The customers demand affordable service, cannot compare the airline companies and their services at once. This might cost them more money.

An airline price consists of a main price and additional service prices. The main price might be lower in a company but the additional services might cost more that mislead the customer. There are many additional services offered by the companies and demanded by the customers. The model should include all the parameters to have a clear cost structure. The

customers check offers from different companies and it is expected that the price offers be different and the customer to select the one with the minimum cost. The companies on the other hand need to model the behavior of the customer and develop pricing strategies that will maximize their revenue while attracting more customers by minimizing their cost. Table A which is given in Appendix shows the notation used in the model. The objective function for a customer can be represented as:

(1) $\text{Min } Z = \sum_{i=1}^{I} \sum_{j=1}^{J} Tji + Xji + Sji + Iji + GjiEji +$

$QjiHQji + LjiPLji + PjiPPji + SMSjiPSMSji + RjiHRji +$

$Mji\,HMji$

The customer aims to minimize its cost. The objective function includes the cost for main price, cost for special equipment, seat selection, pet transportation, reservation, meal service, extra baggage fee and SMS service. A customer can have one or more of these services and the pricing differs within companies. There are also constraints that need to be considered as below.

$$\sum_{j=1}^{J} \mathbf{B}ji \leq \mathbf{C} \quad \{i,j\} \in \{I,J\} \quad (2)$$

$$\sum_{i=1}^{I} \mathbf{B}ji = \mathbf{1} \quad \{i,j\} \in \{I,J\} \quad (3)$$

$G_{JI} + W_{JI} <= 32 \quad \{i,j\} \in \{I,J\} \quad (4)$

$HQ_{ji} - B_{ji} \leq 0 \quad \{i,j\} \in \{I,J\} \quad (5)$

$PL_{ji} - B_{ji} \leq 0 \quad \{i,j\} \in \{I,J\}$

(6)

$PP_{ji} - B_{ji} \leq 0 \quad \{i,j\} \in \{I,J\}$

(7)

$PSMS_{ji} - B_{ji} \leq 0 \quad \{i,j\} \in \{I,J\} \quad (8)$

$HR_{ji} - B_{ji} \leq 0 \quad \{i,j\} \in \{I,J\} \quad (9)$

$HM_{ji} - B_{ji} \leq 0 \quad \{i,j\} \in \{I,J\} \quad (10)$

$HQ_{ji}, \; PL_{ji}, PP_{ji}, PSMS_{ji}, HR_{ji}, HM_{ji}, \; B_{ji} \; э \; (0,1)$
(11)

Eq. (2) ensures that the capacity of the air line company is not violated while Eq.(3) ensures that each customer is assigned to a company. Equation (4) shows that the personal weight limit is 32 kg. Eq.(5-10) shows that a customer might have a different service request. Eq(12) shows that they are binary.

The companies need to analyze the competitors offer and the reaction of the customer to their offers and the price of the competitors. One way to analyze the reactions is to simulate the system using the data for the customers. A set of customer need to be determined for the analysis. Figure 1 shows the pseudocode of the solution approach.

0: Start
1: Set i=1, j=1,
2: For i=1 to I do
3: For j=1 to J do
4: Get $PT_i, X_{ji}, S_{ji}, I_{ji}, W_{ji}, G_{ji}, AW_{ji}, E_{ji}, Q_{ji}, L_{ji}, P_{ji}, SMS_{ji}, R_{ji}, M_{ji}$
5: Calculate offered prices for j, select min
6: Assign customer j to company i
7: Update $PTi, HP_{ji}, HM_{ji}, HR_{ji}, PSMS_{ji}, PP_{ji}, PL_{ji}, HQ_{ji}, B_{ji}$
8: Update filled seats for company i, check capacity limit
9: End for
10: End for
11: Calculate capacity utilization and revenue company i
12: End

Figure 1: Pseudocode for the simulation analysis

It is assumed that the data for the customer is available and a customer chooses the company with the lowest cost offer. Then it is more likely to get a randomly generated set for a simulation analysis. The algorithm first gets the demand for each customer and checks the price offer for each company. The customer is assigned to the company with the lowest cost and the capacity is updated for the next customer. The algorithm assigns all customers to the companies and then capacity utilizations and revenues for the companies are calculated. It is now possible for each company to analyze the results of their pricing strategies. A sensitivity analysis for any changing price strategy will require a new simulation run and the results would be analyzed.

It is also possible to integrate the solution approach to a decision support system. A spreadsheet based decision support system can be developed for a user friendly and easy adaptation.

III. CASE STUDY

This research will focus on pricing strategies of five main airline companies in Turkey: Turkish Airlines, Pegasus, Atlas, Onur and Sun Express. A Boeing 737 800 type aircraft with 180 passenger capacity is considered as it is used by all these five companies in Istanbul-Antalya flight line. The next table shows the relationship between interval of customer number and their prices:

Table 1: The data for the customers and their offered prices

Customer Number From	Customer Number To	THY	Pegasus	Atlas	Onur	Sun Express
0	10	59	39	59	69	59
11	20	59	49	69	77	67
21	30	79	59	79	85	75
31	40	79	69	89	93	83
41	50	99	79	99	101	91
51	60	99	89	109	109	99
61	70	119	99	119	117	107
71	80	119	109	129	125	115
81	90	139	119	139	133	123
91	100	139	129	149	141	131
101	110	159	139	159	149	139
111	120	159	149	169	157	147
121	130	179	159	179	165	155
131	140	179	169	189	173	163
141	150	199	179	199	181	171
151	160	199	189	209	189	179
161	170	219	199	219	197	187
171	180	219	209	229	205	195

There are many services requested by customers and offered by the companies. In order to get the distribution of the demands, a data analysis is performed for the airline data and the probabilities for the service demand are calculated as in Table 2 below. Note that these probabilities might change in each airline and each country. These are the probabilities found based on the available data.

Table 2: The probabilities for the service demand

Service	Distribution
Extra Baggage (kg/passenger)	N(15,4)
Meal selection (/passenger)	P(x<=10%)
Special Equipment (/passenger)	P(x<=3%)
Pet (/passenger)	P(x<=3%)
Reservation (/passenger)	P(x<=5%)
SMS usage (/passenger)	P(x<=10%)

Some companies have higher main price while free additional services. It is also possible to have companies with lower main price and higher service prices. The cost of the services for each airline is summarized in Table 3.

Table 3: Cost of services in each airline

Revenues	THY	Pegasus	Atlas	Onur	Sun Express
Service Price (TL)	43	37.5	36	15.5	38
Extra Baggage (TL/kg)	3	5	3	3	4
Special Baggage and Equipment Charge (TL)	0	38	0	20	20
Seat selection(TL)	0	12	0	12	0
Pets (TL)	0	25	20	0	0
SMS (TL)	0	1.5	0	0	0
Reservation (TL)	0	6	0	0	0
Meal (TL)	0	16	0	12	17
Free Weight (kg)	20	15	20	15	15

We consider Boeing 737 800 type aircraft with 180 seat capacity. We choose Istanbul-Antalya flight for the analysis as each company has a trip for this line. The customer data is generated using a monte carlo simulation methodology based on the past customer performance. It is worth mentioning that probabilities given in Table 2 are used as a cumulative probability limit and they are used to generate the random customer data set. The demand data of each customer are found and then fed to the system as inputs. Table 4 provides a representation for the customer data.

The airline pricing is calculated separately for each company for THY, Pegasus, Atlas, Onur and Sun express As we run the simulation model, the price for each customer from each company is taken into account and the customer is assigned to the one with the minimum cost. Table 5 shows a representation for customers 62-75 of 600 customer data set and the prices the company offered.

Table 4: Service demand from the customers

Customer	Extra Baggage	Special Baggage	Pets	Meal	Reservation	SMS
1	14.26	0	0	0	0	0
2	13.23	0	0	0	0	0
3	21.46	0	0	0	0	0
4	17.83	0	0	0	0	0
5	3.83	0	0	0	0	1
6	16.49	0	0	0	0	0
7	15.15	0	0	0	0	0
8	19.43	0	0	0	0	0
9	1.73	0	0	1	0	0
10	8.38	0	0	0	0	0
11	10.68	0	0	0	0	0
12	20.47	0	0	0	0	0

Table 5: A set of customer data and prices

Passanger	Selection						Price (TL)				
	Minimum price	THY	SUN	Onur	Atlas	Pegasus	THY	SUN	Onur	Atlas	Pegasus
62	102	9	2	20	10	21	102	117	115.5	105	121.5
63	117	10	2	20	10	21	117	137	130.5	140	171.5
64	96.5	10	2	20	10	22	102	97	100.5	105	96.5
65	96.5	10	2	20	10	23	102	97	100.5	105	96.5
66	96.5	10	2	20	10	24	102	97	100.5	105	96.5
67	97	10	3	20	10	24	102	97	100.5	105	102.5
68	108	11	3	20	10	24	108	125	121.5	111	131.5
69	96.5	11	3	20	10	25	102	97	100.5	105	96.5
70	102	12	3	20	10	25	102	105	106.5	105	106.5
71	102	13	3	20	10	25	102	114	112.5	105	118.5
72	96.5	13	3	20	10	26	102	97	100.5	105	96.5
73	101	13	4	20	10	26	102	101	103.5	105	101.5

Once the simulation analysis for 600 customers is completed, then it is possible to see the total capacity utilization of each company. The customers are assigned to companies in such a way that each customer prefers the one with the lowest cost. Table 6 shows the total capacity utilization of each company in 3 simulation replications. Onur, Sun, Pegasus, Atlas and THY has the largest capacity utilizations respectively.

Table 6: Capacity utilization of each company

Replication	Utilization				
	THY	SUN	ONUR	ATLAS	PEGASUS
1	68.89%	81.11%	88.89%	72.22%	77.78%
2	66.67%	83.33%	89.44%	71.67%	77.78%
Average	67.78%	82.22%	89.17%	71.94%	77.78%

Note that it is possible that the closing price of each simulation to be different in each replication. Table 7 shows the closing prices and the average closing prices. Onur, Atlas, Sun, THY and Pegasus has the least cost offers respectively. It is shown that Onur has the price advantage for higher capacity utilization.

Table 7: Closing prices by companies

Replication	Closing price (TL)				
	THY	SUN	ONUR	ATLAS	PEGASUS
1	222.00	226.00	224.50	225.00	232.50
2	222.00	217.00	212.50	215.00	216.50
Average	222.00	221.50	218.50	220.00	224.50

The total revenue made by each company is also expected to be different both within companies and replications. Table 8 shows the revenues of each company in each replication. It is

shown that Onur, Sun, Atlas, Pegasus and THY has the largest revenues respectively. The prices show that lower prices attract more customers and increase revenue. Such analysis can be extended to see the sensitivity of each pricing strategy and companies can update their strategies until they reach to a desired revenue level. A spreadsheet based decision support system is developed to help decision maker simulate their system and do analysis. The menu and data sections are provided in Figure A in appendix.

Table 8: Revenues by companies

Replication	Revenues (TL)				
	THY	SUN	ONUR	ATLAS	PEGASUS
1	19629.00	22186.00	23234.00	20486.00	19877.50
2	18840.00	23042.00	23530.50	20184.00	19930.50
Average	19234.50	22614.00	23382.25	20335.00	19904.00

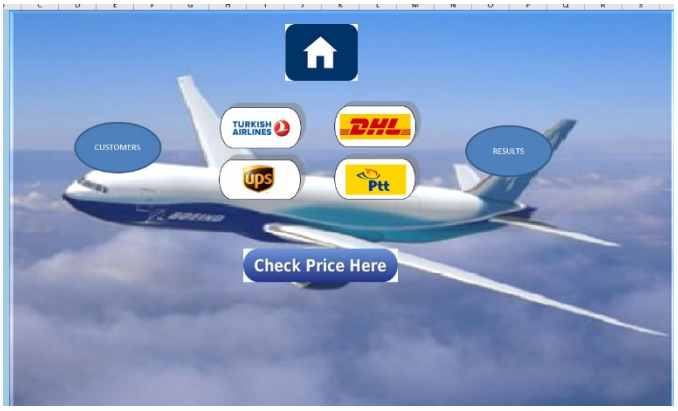
IV. CONCLUSION

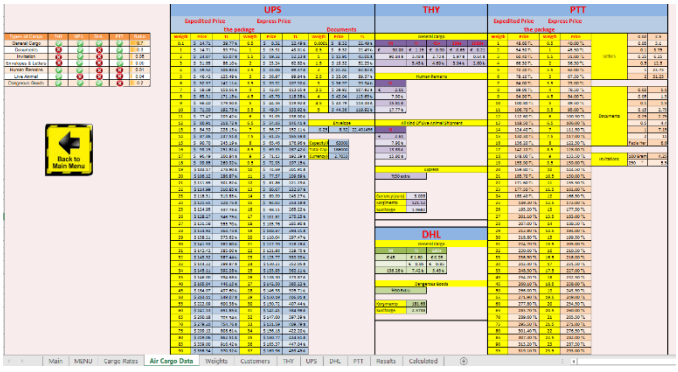
In this paper, a decision support system that simulate the pricing behaviors of both the customer and air line companies is developed. Customers demand different services that also affect the ticket price. In order to analyze the utilization rate of an aircraft, ticket price and revenue for a given destination, total number of customers are randomly generated and the number is simulated in the system. The probabilities for the demand are calculated based on historical demand data. The sensitivity analysis is also provided and the information can be used to recalculate the offers. The simulation process in the decision support system successfully determines the best strategy for the customers and the companies. It is also possible to replicate the simulation methodologies to develop a sensitivity analysis for changing prices. The utilization of the plane and the total revenue for given scenario can also be calculated.

APPENDIX

Table A. Notations used for the analysis

i	:	Index for the airline companies (i ∋ I)
j	:	Index for the customers (j ∋ J)
T_{ji}	:	The ticket price for customer j of company i
X_{ji}	:	Tax price of passenger j of company i
S_{ji}	:	Service price of passenger j of company i
I_{ji}	:	Travel insurance fee of passenger j of company i
W_{ji}	:	Weight of free baggage allowance of passenger j of company i (kg)
G_{ji}	:	Weight of extra baggage of passenger j of company i (kg)
AW_{ji}	:	Weight of free baggage allowance of passenger j of company i (kg)
E_{ji}	:	Extra baggage fee of passenger j of company i (TL/kg)
Q_{ji}	:	Special baggage equipment fee of passenger j of company i (TL/unit)
L_{ji}	:	Seat selection fee of passenger j of company i (TL/seat)
P_{ji}	:	Pets fee (5kg) of passenger j of company i (TL/seat)
SMS_{ji}	:	SMS fee of passenger j of company i (TL/seat)
R_{ji}	:	Reservation fee of passenger j of company i (TL/reservation)
M_{ji}	:	Meal selection(Sandwich, water, tea) fee of passenger j of company i (TL/meal)
HQ_{ji}	:	1 special baggage is required for customer j at company i, 0 otherwise
PL_{ji}	:	1 seat selection is preferred by customer j at company i, 0 O/W
B_{ji}	:	1 If customer j selects company i, 0 O/W
PP_{ji}	:	1 a pet is registered for customer j at company i, 0 O/W
$PSMS_{ji}$	:	1 SMS is requedted by customer j at company i, 0 O/W
HR_{ji}	:	1 a reservation is performed by customer j at company i, 0 O/W
HM_{ji}	:	1 a meal is requested by customer j at company i, 0 O/W
PT_i	:	Carrying capacity of the company i





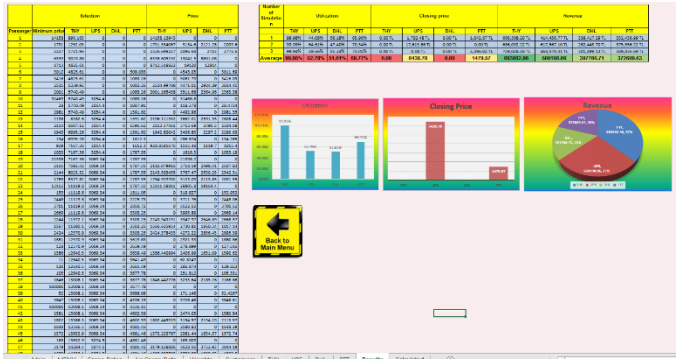


Figure A: Spreadsheet based decision support system

REFERENCES

[1] Smith, Jerald R., and Peggy A. Golden. "Airline: A strategic management simulation." Developments in Business Simulation and Experiential Learning 14 (1987).

[2] Howard, Lee R., and Duane O. Eberhardt. "Airline simulation for analysis of commercial airplane markets." Transportation Science 1.3 (1967): 131-157.

[3] Rosenberger, Jay M., et al. "Air transportation simulation: SimAir: a stochastic model of airline operations." Proceedings of the 32nd conference on Winter simulation. Society for Computer Simulation International, 2000.

[4] Gosavi, Abhijit, Emrah Ozkaya, and Aykut F. Kahraman. "Simulation optimization for revenue management of airlines with cancellations and overbooking." Or Spectrum 29.1 (2007): 21-38.

[5] Bertsimas, Dimitris, and Sanne De Boer. "Simulation-based booking limits for airline revenue management." Operations Research 53.1 (2005): 90-106.

[6] Frank, Michael, et al. "Airline revenue management: A simulation of dynamic capacity management." Journal of Revenue and Pricing Management 5.1 (2006): 62-71.

The Sustainability Indicators in Supplier Selection: The Furniture Industry

B. BALIBAS[1] and C. SEL[1]

[1] Karabuk University, Karabuk/Turkey, barisbalibas@hotmail.com
[1]Karabuk University, Karabuk/Turkey, cagrisel@karabuk.edu.tr

Abstract **- There are many qualitative and quantitative decision-making criteria in supplier selection (e.g., cost, quality, service or lead time). The today's business environment requires choosing the right supplier with the environmental concerns in addition to the well-known selection criteria. We claim that a supplier selection accounting for the environmental concerns increases the supply chain responsibility, e.g. customer satisfaction, cost reduction and a clean environment. Therefore, we introduce a multi-criteria decision-making problem accounting for the environmental and social indicators that might be used to select the supplier of a wood glue which does not contain environmentally hazardous substances. A real-life case encountered in the furniture industry, is analysed with an expert choice software using the analytic hierarchy process decision making methodology.**

Keywords **- Supply chain management, supplier selection, multi criteria decision making, expert choice**

I. INTRODUCTION

Managers need the concepts of conformity and reliability in response to business decisions [1,2,3]. It is possible to make the necessary decisions to consolidate and measure these decisions [4,5]. Decision-making is one of the most important activities in the business world. Firms allocate important calculators and make investments for decision making processes.

The problem of decision making can be defined as a selection of the most appropriate option from a set of options for at least one purpose or measure. Accordingly, the elements of a decision problem constitute the priorities of decision makers, options, criteria, outcomes, environment and decision maker. In the simplest form, a decision problem can be thought as a choice between options according to a purpose or measure [22].

II. SUPPLIER SELECTION

Dickson mentions and defines 23 criteria that must be taken into account in the supplier selection phase. [6]. In addition, for Arbel and Seidmmann [7-9], Beck and Lin [10], Tam and Tummala [11], Ghodsypour and Brien [12], Zviran [13] and Bard [14] the chosen criteria is financial, technical and operational success. Yurdakul and İç [15] has determined the criteria to be considered in selecting suppliers as managerial capabilities, technological capabilities and production facilities and capacities.

One of the important decisions that organizations should make is supplier choice. The responsibility of the procurement function has often been defined as the supply of equipment, materials and materials with adequate quality and quantity, with appropriate prices, with a suitable delivery [11].

The innovations, developments and changes that have emerged in the industry over the last years have led to revisions in production and supply organizations. The reason for this is the increase in customer demand, technological opportunities and innovations, and irregular market trends.

Supplier selection cost, quality, performance, technology has become a crucial concept for companies. In this study, not only material cost, but also operating costs, maintenance, development and support costs are considered on supplier selection.

III. ANALYTIC HIERARCHY PROCESS

Narasimhan [16], Nydick and Hill [17] and Partovi [18] have proposed the use of the analytical hierarchy process (AHP) method. The reason of choosing AHP is that the AHP uses qualitative and quantitative criteria in the electoral process. One of the other reasons is that it is a method that can be used by the business executives to understand. The method is easy to implement, as well as to improve the decision-making process.

AHP is a method that can determine the superiority of criteria and sub-criteria to compare and evaluate systematically. The selection of the best supplier and the effectiveness of the supply systems can be compared based on the superiority. Tam and Tummala have applied the AHP to the telecom industry's problem of supplier selection, in which cost and quality identified as the main factors influencing supplier choice [11].

In AHP, it is the first step to determine the criteria and the sub-criteria in the direction of the decision maker and to create the hierarchical structure. In the use of AHP, the aim is first determined and the criteria that affect the choice are set for this purpose. Then, potential alternatives are determined by considering the criteria. As a result, a hierarchical structure is formed for the decision [20].

After the hierarchical structure has been established, binary comparison decision matrices are constructed to compare the alternatives based on each criterion and to compare the criteria

among themselves. The proposed 1 to 9 significance scale of Saaty [20] is shown in Table 1.

Table 1. Importance scale

Importance Level	Description
1	Equally Important
3	Important in mid-grade
5	Important in strong grade
7	Important in Very High Grade
9	Definite Important

The intermediate values such as 2, 4, 6, 8 does not have an effect in decision-making. For example, if the decision maker is unstable between 1 and 3, it can use the value 2. Binary comparisons are the most important phase of AHP. Relative or absolute measures are used to obtain binary comparisons. According to the information obtained from these, judgments are converted into matrices in AHP. a_{ij} is the feature related to i and j choices. If the feature is shown as a binary comparison value, then $a_{ij} = 1 / a_{ij}$ is obtained. This feature is called a response feature [20].

The second stage of AHP is the creation of normalized matrices. The normalized matrix is obtained by dividing the value of each column separately into the corresponding column sum.; the average of each sequence value is taken in a normalized matrix. These values are the weights for each criterion [20].

The Consistency Ratio (T.O.) must be calculated to determine whether the decision maker is consistent across benchmarks or not. In this calculation, random index numbers are used depending on the number of criteria. If the value found in the result of calculations is less than 0.10, the generated comparison matrix is consistent. Otherwise, the decision matrix should be rearranged [20].

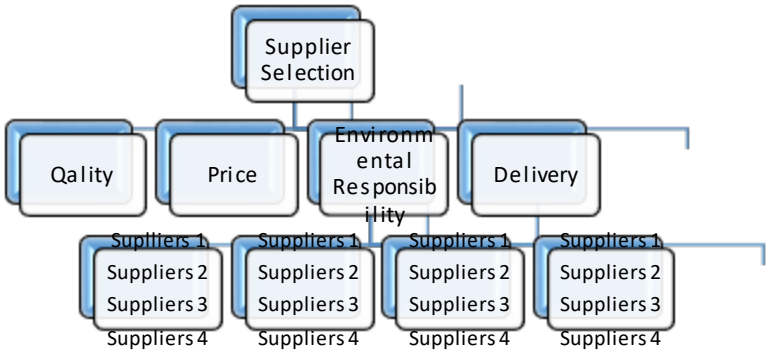
The last step of the AHP is to multiply the importance weights of the criteria with the weights of the alternatives and the priority value of each alternate. The sum of these values equals 1. The highest value alternative is the best alternate for the decision problem.

IV. APPLICATION

AHP is an extensively used decision-making method in the field. Cost, price and quality are used as the basic criteria in supplier selection problems. Hoverer, an environmental criterion should be added along with the criteria regarding to increasing environmental concerns. The evaluation in the supply of the glue material used in the furniture sector is introduced as an industrial application.

In the first stage, we rank the quantitative characteristics as quality, price, environmental responsibility and delivery to ensure. The competence of the company is scored by the AHP method. As a result of the evaluation made, the scores for each alternative firm were obtained.

Figure 1. The hierarchy for supplier selection problem



The second step is to compare the 4 alternative firms determined in the hierarchy according to the importance scale for each criterion. In comparison to the quality criterion, solid product fractions of supplier firms can be taken into consideration. A sample binary comparison matrix generated according to this is given in Table 2.

Table 2. The binary comparison matrix according to quality criterion

Quality	1	2	3	4
1	1	1/3	1/8	1/5
2	3	1	2	1/4
3	8	1/2	1	1/3
4	5	4	3	1

In comparison to the price criterion, it can be considered that suppliers offer lower quality products. An exemplary binary comparison matrix constructed accordingly is given in Table 3.

Table 3. The binary comparison matrix according to price criterion

Price	1	2	3	4
1	1	1	1/3	1/5
2	1	1	1/2	1/4
3	3	2	1	1/6
4	5	4	6	1

In comparison with the Environmental Responsibility Criteria, examples of companies using chemicals and metals that may damage the environment in their products can be considered in the decisions. Accordingly, the sample binary comparison matrix constructed is given in Table 4.

Table 4. The binary comparison matrix according to environmental responsibility criteria

Environmental Responsibility	1	2	3	4
1	1	8	2	4
2	1/8	1	3	1
3	½	1/3	1	¼
4	1/4	1	1/4	1

In comparison with the delivery criteria, it can be considered that the supplier companies make deliveries in time without any mistake. Accordingly, the sample binary comparison matrix constructed is given in Table 5.

Table 5. The binary comparison matrix according to delivery criterion

Delivery	1	2	3	4
1	1	1/2	1/3	1/5
2	2	1	1/8	1
3	3	8	1	1/4
4	5	1	4	1

Matrices created after this phase is evaluated by Expert Choice program. As a result of the evaluation, the Binary Comparison Matrix based on Firm Priorities and Consistency Ratios and Criteria Based Criteria has reached. In the last step in the AHP, the priority of each supplier will be calculated by multiplying the company priorities based on the obtained criteria priorities and criteria. Hereby, we present an ongoing study that the results and their managerial implications will be discussed as a future research.

REFERENCES

[1]Montgomery, D. C., Johnson, L. A., & Gardiner, J. S. (1990). Forecasting and time series analysis. McGraw-Hill Companies.

[2]Kutay, F. (1989). Zaman Serilerinde Tahmin Teknikleri ve Box-Jenkins Modelleri. Ders Notları, Gazi Üniversitesi, Ankara.

[3]Dizdar, E. N. (1998). Üretim Sistemlerinde Olası İş Kazaları İçin Bir Erken Uyarı Modeli, Doktora Tezi. Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara.

[4]Vargas, L. G. (1990). An overview of the analytic hierarchy process and its applications. European journal of operational research, 48(1), 2-8.

[5]Zahedi, F. (1987). A utility approach to the analytic hierarchy process. Mathematical Modelling, 9(3-5), 387-395.

[6]Dickson, G.W. (1966). An Analysis of Vendor Selection Systems and Decisions, Journal of Purchasing, Vol 2, 5-17.

[7]Arbel, A., & Orgler, Y. E. (1990). An application of the AHP to bank strategic planning: The mergers and acquisitions process. European Journal of Operational Research, 48(1), 27-37.

[8]Arbel, A., & Seidmann, A. (1985). Capacity planning, benchmarking and evaluation of small computer systems. European Journal of Operational Research, 22(3), 347-358.

[9]Arbel, A., & Seidmann, A. (1984). Selecting a microcomputer for process control and data acquisition. IIE transactions, 16(1), 73-80.

[10] Beck, M. P., & Lin, B. W. (1981). Selection of automated office systems: a case study. Omega, 9(2), 169-176.

[11] Tam, M. C., & Tummala, V. R. (2001). An application of the AHP in vendor selection of a telecommunications system. Omega, 29(2), 171-182.

[12] Ghodsypour, S. H., & O'Brien, C. (1998). A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming. International journal of production economics, 56, 199-212.

[13] Zviran, M. (1993). A comprehensive methodology for computer family selection. Journal of Systems and Software, 22(1), 17-26.

[14] Bard, J. F. (1986). Evaluating space station applications of automation and robotics. IEEE Transactions on Engineering Management, (2), 102-111.

[15] Yurdakul, M., & İç, Y. T. (2001). AHP ve Hedef Programlama Yöntemlerinin Sağlayıcı Seçimi Probleminde Kullanılması. XXII. Ulusal YA/EM Kongresi, Gazi Üniversitesi, Ankara.

[16] Narasimhan, R. (1983). An analytical approach to supplier selection. Journal of supply chain management, 19(4), 27-32.

[17] Nydick, R. L., & Hill, R. P. (1992). Using the analytic hierarchy process to structure the supplier selection procedure. Journal of supply chain management, 28(2), 31-36.

[18] Partovi, F. Y., Burton, J., & Banerjee, A. (1990). Application of analytical hierarchy process in operations management. International Journal of Operations & Production Management, 10(3), 5-19.

[19] Yurdakul, M., & İç, Y. T. (2000). Üretim Firmalarının Kredibilitesinin Belirlenmesinde Analitik Hiyerarşi Yönteminin Kullanılması. Gazi Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 13(4), 1007-1023.

[20] Saaty, T. L. (1980). The analytic hierarchy process New York. Nachhaltigkeit im Einkauf von Logistikdienstleistungen, 137.

[21] Jones, D. F., Mirrazavi, S. K., & Tamiz, M. (2002). Multi-objective meta-heuristics: An overview of the current state-of-the-art. European journal of operational research, 137(1), 1-9.

[22] Dağdeviren, M., & Tamer, E. R. E. N. (2001). Tedarikçi firma seçiminde analitik hiyerarşi prosesi ve 0-1 hedef programlama yöntemlerinin kullanılması. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 16(1).

Evaluating Research Performance of the European Countries Through Social Network Analysis

İ. F. ŞENTÜRK[1] and B. TAŞ[2]

[1] Bursa Technical University, Bursa/Turkey, izzet.senturk@btu.edu.tr
[2]Bursa Technical University, Bursa/Turkey, burak.tas@btu.edu.tr

***Abstract* - Social network analysis investigates relationships in a networked structure in order to interpret the roles of individuals and evaluate their respective importance. The common approach is employing graph theory that models the network as a graph data structure. Graphs are mathematical abstractions to model pairwise relations between objects. A graph is comprised of nodes connected with edges. While nodes indicate individuals in a network, edges signify relationships or interactions. In this paper, we evaluate the research performance of the European countries considering the research activities within the framework of European Cooperation in Science and Technology (COST). Founded in 1971, COST is the oldest and widest scientific intergovernmental framework in Europe supporting 37 countries including Turkey. COST can be considered as an incubator to set up interdisciplinary research networks since it provides support for network activities such as meetings, training schools, short scientific exchanges, etc. but does not fund the research itself. Therefore, we believe that the research network under COST can be a good indicator to analyze relationships between countries in research activities and evaluate the research performance of the countries. In this paper, we considered research actions funded by COST between 2012-2017 and evaluated research performance of the countries according to their participations and interactions. To assess the performance, we modeled the relationships between countries on a directed graph and applied centrality analysis which is a common approach to evaluate the relative importance of a node within the network. Each action is coordinated by a management committee which is composed of a chair and at most two delegates per participant country. In the graph, each participant country is denoted with a node. We classify the countries according to their roles in the action. Since the proposer of the action becomes the chair, usually, we regard the country of the chair as a gateway to access the action. Therefore, to signify the relationship between two countries, a directed edge is added from the participant country to the respective country of the chair. Note that several projects were considered over a span of 6 years and multiple interactions between two nodes are indicated as the edge weight.**

***Keywords* - Centrality measures, betweenness centrality, closeness centrality, social network analysis, graph theory.**

I. INTRODUCTION

THE term social network denotes social relationships among actors. An actor may refer to a vast array of entities including organizations [1], diseases [2], sensor nodes [3], etc. Relationships imply pairwise connections between two actors. A connection can either express bidirectional link such as friendship between two individuals or a railroad between two cities, or it can also represent unidirectional interactions as well such as an event originating in one place and ending in another one. Transmissions of a pathogen propagating to a new geographical location or email communication among suspects of terrorist activities are some of the examples of directional interactions.

Social network analysis provides a means to study interactions and/or relationships within a social group. Statistical techniques [4] or graph-theoretical approaches can be employed to infer inherent network dynamics to analyze various features of actors and relationships within the network. In this paper, we use graph structure to model the network and exploit graph theory, namely centrality measures. Centrality reveals node level features such as relative prominence of a node within the network. Various approaches exist to assess centrality including degree, eigenvalue, betweenness, closeness, etc. based on the definition of the importance. Accessibility and expected force (influence) are some of the indicators to assess importance of a node [5]. While accessibility considers random walks in the network and measures probability of visiting a certain actor, expected force evaluates the likelihood of the spread of an infection in case of an outbreak of infectious diseases [6].

Walk can be defined as a series of actors visited following the connections among them. According to the considered application, various requirements can be defined on the walk. Visiting an actor multiple times can be restricted to model the application in a more realistic manner. Let us consider an application where a widespread foodborne outbreak is modeled as given in [2]. In this model, each actor represents a country where the obtained pathogens are associated and relationships denote unidirectional links connecting two countries where the isolates crosses towards the given direction. In this model, visiting an actor during the walk more than once can be limited since the corresponding country will be already infected during the first visit. Therefore, one can claim that eigenvalue centrality is not the suitable approach to measure centrality for this application and employ betweenness centrality instead.

Centrality approaches diversify based on various properties of the walk including the length and consideration of the start/end point of the walk [7]. For instance, while degree centrality considers a walk of length one (i.e., the number of immediate connections), eigenvalue centrality requires multiple iterations of walk until reaching convergence. On the other hand, unlike degree and eigenvalue centralities which count walks start/end on a certain actor, betweenness centrality considers the number of walks which passes from the given actor. Another classification is based on whether the volume or the length of the walk is more important. Unlike degree, eigenvalue, and betweenness centralities, closeness centrality considers the length of a walk.

In graph theory, node is the fundamental unit of the graph along with the edges. Nodes denote actors and edges represent relationships. As mentioned earlier, edges can be directed or undirected (bidirectional). Also, weights can be assigned to edges to denote the importance of the link, frequency of an event, or the cost of the transmission, etc. based on the considered application. When the network is modeled on the graph, one of the centrality approaches can be employed to assess relative importance of the nodes.

This paper aims to identify prominent countries in science across Europe by evaluating the research performance of the European countries. Despite availability of various framework programs for research, we selected COST framework program [7] to assess research performance of the countries considering the fact that COST is the longest running European framework supporting transnational cooperation across Europe [8]. COST program does not support the research itself but the networking activities, staff exchange, and training. Therefore, we believe the suitability of this research networking program to analyze research performance by employing social network analysis. We model the research network on a graph structure and exploit graph theory to identify prominent countries in research and their influence on other countries in this network. To understand the model, let us detail the participation procedure to COST actions. Each project funded by COST is referred to as an action. Each action has a chair and a management committee (MC). Action chair is usually the grant proposer and the MC should be represented by at least five countries. Each country can nominate up to two participant researchers. While each country follow a different approach to nominate the participants, TUBITAK, which is the representing institution of Turkey, requires the researcher to have an active project, funded by TUBITAK, related to the COST action [9]. There are 36 member countries of COST and one cooperating state. Considering the fact that a researcher (or a country) cannot participate to an action if there is no funded action, we classify the countries into two groups: providers and beneficiaries. Providers enable an action so that beneficiary countries may join. Corresponding country of the action chair is regarded as provider and the associated countries of the MC members are regarded as beneficiaries.

On the graph, each country is represented as a node and a directed edge is added originating from the beneficiary directed towards the provider. Provider country can be regarded as a gateway to participate COST and access the fund. If a county is represented in an action, there can be one or two members. Since we focus on whether a country is represented or not in an action, even if two participants exist from a country, edge weight is set to one on the graph for a single action. But since there are several actions funded each year, we count the links between countries and set the edge weight accordingly for the corresponding country pairs considering the edge directions. Once the graph is completed, we employ various centrality approaches to analyze the network and assess the research performance of the countries.

The rest of the paper is organized as follows. Previous work is discussed in Section II. Data collection and employed approaches are detailed in Section III. Findings are discussed in Section IV. The paper is concluded in Section V.

II. Related Work

In graph theory, centrality measures can be employed to determine how central a node is for the network, or in other words, to evaluate the importance of a vertex in a graph. However, the term, importance, can be vague which requires elaboration on the network type and application. In this paper, we are concerned with the interaction of countries while collaborating in the research activities. Since the action chair manages the budget, action's fund flows from the country represented by the action chair to other countries in the MC. Thus, in our case, importance infers centrality in accessing grants and relaying the grant money between the nodes.

A variety of works exist in the literature which employs centrality approaches to address different problems in various applications. In [10], trust issue in web-based platforms is considered. To evaluate reputation of the users, two different scores are calculated based on the contribution and centrality. Depending on the platform, contribution may refer to the number of reviews in an e-market platform or answers in a forum. But despite its importance on the success of the platform, calculating contribution can be difficult due to the sparseness of the data. On the other hand, centrality-based score can be calculated based on the who-trusts-whom network. The idea is trying to estimate contribution based score based on the centrality-based reputation. Another study in [11] aims to identify documents that are likely to have higher impact in the future by employing centrality measures. In this work, importance of a document is defined based on the number of citations it receives. [12] considers criminal networks and tries to identify key actors in a drug trafficking network by integrating degree and betweenness centrality measures. Another work presented in [13] evaluates financial institutions and analyzes their roles in financial crisis. It was shown that centrality measures perform well in identifying and monitoring systematically important financial institutions in case of a crash in the banking system. Determining neuronal activity is another

application where centrality measures are used [14]. Activities between neurons in the neural network are investigated and importance of the nodes is identified. In the mentioned work, firing rate of the neuron is used to assess the importance. In another study, betweenness centrality is employed to understand the flow characteristics of the traffic within a city [15]. Closeness Centrality is exploited in [16] In to analyze the complaints. In [17], social network analysis was employed to analyze the image of various brands.

III. METHOD

A. Data Collection

The data that we have considered in this study was collected from the official website of COST [18] using Python 3. We have employed Beautiful Soup [19] for web scraping. Beautiful Soup is a python package to parse HTML and XML documents. Collected data was analyzed using Gephi [20]. Gephi was also employed to visualize the results. In this study we have considered the actions funded between 2012 and 2017. For each year, we counted the number of links between the pairs of countries considering the directionality of the link. To observe the progress, we classified the data into two groups. In the first group, we report the cumulative sum of the performance results for the years 2012-2014. In the second group, the results for the years 2015-2017 are reported.

B. Closeness Centrality

Closeness centrality of a node, u in graph G, is defined as the reciprocal of the farness [21] as given in Equation 1. $d(u,v)$ denotes the distance (farness) between nodes u and v. Farness can be defined as the total length of the shortest paths between node u and the rest of the nodes in the graph. If a node is located closer to other nodes, its reachability is expected to increase.

$$C(u) = \frac{1}{\sum_v d(u,v)} \tag{1}$$

C. Betweenness Centrality

Betweenness centrality considers the number of times a node appears on a shortest path between every pair of other nodes. If the number of shortest paths between a pair of nodes st, is $d(st)$ and node v exists $d(st)_v$ times in these shortest paths, then the betweenness centrality of node v can be expressed as in Equation 2.

$$C_B(v) = \sum_{s,t,v \in V} \frac{d(st)_v}{d(st)} \tag{2}$$

A node is associated with a higher centrality score, if the fraction of the shortest paths passing through the given node is higher.

D. PageRank

PageRank, a variant of the eigenvector centrality, measures centrality of a node based on the number of incoming edges. The idea is similar to a random walk where the probability of ending up on a specific node after a random walk is higher if the PageRank score for that node is higher. PageRank provides a probability distribution for the importance of the nodes and can be computed iteratively where the probability values always sum to one at each iteration. PageRank of node v can be expressed as in Equation 3. d refers to damping factor which is set to 0.85 as recommended [22]. $M(v)$ is the set of incoming edges that link to v. $L(s)$ is the number of outgoing links from node p_S, and N is the total number of nodes.

$$PR(v) = \frac{1-d}{N} + \sum_{s \in M(v)} \frac{PR(s)}{L(s)} \tag{3}$$

IV. FINDINGS

In the rest of the paper, various node/edge colors and node/edge sizes are used in the graphs to improve the visualization in representing the results. Node size and color signify the prominence of the countries. Larger node size and darker color denotes increased importance. On the other hand, thickness of the edges represents frequency of interactions between the respective countries. Higher frequency is denoted with the thicker lines accompanied with a darker line color. To evaluate the performance of the countries, PageRank, betweenness centrality, and closeness centrality measures are used. Obtained results are normalized between 0 and 1 to adjust the values for a fair comparison. To analyze the correlation of our findings with the actual research performance of the countries, we have considered the COST country fact sheets report [23]. In the mentioned report, various facts are given for each member state including the amount of budget transferred to each country for respective years. In the report, the latest available data is from 2015 and therefore we have considered the amount of funds transferred between 2012-2014 and the centrality scores obtained for 2012-2014. The results of the analysis for the correlation between the transferred funds and the centrality scores for the period of 2012-2014 can be found in Table 1. It can be noticed that the betweenness centrality score is highly correlated with the research performance of the countries in terms of the grant money they received. PageRank is the next approach providing the best metric to assess the research performance. Closeness centrality is the least correlated approach in determining the research performance.

Closeness	Betweenness	PageRank
0.5109	0.8676	0.7493

Table 1. Correlation between the amount of transferred funds and the centrality scores for the period of 2012-2014.

A. Results for 2012-2014

The results can be found in Figures 1, 2, and 3 for PageRank, betweenness centrality, and closeness centrality respectively.

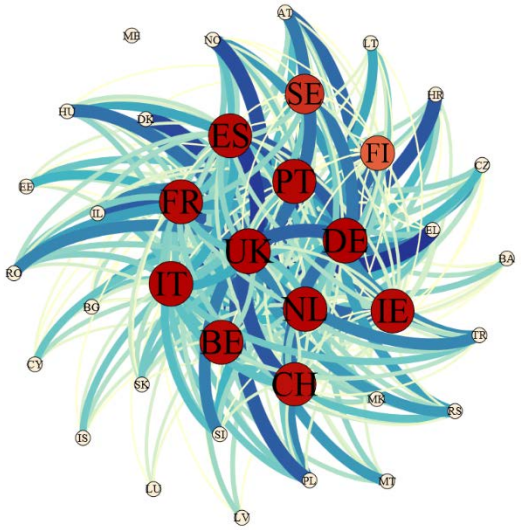


Figure 1. PageRank results for 2012-2014.

According to Figure 1, 11 countries are represented with nodes of almost the same size. Finland is denoted with a slightly smaller and lighter node. The rest of the countries have negligible importance.

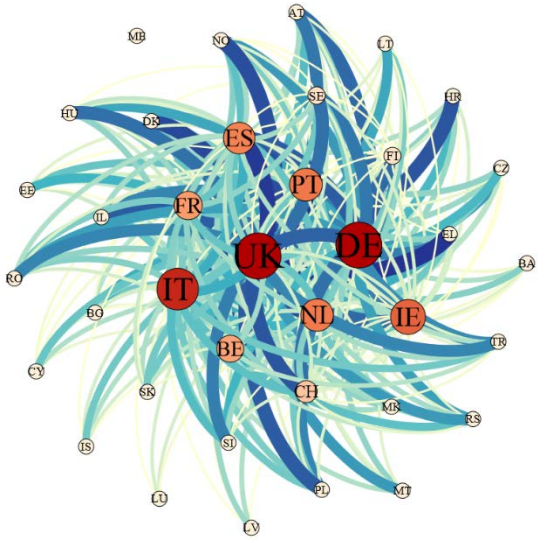


Figure 2. Betweenness centrality results for 2012-2014.

Compared to PageRank, betweenness centrality diversifies the performance of the countries better according to Figure 2. United Kingdom, Italy, and Germany have the highest scientific research performance within the network. Italy is slightly worse than United Kingdom and Germany.

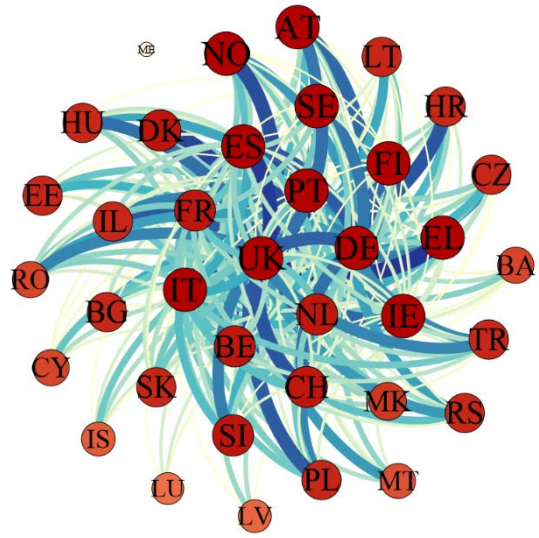


Figure 3. Closeness centrality results for 2012-2014.

According to Figure 3, almost all the countries have similar performance and the difference is negligible. As can be seen from Figures 1-3, three different results are obtained for the same data. The results suggest that betweenness centrality is the best approach to highlight the least number of prominent countries. While presenting the importance of the countries, PageRank emphasizes more countries compared to betweenness. Closeness centrality, on the other hand, provides scores closer to each other which makes it difficult to highlight the countries with the best performance.

B. Results for 2015-2017

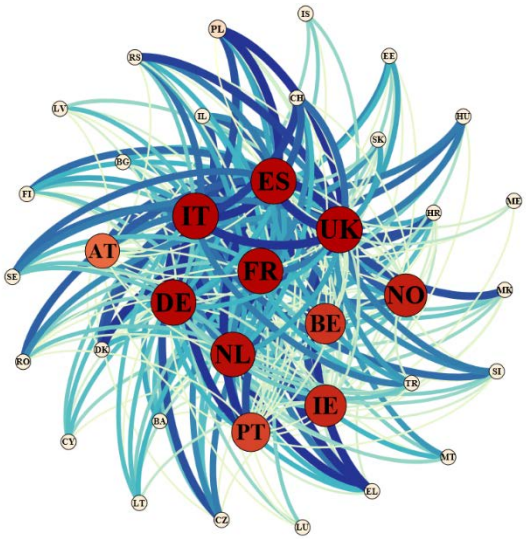


Figure 4. PageRank results for 2015-2017.

Again, 10 countries have similar performance. Austria is slightly worse than these countries. On the other hand, the rest of the countries have negligible importance according to PageRank algorithm as given in Figure 4.

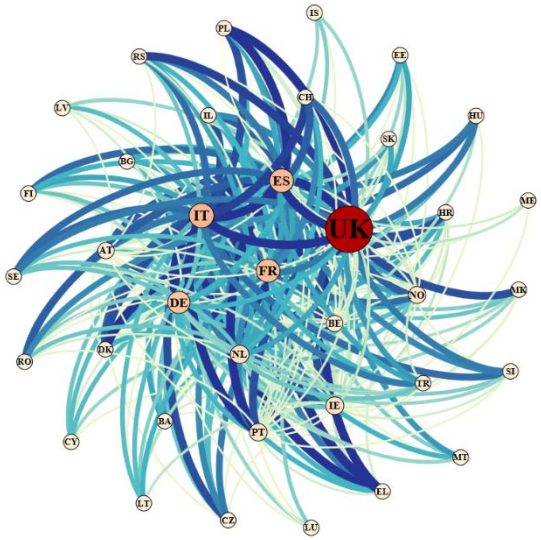


Figure 5. Betweenness centrality results for 2015-2017.

Figure 5 suggests that United Kingdom dominates the network in terms of the scientific performance. It can be claimed that this result avoids the contribution of other countries and may not represent the actual performance of the rest of the countries. But, to highlight only a few countries, betweenness centrality should be applied.

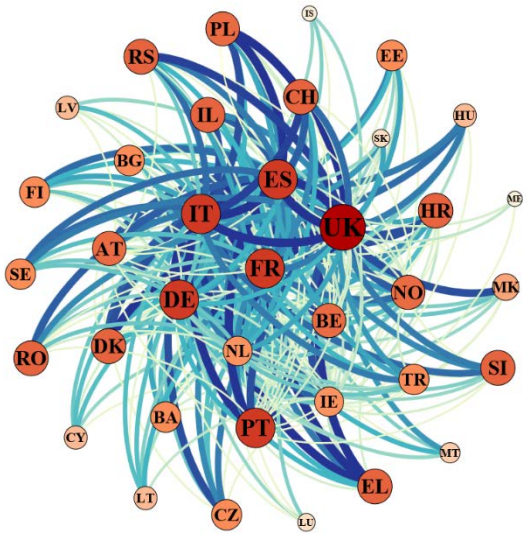


Figure 6. Closeness centrality results for 2015-2017.

According to Figure 6, again, closeness centrality emphasizes the most countries with similar scores.

V. Conclusion

In this study, we have analyzed the research performance of European countries between 2012-2017 through social network analysis. We have employed three different centrality approaches namely, PageRank, betweenness centrality, and closeness centrality to assess relative importance of each country in the COST transnational research network. To analyze the progress, we have considered 2012-2014 and 2015-2017 periods separately. The results are demonstrated on graphs visualized with various node/edge colors and node/edge sizes to represent the significance of respective countries and denote the frequency of interactions between countries. To compare our findings with the actual research performance of the countries, we have inspected the correlation of the obtained results with the amount of funds received by each country in the respective years. The results suggest that betweenness centrality is highly correlated with the actual research performance of the countries in terms of the amount of transferred funds. Centrality analysis show that betweenness centrality should be employed to limit the highlighted prominent countries While PageRank emphasizes more countries compared to betweenness, closeness centrality provides scores closer to each other and makes the results less distinctive.

Appendix

Country codes for 36 member states and 1 cooperating state in the COST framework can be found below.

Country Code	State
AT	AUSTRIA
BE	BELGIUM
BA	BOSNIA AND HER.
BG	BULGARIA
HR	CROTIA
CY	CYPRUS
CZ	CZECH REPUBLIC
DK	DENMARK
EE	ESTONIA
FI	FINLAND
FR	FRANCE
MK	MACEDONIA
DE	GERMANY
EL	GREECE
HU	HUNGARY
IS	ICELAND
IE	IRELAND
IL	ISRAEL
IT	ITALY
LV	LATVIA
LT	LITHUANIAN
LU	LUXEMBOURG
MT	MALTA
ME	MONTENEGRO
NL	NETHERLANDS
NO	NORWAY
PL	POLAND
PT	PORTUGAL
RO	ROMANIA
RS	RUSSIA

SK	SLOVAKIA
SI	SLOVENIA
ES	SPAIN
SE	SWEDEN
CH	SWITZERLAND
TR	TURKEY
UK	UNITED KINGDOM

Table 2. Country codes table.

REFERENCES

[1] C. Rob, S. P. Borgatti, and A. Parker. "Making invisible work visible: Using social network analysis to support strategic collaboration," *California management review*, 44, no. 2, pp. 25-46, 2002.

[2] M. Hoffmann, Y. Luo, S. R. Monday, N. Gonzalez-Escalona, A. R. Ottesen, T. Muruvanda, C. Wang, G. Kastanis, C. Keys, D. Janies, I. F. Senturk, U. V. Catalyurek, H. Wang, T. S. Hammack, W. J. Wolfgang, D. Schoonmaker-Bopp, A. Chu, R. Myers, J. Haendiges, P. S. Evans, J. Meng, E. A. Strain, M. W. Allard, and E. W. Brown, "Tracing Origins of the Salmonella Bareilly Strain Causing a Food-borne Outbreak in the United States," *The Journal of Infectious Diseases*, vol. 213, issue 4, pp. 502–508, 15 February 2016.

[3] I. F. Senturk and K. Akkaya, "Connectivity restoration in disjoint Wireless Sensor Networks using centrality measures," 39th Annual *IEEE Conference on Local Computer Networks Workshops*, Edmonton, AB, 2014, pp. 616-622.

[4] M. S. Handcock, "Statistical models for social networks: Inference and degeneracy," *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, 2003.

[5] Node influence metric, https://en.wikipedia.org/wiki/Node_influence_metric, Accessed: 03/29/2018.

[6] D. A. Janies, L. W. Pomeroy, C. Krueger, Y. Zhang, I. F. Senturk, K. Kaya, and Ü. V. Çatalyürek. "Phylogenetic visualization of the spread of H7 influenza A viruses," *Cladistics*, vol. 31, no. 6, pp. 679-691, 2015.

[7] Centrality, https://en.wikipedia.org/wiki/Centrality, Accessed: 03/29/2018.

[8] COST, http://www.cost.eu/about_cost, Accessed: 03/29/2018.

[9] COST participation rules of Turkey, https://www.tubitak.gov.tr/tr/kurumsal/uluslararasi/cok-tarafli-programlar/cost/icerik-aksiyonlara-katilim, Accessed: 03/29/2018.

[10] P.D. Meo, K. Musial-Gabrys, D. Rosaci, G.M. Sarne, and L. Aroyo, "Using centrality measures to predict helpfulness-based reputation in trust networks," *ACM Transactions on Internet Technology (TOIT)*, 17(1), p.8, 2017.

[11] P. Klimek, A. S. Jovanovic, R. Egloff, and R. Schneider, "Successful fish go with the flow: citation impact prediction based on centrality measures for term–document networks," *Scientometrics*, 107(3), pp.1265-1282, 2016.

[12] D. A. Bright, C. Greenhill, M. Reynolds, A. Ritter, and C. Morselli, "The use of actor-level attributes and centrality measures to identify key actors: a case study of an Australian drug trafficking network," *Journal of Contemporary Criminal Justice*, 31(3), pp.262-278, 2015.

[13] T. U. Kuzubaş, I. Ömercikoğlu, and B. Saltoğlu, "Network centrality measures and systemic risk: An application to the Turkish financial crisis," *Physica A: Statistical Mechanics and its Applications*, 405, pp. 203-215, 2014.

[14] J. M. Fletcher and T. Wennekers, "From structure to activity: Using centrality measures to predict neuronal activity," *International journal of neural systems*, 28(02), p.1750013, 2018.

[15] S. Gao, Y. Wang, Y. Gao, Y. Liu, "Understanding Urban Traffic-Flow Characteristics: A Rethinking of Betweenness Centrality," *Environment and Planning B: Urban Analytics and City Science*, vol. 40, issue 1, pp. 135-153, 2013.

[16] M. T. Argan, "A descriptive study of e-complaining: social network analysis on a website as complaint forum," *Journal of Internet Applications and Management*, vol. 5(1), 2014.

[17] Ç. Basfirinci, "Analysing Brand Image Through Social Network Analysis: A Research on Turkcell and Vodafone Brands," *İGÜSBD*, vol. 3, issue 2, October 2016.

[18] All COST actions, http://www.cost.eu/COST_Actions/all_actions, Accessed: 03/29/2018.

[19] Beautiful Soup, https://www.crummy.com/software/BeautifulSoup, Accessed: 03/29/2018.

[20] Gephi, https://gephi.org/, Accessed: 03/29/2018.

[21] A. Bavelas, "Communication patterns in task-oriented groups," *The Journal of the Acoustical Society of America*, vol. 22, no. 6, pp. 725-730, 1950.

[22] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer networks and ISDN systems*, vol. 30, no. 1-7, pp. 107-117, 1998.

[23] COST country fact sheets, http://www.cost.eu/module/download/58026, Accessed: 03/29/2018.

Complex Network Analysis Of Players In Tennis Tournaments

O.FINDIK[1] and E.ÖZKAYNAK[1]

[1] Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, eozkaynak@karabuk.edu.tr

Abstract **- In parallel with the development of the technology, the storage of the data more easily and quickly and the faster processing on the stored data can make an important contribution to the creation and analysis of networks of coexistence. Complex networking plays an important role in analyzing and revealing common characteristics and structures of connected clusters depending on various characteristics. In this study, a network of association between male tennis player in the Australian Open, the French Open, the US Open and the Wimbledon tennis tournaments, known as four major international tennis tournaments between 2000 and 2017, has been established. While each tennis player is defined as a node in the network of associations created, the competitions of the tennis players with each other are defined as the links connecting these nodes. The universal principles of complex networks such as scale-free, small world, clustering have been examined. Furthermore, through Gephi software, the structural characteristics of networks are visualized by using the data obtained from this association network. As a result of the study, it was seen that the networks among the tennis players struggling in the related tournaments were carrying real world network characteristics and that the data obtained from these networks can be used for network analysis.**

Keywords **– Complex Networks, Social Networks, Scale Free Networks.**

I. INTRODUCTION

People have been constantly communicating and sharing each other throughout history. Over the centuries, the communication and sharing methods have changed as well as the benefits that technology has provided in parallel with improvements. Everything people come together to do is actually a result of this communication and sharing. People who communicate directly or indirectly with each other become part of a complex network. Complex networks are not limited to expressing the relationship of people to each other. Everything that is connected to each other is actually a part of complex networks. Complex network science determines rules and diciplines to understanding and analysing all sorts of system, state and structure that have connection and sharing either directly or indirectly[1]. In the world, complex networks are formed in different structures each passing time, and this complexity is expanding. Complex network analysis, which is constantly expanding and complicated, has become a popular field of study that has received intense attention in recent years in the studying and understanding of every situation that interacts with each other. Sports branches are one of the social, economic and cultural relations that are constantly growing through the matches that individuals, teams or countries play with each other, and which have associations within themselves. In this study, the challenges of the Australian Open, the French Open, the US Open and the Wimbledon tennis tournaments, where players interested in tennis professionally participate extensively throughout the world every year, have been analyzed through a complex network approach. The relation of athletic relations in these tournaments to real world networks is examined and it is also found that these relations have characteristics closer to which complex network approach.

II. METOD

A. Karmaşık Ağlar

Complex networks are a combination of a large number of nodes and links that connecting these nodes. These networks can be formed by associations that have already begun to take shape, as well as a dynamic structure that continues to grow continuously. The connection between the nodes and the links that bring complex networks to the scene was first revealed by Leonhard Euler in the Konigsberg Bridge problem. In the case of the Konigsberg bridge problem, it is aimed to visit the city by passing once only from the bridges in the city of Konigsberg. Eular analyzed this problem using graphs [2]. The theory of complexity states that everything that exists in the world is connected to each other, but that things connected to each other form nodes of the network [3]. Complexity consists of two main elements. The first is that they have a heterogeneous structure compared to the larger ones, and the second is the interactions of the nodes forming the network with each other [4]. Complexity continues to grow continuously over time in the real world. Each new node added to the network affects the expansion of the network [5]. Complex networks can be examined in two groups: real networks and model networks generated from real networks [6]. Model networks are networks created using general principles of real networks [7]. One of the most important features a real network should possess is a small world phenomenon. Stanley Milgram states that there is a six-step path between the two nodes in the small world phenomenon [8]. Small world phenomenology has been observed in networks such as www, online social networks, scientific collaboration networks, cinema player networks [9-14]. In real world networks, the preferential attachment is effective in free-ranging network formation. The fact that the general structure of the network is free-scaled allows the

distribution of degrees to take place in accordance with the power law [15]. Prefential attachment indicates that the probability that new nodes connected to the network are connected to nodes with a higher number of connections in the network is more likely to be connected to nodes with a lower number of connections in the network [16]. Based on this information, it was researched whether the players who struggled in the tennis matches in the world match the real world network characteristics of the networks they created with the matches they have made with each other.

B. Gephi

Complex networks are structured from a large number of nodes and links. Various applications are used in order to be accurate, understandable and interpretable the data to be obtained from these nodes and links. In complex network analysis, it is important to find the shortest paths among the nodes, to obtain the clustered coefficients, to calculate the degree distributions. In addition, visualization of the network in appropriate forms using the obtained data also plays an important role in understanding the structure of the network. Gephi software was used in this study to obtain the analysis data of the tennis competition network and to visualize the network. The choice of Gephi software is free, the performance of getting analysis data and the ability to present detailed options during network visualization.

III. RESULTS AND FINDINGS

To create a complex network of tennis matches, players in the Australian Open, the French Open, the US Open and the Wimbledon tennis tournaments played each other's matches between 2000 and 2017 as used data set. While the players form nodes for the four networks created, the matches that the players made with each other created links between nodes. The connections between the nodes are created in a undirectional way. Node and link information for the created networks is shown in Table 1.

Parameters such as network size, center of gravity, diameter, clustering coefficient, density, average number of links, weighted average of link used in analyzing and interpreting networks are calculated before visualizing the created network. The analysis values calculated for four tournaments are shown in Table 2.

Table 1: Nodes and Edges Created for Tennis Network.

	Australian Open	French Open	US Open	Wimbledon
Nodes	553	526	543	547
Edges	2156	2153	2163	2159

Table 2: Network Measures Calculated for Tennis Network.

	AO	FO	USAO	W
Average Degree	7,797	8,186	7,967	7,894
Average Weighted Degree	8,268	8,692	8,42	8,358
Network Diameter	7	7	7	7
Graph Density	0,014	0,016	0,015	0,014
Modularity	0,333	0,317	0,318	0,319
Average Clustering Coefficient	0,086	0,093	0,099	0,078
Average Path Length	3,3	3,212	3,247	3,295

A. Average Degree

The average degree is used as a parameter to evaluate the effect of strong or weak nodes on the network as a whole. Higher average rates indicate that the nodes in the network have a strong relationship with each other, while lower nodes indicate a weaker linkage among the nodes in the network [3]. When we look at Table 2, it is seen that the average scores of the tennis tournaments we have formed are close to each other for four tournaments. When the degree distribution was examined, it was seen that a small part of the players participating in the tournament played matches with a large number of players This degree distribution shows that it is one of the characteristics of scale-free networks and is similar to the power law which is an important element of real world networks. The degree distributions are shown in Figure 1 and Figure 2.

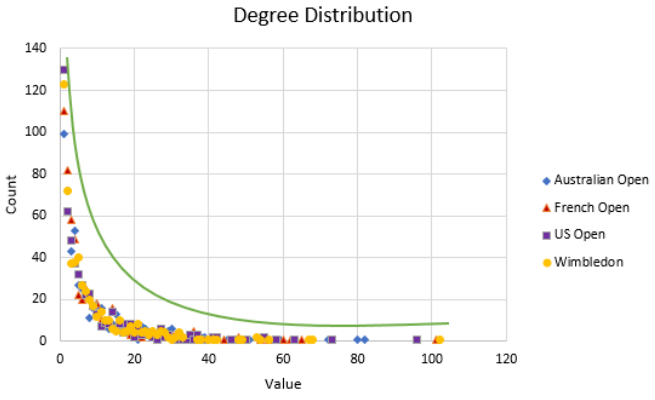


Figure 1: Degree Distribution in Tournament Networks.

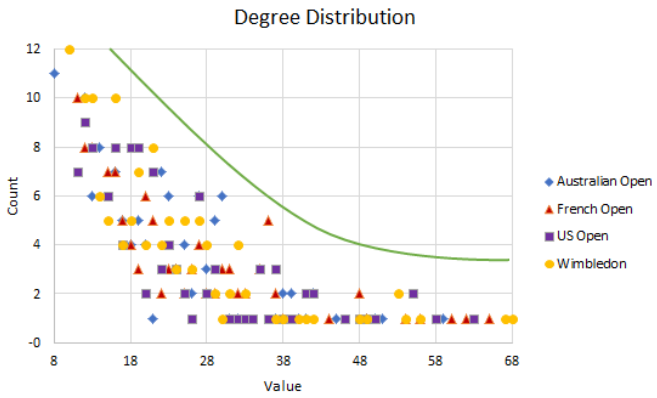


Figure 2: Degree Distribution in Tournament Networks with Zoom

.

Given the degree distributions calculated for the four tournaments and the resulting graphs, they show a degree distribution that is in line with the small world phenomenon, one of the features of real world networks of tennis competitions. The fact that fewer athletes have so many matches with new players attended tournaments every year has led to the distribution of degrees of strength in accordance with the power law.

B. *Network Diameter*

The diameter of the network is a parameter that indicates the shortest path between two nodes in the network. This parameter is important to examine the centrality of networks. When the network diameter is low, the centrality is high while the high diameter of the network indicates that the network is away from centrality [17]. Table 2 shows that the network size for the networks we have created is the same for all four networks. This indicates that the centrality of networks has similar characteristics. Similar features of the diameter of the networks indicate that there is a cluster in certain players on the network. Who have been successful on tournaments and who have played more such as Federer, Djokovic, Nadal, Murray, Ferrer have formed a cluster around themselves is becoming more evident in the visualization of networks. When we look at the values of the network diameter, it seems to have similar characteristics to the small world phenomenon.

C. *Graph Density*

The network density parameter refers to the ratio of all connections in the network to the maximum number of connections that can subsequently occur. As the network density ratio approaches zero, the connections between the nodes are reduced and the result that the nodes are isolated from each other can be uncovered, and the result that the nodes in the network connect to each other can be deduced as the network density ratio approaches one [18]. Table 2 shows that the network densities for the four networks we created are similar to each other and close to zero. Depending on this finding of the general structure of the networks, the diversity of players struggling in tournaments is constantly changing, but this variability decreases in the upper tours and players such as Federer, Djokovic, Nadal, Murray, Ferrer Berdych, Roddick, Wawrinka, Davydenko etc. who had previously struggled with each other have played more matches.

D. *Modularity*

The modularity parameter is an important parameter used to evaluate the clustering within the network. The values of [-1, +1] give information about the density of the relations of the nodes in the network to each other. If all of the links in the network occur between nodes similar to each other, the modularity parameter is calculated as 1, modularity is computed as -1 as a result of the fact that the links between the nodes in the network occur entirely between nodes that do not resemble each other [3]. Table 2 shows that the calculated modularity values for the networks we have created are close to each other, but especially in each tournament, there are players such as Federer, Djokovic, Nadal, Murray, Ferrer who have encountered these densities in relation to each other.

E. *Average Clustering Coefficient*

The clustering coefficient is a parameter used to measure the density of links around a node. The average clustering coefficient is calculated by taking the average of the cluster coefficients of each node in the network [17]. Table 2 shows that the cluster coefficients in our networks are calculated as near zero values. The reason for this is that the clusters on the network are in the upper laps in the tournaments and in the more successful players. Players like Federer, Djokovic, Nadal, Murray, Ferrer get through to next round in every tournament and compete different opponent and with each other to increase their clustering around them. Because of this situation, the cluster coefficients of these players are calculated higher than the average cluster coefficients.

F. *Average Path Length*

The average path length is a parameter that indicates the average of the shortest inter-node distances of all pairs of nodes in the network. The low average distance of the paths indicates that the nodes that are not directly connected to each other in the network are more likely to establish direct connections with each other [3]. Table 2 shows that the average path lengths in the networks we created are calculated at similar values and the pairs of nodes can reach each other in 3-4 steps. This shows that players who have never competed with each other have a high probability of having a match with each other in the next tournaments.

G. *Network Visualization*

Visualization helps to understand complexity and analyze network structure in complex networks with many nodes and

connections. Gephi software has been used for the visualization for tennis matches Networks. In Gephi software, separate networks were created and visualized for the Australian Open, French Open, US Open and Wimbledon tournaments. The eigenvector centrality values of networks have been exploited to reveal the most efficient, connected nodes in the network during visualization. Modularity values are also calculated to visualize different clusters formed in networks. In Gephi software, Modularity and eigenvector centrality calculations of networks are used. Depending on these parameters, the overall visual appearance of the networks is given in Figure 3 for the Australian Open, Figure 4 for the French Open, Figure 5 for the US Open and Figure 6 for Wimbledon.

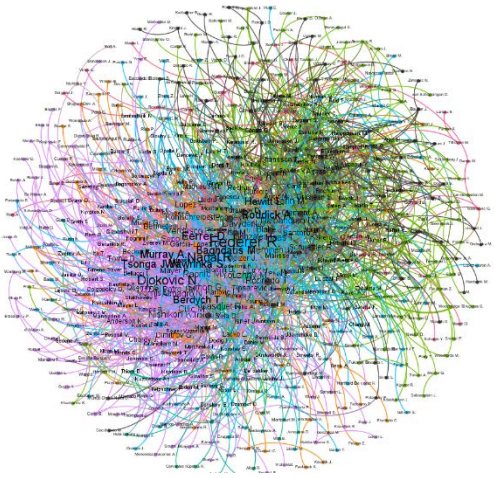


Figure 3: Network Created for Australian Open

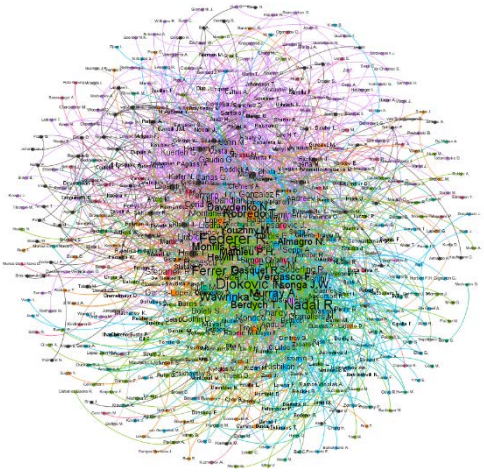


Figure 4: Network Created for French Open

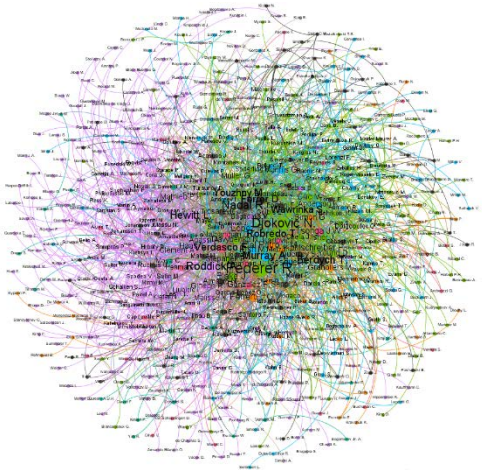


Figure 5: Network Created for US Open

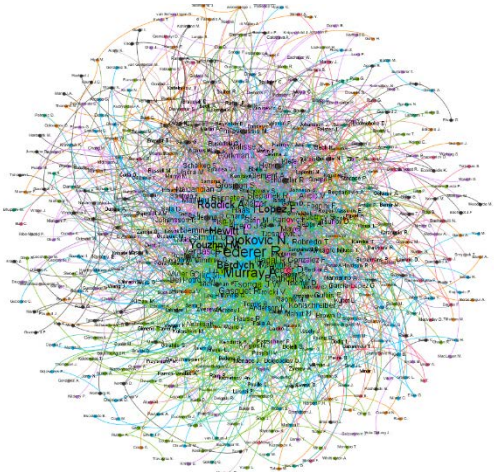


Figure 6: Network Created for Wimbledon

The variety of colors in the network indicates different clusters. In addition, according to the number of competitions performed by the players in other words according to their grades, the nodes are displayed at different diameters. The fact that the connections between some nodes are shown thicker is due to the density of the link between the two nodes. The intensity (weight) between the two players according to the number of matches they have made is visualized in the networks. Nodes with excess grades and weights are gathered at the center of the networks. This nodes which are located in the center belongs to the players who play most in tournaments.

The network has spread out towards from the most competitive to the less-competitive players. While the weakly connected nodes created by a few players in this way are located farther away from the center of the network, the nodes which are strong connections formed by a large number of players are at the center of the network. Players at the center of the network and clustered are both clustered with players who match only themselves, as well as with high-clustered players like themselves. As the number of players struggling up to the nearest round to the finals each year is high, the closeness to the network center changes accordingly. It can be said that for the nodes in the centers of the networks there are players who have

consistently succeeded in the tournements. In Figure 6, Federer, who is at the network hub for the Australian Open, shows both the clusters he has created within himself and the links with different clusters.

The clusters at the centers of the networks can vary in each tournament. Figure 7 shows the Federer at the network center for the Australian Open, Figure 8 shows in the center of the network for the French Open, Nadal, Djokovic and Federer are close to each other. Looking at the data of two tournament, for the Australian Open the Federer's grade is 101, Nadal's grade is 62, Djokovic's grade is 61, while the French Open is Nadal's grade 82, Federer's grade 80 and Djokovic's grade 79. This data confirms that for both networks, the cluster at the center is shaped according to the number of degrees of nodes.

Moreover, when the structure of the networks is examined, it is seen that tournament win is not enough to become popular in the network, and even if there are not many championships, it can become a popular node in the network. For example, while the Federer, the second most popular node in the French Open, has only one championship, the most popular node, Nadal, has 10 championships. In the same tournament, Kuerten's populace, which had 2 championships between 2000-2017, ranks 26th.

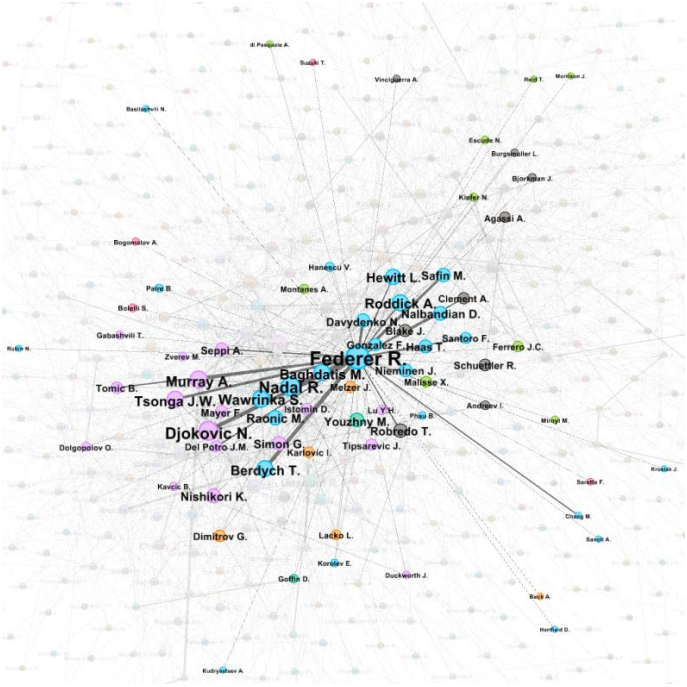


Figure 7: Clustering in the Network Center on Australian Open

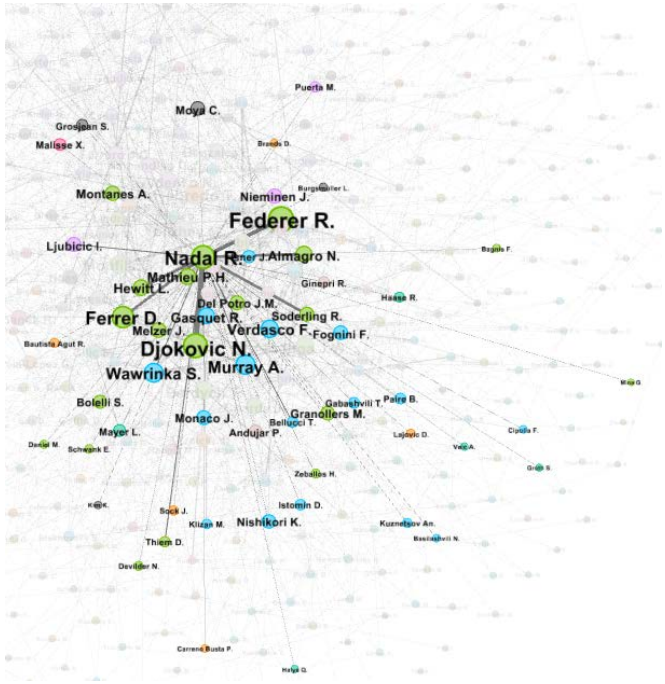


Figure 8: Clustering in the Network Center on French Open

IV. CONCLUSION

In this study, the analysis and visualization of the players in the Australian Open, the French Open, the US Open and the Wimbledon tennis tournaments between 2000-2017 in terms of the complex networks of the players and their competitions were conducted. It was observed that the encounters in the tennis tournaments according to the obtained network diameter, network density, modularity, average cluster coefficient, average path length values have real world network characteristics because they are similar to small world phenomenology and scale-free theories. When we look at the networks belonging to four created tournaments, they are structurally similar to each other. The networks and analysis results obtained in this study showed that the associations in tennis competitions can carry similar features to real world networks through complex network structures. In addition, it has been seen that the results obtained by these sports networks can be used to create similar networks in real world networks. The data obtained from the complex networks of tennis competitions can be used to calculate the likelihood of players who have never met each other in the tournaments to be played in the following years by using link prediction methods.

REFERENCES

[1] S. H. Strogatz, *Exploring Complex Networks,* Nature, 410, (6825), 268-276 (2001).
[2] M. E. J. Newman, *SIAM Rev.* 45, 167 (2003).
[3] Ş. Akal, "*Gerçek ve model ağların karakteristik özelliklerinin karşılaştırılması*", Marmara Üni. Öneri Dergisi., 11 (41): 251-272 (2014).
[4] G. Caldarelli, A. Vespignani, *Preliminaries And Basic Definitions In Network Theory,* Large Scale Structure And Dynamics Of Complex

Networks From Information Technology To Finance And Natural Science, World Scientific Publishing Co., Singapore, 2, 3, 5, 6 ,12-15, 205-207 (2007).

[5] M. Van Steen, *An Introduction to Graph Theory and Complex Networks*, 1st ed., 11,12, 166, 167 (2010).

[6] E. E. Sulak, "*Sosyal ağlarda etiket ağı analizi*", Yüksek Lisans Tezi, Karabük Üniversitesi Fen Bilimleri Enstitüsü, Karabük, 4, 12 (2016).

[7] İ.Türker, A. T. Karadeniz, S. O. Tan, *A Computer Software For The Education Of Complex Network Analysis*, The Eurasia Proceedings of Educational & Social Sciences (EPESS), 2, 57-62 (2015).

[8] S. Milgram, *Psychol*, Today 1, 60 (1967).

[9] A. –L. Barabasi, R. Albert, ,*Emergence Of Scaling In Random Networks*, Science, 286, 509 (1999).

[10] J. Leskovec,E. Horvitz, Proceeding of the 17th international conference on World Wide Web ACM, New York, 915 (2008).

[11] A. –L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, T.Vicsek, *Evolution Of The Social Network Of Scientific Collaborations*, Physica A, 311, 590(2002).

[12] M. E. J. Newman, *Scientific Collaboration Networks. I. Network Construction And Fundamental Results*, Physical Review E, 64, 016131 (2001b).

[13] A. Çavuşoğlu, İ. Türker, *Scientific collaboration network of Turkey*, Chaos, Solitons & Fractals 57, 9-18 (2013).

[14] L. A. N. Amaral, A. Scala, M. Barthelemy,H. E. Stanley, *Classes Of Small-World Networks*, Proceedings of the National Academy of Sciences of the United States of America, 97, 11149 . (2000).

[15] A. Clauset, C. R. Shalizi, M. E. J. Newman, *Power-Law Distributions In Empirical Data*, SIAM Review, 51, 661 (2009).

[16] A. –L. Barabasi, R. Albert, *Emergence Of Scaling In Random Networks*. Science, 286, 509 (1999).

[17] A.-L. Barabási, R. Albert, H. Jeong, *Diameter of the World-Wide Web*, Nature, 401, 398–399 (1999).

[18] M. E. J. Newman, *The Structure and Function of Complex Networks*, SIAM Rev., 45, 2, 167–256, (2003).

Complex Network Analysis of UEFA Europe League Competitions

E. E. SULAK[1], H. YILMAZ[1] and E. OZKAYNAK[1]

[1] Karabuk University, Karabuk/Turkey, ekmelsulak@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, hakanyilmaz@ karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, eozkaynak@ karabuk.edu.tr

Abstract **- Today, with the development of information technologies, studies on the evaluation of relations between people, objects and events to put it simply relations between creatures and facts have taken place among general topics. As the data related to the relationships between creatures and facts continue to increase and become storable together with developing technology, new methods are being developed for analyzing and evaluating these data. There are both semantic and structural relations within the obtained data stacks. Complex network analysis is one of the most common methods used to reach semantic relationships as a result of analyzing and evaluating these relations. In this study, a network was formed among the teams who competed in the UEFA Europe League between 2004-2017 by using the data of football competitions played after the groups. As a result of the evaluation, different categories such as last 32, last 16, quarter final, semi final and final matches are analyzed and the similarities of the networks to the real world networks are compared with the data obtained about the network structures. As a result of the study, it is observed that the teams that compete in the UEFA Europe League are similar to the real world networks.**

Keywords **- Complex Networks, Scale-Free, Small World.**

I. INTRODUCTION

In real life, people, objects and events are constantly interacting. This interaction that takes place in daily life brings a network structure. In recent years, complex networking has become a scientific discipline that has been extensively studied. On the basis of this increasingly intense interest in complex network science, there is an increase in interesting applications developed in various fields of biology, sociology, technology and communication, and a large amount of real data corresponding to complex network analysis [1-4]. It seems that social networks are constituted by some factors such as friendship, work and sexual partner. Given the work done in this area, with Leonhard Euler's Konigsberg bridge problem, the concept of Graf emerged for the first time [1]. Milgram's small world theory is noteworthy in his studies of complex network science. In this study, Milgram has shown that the average distance between two nodes is six steps [5]. In a genetic regulatory network in the field of biology, genes, blood vessels, food networks and the relationships of metabolic pathways have been analyzed [6-9]. Complex network studies in the field of technological and communication such as electric network, airways, railways, internet network and www network have been important work in this area [4,10-13]. The size of a network is directly proportional to the number of nodes and links. Networks with fewer nodes and links are easier to visualize and analyze. As the network grows, some parameters are needed for visualization and analysis. These parameters are average degree, average weighted degree, network diameter, graph density, modularity, connected components, average cluster coefficient and average path length. These parameters are discussed in detail in Chapter II. In this study, during the UEFA Europe League between 2004 and 2017, team-based nets were created and analyzed using the data of the competitions among the teams from the groups [14]. The last 32, the last 16, the quarter final, the semi-final and the final matches played after the group matches were analyzed in separate categories and the results of the networks were visualized and the results obtained were interpreted.

II. COMPLEX NETWORKS

The structure that many nodes and the links between these nodes form is called a complex network. The concept of network is mathematically indicated and its elements are expressed by graphs [15]. Networks are more understandable using graph structures. In the analysis of complex networks, it is important to know some basic concepts and definitions in complex network science so that the obtained analysis data can be understood. It is possible to analyze as a real or model network by looking at the topological properties of the network. The basic concepts in network science, both topologically informing about the structure of the network, give an idea of the state of the network. Between real networks and model networks, they may have common characteristics, and these networks may show different characteristics [16].

A. Degree of Nodes

The relationship between node and link, which are the two basic concepts that make up the structure of networks, can be explained by the notion of degree. In a network composed of football competitions, teams express nodes while matches between teams express links. The degree is the number of links in a network of a node. The concept of degree is an important parameter used to identify popular nodes in a complex network. In a network composed of football competitions, the team with

the highest rating is not only the team that plays the most matches but also the knowledge of the most popular node of the network [17].

B. Degree Distribution

In complex networks, the number of repeats of each node's degree values on the network is an important parameter that gives the degree distribution of the network. The degree distribution is the histogram of the network. It is possible to have an idea of the network structure by looking at the data obtained from the degree distribution [15,18]. Networks are random networks if the degree distribution is poisson as shown Figure 1. Since links are randomly distributed in random networks, the average ratings of the nodes are close to each other [19].

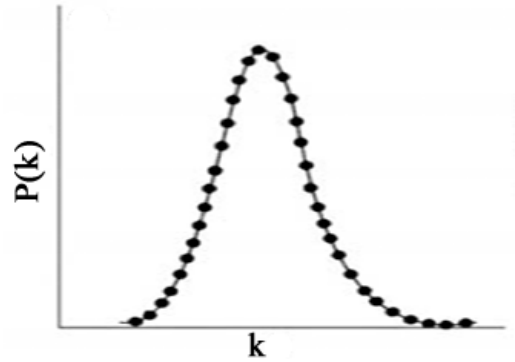


Figure 1: Poissonian degree distribution for random networks [20].

Networks which have a heterogeneous degree distribution as shown in Figure 2 are real world networks. Real world networks are networks that show power-law feature. In wide world networks, the degree distribution does not concentrate in a particular region but spreads over a large scale [18,19,21].

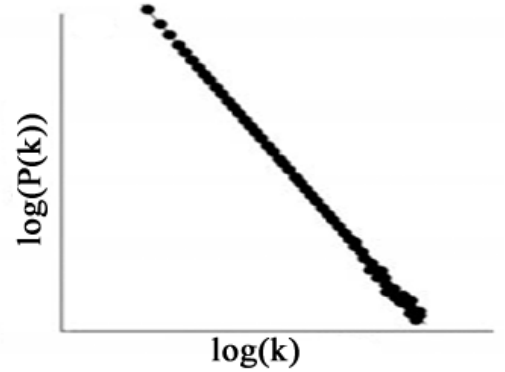


Figure 2: Power-law degree distribution in scale-free networks [20].

C. Clustering Coefficient

In cases where there are many nodes that are popular in the network, small clusters formed with the nodes connected to it can be explained by the concept of clustering in complex networks. Neighborhood relationships between adjacent nodes connected to a popular node allow the node to show the cluster feature within the network. The matches the teams against each other who matched teams that are popular in the network by making the most matches is effective in seeing the effectiveness of the central nodes in the network structure by creating clusters within the network [19,22].

D. Shortest Path

The closest distance in the networks connecting the two nodes gives the shortest path parameter in the complex networks. While shortest path information can be easily found in complex networks where all nodes are interconnected, in some complex networks the shortest path information is not readily available due to the presence of isolated nodes or clusters. In Milgram's letter experiment, the fact that the nodes are very close together, that is, the average length of the shortest path varies from 3 to 6 steps, is a feature that indicates that the network can be a real world network [19,23].

E. Diameter

The minimum distance between two nodes that are not connected to each other by node pairs in a complex network is the information that gives the diameter of the network. Diameter is a condition that contains information about the structure of the network topologically. In the small world theory, the average diameter of the network is assumed to be 6. Nets determined in accordance with the small world theory of diameter are similar to real world networks [16,24]. Graph density is called the ratio of links and possible links on the network Graph density is important because of the consideration of the nodes that have not yet connected, but possibly connect.

F. Modularity

The multiplicative constant of the links in the network and the number of links in an equivalent network that occurs in random links is called the modularity [25]. In other words, modularity is likely to be divided into subgroups of the network. As the diameter of the network increases, the probability of clustering nodes and links increases, so the modularity of the network will also increase. Especially if the diameter is short, the small change in the diameter will cause a large increase in the modularity.

III. Method and Results

A. Data From Networks

In the network formed by UEFA Europe League matches, the matches played after UEFA Europe League group matches between 2004-2017 were taken in tours and in whole. Teams matching on established networks form nodes. The matches that teams have made with each other have created links. The generated network consists of undirected links. The node and link information created for the UEFA Europe League network is shown in Table 1.

Table 1: Nodes and links for UEFA Europe League networks.

	All Rounds	Round 2	Round 3	Quarter Final	Semi Final	Final
Nodes	154	154	96	64	37	20
Edges	388	203	101	51	26	13

The visualization of the network and the understandability of the data to be obtained can become so complicated if the generated networks contain the more nodes and links. Therefore, various applications are used to make the data obtained from nodes and links accurate, understandable and to visualize the network. Gephi software is preferred for obtaining data from the network created in this study and visualizing the network. Gephi software is one of the most preferred software, both because it is free and because the analysis data is accurate, understandable and the performance of network visualization is high.

The parameters used in the analysis, interpretation and visualization of network such as network centrality, average number of connections, size, density, diameter, cluster coefficient, weighted average degree in UEFA Europe League matches are obtained from the networks created for each round of matches played after the group matches and are given in Table 2.

Table 2: Network measures calculated for UEFA Europe League networks.

	All Rounds	Round 2	Round 3	Quarter Final	Semi Final	Final
Average Degree	5,039	2,636	2,104	1,594	1,405	1,3
Average Weighted Degree	5,234	2,701	2,167	1,625	1,405	1,3
Network Diameter	7	9	14	14	7	3
Graph Density	0,033	0,017	0,022	0,025	0,039	0,068
Modularity	0,367	0,647	0,755	0,807	0,826	0,737
Average Clustering Coefficient	0,077	0,016	0,028	0	0	0
Average Path length	3,143	4,6	6,401	5,621	2,716	1,69

In Table 2, it is seen that the average cluster coefficients are zero in the quarter final, semi final and final networks. The average clustering coefficient approaches zero is due to the absence of clusters in the network or the emergence of isolated clusters. In UEFA Europe League matches, different teams appear in the upper tours every year and these teams have not struggled before in these tours has caused the cluster to be less common in the network. It seems that the average path length values in the networks have fallen as the upper tour pass. There are two reasons for this. The first reason is that the network diameter is low due to the reduced number of teams matching in the upper tours. The second reason is that the teams matching in the upper tours have not struggled or struggled less in the same level before. The diameter of the network is a parameter that indicates the shortest path between two nodes in the network. When looking at the diameter of the networks, it appears that they have moved away from the centrality by the second, third and quarter final tours. The reason for this is that the teams that go up to the top rounds differ every year. The decrease in the network diameter of the other tours is proportional to the decrease in the number of teams matching. Teams like Sevilla, Benfica have to match more in the last tours has been influenced by the decline of the network diameter and the convergence of centennials. When we look at density data in networks, they are calculated as near zero values. When the density information obtained according to the general structures of the networks is examined, although the teams struggling each year show a large variation, Sevilla, Benfica, Villereal, who have been successful in the top tours every year, have shown that the matches they have with each other have influenced the network density in the top tours. When the modularity data of the networks is examined, similar values are seen in the middle of the networks created for other tours except for the network created for all matches. The reason for this is similar to the situation in the network density, which has been effective in raising the modularity of networks in the upper tiers of the challenges that teams such as Sevilla, Benfica, Villereal. When the average degrees are seen in the generated networks, it is seen that the average degrees are decreased as they go up the upper tours. The reason for this is that a small part of the team that competes at every level of the match every year has a variety of matches with the teams and the teams who are matching at the top tours are changing every year. The degree distributions for these networks show that it is one of the characteristics of scale-free networks and is similar to the force law, which is an important component of real world networks. The degree distributions are shown in Figure 1.

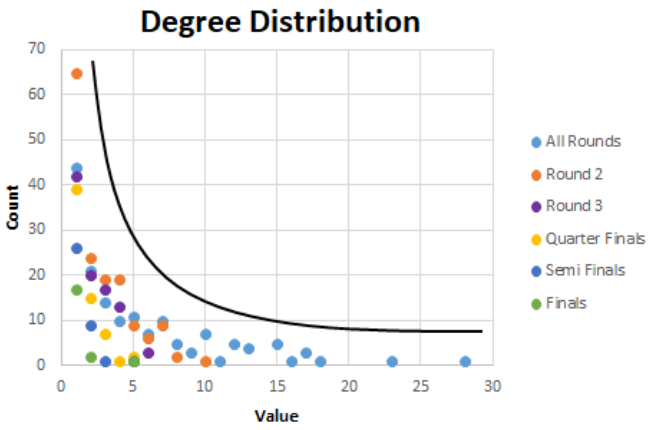


Figure 1: Degree distribution in UEFA Europe League networks.

It is seen that UEFA European League matches show a degree distribution in accordance with the small world phenomenon which is one of the characteristics of real world networks, considering the degree distributions calculated for all matches and the graphs formed. Each year, with different teams participating in the championship, fewer teams participated in

the championship more and more, causing the distribution of grades to fit the power law.

B. Visualization of the Network

Visualization is important for easier understanding of networks with too many nodes and links in the structure. Gephi software has been preferred for visualization of networks created for UEFA Europe League matches. Second round, third round, quarter final, semi final, final and a combination of all tours networks are visualized separately in Gephi software. Modularity and eigenvector centrality values are calculated to better understand the clusters in the network during visualization. Generated networks are given in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7.

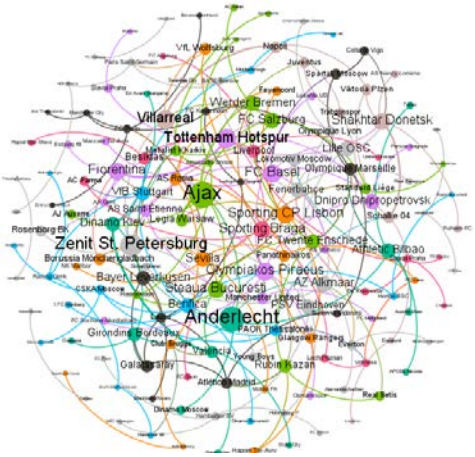


Figure 2: UEFA Europe League 2nd Tour Network.

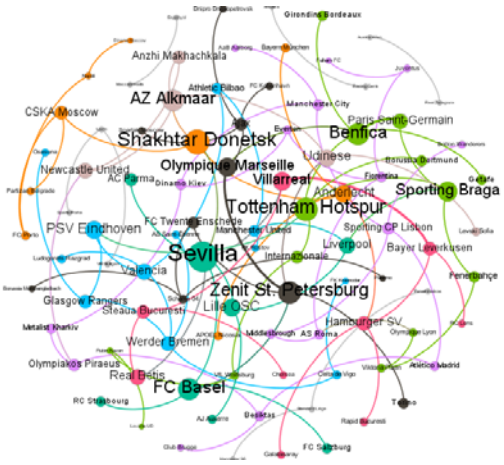


Figure 3: UEFA Europe League 3rd Tour Network.

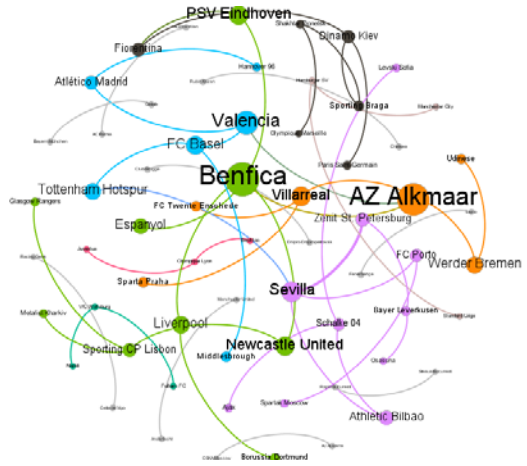


Figure 4: UEFA Europe League Quarter Final Network.

Figure 5: UEFA Europe League Semi-Final Network.

Figure 6: UEFA Europe League Final Network.

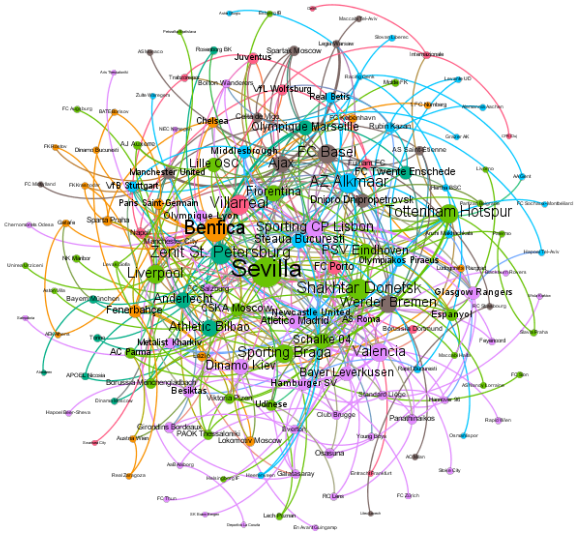


Figure 7: UEFA Europe League Network.

When the results obtained after visualization of the networks are taken into consideration, it can be seen that there are many different clusters of tours. The resulting clusters are expressed in different colors. In the centers of the networks, the teams that played the most in those tours seem to be involved because of their high degree. The thickness of some links between the teams is due to the greater number of matches between them. For this reason, teams with high degrees and weights are gathered at the center of the networks. Teams playing few games in the championship have spread around the network due to poor network connections. The fact that different teams take place in the center of the network in every round of the championship is due to the fact that different teams increase their profits by making more matches. While the Ajax team was in the center of the network in the second round, it was seen that Sevilla took center stage in the 3rd, semi-final, final and all matches networks. In the quarter finals, Benfica's presence in the center of the network shows that Benfica is the most-played team in this round. The isolation of some of the pairs of nodes seen in all of the networks shows that these teams played only one match on these tours and did not match in those tours afterwards.

IV. CONCLUSION

In this work, which was performed from the second round, third round, quarter-finals, semi-finals and final matches of the UEFA Europe League matches between 2004-2017, the relationship between the teams matching in the championship and the matches between them was evaluated and visualized in terms of complex network analysis. It has been seen that the results obtained are similar to real world networks because of the properties suitable for power law in terms of distribution of networks. Variations in the diameters and cluster coefficients of the networks were found to be due to the reduction of team numbers in the upper tour. According to the analysis results obtained in this study, UEFA European League matches have been shown to be similar to real world networks in some tours. As a result of the obtained data, it is thought that relations in different sports branches may have similar characteristics to real world networks.

REFERENCES

[1] M. E. J. Newman, SIAM Rev. 45, 167 (2003).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *in Evolution of Networks* (Oxford University Press, Oxford, 2003).
[3] P. S. Dodds, R. Muhamad, and D. J. Watts, Science 301, 827 (2003).
[4] D. J. Watts and S. H. Strogatz, Nature (London) 393, 440 (1998).
[5] S. Milgram, Psychol. Today 1, 60 (1967).
[6] T. I. Lee et al., Science 298, 799 (2002).
[7] G. B. West, J. H. Brown, and B. J. Enquist, Nature (London) 400, 664 (1999).
[8] J. Camacho, R. Guimerá, and L. A. N. Amaral, Phys. Rev. Lett. 88, 228102 (2002).
[9] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L Barabási, Science 297, 1551 (2002).
[10] L. A. N. Amaral, A. Scala, M. Bartheélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. 97, 11149 (2000).
[11] P. Sem et al., Phys. Rev. E 67, 036106 (2003).
[12] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. 29, 251 (1999).
[13] A.-L. Barabási, R. Albert, and H. Jeong, Physica A 272, 173 (1999).
[14] Internet: "*UEFA European Cup Coefficients Database*", https://kassiesa.home.xs4all.nl/bert/uefa/data/index.html (2018).
[15] A. Barrat, M. Barthélemy and A. Vespignani, "*Dynamical Processes on Complex Networks,1st ed.*", Cambridge University Press, New York, 1-3, 11-13, 26, 27 (2008).
[16] E. E. Sulak, "*Sosyal ağlarda etiket ağı analizi*", Yüksek Lisans Tezi, Karabük Üniversitesi Fen Bilimleri Enstitüsü, Karabük, 4, 12 (2016).
[17] S. Demir, "*Türkiye'de mühendislik alanındaki bilimsel işbirliği ağı haritasının analizi ve görselleştirilmesi*", Yüksek Lisans Tezi, Karabük Üniversitesi Fen Bilimleri Enstitüsü, Karabük, 5, 8, 10-11 (2015).
[18] E. Estrada, "*Graph and Network Theory,Mathematical Tools for Physicist*", 2nde d., Grinfeld, M., Wiley-Vch, Weinheim, 134, 135, 137-139 (2015).
[19] İ. Türker, "*Türkiye'deki bilimsel işbirliği ağı ve dinamikleri*", Doktora Tezi, Karabük Üniversitesi Fen Bilimleri Enstitüsü, Karabük, 11-15 (2013).
[20] L. Costa, F. Rodrigues and A. Cristino, "*Complex networks: the key to systems biology*", Genetics and Molecular Biology, 31 (3): 591-601 (2008).
[21] Ş. Akal, "*Gerçek ve model ağların karakteristik özelliklerinin karşılaştırılması*", Marmara Üni. Öneri Dergisi., 11 (41): 251-272 (2014).
[22] R. Van der Hofstad, "*Random Graphs and Complex Networks, Vol. I*", Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, Netherlands, 1, 2, 16, 17, 21, 22, 27, 28, 33, 34 (2014).
[23] M. Van Steen, "*An Introduction to Graph Theory and Complex Networks,1st ed.*", Maarten van Steen, 11,12, 166, 167 (2010).
[24] Z. Zhang, J. Zhang, "*A big world inside small-world Networks*", Plos one, 4 (5): e5686 (2009).
[25] M. Newman, "*Modularity and community structure in networks*", Proceedings of the national academy of sciences, 103 (23): 8577-8582 (2006).

Implementation of Clarke & Wright Savings Algorithm in Social Network Analysis: An Exemplary Event Planning System

L.SABAH[1] and M.ŞİMŞEK[1]

[1]Düzce University, Düzce/Turkey, leventsabah@duzce.edu.tr
[1]Düzce University, Düzce/Turkey, mehmetsimsek@duzce.edu.tr

Abstract **- Today, there are some difficulties of planning a social activity. One of the biggest challenges is that people have more mobile lifestyles in the last few decades. For this reason, the possibility of participating of people in a particular event varies according to the person's current location. However, the fact that today's people are mobile as well as online, this makes it possible to reach them virtually even if they are not physically reachable. In this study, an approach has been developed to identify and recommend the most suitable candidates according to the compliance and location of the people for a social activity that is planned at a certain capacity. In the developed approach, people are weighted according to their compliance and physical distance to the event. Then, people are sorted by the location of the event. The most suitable people are identified, not more than the capacity of the event. The proposed approach uses person characteristics (gender, age, areas of interest), the Euclidean distance of the people to the event and the location where the event is to be done. The proposed approach also uses the Clarke & Wright Saving Algorithm, which is a heuristic algorithm and is mostly used in vehicle routing problems in the logistics area. Thus, the most appropriate persons can be identified that can be invited to an event with minimal cost. Furthermore, in addition to the use of this algorithm, if people invited to the event are taken by car, it is possible to determine the most suitable routes and to ensure that people are collected according to these routes and participate in the event.**

Keywords **- Social Network Analysis, Spatial Data Analysis, Geographic Information Systems, Clarke & Wright Saving Algorithm, Event Planning System, Analytic Hierarchy Process**

I. INTRODUCTION

Recent years have witnessed the popularity of event-based social networks (EBSNs) acting as a bridge between the cyber world and the physical world. With the help of EBSN platforms, users could be easily informed about upcoming events from the cyber space and decide to attend offline events in the physical world according to their preferences [1]. Cultural events and activities have always been an excellent source of revenue for cities around the world. Many public authorities try to organize and establish periodical cultural events in order to attract more and more visitors to their areas and stimulate commercial activity for local markets [2]. Planned events in tourism are created for a purpose, and what was once the realm of individual and community initiatives has largely become the realm of professionals and entrepreneurs. Four main categories of planned events within an event-tourism context, including the main venues associated with each. Business events (or the MICE (meetings, incentives, conferences, and exhibitions) sector) require convention and exhibition centres, including numerous, smaller private parties and functions held in restaurants, hotels, or resorts. Sports also require special-purpose facilities including athletic parks, arenas and stadia. Festivals and other cultural celebrations are less dependent on facilities and can use parks, streets, theatres, concert halls and all other public or private venues. Entertainment events, such as concerts, are generally provided by the private sector and utilize many types of venue [3]. An individual does not only have the opportunity to search for information or use a service but she/he is also able to generate and contribute self-generated information and disseminate it to the broad public, thus becoming a vivid member of a global community [4]. The location of a social event also plays a role in event recommendation [5]. The increasing volume of information received by people in their daily lives usually presents the challenge of deciding what information is useful for them, and which does not. Recommender systems are tools that can be used to suggest items that may not have been found by users themselves [6]. In this context, the advent of mobile devices has allowed the use of location information to provide context-aware recommendations by considering the distance between users and items, as well as their subsequent movements. The ability to combine users' location and movements, together with other aspects like users' preferences, items' properties, or users' ratings provides more valuable information that can help to suggest more accurate items of potential interest to users [7]. Recently researchers have investigated the possibility of using recommender systems to identify useful contents also in the opportunistic environment (mainly based on content filtering and tag expansion). Recommender systems perform better than publish/subscribe mechanisms by mainly relying on information about users' past actions (e.g., past purchases in e-commerce or past visualizations of multimedia content in video-on-demand services) [8]. Human-generated knowledge with efficient automated techniques when solving hard computational tasks opens new possibilities [9]. Events are increasingly becoming a focus for target marketing organizations because of the tourist pay and expenses that they attract. Consequently, an event has emerged tourism phenomenon that seeks to use events as

tourism assets for the growing tourism. Such practices can have significant implications for local communities [10]. In this study, as a new approach to event planning and recommendation; interests, gender and age information of people are assessed through the AHP. Afterwards, it is aimed to determine the suitability of these persons according to their distance to the event by using the Clarke & Wright algorithm. Thus, it is aimed to select the most suitable community for an activity by using geographical distance from planned event and personal information of people.

II. MATERIALS AND METHODS

A. *Clarke & Wright Algorithm*

One of the methods used for solving vehicle routing problems is the heuristic methods. Heuristic methods; it produces solutions with good performance, close to the best solution with less processing and calculation time for big problems. It is often preferable to have approximate results for such problems. Thus, satisfactory results for the general purpose are quickly found. The heuristic saving algorithm proposed by Clarke and Wright in 1964 [11][12].

Clarke-Wright heuristic, it is constructed by an algorithm based on the saving value which is revealed by returning to the starting point at both of these points, instead of going from one starting point to two different points separately [13].

The Clarke & Wright algorithm is an easy to implement, fast and simple solution-building heuristic algorithm. The Clarke & Wright algorithm based on the "savings" principle, which can be achieved by combining two separate routes, can be applied both directionally and non-directionally. There are two different types of Clarke & Wright algorithms, parallel and serial in practice. In the parallel Clarke & Wright algorithm, multiple routes are created at the same time, while in the serial Clarke & Wright algorithm a route is created at each time. There is no significant difference in quality of solution between serial and parallel algorithms [14]. In our work, the serial version was used in the selection of the persons who are suitable for the activity. In addition to preparing routes with Clarke & Wright, also how many cars will need to event location [15].

In general, the steps of the Clarke & Wright algorithm are as follows.

Let:

- d: shortest distance of x_1y_1 and x_2y_2 coordinates.
- 0 : event place location.
- s_{ij}: savings value of i and j location.
- c={1,2,…n} : set of individuals location.
- c_{ij}: distance between i and j location.

Step 0 : Distances of the points to be used on the route to the event place and the distances of the points according to each other are calculated by the shortest Euclidean relation (Eq. 1) (Fig. 1).

$$d=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \quad (1)$$

Step 1 : $s_{ij}=c_{0i}+c_{0j}-c_{ij}$ formula is used to calculate savings (Fig. 2).

Step 2 : Calculated saving values are sorted by descending order.

Step 3 : Route generation is begin. In this step, (i, j) pairs are compared. If they are present on the existing route, they are not evaluated and the iteration is continued.

Step 4: The process is terminated after all the pairs have created a route based on their savings value.

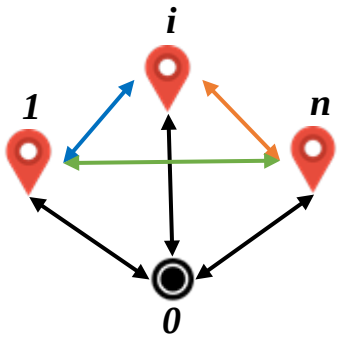


Figure 1. Step 0: Initialization of distance of individuals and event place

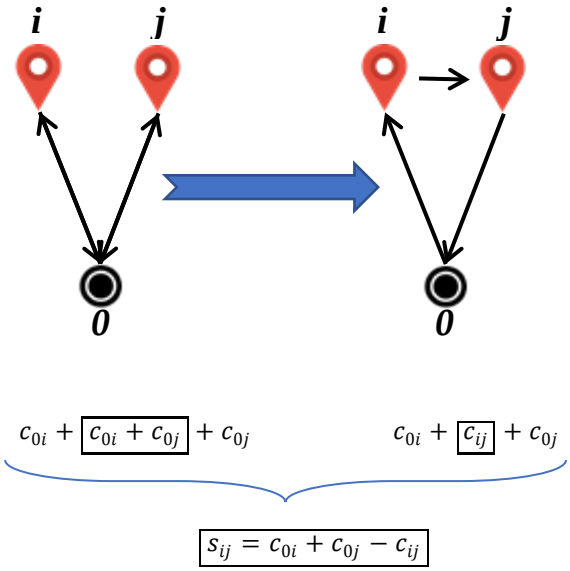


Figure 2. Step 1: Generating savings

B. *Adaptation of Clarke & Wright Algorithm to Event Planning System*

A data set was created for the implementation of the proposed approach (Fig. 3). There are x and y coordinates, age, gender and 5 points of interest information of the person who has the data of 50 people. This information is randomly generated. Age range is 13-70, gender is 1 for female, 2 for male. 5 areas of interest were created between 0-10. 0 indicates the lowest degree of interest, 10 indicates the highest degree of interest.

id	x	y	age	gender	interest1	interest2	interest3	interest4	interest5
1	30.5453977	40.7768876	23	2	5	9	0	1	10
2	30.5072893	40.7255241	66	2	0	9	3	5	3
3	30.2852661	40.7603187	32	1	6	5	9	2	3
4	30.1411167	40.4985302	26	1	1	10	10	3	9
5	30.3051487	41.0121659	70	2	3	6	3	0	0
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
49	31.5824034	40.9181375	30	2	7	1	4	1	10
50	30.4437120	40.8835207	55	2	6	10	7	10	2

Figure 3. Dataset properties

It is important to choose which of these persons are to be selected before being evaluated by Clarke & Wright from this randomly generated data set. Various methods can be used at this step. For example, only 25-35 year-olds, women with 1 and 3 interest area at 5 or higher degree can be filtered and using the location of these selected people in the Clarke & Wright algorithm. In our approach, Analytic Hierarchy Process (AHP) method, which is a widely used method for selection processes consisting of many alternatives, has been used. AHP was first introduced by Saaty (1980) as an approach to assign the relative importance of score points based on criteria of different weights. AHP helps decision makers understand the complexity of a problem and make informed decisions. It thus encompasses both objective and subjective considerations. AHP includes six phases that are common to all applications: (a) identify the criteria that describe the alternatives in a decision problem and organize them in a hierarchy, (b) compare pairwise criteria according to user preferences and receive criteria weights, (c) rate or obtaining the performance of each alternative with respect to each criterion, (d) scaling the criteria, (e) synthesizing and ranking alternatives, and (f) selecting the best alternatives [16]. With AHP, people are sorted by interests. In order to do this, the criteria matrix has been created with 5 areas of interest relative to each other. Subsequently, alternative matrices were created using the ratings of the 50 people in 5 interests. As last step, a 50x1 result matrix generated by multiplying 50x5 alternative matrix with 5x1 criteria matrix (Eq. 2). With this generated matrix, the most suitable persons are sorted in decreasing order according to the percentage value obtained according to the combination of the grades of all the interest areas instead of the direct selection of the persons according to their specific interests. Individuals who are ranked according to their interests and then those who may be eligible for the event are filtered by gender and age range. The Clarke & Wright algorithm is applied after people who meet the interest but do not comply with other criteria are removed from the list (Fig. 4). Black colored individuals are selected in Fig. 4.

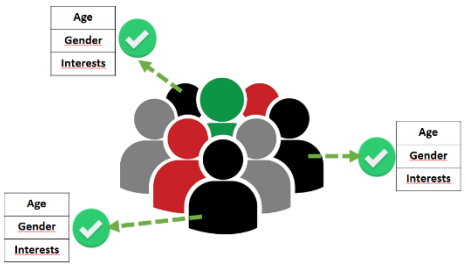


Figure 4. Filtering dataset with individual's properties

$$\underset{50x5}{\begin{bmatrix} 0{,}02 & \cdots & 0{,}04 \\ \vdots & \ddots & \vdots \\ 0{,}02 & \cdots & 0{,}01 \end{bmatrix}} X \underset{5x1}{\begin{bmatrix} 0.40 \\ \vdots \\ 0{,}10 \end{bmatrix}} = \underset{50x1}{\begin{bmatrix} 0{,}027016 \\ 0{,}024827 \\ \vdots \\ 0{,}010224 \end{bmatrix}} \quad (2)$$

For the implementation of the Clarke & Wright algorithm, 50 people are selected for the event, and then a list is made of the x and y coordinates of those persons and the coordinates of the place where the event is to be performed (Fig. 5).

Figure 5. Weighted individuals with distance to event place

With the desktop software that we have created for the proposed approach, this list is used as a parameter and then the distance matrix is created respectively (Fig. 6). The savings matrix is created based on the Clarke & Wright formula using the values in this generated matrix (Fig. 7). The savings values that are obtained here are sorted by descending order (Table 1). Starting from the person with the highest savings value, the number of people set for the event is selected in order (Fig. 8). If there are fewer people than the event capacity, all individual are selected. If there are more people than event capacity, the capacity is selected. According to the load capacity of the Clarke & Wright algorithm, vehicle filling is performed. In our approach, 1 person is designated as 1 unit for each vehicle. Therefore, the number of people not capacity is based. The distance relation between the coordinates of the determined points is calculated according to the shortest Euclidean relation (Eq. 1). Generated distance matrix is seen in Fig. 6. At the next step, savings for all (i,j) pairs are calculated for $1 \leq i \leq j < 50$. In Fig.7 generated savings matrix is seen. For; $s(1, 2) = c_{01} + c_{02} - c_{12} = 0{,}6472 + 0{,}6960 - 0{,}0640 = 1{,}2792$ is calculated. After the other savings are calculated, a savings matrix consisting of aggregated results is created.

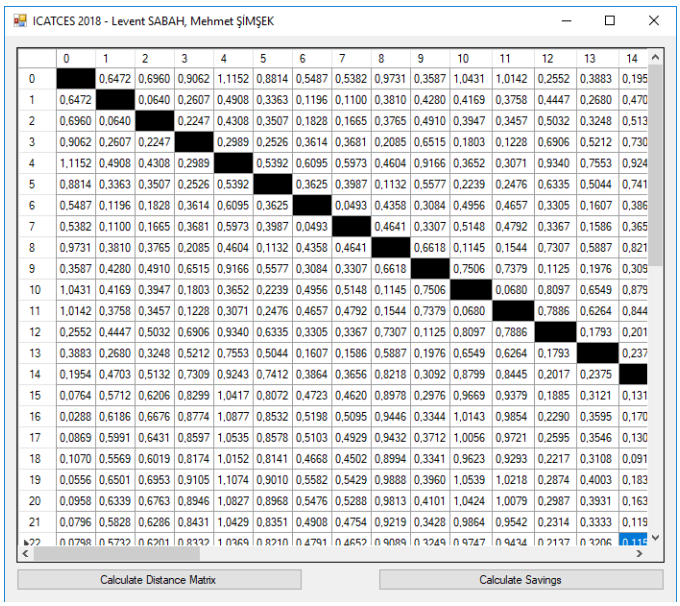


Figure 6. Calculating Distance Matrix

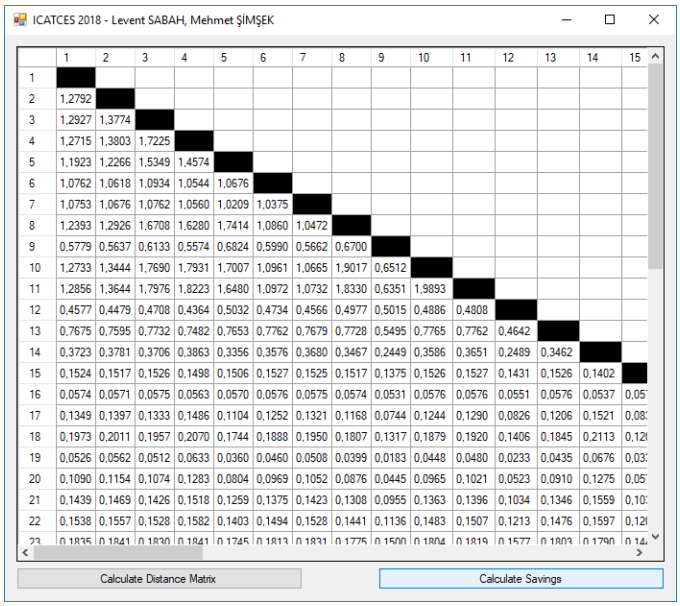


Figure 7. Calculating savings matrix

Then, the routes are created by sorting these values with descending order. Savings values are sorted from highest to lowest, and after the savings matrix is created, it is first calculated from the biggest savings value. The rankings of the same valuable savings are made arbitrarily and do not cause any changes in the calculation. Pairs in the same route are not evaluated and continue to be iteration. The iteration will continue until all demands are met.

Table 1. Sorting savings values with descending order

Pair	Value
11-10	1,9893
10-8	1,9017
11-8	1,833
11-4	1,8223
11-3	1,7976
10-4	1,7931
10-3	1,769
8-5	1,7414
.	.
.	.
50-49	0

```
11→10→8→4→3→
5→41→50→2→34→
1→6→7→37→13→
45→38→9→44→42→
35→36→40→12→49→
32→14→43→33→31→
39→28→27→30→48→
26→18→46→47→24→
20→23→17→22→21→
25→15→29→19→16→0
```

Figure 8. Generated route

The coordinate points and ID information for the people evaluated in the event planning are shown in Fig. 9. The view of the coordinate points on the map is shown in Fig. 10.

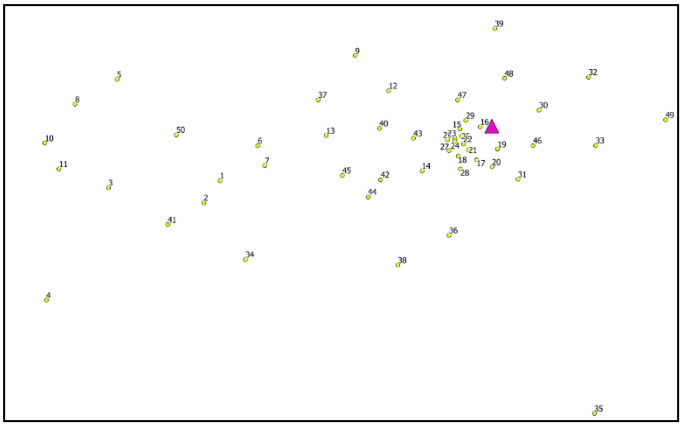


Figure 9. Locations of individuals are yellow markers, location of event place is purple marker.

III. RESULTS

The Clarke & Wright algorithm, a commonly used method in vehicle routing problems, has been applied to an event planning system.

In order for the global world to keep pace with the competitive conditions, while delivering the services and products produced in accordance with the developing production and service technologies to the demand points, at the same time it is

necessary to keep the customer satisfaction at the highest level and also to reduce the transportation costs and the usage time.For this reason, the company that wants to reach the leading position in the market must develop appropriate logistics plans [17].

With the proposed approach, an event is planned and evaluated by the Clarke & Wright algorithm of the distances to the activity of the people who are thought to participate. In addition, the results are obtained by filtering out the relevance of the persons to the activity, the age and gender of the people with a high interest score after the evaluation of their interests by the AHP. Thus, depending solely on the distance or with data such as interests, age and gender, the turnout rate is higher and the cost of participation (transportation distance to the event, time of transportation) can be planned to be optimal.

Figure 10. Map of locations

IV. CONCLUSION

In the approach we recommend, distance matrix created in the Clarke & Wright algorithm, more precise results with the use of avenues and streets in calculating inter-node distances from map services such as Google, Bing instead of the Euclidean distance of the nodes can be taken.

Furthermore, the distance between the nodes in our recommended method is handled without direction. With the use of one-way streets with the help of map services, practical applications can be done more accurately. The use of map services in large-scale applications is suitable for performance

In addition to our proposed approach, time-limited routes can also be made, taking into consideration the special conditions of persons (being handicapped, wheelchair use, and companion). Thus, when planning, an event that appeals to every part of the community can be organized.

REFERENCES

[1] Y. Mo, B. Li, B. Wang, L. T. Yang, and M. Xu, "Event recommendation in social networks based on reverse random walk and participant scale control," *Future Generation Computer Systems*, vol. 79, pp. 383–395, 2018.

[2] B. García, "Cultural policy and urban regeneration in Western European cities: lessons from experience, prospects for the future," *Local Economy*, vol. 19, no. 4, pp. 312–326, Nov. 2004.

[3] D. Getz and S. J. Page, "Progress and prospects for event tourism research," *Tourism Management*, vol. 52, pp. 593–631, 2014.

[4] Z. Koukopoulos and D. Koukopoulos, "Smart dissemination and exploitation mobile services for carnival events," *Procedia Computer Science*, vol. 110, pp. 24–31, 2017.

[5] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: a survey," *GeoInformatica*, vol. 19, no. 3, pp. 525–565, 2015.

[6] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Boston, MA: Springer US, 2011, pp. 1–35.

[7] A. H. Celdrán, M. G. Pérez, F. J. García Clemente, and G. M. Pérez, "Design of a recommender system based on users' behavior and collaborative location and tracking," *Journal of Computational Science*, vol. 12, pp. 83–94, 2016.

[8] V. Arnaboldi, M. G. Campana, F. Delmastro, and E. Pagani, "A personalized recommender system for pervasive social networks," *Pervasive and Mobile Computing*, vol. 36, pp. 3–24, 2017.

[9] I. Cenamor, T. de la Rosa, S. Núñez, and D. Borrajo, "Planning for tourism routes using social networks," *Expert Systems with Applications*, vol. 69, pp. 1–9, 2017.

[10] F. Higgins-Desbiolles, "Event tourism and event imposition: A critical case study from Kangaroo Island, South Australia," *Tourism Management*, vol. 64, pp. 73–86, 2018.

[11] D. Cilt, M. Talep, O. Problemi, Y. Bir, and M. Algoritma, "Talep ve kapasite kisitli optimizasyon problemi için yeni bir melez algoritma," pp. 1–2.

[12] S. M. Application, V. R. Problem, T. Windows, and M. Dem, "Çukurova Üniversitesi İİBF Dergisi Zaman Pencereli Araç Rotalama Problemine Tasarruf Yöntemi ile Bir Uygulama Saving Method Application for Vehicle Routing Problem with Time Windows," *Çukurova Üniversitesi İİBF Dergisi*, pp. 189–205, 2013.

[13] N. Özçakar, A. Görener, and V. Arıkan, "Depolama Sistemlerinde Sipariş Toplama İşlemlerinin Genetik Algoritmalarla Optimizasyonu," *Istanbul Management Journal*, vol. 23, no. 71, pp. 118–144, 2012.

[14] B. Keçecİ, F. Altiparmak, and İ. Kara, "Heterojen Eş-Zamanli Topla-Dağit Araç RotalamaProblemi :Matematiksel Modeller VeSezgiselBirAlgoritma Heterogeneous Vehicle Routing Problem With Simultaneous Pickup and Delivery : Mathematical Formulations and a Heuristic Algorithm," vol. 30, no. 2, pp. 185–195, 2015.

[15] V. Erol, "Araç Rotalama Problemleri için Popülasyon ve Komşuluk Tabanlı Metasezgisel bir Algoritmanın Tasarımı ve Uygulaması," pp. 1–161, 2006.

[16] H. Ma, S. Li, and C.-S. Chan, "Analytic Hierarchy Process (AHP)-based assessment of the value of non-World Heritage Tulou: A case study of Pinghe County, Fujian Province," *Tourism Management Perspectives*, vol. 26, no. January, pp. 67–77, 2018.

[17] U. DARCAN, "Stokastik Araç Rotalama Algoritmalarının Karşılaştırmalı İncelenmesi," Yıldız Teknik Üniversitesi, 2007.

Analysis of Transportability of Causal Effects in Wireless Sensor Networks

P.C.KARTHIK[1] and E.POOVAMMAL[1]

[1]SRM Institute of Science and Technology, Kattankulathur, Kancheepuram district, Tamilnadu, India, PIN-603203, pc.karthik@gmail.com

[1]SRM Institute of Science and Technology, Kattankulathur, Kancheepuram district, Tamilnadu, India, PIN-603203, poovammals@gmail.com

***Abstract* - In the empirical sciences, experiments are carried out in the laboratory so that the results may be applied in other locations where the conditions are the same as in the laboratory. Hence, it should be possible to establish certain causal statements in the laboratory, which can be transported to real-world environment. In the present paper, experiments were carried out on wireless sensor networks in the laboratory and the results were transported to environments where experimentation is impossible. The tool Tetrad was used to arrive at the causal directed acyclic graph for the wireless sensor network in the laboratory. The concepts of selection diagram, the principle of transportability and the condition for failure of transportability are discussed. The applicability of transportation algorithm for typical environment and also the failure of the same are analyzed. Simulations are carried out using the simCausal package in R programming and the validity of the probability expressions transported by the transportation algorithm are shown to be correct.**

***Keywords* - Transportability, do-calculus, selection diagrams, directed acyclic graphs, causal modeling, s-bow graph, wireless sensor networks, tetrad**

I. INTRODUCTION

Science involves transporting results from a laboratory setting where we perform experiments and make observations, to a real-time environment where experiments are not possible and only observations are carried out. This scenario is called the transportability scenario. Transportability is concerned with two separate environments which have same cause-effect relationships. Even though experimentation is not possible in one of the environment, the cause and effect relationship of that environment can be calculated, from the other environment, where experimentation is feasible. As an example, a doctor carries out experiments in a source environment on rats and wishes to extend the research results to people, the reason being both mice and human beings are living beings having same cause-effect relationship. To transport results to the target environment, it is necessary to find out the segments of the target and source environments that are similar and those that differ.

In [1], the authors use a formal representation for defining similarities and differences between the source and target environments. In [2], the authors convert the transportability problems to operations on symbols. The authors provide methods for obtaining simple transportations. As an example, they provide conditions for transportability when the cause-effect relationships are equal in the source and destination environments. In [3], the authors define an algorithm for carrying out simple transportations. These transportation problems can be extended to an environment where wireless sensor networks are in operation. The reason for choosing the application domain as wireless sensor networks is that normally wireless sensors are placed in dangerous environments such as dense forest areas, in volcano-prone areas, where the entry of human beings is not recommended. Hence experimentation is also not possible. The results are transported to wireless sensor networks which are at remote locations. As far as the authors know, the application of the concept of transportability to the wireless sensor network domain has not been attempted anywhere else.

II. TRANSPORTABILITY

Transportability implies the ability to carry over cause-effect relationships obtained from an environment where we can conduct experiments to a new environment where we can only observe and cannot do experiments. We make use of the concept called "selection diagrams" for providing information on similarities and differences between the two environments [1]. We can change the problem of transportability to a problem in do-calculus [4]. Based on this method, we find algorithms and graph-theoretical conditions for finding out the feasibility of transferring cause-effect relationships to the target environment based on experiments carried out in the source environment.

III. SELECTION DIAGRAMS

To describe the similarities between the source and target environments, we should include the cause-effect mechanisms and describe the differences between the two environments by changes in the mechanisms. A set of variables called the selection variables are used, in which each selection variable denotes a difference between the source and target environments. To move between the two environments, we use conditional probabilities on the selection variables.

Let P(v|do(a)) represent the probability of set of variables V which corresponds to experiments carried out in the source environment. P*(v|do(a)) is the probability of V when the

experiments are done in the target population $\prod^*$ rather than the source population $\prod$. The difference between the two populations may be shown as in Eqn. (1).

$$P^*(v \mid do(a)) = P(v \mid do(a), s^*) \quad (1)$$

Definition 1 <A,A*> represents two causal models on domains $< \prod, \prod^* >$. Both the causal models have a common causal directed acyclic graph G. <A, A*> produces a selection diagram B when we can produce B as given in lines 1 and 2
.

1. Every edge in G is present in B.
2. B has another edge $S_i \rightarrow V_i$ if there is a difference $f_i \,!= f_i^*$ or $P(U_i) \,!= P^*(U_i)$ between A and A*.

The selection variables represent the ways where differences in the graphs between the two environments exist.

IV. DEFINITION OF TRANSPORTABILITY

Definition 2 (Transportability**):** B is a selection diagram on domains $< \prod, \prod^* >$. <P,D> is the probability distributions from observation and experiment on $\prod$ and P* is the probability distribution obtained from observation on $\prod^*$. The causal equation $T(\prod^*) = P^*(b \mid do(a), c)$ can be transported to $\prod^*$ from $\prod$ in B if $T(\prod^*)$ can be calculated uniquely from P, P*, D in a causal model that creates B.

Theorem 1 provides a mechanism for proving transportability**.**

Theorem 1**:** B is a selection diagram corresponding to the environments $\prod$ and $\prod^*$. S represents the selection variables set in B. Equation D=P*(b|do(a),c) can be transported to $\prod^*$ from $\prod$, if we can reduce P(b|do(a),c,s) to an equation where the selection variables set S occurs in conditional probabilities in expressions that do not contain the do() operator.

Definition 3 (Trivial Transportability): A causal expression T can be transported trivially to $\prod^*$ from $\prod$, when T($\prod^*$) can be identified from (G*,P*).

Example 1: T is the cause-effect relationship P(b|do(a)). $\prod$ and $\prod^*$ have the selection diagram A→B←S. T can be transported trivially in this case, because P*(b|a) = T($\prod^*$).

One more situation in which transportation can be easily achieved is when the causal-effect relationship is the same in the source and target environment.

Definition 4 (Direct Transportability)

If T($\prod^*$)=T($\prod$), we can directly transport the cause-effect relationship T to $\prod^*$ from $\prod$.

The do-calculus implies that one can make use of a graph-theoretic condition for testing if transportation directly is possible. The condition is $(S \perp\!\!\!\perp B \mid A, C)_{G_{\overline{A}}}$. This condition means that A should stop all paths to B from S when the edges ending on A are deleted and conditioned on C.

V. GRAPHICAL CONDITION FOR TRANSPORTABILITY

There are two theorems which help find out if a cause-effect relation can be transported from a source to a target environment, when a selection diagram is provided.

Theorem 2**:** B is a selection diagram for the source and target populations, $\prod$ and $\prod^*$. The selection variable set in B is S. The cause-effect relationship P*(b|do(a),c) can be transported to $\prod^*$ from $\prod$, when C d-separates B from the set S in B, which is A-manipulated. In other words, C corresponds to $(B \perp\!\!\!\perp S \mid C, A)_{D_{\overline{A}}}$.

Definition 5: (S-admissibility)

An S-admissible set is a set X of variables that satisfy the condition $(B \perp\!\!\!\perp S \mid X, A)$ in $D_{\overline{A}}$.

Corollary 1: Assume that a set C of observable variables are S-admissible. Then the cause-effect relationship P*(b|do(a)) can be transported to $\prod^*$ from $\prod$.

Theorem 3: The cause-effect relationship P*(b|do(a)) can be transported to $\prod^*$ from $\prod$ when the conditions given in points 1 to 3 are true:

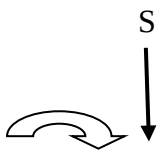
1. P*(b|do(a)) can be transported trivially.
2. C is a set of random variables. C is S-admissible and P*(c|do(a)) can be transported.
3. F is a random variables set. F satisfies the condition $(A \perp\!\!\!\perp B \mid F, S)_{D_{\overline{A(F)}}}$. Also transportability applies to P*(f|do(a)).

VI. CHARACTERISTICS OF TRANSPORTABILITY

In this section, we describe conditions under which cause-effect relationships are non-transportable.

Lemma 1 : A and B are two sets. A and B contain disjoint variables, in the source and the target environments, $\prod$ and $\prod^*$. B is the selection diagram. $P_a^*(b)$ cannot be transported to $\prod^*$ from $\prod$, provided there are two causal models M and N that satisfy B such that $P_1(V) = P_2(V)$, $P_1^*(V) = P_2^*(V)$, $P_1(V \setminus T \mid do(T)) = P_2(V \setminus T \mid do(T))$, given any set T, and $P_1(b \mid do(a)) \,!= P_2(b \mid do(a))$.

The easiest example of a graph which cannot be transported is given in Figure 1. This graph is called the s-bow arc. The s-bow arc consists of an exogenous parent U which is hidden and shared by endogenous nodes A and B.



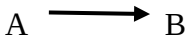


Figure 1. S-bow graph with S as the selection variable

Theorem 4**:** $P_a^*(b)$ cannot be transported in a s-bow arc graph.
Definition 6: D is a selection diagram. A subset of the edges in D that are bidirectional, create tree spanning across all the vertices present in D. D is then called the sC-component.
Definition 7**.** D is a selection diagram. C(D)={D}. There exists a node A such that all the other nodes in the selection diagram are its ancestors. All the observable nodes have a maximum of one child. A selection node points to A. D is said to be a A-rooted sC-tree (selection confounded tree).
In Figure 1, the selection diagram is a A-rooted sC-tree and hence $P_a^*(b)$ cannot be transported.
Theorem 5: Let D be a selection diagram. Given a node A, the cause-effect relationship $P^*_{Pa(A)}(b)$ can be transported provided no subgraph of D is a sC-tree that is A-rooted.
Theorem 6: D is a sC-tree that is A-rooted. The causal effects due to sets of nodes belonging to D on the node A cannot be transported.
Definition 8(sC-Forest). D is a selection diagram and the maximal root set is A. It follows that D is a sC-forest that is A-rooted, when all the nodes that can be observed, have a maximum of one child, D is a sC-component and a selection node has a directed edge towards any vertex in D.
Definition 9(s-hedge). A and B are two sets containing variables from D. G, G' are T-rooted sC-forests with the condition that $G \cap A \neq 0, G' \cap A = 0, G' \subset G, T \subset An(B)_{G_{\overline{A}}}$.
It follows that G and G' are a s-hedge for $P_a^*(b)$ in D.

Theorem 7: Let us suppose that G, G' form a s-hedge for $P_a^*(b)$ in $\prod^*$ and $\prod$. It follows that $P_a^*(b)$ cannot be transported to $\prod^*$ from $\prod$.
Based on the theorems and definitions provided in this section, an algorithm called sID shown in Figure 3 has been developed [5,6]. The algorithm decomposes the selection diagram into c-components. Thus the algorithm divides the original graph into smaller blocks. The algorithm divides in this fashion until the entire formula can be transported or a s-hedge structure results. When the algorithm ends up in a s-hedge structure, it terminates saying that the causal effect is not transportable.

VII. TRANSPORTABILITY IN WIRELESS SENSOR NETWORKS

Wireless Sensor Networks (WSN) is one of the areas where the transportability principle can be applied. In the laboratory setting, we can do experiments as well as observe the variables in the WSNs while in the real-time environment, we can only observe the behavior of variables in the WSN. We wish to find out the effect of a WSN variable say Energy in real-time environment when some other variable in the same environment say Time is set externally. We can transport causal effect relationship from the laboratory environment to the real-time environment by using the transportation algorithm in Figure 3. Hence we have to model the WSN in the laboratory using causal models.

We have done a causal modeling of a particular WSN in our lab using the tool Tetrad [7]. The causal model is shown in Figure 2. The Tetrad tool takes the values for the various variables as input and creates a causality graph of the variables as output. We wish to find the probability Eqn. (2).

$$P^*(Energy \mid do(Type)) \qquad (2)$$

where P* indicates the probability calculated in the real-time environment. The 'do' operator shows that the variable Type is set by intervention to a particular value. The selection node S is Type. The transportation algorithm simplifies the Eqn. (2) to Eqn. (3)

$$P^*(Energy \mid do(Type)) = \sum_{NumPackets} P^*(NumPackets, Energy) \qquad (3)$$

The transport algorithm in the package causaleffect of R- a programming tool is made to run. A step by step trace of the execution of the algorithm sID [5,6] is given. The selection node is chosen as the variable "Type". The input graph is G as shown in Figure 2.

The line-by-line trace of the algorithm sID for the transportation of probability given in Eqn. (3) is derived.

Trace Algorithm sID
Input parameters
y = Energy, x = Type, P*="", D = G where G is the graph in Figure 2., S = Type
BEGIN
x = Type and hence x is not empty.
V corresponds to union of the vertices in sets x and y.
V= { Energy, Type}
y = Energy
From the graph G in Figure 62the ancestors of the node y are {Energy, NumPackets }.
The asymmetric set difference between the sets V and y is {Type}.
The asymmetric set difference is not NULL.
Hence the algorithm sID is recursively called with the parameters.
y = Energy, x = intersection between x and the ancestors of y = NULL, P* = P*, D = graph containing ancestors of y in D = { NumPackets --> Energy } where --> denotes edge between two nodes.
The new call is made to sID algorithm with the parameters
x (for the new call to sID) = NULL
Hence the first line of the sID algorithm is executed and the expression assigned to P* is summed over asymmetric set difference between the vertices in the new graph D and the set y, which is {NumPackets}. The expression within P* is V which is {NumPackets, Energy}.

$$\sum_{NumPackets} P^*(NumPackets, Energy)$$

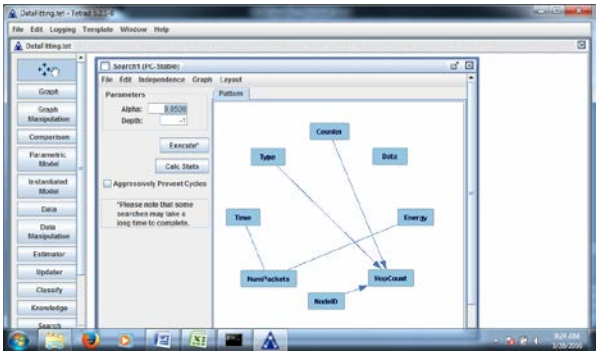


Figure 2. Causality graph for the variables in WSN in lab setting

"The expression returned is shown in Eqn. (3).

Algorithm sID(y, x, P*, I, D, S)

INPUT: x, y value assignments, P* observational distribution in $\prod^*$. I set of interventional distributions in $\prod$. D a selection diagram, S selection variables.

OUTPUT: Expression for $P_x^*(y)$ in terms of P*, I of FAIL(F,F').

1 if x = ϕ return $\sum_{V \setminus Y} P^*(V)$

2 if $V \setminus An(Y)_D \neq \phi$

return $sID(y, x \cap An(Y)_D, \sum_{V \setminus An(Y)_D} P^*, An(Y)_D)$

3 Set $W = (V \setminus X) \setminus An(Y)_{D_{\overline{X}}}$

if $W \neq \phi$, return $sID(y, x \cup w, P^*, D)$

4 if $C(D \setminus X) = \{C_0, C_1, ..., C_k\}$

return $\sum_{V \setminus \{Y,X\}} \prod_i sID(c_i, V \setminus c_i, P^*, D)$

5 if $C(D \setminus X) = C_0$

6 if $(S \coprod Y \mid X)_{D_{\overline{X}}}$, return $P(y \mid do(x))$

7 if $C(D) = \{D\}, FAIL(D, C_0)$

8 if $C_0 \in C(D)$, return $\sum_{C_0 \setminus Y} \prod_{i \mid V_i \in C_0} P^*(v_i \mid V_D^{(i-1)})$

9 if $(\exists C')C_0 \subset C' \in C(D)$ return $sID(y, x \cap C',$ $\prod_{i \mid V_i \in C'} P^*(V_i \mid V_D^{(i-1)} \cap C', v_D^{(i-1)} \setminus C', C')$ "

Figure 3. sID algorithm for transportation

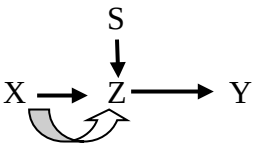


Figure 4. Causal graph with failing transportability

Trace of Algorithm sID

Input parameters

y = Y, x = X, P*="", D = graph in Figure 4. (Note : the thick arrow between nodes X and Z denotes an unobserved bi-directional edge), S = Z

BEGIN

The size of x is one since it contains only X. Hence the first line of the algorithm is not executed.

The ancestors of Y taken from the graph D are{X, Y}. V is the set {X, Y}.

The asymmetric set difference between V and the ancestors of Y is the NULL set. Hence the second line of the algorithm is not executed.

The asymmetric set difference between V and X is {Y}. The ancestors of Y after deletion of the edges pointing to X are {Y, Z}. Hence the value of W is NULL.

The components of the asymmetric set difference between D and X is {Y}

Since the components of the asymmetric set difference between D and X contains a single element, the algorithm moves to the next line.

The set S is not d-separated from Y conditioned on X, when the edges pointing to X are deleted. There is only one path from node S to node Y in Figure 4 which is a chain. However, the middle node Z has a descendant node X which belongs to the set X. Hence the d-separation condition is not met.

The sC-components of D = {D} since the graph in Figure 4 contains a tree when the unobserved edges are included. Hence the FAIL message is displayed and the sID algorithm stops.

VIII. RESULTHS AND ANALYSIS

The simCausal package from R has been used for simulating the real-world environment for wireless sensor networks. The simCausal package generates values for the random variables when the probability distribution for the random variables, the structural equation model and the underlying causality graph are given. These details are provided to the simCausal package from the experiments that we carried out in our wireless sensor networks lab. Different probability distributions have been used for the laboratory and real-world environment only for the selection variables. Selection variables are random variables whose probability distributions vary between the laboratory and the real-world environment. The probability distributions for the rest of the random variables are the same in the two environments.

The structural equation model generated by the tool Tetrad and used in the simCausal package is given in Eqn. (4) to (6).

$$Energy = 0.2374 * NumPackets \quad (4)$$

$$Time = 0.7418 * NumPackets \quad (5)$$

$$HopCount = 0.7752 * Type + 0.0028 * Counter + 0.0705 * NodeID \quad (6)$$

The causality graph used for the simCausal package is from Figure 2. The probability distributions used for the variables in the wireless sensor networks are in Figure 5.

The Type variable denotes whether the packet type is control packet or data packet. The NumPackets variable denotes the number of packets that are sent. The Counter variable is a counter that increases with each packet that is transmitted. The NodeID variable is an identification value for the particular node that is processing the packet. The Energy variable describes the energy in millivolts that is spent in the node while transmitting the packet. The Time variable denotes the time spent in processing the packet.

(Eqn. 16)

Type: Binary distribution, values = 0,1
NumPackets: Uniform distribution, Minimum value = 0, Maximum value = 100
Counter: Normal distribution, Mean = 10, Standard deviation = 10, Minimum value = 1
NodeID: Normal distribution, Mean = 10, Standard deviation = 10, Minimum value = 1

Figure 5. Probability distributions for the WSN variables

The Type variable has been assigned the binary distribution in Figure 5 since it takes only two values namely 0 and 1. The NumPackets variable has been given the uniform distribution as its values increase by 1 for each row starting from the value 1. The rest of the variables have been assigned the normal distribution.

The Energy, Time and HopCount variables have formulas assigned to them for their mean values since these formulas correspond to the structural equation model in Eqn.(4) to (6), which were derived using the Tetrad tool.

A table of 100 rows that contain values for the various variables listed in Figure 5 were generated using the simCausal package in R programming language as shown in Figure 6 and Figure 7. The variable Type has been selected as the selection variable and the experiment is done in both the real world and laboratory environments. The corresponding probability expression is evaluated in both the scenarios using Eqn. (6) and it is found that the expected value for the expression in LHS of Eqn. (6) is 14.02 while the expected value for the expression in RHS of Eqn. (6) is 14.13. The selection variable is changed from Type to the other variables such as Counter, NodeID and HopCount and the laboratory experiments are done. The expected values of the LHS and RHS of Eqn. (16) were evaluated as 12.56 and 12.64, 12.38 and 12.45 and, 10.58 and 10.67.

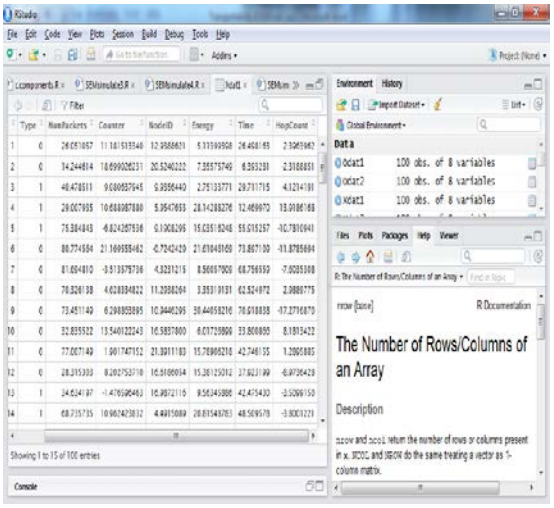


Figure 6. Table of 100 rows generated for the variables in laboratory environment

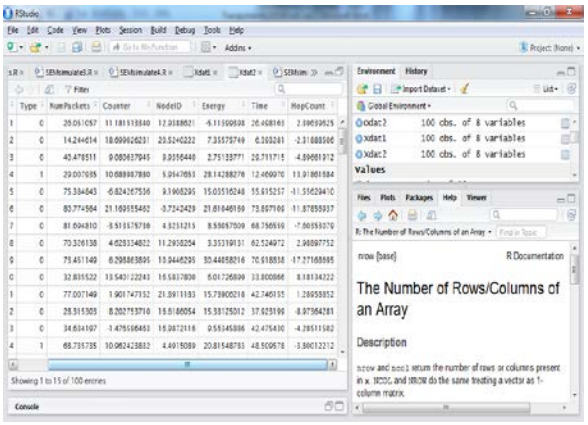


Figure 7. Table of 100 rows generated for the variables in real-world environment

From the expected values for the variables, it can be inferred that the probability expressions in the laboratory environment and the transported probability expression in the real-world environments are the same with an error percentage of less than 2%. Thus, it is possible to predict accurately what will be the effect of carrying out an experiment in the laboratory environment and transporting it to a real world environment, where some of the probability variables might have different values and probability distributions.

IX. PERFORMANCE OF THE TRANSPORTATION ALGORITHM

The performance of the transportability algorithm has been evaluated by generating bipartite random graphs with number of nodes in the graph varying between 50 and 1000 and transporting probability expressions from source graph to the destination graph. A bipartite graph is a graph in which the vertices are split into two sets and there are edges between the vertices in the two sets. A 2-dimensional plot is drawn with number of nodes on the x-axis and the time taken (in seconds) by the transportation algorithm along the y-axis.

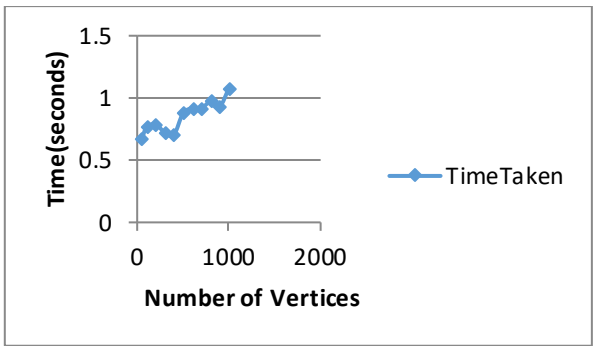


Figure 8. 2D plot of number of vertices in the graph and time taken in seconds for transportation algorithm

It can be inferred from Figure 8 that the time taken for the transportation algorithm increases almost linearly with increasing number of nodes in the random graph that is generated. Also, the time for processing the nodes by the transportation algorithm is under two seconds even for a 1000 node graph which is adequate for most causal graphs

X. CONCLUSION AND FUTURE WORK

We have modeled a Wireless sensor network in our lab using the Tetrad tool to arrive at a causal directed acyclic graph for the variables in the wireless sensor network. We have considered the scenario for transportation. In this scenario we have transported the probability for a set of variables to the real-time environment using the transportation algorithm and arrived at an expression for the transported probability. We have shown that the probability expressions generated by the transportation algorithm sID are correct through simulations conducted using the simCausal package in R. We have done a performance analysis of the transportation algorithm by carrying out the algorithm on bipartite graphs with varying vertex count.

As part of future work, experiments in the wireless sensor networks can be carried out in controlled environments such as laboratory and the results transported to dangerous environments such as a volcano or a forest without actually carrying out the experiments in the real time environment. Thus, it can be predicted as to what would be the values of the variables in the wireless sensor networks in the real-time environments.

REFERENCES

[1] Pearl, J., Bareinboim, E.. Transportability across studies: a formal approach, Technical Report R-372, Cognitive Systems Laboratory, Department of Computer Science, UCLA, 2011

[2] 2. Pearl, J., Bareinboim, E. External Validity and Transportability: A Formal Approach. JSM Proceedings, Miami Beach FL, 2011, 157–171.

[3] 3. Pearl, J., Bareinboim, E. Transportability of causal and statistical relations: a formal approach. In: Burgard W, Roth D. eds. AAAI Press, Menlo Park, CA : Proceedings of the Twenty-Fifth National Conference on Artificial Intelligence, 2011, 247—254

[4] 5. Pearl, J. (2009). Causality: models, reasoning, and inference. (Second edition). New York : Cambridge University Press

[5] 9. Bareinboim, E., Pearl J.. A general algorithm for deciding transportability of experimental results. Journal of Causal Inference, 1, 2013, p. 107-134, doi: 10.1515/jci-2012-0004.

[6] 10. Bareinboim, E., Pearl, J. Causal transportability with limited experiments. Technical Report R-408, Cognitive Systems Laboratory, Department of Computer Science, UCLA, 2013.

[7] 11. Glymour, C., Scheines, R. Causal modeling with the TETRAD program. Synthese, International Journal for Epistemology, Methodology and Philosophy of Science, 68, 1986, 37-73.

Three Axis Gimbal Design and Its Application

E. DERE[1], M. OZCAN[2] and S. CANAN[1]

[1] Elfatek Elektronik Ltd. Sti., Konya/Turkey, eminecnby@gmail.com
[2]Necmettin Erbakan University, Konya/Turkey, mozcan@konya.edu.tr
[1] Elfatek Elektronik Ltd. Sti., Konya/Turkey, suleyman.canan@gmail.com

Abstract **- Nowadays everyone is trying to record the moment everywhere and wants it to be perfect. Beyond resolution, there is a desire to get steady shots regardless of the environmental conditions. The gimbal stabilization system ensures a stable image by blocking motion-related vibrations before they are transferred to the camera lens axes. Thanks to the Three Axis Gimbal, perfect images can be achieved by minimizing the vibrations while jogging, climbing or coming down stairs, cycling, or using any kind of vehicle. In short, a three-axis gimbal can be integrated everywhere a fixed image is needed. It is envisaged that gimbal stabilization system will be needed in many scientific studies in the following periods.**

The aim of this study is to present the Three Axis Gimbal mechanism. Three separate brushless servo motors are installed on each axis for absorbing unwanted movements. The gimbal is also equipped with an inertial measurement unit consisting of a gyroscope and accelerometer close to the camera mount point. The general control system and PID controller are simulated by using MATLAB and the results are shown graphically.

Keywords **- Inertial Measurement Unit, Brushless Servo Motor, Gimbal System, PID Controller.**

I. INTRODUCTION

Gimbal is the system used to prevent the shaking, one of the biggest problems in video recordings. Figure 1 illustrates a simple block diagram of the gimbal assembly. There are two or three engines on the systems called as gimbal and they aim to prevent or eliminate vibration [1]. The basic logic of this system which can minimize the vibration in video recording devices is to create a reverse motion in the opposite direction of the vibration. This reverse motion is provided by the Inertial Measuring Unit (IMU) Sensor which is placed on the camera. The IMU Sensor detects the camera movements and reports motion to three brushless servo motors positioned in line with the camera lens. The sensor detects the relative position of the camera according to the ground. Based on the predetermined optimum position, it is detected how much the optimum position defined in each movement of the camera deteriorates. The main aim is to protect this optimum position. The information received from the sensor is processed on the electronic board and transmitted as a command to the brushless servo motors, which provide smooth motion. Thus, the brushless servo motor that produces the opposite movement of the camera allows to obtain a smooth image.

Thanks to the Three Axis Gimbal; a cameraman shooting on the baseline of the field in a football match can record smooth images while running in order not to miss the event; in unmanned armed vehicles used in the defense industry, cameras mounted at the barrel level can produce smooth images which helps to achieve accurate targeting, and perfect images can be obtained in many land and air shots as well [2].

In this study, a three-axis gimbal system which uses PID controller as the general control system is presented [3,4]. In the second chapter, the components of the moving platform system are detailed. In the third chapter, the filtering of the values read from the sensor, and the control system are referred. Finally, in the fourth chapter, proposals and conclusions are submitted.

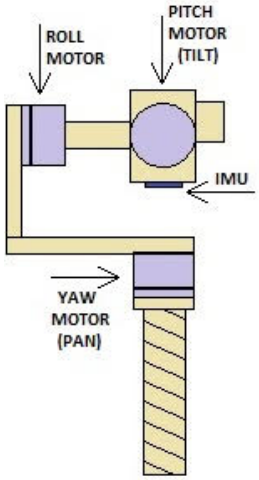


Figure 1: Gimbal block diagram.

II. MECHANICAL DESIGN

The frame carrying the system should be enduring enough to carry the camera and light enough to provide ease of use. Carbon fiber pipe is the material qualified for the conditions that we are looking for.

A six-axis IMU sensor card (which is often used in multicopter and robotic projects) which has a three-axis gyroscope and a three-axis angular accelerometer is often needed to detect camera movements. Thus, we can obtain the information such as orientation, speed, and position from a single unit.

The three axes mentioned in the Three-Axis Gimbal are shown in Figure 2. These three axes are called pitch, yaw, and roll, which carry the same name as the axes of the movement of a plane. In order to absorb the unwanted movements of these three axes, three separate brushless servo motors are mounted on these axes corresponding to the camera lens. The brushless servo motor mounted on the pitch axis absorbs the unwanted up-down movement of the camera lens, undesired right-left motion is absorbed by the brushless servo motor camera lens mounted on the yaw axis, and undesired rolling motion from one edge to the other is absorbed by the brushless servo motor

mounted on the roll axis. The biaxial gimbal does not have yaw axis, so that the recorded videos are shakier than the three-axis gimbal. Because, there is no absorption in the sudden and involuntary turns towards right or left.

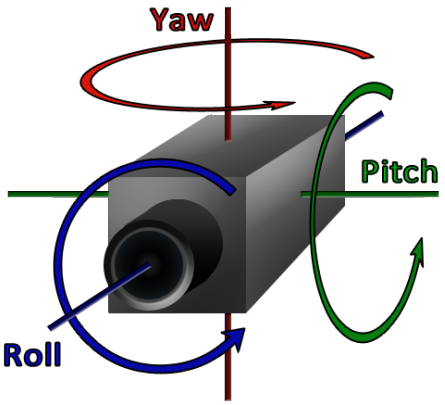


Figure 2: Gimbal axes.

Motors mounted in the line with the camera lens receive feedback from the sensor and are used to provide a rotational motion in the opposite direction of the movement to keep the camera lens steady. In this study, a brushless servo motor is preferred because of its ability to tolerate the fault quickly and smoothly.

III. SOFTWARE DESIGN

A. IMU Data Fusing

Gyros filter out accelerometer outputs to make a more accurate measurement. There are various algorithms for filtering. One of the most commonly used is the Kalman filter [5]. But it has a complicated algorithm. It makes a calculation by variable weighted average ratio and can use many different methods to calculate this ratio, but it is difficult to understand. The system works to predict new outputs by using previous outputs and measurements. In sum, the Kalman filter predicts the best value of the next output by monitoring the constantly changing inputs of the system. This filter is used in many areas such as image processing, orientation, motion tracking.

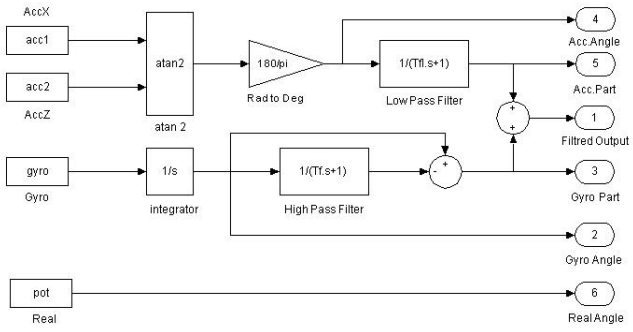


Figure 3: Complementary filter block diagram.

Block diagram of the Complementary Filter is illustrated in Figure 3. This filter is a method of taking averages of the reference data with a constant weighted average ratio [6]. It is the simplest algorithm. It is generally used for hobby purposes and in the applications to understand the operating logic. It simultaneously manages both high-pass and low-pass filters. Low pass filter filters high frequency signals (such as accelerometer in the case of vibration) and low frequency signals (such as gyroscope drift). By combining these filters with a complementary filter, a good signal can be obtained without the complications of the Kalman Filter.

Consequently, the Complementary Filter can be used instead of the Kalman Filter. Smoothing is better, and its algorithm is much simpler than Kalman. Therefore, The Complementary Filter is preferred in this study. An example of how the system is used with the Complementary Filter is shown in Figure 4. A graphical comparison between the signal filtered by the Complementary Filter and the true angle is shown in Figure 5.

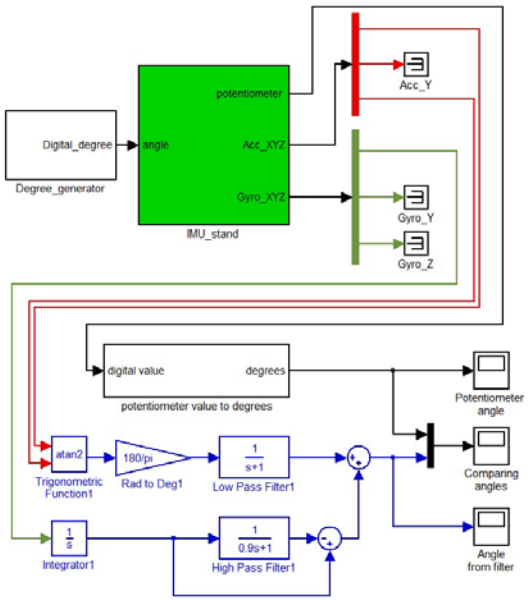


Figure 4: Complemenraty filter with IMU.

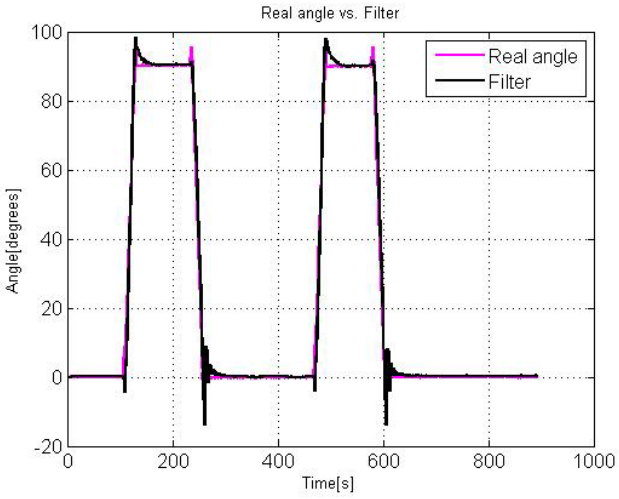


Figure 5: Complemenraty filter output and comparison of actual value.

B. Controller Design

Although there are a variety of modern techniques for controlling gimbal systems, PID controllers are preferred because of their low cost, ease of implementation and high performance.

The block diagram in Figure 6 is a feedback mechanism controller commonly used in PID (Proportional-Integral-Derivative) industrial control systems. In 1942, Ziegler and Nichols presented two classical methods for determining the parameters of the PID controller [7]. These methods are still widely used in the original form or in some modifications. The methods are based on determining some properties of the process dynamics.

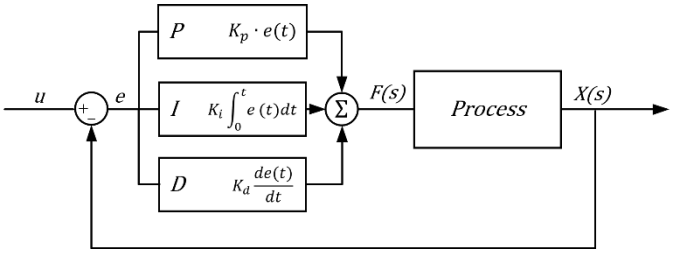


Figure 6: PID block schema.

The PID control system continuously calculates the error value as the difference between the desired value and the measured process variable. The controller tries to minimize the error over time. In doing so, sets a new value (determined by a weighted sum) for a control variable, such as power fed to a control valve, a dashpot, or a heating element.

Transfer function of PID controller:

$$G(s) = K_P + \frac{K_i}{s} + K_d . s \qquad (1)$$

K_P : proportional gain,
K_i : integral gain,
K_d : derivative gain,

As shown in Table 1, the proportional gain (Kp) reduces the rise time and the steady-state error (but never removes it). The integral gain (Ki) removes the steady-state error but can worsen the transient response. Derivative gain (Kd) increases the stability of the system, reduces overshoot and improves the transient response.

Table 1 : Effect of PID gains on system

Controller Response	Rise Time	Overshoot	Settlement Time	S-S Error
K_p	Decreases	Increases	Changes Slightly	Decreases
K_i	Decreases	Increases	Increases	Removes
K_d	Changes Slightly	Decreases	Decreases	Changes Slightly

Figure 7 shows the PID controller designed with MATLAB Simulink, where all controller coefficients are set. The program is started and the necessary parameters are obtained by PID tuning [8]. Once Simulink is running, we can see the signal shown in Figure 8.

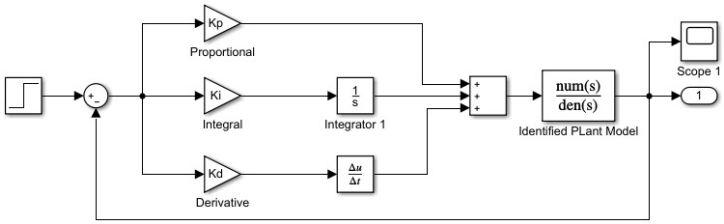


Figure 7: Design of PID control system in MATLAB Simulink.

Figure 8: Output signal in case of PID

IV. CONCLUSION AND RECOMMENDATION

In this study, three-axis gimbal system is introduced. Under the title of mechanical design, the basic elements of the gimbal system are introduced. Under the software design title, gimbal control system, sensors and sensor information filtering are mentioned. PID control is used to stabilize three axes of this platform model. The PID control system is simulated in the MATLAB / Simulink environment and the results are shown in the diagram. PID control is often preferred because of its high performance, but control methods such as PI, PI2, optimal control, and compensator can also be used in two-axis and three-axis gimbal systems. Brushless servo motors are used in the system as motion providers. Step motor or servo motor may be preferred instead of brushless servo motor. The error signal is detected by using the IMU sensor while the Complementary Filter is preferred for sensor filtering. Besides Complementary and Kalman filters; the Mahony Madwick Filter, which is used to calculate the quaternion with the information received from the sensor, may also be preferred. This system is being implemented and it is targeted to be used in military applications carried out within our company.

REFERENCES

[1] MD. Shuster and SD. Oh, Three-axis attitude determination from vector observations, J Guidance Control 1981; 4: 70-77

[2] J.J. Xiong, E.H. Zheng, "Position and Attitude Tracking Control for a Quadrotor UAV", ISA Trans., vol. 53, no. 3, pp. 725-731, 2014

[3] D. Stojcsics, A. Molnar, "Fixed-Wing Small-Size UAV Navigation Methods with HIL Simulation for AERObot Autopilot", IEEE 9th Int. Symp. on Intel. Sys. and Inf. (SISY), pp. 241-245, 2011

[4] J.M. Sullivan, "Evolution or Revolution? The Rise of UAVs", *IEEE Tech. and Soc. Mag.*, vol. 25, no. 3, pp. 43-49, 2006

[5] J.L. Marins, X. Yun, E.R. Bachmann, R.B. McGhee, M.J. Zyda An Extended Kalman Filter for Quaternion-Based Orientation Estimation Using MARG Sensors. International Conference on Intelligent Robots and Systems, Maui, Hawaii, USA, pages 2003{2011, 2001

[6] G. Baldwin, R. Mahony, J. Trumpf, T. Hamel, T. Cheviron Complementary Filter Design on the Special Euclidean Group SE. 9th European Control Confrerence on behalf of the EUCA, Kos, Greece, 2007.

[7] J. G. Ziegler and N. B. Nichols: Optimum Settings for Automatic Controllers, Trans. ASME, Vol. 64, 1942, s. 759-768.

[8] P.M. Meshram„ and R.G. Kanojiya, Tuning of PID controller using ZieglerNichols method for speed control of DC motor, Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on (pp. 117-122). IEEE.

Training ANFIS using The Whale Optimization Algorithm

M.CANAYAZ[1] and R. ÖZDAĞ[2]

[1] Van Yuzuncu Yil University, Van/Turkey, mcanayaz@yyu.edu.tr
[2] Van Yuzuncu Yil University, Van/Turkey, rozdag@yyu.edu.tr

***Abstract* - Nowadays, it is tried to predict the future events through the data. Practical areas such as deep learning are primarily trying to regulate data, and then these data are used for estimation. There are many algorithms used in this area. Besides these algorithms, artificial neural networks are also widely used in this field. ANFIS is a special network that uses artificial neural network and fuzzy classifier. It computes the output by distributing the input data blurred by the membership functions with the fuzzy rules on the network. Some parameter values need to be set in ANFIS. In this study, ANFIS networks will be trained with the Whale Optimization Algorithm, one of the current swarm-based meta-heuristic algorithms to find suitable parameter values and evaluation will be made on sample problems.**

***Keywords* – ANFIS, Whale Optimization Algorithm, Nonlinear System Identification, Prediction Time Series**

I. INTRODUCTION

Nowadays, big data emerge as a concept used to express the size of the data obtained from all kinds of sources. Great successes are achieved both from economic and scientific aspects along with the efficient use of the applications that will process this huge data size. Deep learning algorithms provide improved tools for the processing of big data. Big companies use artificial intelligence in most of deep learning applications. Artificial intelligence applications have gained great momentum along with the developing technology. It is seen that the studies in this regard have increased both in universities and in the private sector in recent years. As it is known, the development of these applications requires high costs. The processing of high-sized data used in these applications sometimes lasts for days or sometimes for weeks. When technological developments are examined, it is seen that autonomous tools are the most important of these applications. Apart from this, artificial intelligence is significantly used in computer vision applications, in the biomedical field. When we look at literature in the field of artificial intelligence, it is possible to see that it has developed in many fields. Artificial neural networks that are created by being inspired by the human nervous system are mainly used while developing these applications. The use of these networks in various forms is available. ANFIS networks are a special networking model composed of a combination of these neural networks and fuzzy logic tools. ANFIS has been used in many estimation problems and engineering problems. ANFIS networks need to be trained so that they can yield productive results. The purpose in training is to find the appropriate parameter values. These networks have been trained using several heuristic methods such as PSO[1], ABC[2], Genetic Algorithm[3] and applied to various problems.

The aim of this study is to demonstrate the performance of the Whale Optimization Algorithm (WOA)[4], which is one of the current meta-heuristic algorithms to be used in the training of ANFIS networks for the first time, on various problems. In this context, the study will consist of Whale Optimization Algorithm, ANFIS structure, proposed method, and results.

II. THE WHALE OPTIMIZATION ALGORITHM

The Whale Optimization Algorithm is one of the current swarm based meta-heuristic algorithms developed by being inspired by the hunting strategies of humpback whales [4-6] The unique hunting strategies of these whales were modeled in three steps. These are encircling the prey, moving towards the prey and searching for the prey.

A. Encircling the Prey

The whales know the location of prey and encircle them. Initially, they think that the current best candidate solution is the best solution reached and the closest optimal solution. After the best candidate solution is determined, other members of the population update their positions using this best solution. This is represented as in equation (1,2).

$$\vec{D} = \left|\vec{C}.\vec{X}^*(t) - \vec{X}(t)\right| \quad (1)$$

$$\vec{X}(t+1) = \left|\vec{X}^*(t) - \vec{A}.\vec{D}\right| \quad (2)$$

t represents the current iteration, A and C represent the coefficient vectors, X* represents the best solution vector.
The coefficients vectors A and C are formulized as in equations (3) and (4), respectively.

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2.\vec{r} \tag{4}$$

r- represents a random vector, *a*- represents a vector decreasing from 2 to 0 throughout the iterations.

B. Moving Towards the Prey

Humpback whales approach their prey with a strategy called bubble-net feeding method. In this method, they move towards their prey by blocking the view of their prey with the bubbles they create. They move towards their prey by 2 methods that are called narrowing the circle and spiral movement method. The method of narrowing the circle is possible by decreasing the value of "a" in equation (3).

In the spiral movement method, the distance between the whale and prey is first calculated as in equation (5). This value is then used to calculate the spiral movement method in equation (6). Spiral movement is presented in Figure 1.

$$\vec{D}' = \vec{X}^*(t) - \vec{X}(t) \tag{5}$$

$$\vec{X}(t+1) = \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{X}^*(t) \tag{6}$$

b-represents the logarithmic spiral constant, l-represents a random number in the range of [-1,1].

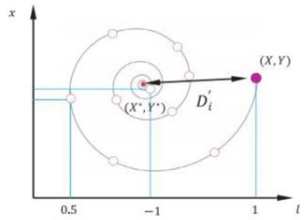


Figure 1. Spiral Hareket

In the algorithm, which one of them, spiral movement or linear movement, will be employed is determined by 50% probability as it is shown in Equation 7.

$$\vec{X}(t+1) = \begin{cases} \left|\vec{X}^*(t) - \vec{A}.\vec{D}\right|, & p < 0.5 \\ \vec{D}'.e^{bl}.\cos(2\pi l) + \vec{X}^*(t), & p \geq 0.5 \end{cases} \tag{7}$$

p-represents a random number in the range of [0,1].

C. Searching for the Prey

During this searching for prey, the whales search for prey randomly and change their position according to the positions of other whales. The mathematical model of this search is presented in Equation 8 and Equation 9.

$$\vec{D}' = \vec{C}.\vec{X}_{rand} - \vec{X} \tag{8}$$

$$\vec{X}(t+1) = \vec{X}_{rand} - \vec{A}.\vec{D} \tag{9}$$

X_{rand} represents a randomly selected solution vector.
Which one of them, global or local searches, will be performed is decided based on the value of vector *A*. Since the point which is more distant than the best point can be selected when it is $A > 1$ or $A < 1$ for vector A, these cases are regarded as global search.

The steps of the WOA are presented below.

Steps of the Algorithm

Step 1: Population n, a parameter, A, C coefficient parameters, Maximum iteration are set.
Step 2: Candidate solutions with n populations are created randomly and the fitness function value of each candidate is calculated.
Step 3: Initially, the best solution X* is accepted as the global best.
Step 4: A,C, a, p parameters are updated throughout the algorithm.
Step 5: The movements of the whale are updated according to the value of p.
Step 6: The best solution is offered when maximum iteration is reached.

III. ANFIS

Adaptive network-based fuzzy inference system (ANFIS) is an artificial neural network developed by applying the Sugeno fuzzy model. ANFIS has an extensive area of usage in the literature as special network structures that combine the learning ability of artificial neural networks with the inference-making ability of fuzzy systems. "If Then" rule structure is frequently used in estimation problems that require decision making mechanisms because it uses the input, output values. ANFIS has five-layer architecture [7-12].

Assume - two inputs X and Y and one output Z

Rule 1: If x is A_1 and y is B_1,
then $f_1 = p_1x + q_1y + r_1$
Rule 2: If x is A_2 and y is B_2,
then $f_2 = p_2x + q_2y + r_2$

Layer 1

Each node in this layer represents a fuzzy set such as A_i and B_i. The membership grades depending on the input samples and the membership function used are used as the output of the nodes. In other words, it is the layer where input values are made fuzzy. The node outputs are presented in equation (10).

$$u_i^2 = \mu_{A_i}(x) \text{ i=1,2} \tag{10}$$

$$u_{i+2}^2 = \mu_{B_i}(y)$$

Here, x, y and i represent the deterministic input values in the node, A_i and B_i represent the fuzzy terms, and μ_{Ai} and μ_{Bi} represent the membership functions. There are 2 different output values and a total of 4 nodes in the network with two different inputs such as x and y, as it is shown in equation (10). For each node, membership functions with a maximum of 1 and a minimum of 0 are used. In general, the membership function of a fuzzy set can be any parameterized membership function, such as triangle, Gaussian, or the generalized Bell function. The parameters in this layer are called Premise Parameters.

Layer 2

This layer is called the rule layer. The nodes in this layer represent the rules and numbers created using the Sugeno fuzzy logic inference system. The output of the rule nodes, μ_i, is the multiplication of membership grades from the previous layer. The node output shows the firing strength of a rule. μi values, where it is (i=1,….,n), are obtained as follows.

$$\omega_i = \mu_{A_i}(x) * \mu_{B_i}(y) \tag{11}$$

Layer 3

In this layer, the firing strengths from the rule layer are normalized. For this reason, this layer is called the normalization layer. The firing strength normalized for i_{th} node is calculated as follows.

$$\varpi_i = \frac{\omega_i}{\omega_1 + \omega_2} \tag{12}$$

Layer 4

In this layer that is called the defuzzification layer, the weighted output values of each rule are calculated. This calculation is achieved by the multiplication of the normalized firing strength in the figure below with the $\{p_i, r_i, q_i\}$ values, that are the output parameters of the fuzzy inference system. The parameters in this layer are referred to as consequent parameters.

$$\varpi_i f_i = \varpi_i (p_i x + q_i y + r_i) \tag{13}$$

Layer 5

This layer is called output layer. The output of ANFIS is achieved with the summation of the outputs obtained for each rule in the previous layer. In this layer, a single number is produced by the defuzzification of fuzzy rules.

$$f(x,y) = \sum_i^4 \varpi_i f_i = \frac{\sum_i^4 \omega_i f_i}{\sum_i^4 \omega_i} \tag{14}$$

ANFIS's learning algorithm is a hybrid learning algorithm formed by the combined use of the least squares method and the backpropagation learning algorithm.

IV. TRAINING OF ANFIS USING THE WOA

In ANFIS, parameters are needed for the processes of fuzzification in layer 1 and defuzzification in layer 4. These are called premise and consequent parameters [2]. These parameters are provided by the WOA in the proposed method. These parameter values are expressed by the position values of the whales in the algorithm. As the position of the whales is updated according to the leading whale, these parameter values are updated and the ANFIS is trained with the updated parameter values. For the fitness function that determines the suitability of the solution in the optimization algorithms, the RMSE error function shown in Equation (15) is used.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \overline{y_i})}{N}} \tag{15}$$

$\overline{y_i}$ used in Equation (15) represents the output obtained by ANFIS in time I, and y_i represents the actual output of the system. N represents the number of samples used in the application.

V. EXPERIMENTS

Two types of problems were addressed in the applications developed for the training of ANFIS. These are the application for the identification of nonlinear systems and the chaotic time series estimation problems. For the identification of non-linear systems, a dynamic system consisting of an input and an output (SISO) was used in the given problem. Furthermore, 2 input values and 1 output value are used for the ANFIS structure. Gaussian functions are used as membership functions for the inputs in the ANFIS structure. The two inputs of ANFIS are defined as the previous input u(t) and the previous output y(t) of dynamic systems. y(t+1) is accepted for the output value. Equation (16) is used to obtain the input value u(t).

$$u(k) = \begin{cases} \sin(\pi k/25) & k < 250 \\ 1 & 250 \le k \le 500 \\ -1 & 500 < k \le 750 \\ 0.3\sin(\pi k/25) + & 750 < k \le 1000 \\ 0.1\sin(\pi k/32) + 0.6\sin(\pi k/10) & \end{cases} \tag{16}$$

25 populations, 1000 maximum iterations were used for the WOA. The same values are used for GA.

Another problem is the chaotic time series estimation problem. The dataset used for this application is the dataset called "mdata" which is prepared by Matlab program. The purpose of the problem is to estimate the x(t+6) value using the x(t-18), x(t-12), x(t-6), and x(t) data. The data set is divided into 500 training and 500 test data.

The application was performed in the MATLAB program on a computer with I7 processor and 8 GB Ram. The algorithms were run independently of each other for 30 times and the ANFIS was trained, and the average of the RMSE values obtained at the end of these trainings is presented in Table 1 for Example 1 and in Table 2 for Example 2.

Example 1: Nonlinear System Identification

Equation (17) is used in the application to be performed for the non-linear dynamic system. 1000 input and output values are generated using this system. These are divided into 70% training and 30% test data and used in ANFIS training. The results of this problem are presented in Table 1. In addition, the values taken from ANFIS(GA) is also included in the table.

$$y_p(k+1) = f(y_p(k), y_p(k-1), y_p(k-2), u(k), u(k-1)) \quad (17)$$

$$f(x_1, x_2, x_3, x_4, x_5) = \frac{x_1 x_2 x_3 x_5 (x_3 - 1) + x_4}{1 + x_2^2 + x_3^2}$$

Table 1. Example 1 results

Algorithms	Number of parameters	RMSE Train	RMSETest
GA	24	0.05342	0.044854
WOA	24	0.047588	0.059755

For the example 1, the Training error value obtained from the WOA RMSEtrain is 0.047588 and the test error value RMSEtest is 0.059755. The standard deviation value is 0.0024313, 0.0019072, respectively. The training and test result graphic obtained at the end of the algorithm is presented in Figure 2 and Figure 3, and the comparison values with other algorithm are presented in Table 1.

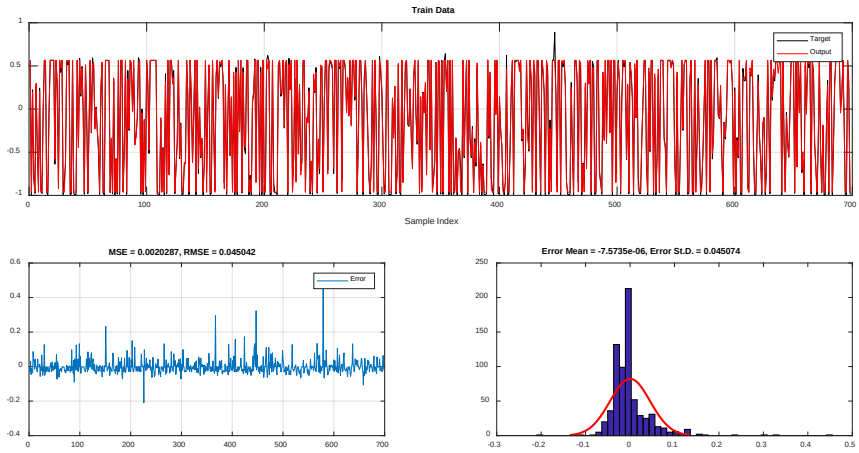


Figure 2. Example 1 Train Data Results

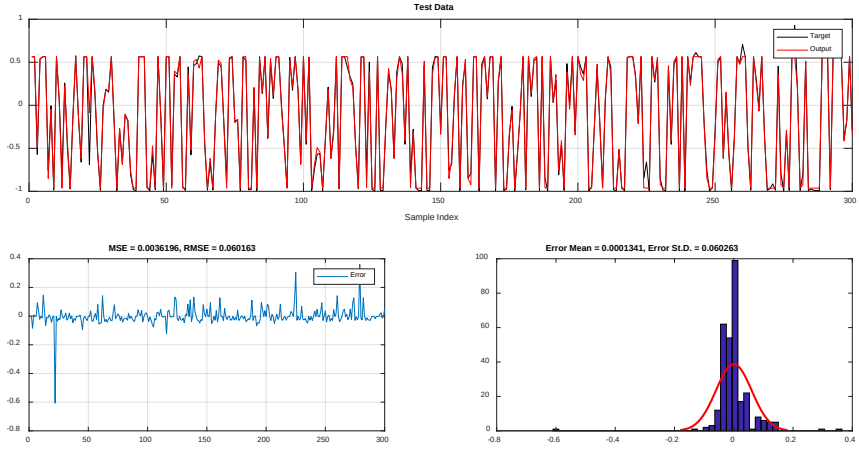


Figure 3. Example 1 Test Data Results

Example 2:Predict of Caotic Time Series Mackey-Glass

The data is generated from the Mackey-Glass time-delay differential equation which is defined by

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+(x(t-\tau))^{10}} - 0.1x(t) \qquad (18)$$

When x(0) = 1.2 and = 17, we have a non-periodic and non-convergent time series that is very sensitive to initial conditions. (We assume x(t) = 0 when t < 0.)

The time series graphic used for Mackey-Glass is presented in Figure 4. For the example 2, the Training error value obtained from the WOA RMSEtrain is 0.042904 and the test error value RMSEtest is 0.04227. The standard deviation value is 0.0034821, 0.0034207, respectively. The training and test result graphic obtained at the end of the algorithm is presented in Figure 5, and in Figure 6 and the comparison values with other algorithm are presented in Table 2.
When we look at the result figures for Example 2, although the minimum result is obtained, the target and output values are not exactly matched. This shows that ANFIS trained with WOA in the time series estimation problem is weak.

Table 2. Example 2 results

Algorithms	Number of parameters	RMSE Train	RMSETest
GA	208	0.056788	0.056184
WOA	208	0.042904	0.04227

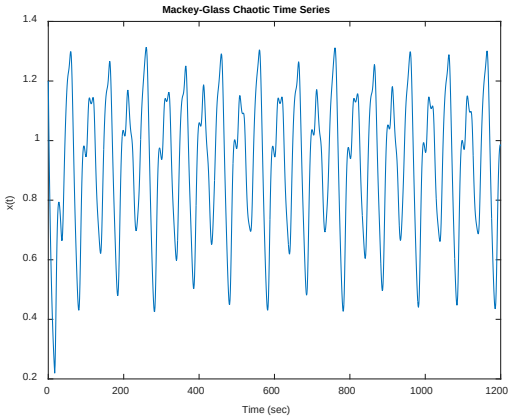


Figure 4. Mackey-Glass time series

VI. CONCLUSION

In this study, the WOA, which is one of the current swarm based algorithms, was first used for the training of ANFIS networks. ANFIS networks have a special network structure that is the combination of fuzzy logic and artificial neural networks. The premise and consequent parameters in this network are updated by means of the WOA. When the results are examined, it is seen that the WOA had the values with lower error rates compared to GA. In conclusion, it is demonstrated that the WOA can be used in the training of ANFIS.

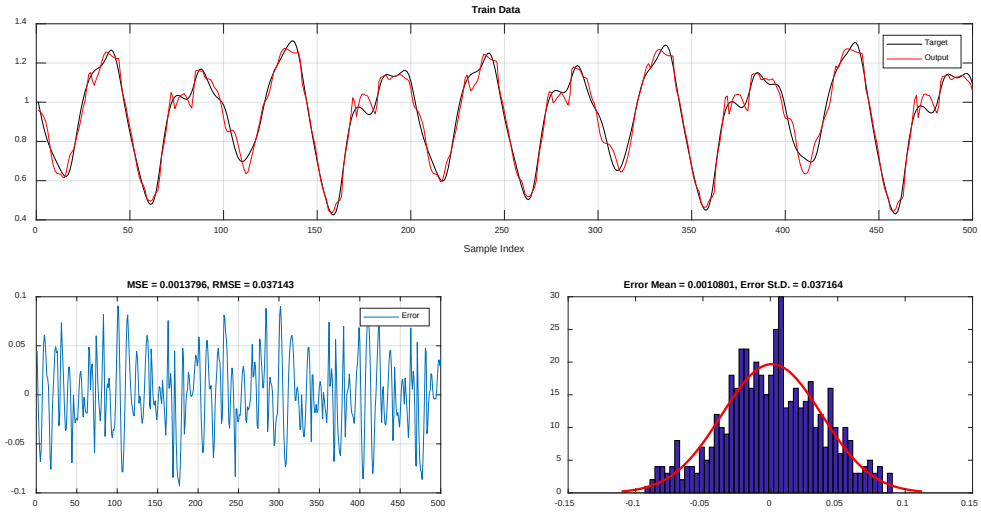


Figure 5. Example 2 Train Data Results

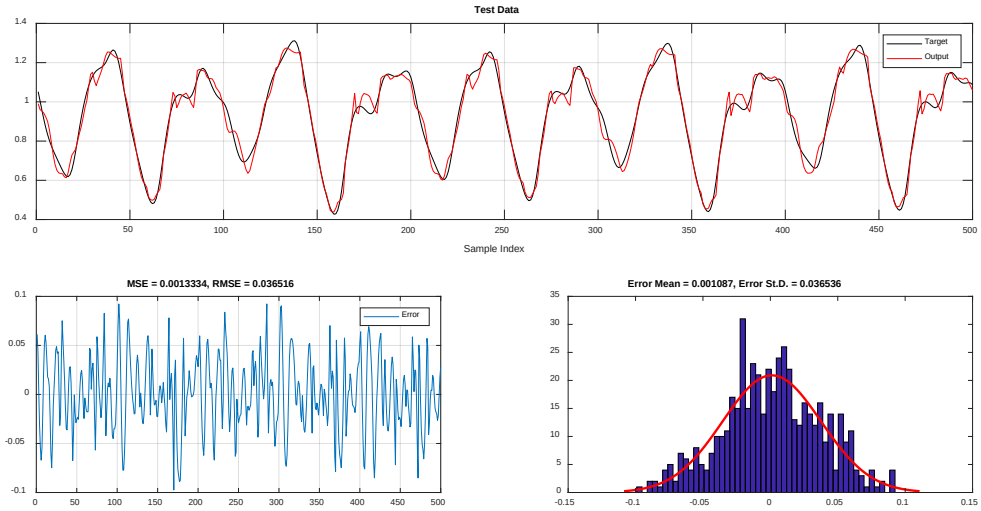


Figure 6. Example 2 Test Data Results

REFERENCES

[1] V. S. Ghomsheh, M. A. Shoorehdeli and M. Teshnehlab, "Training ANFIS structure with modified PSO algorithm," in 2007 Mediterranean Conf. on Control & Automation, Athens, pp. 1-6.

[2] D. Karaboga and E. Kaya, "Training ANFIS using artificial bee colony algorithm for nonlinear dynamic systems identification," in 2014 22nd Signal Processing and Communications Applications Conference (SIU), pp. 493–496.

[3] B. Haznedar, and A. Kalınlı, "Training ANFIS Using Genetic Algorithm for Dynamic Systems Identification," Int. J. of Intell. Sys. and Appl. in Eng., vol. 4, no. 1, pp. 44-47, 2016.

[4] S. Mirjalili, and A. Lewis, "The Whale Optimization Algorithm," Adv. in Eng. Soft, vol. 95, pp. 51-67, 2016.

[5] E. Tanyıldızı, and T. Cigal, "Kaotik Haritalı Balina Optimizasyon Algoritmaları", Fırat Ünv. Müh. ve Bilim Dergisi, vol 29, no.1, pp. 309-319, 2017.

[6] M. Canayaz and M. Demir, "Feature selection with the whale optimization algorithm and artificial neural network," in 2017 Int. Artificial Intell. and Data Processing Symposium (IDAP), pp. 1-5.

[7] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," IEEE Trans Syst Man Cybern, vol. 23, no. 3, pp. 665–685, 1993.

[8] J.S.R, Jang. C.T., Sun, and E., Mizutani, Neurofuzzy and soft computing, Prentice Hall, Upper Saddle River, 1997

[9] R. Tür, C. E. Balas, "Belirgin Dalga Yüksekliklerinin Neuro-Fuzzy Yaklaşımı İle Tahmini: Filyos Deniz Yöresi Örneği", J. Fac. Eng. Arch. Gazi Univ. vol. 25, no. 3, pp. 505-510, 2010

[10] D, Karaboğa, E., Kaya, "Training ANFIS by using the artificial bee colony algorithm", Turk J Elec Eng & Comp Sci vol. 25, pp. 1669-1679, 2017.

[11] Ö. Demirel, A. Kakilli and M. Tektaş, "ANFIS ve Arma Modelleri İle Elektrik Enerjisi Yük Tahmini", J. Fac. Eng. Arch. Gazi Univ., vol. 25, no 3, pp. 601-610, 2010.

[12] Ç. Elmas, Yapay Zeka Uygulamaları, Seçkin Yayıncılık, 2016.

Usage of the A* Algorithm to Find the Shortest Path in Transportation Systems

E.DERE[1] and A.DURDU[1]

[1] Efta Enerji Ltd. Sti., Konya/Turkey, esat.dere@eftaenerji.com
[1]Selcuk University, Konya/Turkey, durdu.1@selcuk.edu.tr

***Abstract* - In this study, we presented the use of the A-Star algorithm to find the shortest path between a starting-point and ending-point on the map which is taken from Google Maps and segmented as grid-cells. The active paths on Google Maps are specified in the Algorithm Map by dividing into grids, and it is desirable to find the shortest path from the A* algorithm. In addition, the traffic intensity of various roads is shown on the Google Maps. This information is processed on the Algorithm Map so that the algorithm can find the shortest route by considering traffic density.**

***Keywords* - A Star Algorithm, Path planning, Transportation Systems**

I. INTRODUCTION

Throughout history, transportation has become the concept occupying a great space in the life of the human. The first men in the history addressed the need of transportation by walking. Later, this requirement has been satisfied by domestication of animals. The concept of transportation has evolved by invention of the wheel. Today, transportation has come to the forefront because of the growing population. Additionally, the requirements of automobiles, airplanes, high-speed trains and other vehicles have been increased. At the same time, more vehicles need more roads however building of the roads is not possible everytime. Therefore, the problems have emerged such as heavy traffic and way-finding. Various technologies have been developed to minimize these problems. GPS-assisted navigation systems were built to find the shortest and fastest routes to the destination. In this study, it is aimed to find the shortest path to the target of a vehicle by using the A* algorithm which is a road planning algorithm commonly used in robotics systems.

Way-finding and navigation systems based on GPS are available for existing vehicles. These systems generally use the offline informations of the maps. Path planning algorithms in robotic systems and computer games have been used in the shortest path between the current place and target-place on the current map. In this work, a part of a map is simply converted to an algorithm map, and the shortest path from the starting-point to the destination is found by using online information such as traffic intensity. The main contribution of this work is the study which uses the online traffic informations for smart cities.

In the second section, the A* algorithm will be introduced. Applications will be explained in the third sections and in the fourth sections, proposals and conclusions will be submitted.

II. THE A* (ASTAR) ALGORITHM

Peter Hart, Nils Nilsson and Bertram Raphael of Stanford Research Institute first described the algorithm in 1968 [1]. The A* algorithm is one of the algorithms that is used to find the shortest path in computer science. For instance, it can be used to solve a problem such as Travelling Salesman Problem (TSP) [2]-[3]. Similarly, it is an algorithm which is frequently used to find the shortest path to the target (or destination) for the game players in game programming.

Briefly, it is a "best fit" algorithm that finds out which nodes are the best way for reaching to the target node from any given node. The A* algorithm can be classified as an admissable heuristic algorithm [4]. The algorithm uses the equation (1) to calculate the distance;

$$f(n) = g(n) + h(n) \quad (1)$$

where;

f(n): computational heuristic function,

g(n): cost of reaching to the current node from the starting node,

h(n): estimated distance to arrive at the target node from the current node.

As will be noted, the reason why f(n) is intuitive is that it is an intrinsic function h(n) which is based on the prediction.

A. Operation of the Algorithm:

The algorithm has a fairly simple structure that uses the addition above. In the algorithm that uses a priority queue as a data structure, the node with the highest priority is the node with the lowest value of f(n).

1. The algorithm takes the lowest value (and therefore the most important) node in each step (that is, it goes to the node) and gets the node out of the queue.

2. The values of all nodes that are adjacent to this node are updated (there is now a cost to come to this node

and as will be noted, this value takes place in the function f(n)).

3. The algorithm repeats the above steps until the target is reached (i.e., until the target node is in the priority queue (priority queue) or until no node remains in the queue).

III. RESULTS AND DISCUSSION

In most pathfinding algorithms, the navigation performance is based on the attributes of grid representation [5]. The map part taken from Google Maps with the jpeg-file extension is processed on the algorithm map by dividing it into grids. Figure 1 shows the map taken from Google Maps and Figure 2 shows gridded map.

Figure 1: The map part taken from Google Maps

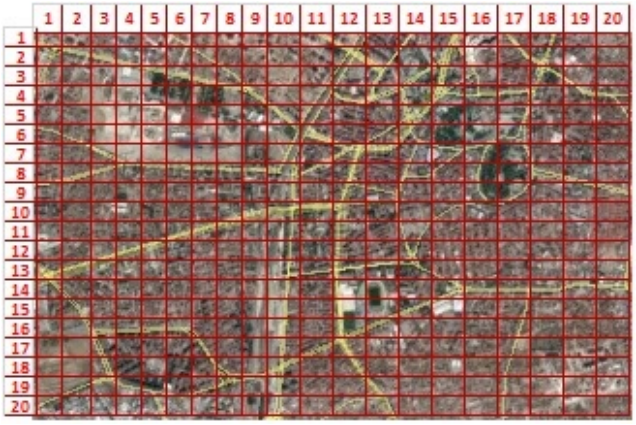


Figure 2: Part of the gridded map

All of the active paths on the map segment, which is composed of 20x20 cells, are processed on the 20x20 size algorithm map. The map processed from Google Maps to Algorithm Map is illustrated in Figure 3.

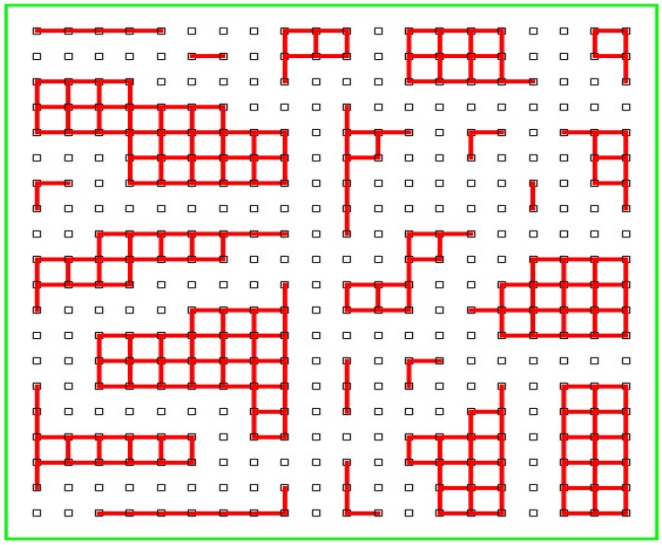


Figure 3: Map processing from Google Maps to Algorithm Map

The Google Maps part was processed on the Algorithm Map, and then the algorithm has been runned by specifying start and target nodes, and the shortest path was found. In Figure 4, the starting-point and the target-point are given at the Google Map. For this problem, the A* algorithm is prosessed and shown the shortest path in Figure 5.

Traffic density information for the active roads on the Google Maps may be provided. According to this information, when the Algorithm Map is pointed, the algorithm will not include the routes in the regions where the traffic is heavy and will find the shortest path accordingly. Traffic intensity is added to the problem given above in Figure 6. The solution of the shortest path with traffic intensity is illustrated in Figure 7 and Figure 8.

Figure 4: Processing the result of the algorithm on Google Maps

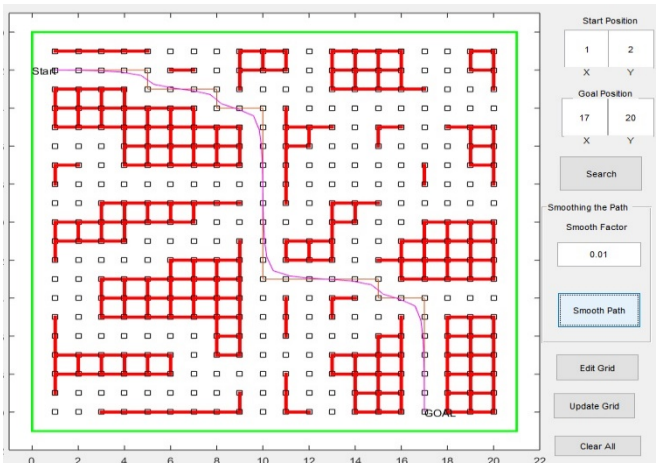


Figure 5: The shortest path between start and target nodes

Figure 6: Part of Google Maps showing traffic intensity

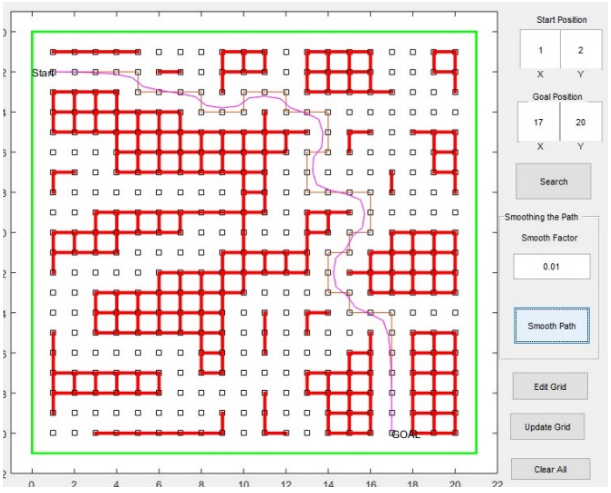


Figure 7: Find the shortest route by traffic intensity

Figure 8: Processing the route calculated according to traffic density on Google Maps

IV. PROPOSALS AND CONCLUSIONS

In this study, map parts taken from Google Map or any other map package source are processed on the algorithm map by dividing the grid number of the appropriate cell number and accordingly the shortest path between the starting node and the target node is presented. The processing of the map part on the algorithm map and the processing of the shortest path calculated by the algorithm on the map are performed manually. This operation can be done digitally by using image processing algorithms and developing necessary software. In the parts that will be taken from the map sources giving instant traffic density information, the shortest path calculation can be made according to the roads with the least traffic volume. With this process, we can send an itinerary to autonomous devices by transmitting the map parts with jpeg-file extension. Algorithms such as Q-Learning [6] on the same grid can also be tried except for A*, to compare results.

ACKNOWLEDGMENT

Authors are thankful to RAC-LAB (www.rac-lab.com) for providing the trial version of their commercial software for this study.

REFERENCES

[1] P.E. Hart, N.J. Nilsson, and B. Raphael (1968). "A Formal Basis for the Heuristic Determination of Minimum Cost Paths". IEEE Transactions on Systems Science and Cybernetics SSC4. 4 (2): 100–107.

[2] D. Delling, P. Sanders, D. Schultes and D. Wagner (2009). "Engineering route planning algorithms". Algorithmics of Large and Complex Networks: Design, Analysis, and Simulation. Springer. pp. 117–139.

[3] W. Zeng and R.L. Church (2009). "Finding shortest paths on real road networks: the case for A*". International Journal of Geographical Information Science. 23 (4): 531–543.

[4] R. Dechter and J. Pearl (1985). "Generalized best-first search strategies and the optimality of A*". Journal of the ACM. 32 (3): 505–536.

[5] Z. Algfoor, M. Sunar and H. Kolivand, "A comprehensive study on pathfinding techniques for robotics and video games". Int. J. Comput. Games Technol. (2015).

[6] H. Cetin and A. Durdu, "Path planning of mobile robots with Q-learning", IEEE 22nd Signal Processing and Communications Applications Conference (SIU 2014), pp 2162-2165, 23-25 (April 2014). Trabzon, Turkiye.

Feature Selection for Gender Classification in TUIK Life Satisfaction Survey

A. ÇOBAN[1] and İ. TARIMER[2]

[1] Muğla Sıtkı Koçman University, Muğla/Turkey, adilcoban@mu.edu.tr
[2]Muğla Sıtkı Koçman University, Muğla/Turkey, itarimer@mu.edu.tr

***Abstract*— As known, attribute selection is a method that is used before the classification of data mining. In this study, a new data set has been created by using attributes expressing overall satisfaction in Turkey Statistical Institute (TSI) Life Satisfaction Survey dataset. Attributes are sorted by Ranking search method using attribute selection algorithms in a data mining application. These selected attributes were subjected to a classification test with Naive Bayes and Random Forest from machine learning algorithms. The feature selection algorithms are compared according to the number of attributes selected and the classification accuracy rates achievable with them. In this study, which is aimed at reducing the dataset volume, the best classification result comes up with 3 attributes selected by the Chi2 algorithm. The best classification rate was 73% with the Random Forest classification algorithm.**

***Keywords* — Data, algorithms, attribute selection, data mining, Orange program, machine learning.**

I. INTRODUCTION

The concept of happiness is defined as "a state of prosperity from complete and continuous attainment of all wishes" [1]. Definitions of happiness based on philosophy and religion are used to mean a better life than emotion [2]. Corporations and public institutions want their customers and employees to be happy. Governments and societies want citizens to be happy also. They protect and evolve their assets in this regard.

New methods and approaches have been developed for how to do the most accurate measurement of economic, biological, religious, and psychological areas, along with increased work on the happiness measurement [3]. In this context, various scientific fields such as positive psychology and happiness economics have been derived [4] [5]. Until the measurement of happiness is measured, many subjects are included in the research [2].

Although the idea of happiness has been wondered by thinkers for thousands of years, it has only recently been possible to search and investigate systematically. A number of studies have been carried out since 1980 onwards [6], [7]. In addition to the concept of happiness, life satisfaction is considered as a more measurable concept. Happiness concepts representing lifelines, emotions and human consciousness have been used in [8] [9] [10].

Life satisfaction is a cognitive and judicial situation which expresses the evaluation of life as a whole. Happiness on the other hand is conceived as an emotional state produced by positive and negative events and experiences in the life of the individual. Although there are some correlations between happiness and life satisfaction at different levels, these concepts are still different [11].

The concept of subjective well-being, which we cannot separate from the concept of happiness, is defined as people's evaluations of their quality of life [6] [7]. Researches and surveys on life satisfaction and happiness have been used as subjective well-being indicators in [12].

Many internationally organizations collect data for life satisfaction surveys. The Nordic countries are the seven top ones amongst the best 10 countries in life satisfaction concept from globally scaled researches' reports. This result was reported in World Data Base of Happiness that archived by Erasmus University Rotterdam [13] and the World Happiness Report [14] prepared by the United Nations. Turkey has ranked 69th happiest country regarding to the World Happiness Report's research.

International happiness survey, organized by GALLUP Research Company, is also an example of global-scale research. According to the company's publication on International Happiness Day, the happiest countries are Latin American countries. These results, however, appear to be very different from those reported by the United Nations [15]. This difference suggests that the basic concept used in the two separate studies differs. The survey by the United Nations has focused on "life satisfaction" while GALLUP researches the concept of "happiness". According to research on positive experience made by GALLUP, Turkey took place in the bottom row [16].

There are institutions and organizations that conduct research on happiness in Turkey. Some of these are Zenna [17], Futurebright [18], Turkey Statistics Institute [19] and the Habitat Association [20]. Apart TSI, the organizations mentioned above offer services to the firms that want to develop marketing and sales strategies.

The first Life Satisfaction Survey made by Turkey Statistical Institute (TSI), is conducted as an additional module to the Household Budget Survey in November 2003. Since 2004, the research has been conducted regularly every year and has been updated with different additions and developments until 2012. In 2013, the sample volume of the Life Satisfaction Survey was increased to give an estimate at provincial level in order to see the difference between provinces and regions and to create resources for other studies at province level. TSI has carried out the research mentioned above at the provincial level in 2013 and 2016 [21].

We focused on applying data mining techniques on TSI life satisfaction surveys and classifying the gender with

variables of satisfaction in this study. For this purpose, a new data set was created with 21 attributes indicating satisfaction levels. The strength of the association of these attributes with the class was rated by attribute selection algorithms. Then the rated attributes were classified by Naive Bayes and Random Forest classification algorithms. Finally, feature selection algorithms were compared according to classification accuracy ratios.

II. MATERIAL AND METHOD

In this study, Orange Data Mining program [22] is preferred to use. The Orange program is an open source software package that can also be run as a visual programming tool or as a script in the Python programming language platform. In the orange library, data is filtered with hierarchically arranged components such as probability assessment, attribute rating and classification [23].

The Rank widget scores the attributes according to their correlation with the class. Attribute scoring methods that can be used in Rank widget are Information Gain, Information Gain Ratio, Gini Impurity Index, Chi-Square-Chi-Square, ReliefF, Fast Correlation Based Filter (FCBF) [24].

The feature selection algorithms used in the study are briefly introduced:

Information Gain: The ID3 algorithm uses information gain when constructing decision trees. When choosing an attribute, it is used to reduce the variance of multi-valued attributes taking into account the number and size of branches. It shows how close we are to the classification results given by the sub-values of the tested object. The concept of information acquisition is explained by the concept of entropy. Entropy is a concept that expresses irregularity. The lower the entropy value in the nodes of decision tree, the better the distinction is made. The entropy gets value between 0 (lowest value) and 1 (highest value) [25] [26].

Information Gain Ratio: An improved version of the Information Gain algorithm is information gain ratio. It is used because it solves the disadvantage of Information Gain. Although Information Gain is often a good measure to determine the degree of importance of qualification, a considerable bias problem arises. Information Gain Ratio is used to overcome this problem [27].

Gini Impurity Index: Used by the CART algorithm. A decision tree finds the attributes to be used when branching the structure. Attributes that can make the most discrimination are preferred because they get higher scores. If there is only one class label in the branch, no new branching occurs. As in the concept of entropy, the lower the purity index of the nodes is, the better the discrimination is made [28].

Chi2 (Chi-square distribution): A statistical tool used for rejecting derived hypotheses that data are independent. It is a non-parametric technique used to determine whether the distribution of observed frequencies differs from the expected theoretical frequencies [29].

ReliefF: This is an algorithm which is not dependent on heuristic methods but works on low-order polynomial time, and is based on the feature selection algorithm which is resistant to noise and feature interactions and is used in binary classification [30].

Fast Correlation-Based Filter (FCBF): It has been developed with the enhancement of the correlation-based (CBF) algorithm. Without a bidirectional correlation analysis, it is a fast algorithm that can determine related attributes and the redundancies between them [31].

Within the scope of the study, 196,203 records were used which are all the raw data of TSI Life Satisfaction Survey 2013. 83074 of them are male and 113129 of them are female. Expressing general satisfaction attributes and gender attribute from the data were selected to create a new data set through Microsoft Excel 2016. Descriptions of attributes are given in Table 1. Figure 1 shows the steps followed in the experimental works.

Table 1 Questions and Variables

	Attributes	**Questions**
1	Personal Health	B12.1 Are you satisfied with your health?
2	Marriage	B12.2 Are you satisfied with your marriage?
3	Personal Education	B12.3 Are you satisfied with education you have received so far?
4	Housing	B12.4 Are you satisfied with the housing you are in?
5	District	B12.5 Are you satisfied with your neighborhood or district?
6	Job	B12.6 Are you satisfied with your job?
7	Job Income	B12.7 Are you satisfied with your income earned from your job?
8	Household Income	B12.8 Are you satisfied with your monthly household income?
9	Social life	B12.9 Are you satisfied with your social life (such as entertainment, cultural and sports)?
10	Self-care	B12.10 Are you satisfied with the time set aside for yourself?
11	Time spent on traffic to and from work	B12.11 Are you satisfied with the time you spend in traffic to go to work?
12	Relative	B13.1 Are you satisfied with your relationships with your relatives?
13	Friend	B13.2 Are you satisfied with your relationships with your friends?
14	Neighbor	B13.3 Are you satisfied with your relationships with your neighbors?

15	Workplace Relations	B13.4 Are you satisfied with relationships with people related to your work?
16	General Health Services	B14.1 Are you satisfied with your health services?
17	Public order	B14.2 Are you satisfied with public security services?
18	Judicial	B14.3 Are you satisfied with judicial services?
19	General Education	B14.4 Are you satisfied with the education services?
20	SII Services	B14.5 Are you satisfied with Social Insurance Institution Services?
21	Transportation	B14.6 Are you satisfied with the transportation services?

Answers to the satisfaction scale questions are evaluated by five scores:

1 – Strongly satisfied 2 – Satisfied 3 – Undecided 4 – Disagreed 5 – Strongly disagreed

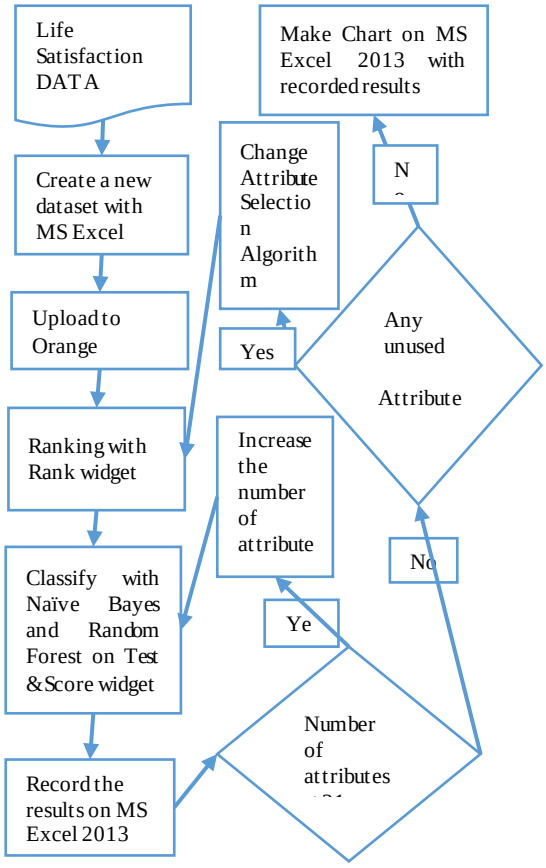


Figure 1: Experiment Steps

Figure 1 shows the experiment steps proceeded in the study. The data cleared through Microsoft Excel 2013 was uploaded to Orange with the File widget in the data mining program. The class attribute is selected as target in the interface of the File widget and selections are made for each attribute to evaluate them as nominal. The attributes were scored attribute selection algorithm in the Rank widget. Once ranked according to their score in the Rank interface, two of attributes selected to send Test and Score widget. The number of attributes was increased in each round and sent to the Test and Score widget. In Rank widget, each ranking method was performed separately and the classification test was repeated. In the Test and Score interface, Cross Validation was selected as the sampling method and 10 folds were used. The classification methods used during the test and score phase were Naive Bayes and Random Forest. The accuracy ratios obtained after each test were transferred to in a MS Excel table and visualized by charts. The data processing model generated in the Orange data mining program interface of this model is shown in Fig 2. The interfaces of the widgets used in the data processing model are shown in Fig 3, 4 and 5.

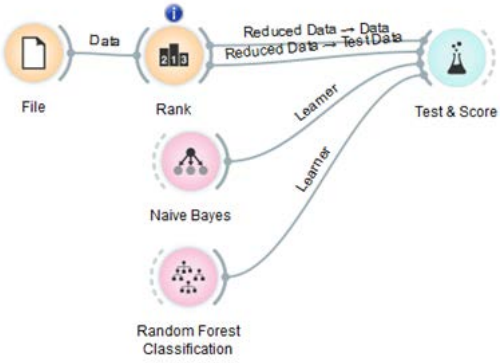


Figure 2: Model used on Orange Data Mining

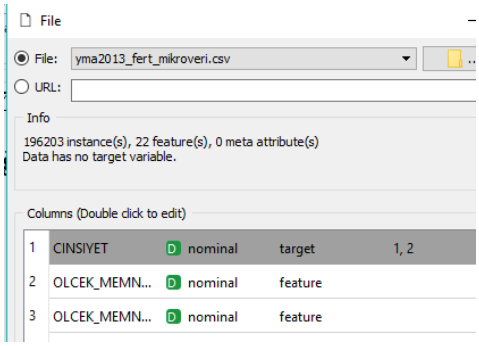
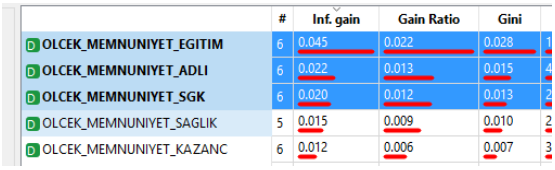


Figure 3: File

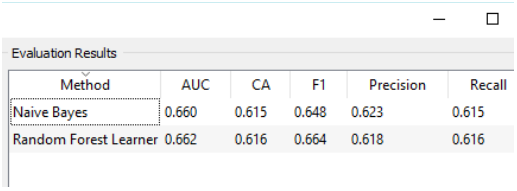
In Figure 3 shows that selecting the file to be worked on, determination of the type of attributes (nominal) and determination of target attribute are done.



	#	Inf. gain	Gain Ratio	Gini	
OLCEK_MEMNUNIYET_EGITIM	6	0.045	0.022	0.028	1
OLCEK_MEMNUNIYET_ADLI	6	0.022	0.013	0.015	4
OLCEK_MEMNUNIYET_SGK	6	0.020	0.012	0.013	2
OLCEK_MEMNUNIYET_SAGLIK	5	0.015	0.009	0.010	2
OLCEK_MEMNUNIYET_KAZANC	6	0.012	0.006	0.007	3

Figure 4: Rank

In Figure 4, the scored attributes can be seen as ranked.



Evaluation Results

Method	AUC	CA	F1	Precision	Recall
Naive Bayes	0.660	0.615	0.648	0.623	0.615
Random Forest Learner	0.662	0.616	0.664	0.618	0.616

Figure 5: Test and Score

In Figure 5, classification test result is shown.

III. FINDINGS AND RESULTS

In this section, the results produced from the experiments in the study were obtained and these results are converted into tables and charts to be compared with each other. In this context; attributes according to result obtained from the feature selection algorithms; the number is increased every time and subjected to classification test for each algorithm separately. The scores obtained by attribute scoring methods are shown in Table 2.

Table 2 Scores by Attribute Selection Algorithm

No.	Attributes	Attribute Selection Algorithms					
		Inf. gain	Gain Ratio	Gini	Chi2	ReliefF	FCBF
3	Personal Education	**4,523E-02**	**2,245E-02**	**2,815E-02**	**1,101E+04**	**5,000E-02**	**3,157E-02**
18	Judicial	**2,233E-02**	**1,272E-02**	**1,493E-02**	**4,393E+03**	1,000E-02	**1,653E-06**
20	SII Services	**1,956E-02**	**1,218E-02**	**1,271E-02**	2,894E+03	-2,200E-02	1,553E-06
1	Personal Health	1,522E-02	9,483E-03	1,010E-02	2,148E+03	-1,000E-02	1,212E-06
7	Job income	1,158E-02	5,767E-03	7,261E-03	**3,848E+03**	9,425E-04	1,900E-07
2	Marriage	7,142E-03	7,001E-03	4,598E-03	3,988E+02	1,988E-02	1,233E-07
17	Public order	5,672E-03	4,854E-03	3,882E-03	2,708E+02	-4,000E-02	5,263E-07
21	Transportation	5,630E-03	4,032E-03	3,677E-03	1,938E+02	**3,400E-02**	4,850E-07
4	Housing	2,755E-03	2,127E-03	1,823E-03	3,518E+02	**4,000E-02**	2,487E-07
19	General Education	2,692E-03	1,854E-03	1,836E-03	1,287E+02	-4,800E-02	2,227E-07
16	General Health Services	2,254E-03	1,588E-03	1,541E-03	1,995E+00	3,197E-17	1,888E-07
12	Relative	2,024E-03	1,801E-03	1,383E-03	1,533E+02	1,600E-02	1,930E-07
14	Neighbor	9,777E-04	9,081E-04	6,665E-04	6,314E+01	-3,200E-02	9,578E-08
10	Self-care	8,347E-04	5,185E-04	5,697E-04	3,313E+00	-1,600E-02	6,502E-08
9	Social Life	6,866E-04	3,914E-04	4,692E-04	2,790E+01	-2,000E-03	5,067E-08
8	Household Income	6,678E-04	3,497E-04	4,532E-04	6,847E+00	2,000E-03	4,692E-08
13	Friend	5,658E-04	6,102E-04	3,858E-04	9,185E+00	-4,000E-03	5,972E-08
5	District	3,728E-04	3,316E-04	2,498E-04	2,662E+01	-2,000E-03	3,610E-08
11	Time spent on traffic to and from work	2,071E-04	1,372E-04	1,199E-04	1,142E+03	2,171E-03	3,698E-07
6	Job	1,785E-04	1,334E-04	9,802E-05	5,948E+02	1,794E-02	4,410E-07
15	Workplace Relations	6,071E-05	6,179E-05	3,474E-05	2,803E+01	8,533E-03	**6,046E-03**

*The numbers marked in bold are the highest three values in the sequence for each algorithm.

The results shown in Table 2 demonstrates that; education, judicial and SII services attributes get the highest score by three of the algorithms (Inf. Gain, Gain Ratio and Gini). The Chi2 algorithm gives the highest score in terms of the income attribute in addition to the training and forensic attributes. The ReliefF algorithm gives the highest score to residential attribute besides education attribute and transportation attribute. FCBF, on the other hand, gives the highest scores to education attribute, judicial and business relation attributes.

In Figure 6, Naive Bayes classification test results are shown by a chart.

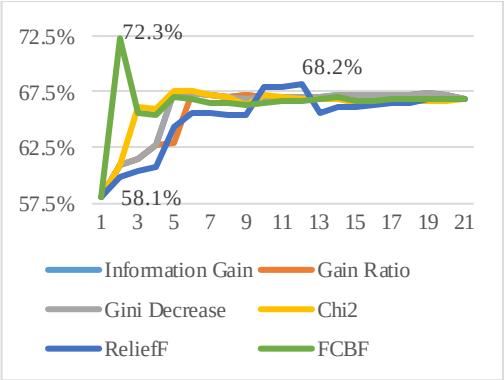


Figure 6 Naive Bayes classification results

From Figure 6, the highest value of the feature selection algorithms tested with Naive Bayes is seen as 72.3%. This value was obtained by selecting two attributes (3 and 15) with the highest score selected by FCBF algorithm. Information gain, Gini Decrease and Chi2 were obtained with the highest score of 67.5 with five attributes. The highest score (67.4) obtained by Gain Ratio was reached with six attributes. The highest score of ReliefF, 68.2, was reached with twelve attributes. As the number of attributes is increased, the scores obtained are retained same (no increase). The classification score value does not increase further because the contribution scores of the subsequently added attributes are low.

In Figure 7, test results of the Random Forest classification are shown by a chart.

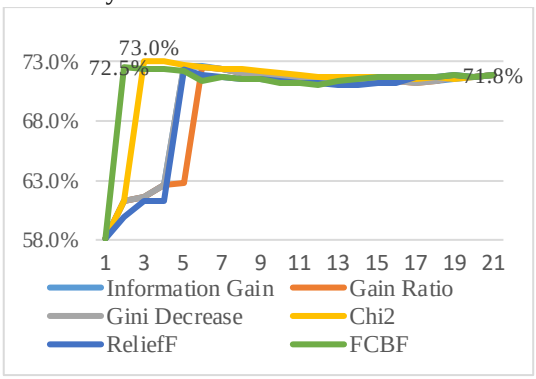


Figure 7 Random Forest classification results

As shown in Fig. 7, the highest value of the feature selection algorithms tested with Random Forest is seen as 73.0%. This was obtained when 3 attributes (3, 18, and 7) with the highest score were selected by the Chi2 selection algorithm. The highest score of 72.6 obtained with information gain and gini decrease was reached with five attributes. The highest score of 72.5 was reached with six attributes selected by Gain Ratio. The highest score of ReliefF, 72.2, was reached with five attributes. The highest score of 72.5, was reached with two attributes (3 and 15) selected by FCBF algorithm.

The highest classification score was obtained with personal training, judicial and job income attributes. The best result has been achieved using only these three attributes without having to use other attributes.

IV. DISCUSSION AND CONCLUSIONS

In this study, it is stated that data mining techniques can be applied on life satisfaction questionnaires and gender classification can be done with satisfaction variables. Therefore, at this point, it has been determined whether there is a difference of satisfaction among the genders. The accuracy of 73% in the classification confirms that the satisfaction levels differ between men and women for different attributes.

The satisfaction of personal education was selected among the first three most valuable attributes by all algorithms. Judicial satisfaction was also selected among the first three most valuable attributes by five algorithms. SII Services was also selected as previous mentioned attributes by four algorithms. Since these attributes are good at distinguishing gender, it has been seen that men and women participants have different thoughts in these matters. This suggests that the services received at these topics are assessed differently by men and women participants.

Another consequence of this work is that it is difficult to apply data mining methods to datasets that have large volume of data. In such a case, it is necessary to decrease variables by subsetting most valuable attributes, so irrelevant features are dismissed and classification accuracy is increased.

The disadvantages (class imbalance and subjectivity in the answers to the questions) in this dataset obtained from the life satisfaction survey were confronted as a problem when analyzing the dataset. For this reason, if the data set we have studied is also handled with algorithms developed for imbalanced data, higher accuracy percentages can be achieved.

V. ACKNOWLEDGEMENTS

The authors of this work would like to thank that Turkey Statistical Institute have given the life satisfaction data us to use in our thesis.

VI. REFERENCES

[1] TDK, "Güncel Türkçe Sözlük," 26 September 2006. [Online]. Available: http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.59554633ec65c9.96488763.

[2] "Science of Happiness," Pursuit of Happiness Inc., [Online]. Available: http://www.pursuit-of-happiness.org/.

[3] S. Büyükdüvenci, "Aristotales'te Mutluluk Kavramı," *Felsefe Dünyası,* no. 3, 1993.

[4] Ö. KARAIRMAK ve R. SİVİŞ, "Modernizmden Postmodernizme Geçiş ve Pozitif Psikoloji Özlemi," *Türk Psikolojik Danışma ve Rehberlik Dergisi,* vol. 30, no. 3, pp. 102-115, 2008.

[5] M. ŞEKER, "Mutluluk Ekonomisi," *Sosyoloji Konferansları Dergisi,* no. 39, pp. 115-140, 2009.

[6] S. SELİM, "Life Satisfaction and Happiness in Turkey," *Social Indicators Research,* vol. 3, no. 88, pp. 531-536, 2008.

[7] Z. CİHANGİR ÇANKAYA, "Özerklik Desteği, Temel Psikolojik İhtiyaçların Doyumu Ve Öznel İyi Olma: Öz-Belirleme Kuramı," *Türk Psikolojik Danışma ve Rehberlik Dergisi,* vol. 4, no. 31, pp. 23-31, 2009.

[8] F. M. Andrews ve R. Crandall, "The validity of measures of self-reported well-being," *Social Indicators Research*, vol. 3, no. 1, pp. 1-19, 1976.

[9] A. Campbell, The Sense of well-being in America:Recent Patterns and Trends, New York: McGraw-Hill, 1981.

[10] R. A. Emmons ve E. Diener, "Factors predicting satisfaction judgments: A comparative examination," *Social Indicators Research*, vol. 16, no. 2, pp. 157-167, 1985.

[11] M.-W. TSOU ve J.-T. LIU, "Happiness and Domain Satisfaction in Taiwan," *Journal of Happiness Studies*, vol. 2, no. 3, pp. 269-288, 2001.

[12] M. G. Royo ve J. Velazco, "Exploring The Relationship Between Happiness, Objective And Subjective Well-Being: Evidence From Rural Thailand," Wellbeing in Developing Countries ESRC Research Group, WED Working Paper 16, Bath, 2006.

[13] "World Database of Happiness," Erasmus University Rotterdam, Happiness Economics Research Organization. [Online]. Available: http://www.worlddatabaseofhappiness.eur.nl/.

[14] J. Helliwell, R. Layard ve J. Sachs, "World Happiness Report," United Nations, [Online]. Available: http://worldhappiness.report/. [Erişildi: 5 May 2017].

[15] J. Clifton, "The Happiest People in the World?," GALLUP Inc., 16 March 2016. [Online]. Available: http://www.gallup.com/opinion/gallup/189989/happiest-people-world.aspx?g_source=happiness&g_medium=search&g_campaign=tiles. [Erişildi: 18 May 2017].

[16] J. Clifton, "The Happiest and Unhappiest Countries in the World," Gallup, Inc., 2017.

[17] ZENNA Kurumsal Marka Yönetim Araştırmaları ve Danışmanlığı, [Online]. Available: http://www.zennadanismanlik.com/.

[18] Futurebright Research, [Online]. Available: http://futurebright.com.tr/tr/.

[19] Türkiye İstatistik Kurumu, [Online]. Available: http://www.tuik.gov.tr/.

[20] Habitat Derneği, [Online]. Available: http://habitatdernegi.org/.

[21] "Bilgi Talebi," Türkiye İstatistik Kurumu, [Online]. Available: http://www.tuik.gov.tr/ UstMenu.do?metod=bilgiTalebi.

[22] Orange Data Mining, [Online]. Available: https://orange.biolab.si/.

[23] D. Janez, C. Tomaz, E. Ales, G. Crt, H. Tomaz, M. Mitar, M. Marchin, P. Matija, T. Marko, S. Anze, S. Miha, U. Lan, Z. Lan, Z. Jure, Z. Marinka ve Z. Blaz, "Orange: Data Mining Toolbox in Python," *Journal of Machine Learning Research* , no. 14, pp. 2349-2353 , 2013.

[24] "Scoring methods (scoring)," Orange Data Mining, 2015. [Online]. Available: https://docs.orange. biolab.si/3/data-mining-library/reference/evaluation .cd.html.

[25] H. Föllmer, On entropy and information gain in random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 26(3), 207-217, 1973.

[26] C. E. Shannon, «A Mathematical Theory of Communication,» *The Bell System Technical Journal*, cilt 27, p. 623–656, 1948.

[27] GmbH, RapidMiner, "Weight by Information Gain Ratio," RapidMiner Inc., [Online]. Available: https://docs.rapidminer.com/studio/operators/modeling/feature_weights/weight_by_information_gain_ratio.html.

[28] U. M. Fayad., and B. I. Keki. "The attribute selection problem in decision tree generation." *AAAI*. 1992.

[29] GmbH, RapidMiner, "Weight by Chi Squared Statistic," RapidMiner, Inc., [Online]. Available: https://docs.rapidminer.com/studio/operators/modeling/feature_weights/weight_by_chi_squared_statistic.html.

[30] I. Kononenko, E. Šimec ve M. Robnik-Šikonja, "Overcoming the Myopia of Inductive Learning Algorithms with RELIEFF," *Applied Intelligence*, vol. 7, no. 1, pp. 39-55, 1997.

[31] L. Yu ve H. Liu, "Feature Selection for High-Dimensional Data A Fast Correlation-Based Filter Solution," *Proceedings of the 20th international conference on machine learning (ICML-03)*, 2003.

Automatic Author Detection in Turkish Books Using N-Gram and Naïve Bayesian Approach

S. KAYA[1] and A. GUNES[2]

[1] Fatih Sultan Mehmet Vakif University, Istanbul/Turkey, skaya@fsm.edu.com
[2]Istanbul Aydin University, Istanbul/Turkey, aligunes@aydin.edu.tr

Abstract **- Text classification, which is a sub-field of the natural language processing, is utilized for the solutions of problems in various areas. One of these areas is the author detection in written texts. When a person writes a text, he or she makes several marks due to the spelling characteristics. Author detection or recognition means that ownership of the text is questioned by comparing these spelling features. Different features belonging to the author can be extracted from texts and many comparisons can be made. In this study, author recognition was performed using bigram, trigram and quadrigram frequency property with the Naïve Bayesian approach for decision making. 120 different Turkish books written by 20 Turkish authors in different distributions were studied. Initially, the authors' bigram, trigram and quadrigram frequency properties were extracted from the books. Then, the comparison of the n-grams performances attained by Naïve Bayesian method is examined through this paper.**

Keywords **– Text classification, Author detection, Naïve Bayesian approach, N-gram.**

I. INTRODUCTION

HUMANITY has produced countless documents since the beginning of history. Who wrote the document, the states of citation, similarities with other documents have become the information desired to be known. Author detection is basically the process of identifying the author of a given text automatically by just looking at the text itself. If we explain more broadly, author detection is the process of finding the author of a given document among a group candidates who were previously introduced to the system [1]. Authorship can be beneficial for many areas: literary science, sociolinguistic research, language psychology, social psychology, forensics, medical diagnosis [2].

The techniques used to detect the author vary and they are quite complex. As people's thoughts are like their fingerprints, written texts are like the reflections of their way of thinking. Known things, unknown things, misunderstandings, thesaurus, writing rules, word usage rates, speaking and thinking habits affect writing styles. Even if one extract features of an author's text, analyzing the text and detecting author is very difficult for a person, however computers can do it easily.

Before running the computers, we must ask two question:

- What is the special features of author on an article?
- How to evaluate these features?

We firstly give answers in the literature to these questions in this article. Following, we will show comparative results by evaluating the author's bigram, trigram, and quadrigram frequency with the Naïve Bayesian approach.

Rest of the Paper is organized as follows. Section 2 reviews existing techniques used for Authorship Analysis. Section 3 answer that what is author features on the text. Section 4 include text pre-processing, how we evaluate results and our outputs followed by section 5 which is conclusion.

II. RELATED WORKS

Augustus de Morgan at 1851, ask her friend about author of a text whether it can be detected by word lengths[3]. Following this, first study on author recognition was done by a physicist, T.C. Mendenhall. In this work, extract graphical representation of the author with taking length of words and frequency of their occurrence. This features of author is named "word-spectrum" or "characteristic curve". In the article, some books of some authors have charts drawn according to the features indicated and graphical comparisons are made. The results of the experiments are quite interesting but it was inadequate [4]. Other works were done with word length by C. Moscol in the same years [5], [6].

In 1939, Yule published a good study on author recognition. This study included length of sentences against Mendenhall's length of word. Yule compared Bacon and Coleridge by using sentence length distributions [7]. Yule added 1944 innovations to his work. This innovation is vocabulary richness. Yule proposed a formula using k-statistic that contains pair of word frequency [8].

The next remarkable work was done by a linguist, George K. Zipf. "Human Behavior and The Principle Of Least Effort" book written by Zipf contains some important laws for natural language processing. Although these laws are not entirely on author recognition but also many paper have referred [9]. Zipf say that people express themselves with minimal effort. Zipf proves it to his human language works. People use the words they learn from others without trying to learn new words. In a similar way, this means that a person's words and style will not change much from one paper to another. A people uses similar words and styles in his articles [10].

Another effective work was Mosteller and Wallace on Federalist papers in 1964. Federalist papers are a collection of article and essays which were written by Alexander Hamilton,

John Jay, James Madison in 1787. The number of essays is eighty-five. Although it is known that who wrote the article in general, some of these not clear to known. Mosteller and Wallace used a novel technique using function words to determine which paper was written by who among three writers. Function words also called functors are a word to express a grammatical or structural relationship with other words or sentences. These words can be auxiliary verbs, prepositions, conjunctions, determiners, pronoun etc... . In the work, frequency of function words was taken to determine author of papers by using Bayesian analysis. After this work, the function words have begun to be used in author detection [11][12].

When it came to the 1990s, many features were considered to be singular characteristics belonging to the author. All of these series of studies were called "stylometry". Holmes compiled author recognition features which were suggested such as sentence length, word length, word frequencies, character frequencies, and vocabulary richness [13]. And then Rudman tried to put together problems and solutions by doing a study on the state of authorship attribution studies [14].

Increase of electronic documents like computer archives, online documents (email, blogs) made author recognition studies more important in 1990s. At the same time, developments in computer science changed the nature of authorship studies. Improvement in some areas such as artificial intelligence, machine learning, natural language processing, etc. and the increase in the capacity of computers has made authorship a little more interesting [15]. Works before computer revolution focuses on few features that were low-dimensional and statistical.

III. Stylometric Features Of An Author

Since the introduction of author recognition as a problem, many scientific studies have been conducted. In each study, either the previously used features were implemented, or a new logic was put forward. We can categorize these attributes as character-based, lexical, syntactic, and structural.

A. Character-Based

Character is the smallest unit of language. All signs belonging to a language are a character. Characters can be alphabetic, numeric, punctuation, and even spaces. Words, sentences are formed by characters being come together and articles take shape by their union. It is named character based on the author's character usage style. Location, frequency, presence, sequence of characters can be taken to identify the author feature vector[4]–[6].

Another impressive character-based study is the character n-gram method. In this method, the frequency distribution is calculated over the character sequences and the author feature is extracted. N indicates how many character lengths a gram-set. The method has proven to be successful in many cases [16], [17].

B. Lexical

A text is allocated to its tokens before being processed by natural language processing tools. A token is generally a word and the word is the most important parameter when you think about it from this point of view. Lexical is used in the meaning of the correspondences, forms in the dictionary of the language to which the word belongs.

Lexical or vocabulary richness are used as a feature. The author's lexical richness is a special case of writing. It can be thought that the number of different words is related to the text length. However, Yule has done a study on the fact that the vocabulary is not related to the text length [18]. Each writer has his own thesaurus, and the presence or absence of a word in an article can provide the author of the article [19].

Word n-gram property can be categorized into lexical. Word n-gram is sequential word groups. It has increased the success rate in studies using this feature [17], [20].

C. Syntactic

Syntax is set of rules, principles for a language. The analyze based of syntax called syntactic. Although a language has a standard syntax, not every person has the same habit of using them. A trace can be made for an author by looking at the properties of adjectives, pronouns, verb etc. sequences. The main idea is that a writer tends to consistently produce the same trace. If we think about this method a little, it is understood that this technique is language dependent [15].

It is necessary to perceive the types of words in this technique and reconstruct the text according to the keywords. To give a few examples of key words; noun-N, preposition-P, noun phase-NP. It can translate all text with similar logic. And then, the relations between these structures are tried to be found [21].

There are also studies where the distances between frequencies of the syntactic sequentially are used as properties. The logic in these studies, a writer will use similar syntactic groups in all his articles. This technique can be used not only vocabulary type but also which word [16]. The process is called word tagging. Tagged words can also be used for n-gram based feature extraction [22].

D. Structural

Structural features can be font, style, shape, size of the text and content-specific, organizational some attributes. Anderson used the typographical features in a study [23]. Structural characteristics are generally applied on electronic documents [24].

IV. Method and Experimental Results

In this study, we gather 120 Turkish books. These books are written by 20 different authors in several proportions. The book of 20 writers were selected randomly for extracting stylometry from 120 books, so there are 100 books left for the test.

```
1  TrBooks ← PreProcessBook(TrainingBooks)
2  AuthorsList←ExtractBigram(TrBooks)
3  For Each Author ← AuthorsList
4      StylometryClasses ← {Author.Name, Author.Grams}
5  TeBooks← PreProcessBook(TestBooks)
6  BookList ←ExtractBigram(TeBooks)
7  For Each TheBook ← BookList
8      TheClass ←NBClassify(TheBook, StylometryClasses)
9      Print(Book.Name, TheClass)
```

Figure 1: Pseudo Code for Bigram

Figure 1 shows that pseudo code for a bigram and trigram and quadrigram are similar.

A. Text Pre-Processing and Extracting Author's Stylometry

First of all, whole the books was passed a text pre-processing. Numbers, spaces, marks and other characters is non-alphabetic was removed from the text. Besides, some stop words like "ve", "veya", "ile" replace unique single character. The reason we do this, stop words taken palace too often in a text [25] and we think this affects the reliability of work. Figure 2 illustrated that how to appear a raw book after pre-processing operation.

…
güçmeydanaçıkmakolduğunubilirlerdibirkaçsenedenberik
ontratıyenilemekkiralarıalmakgibiişleriyüklenenmümtazo
nuhattadükkanında>karşısındaikenbilegörmeninnekadargü
çolduğunubilirdidahagençadamdükkanagirergirmezsiyahg
özlüğünübirkudrettılsımbüyülübirsilahgibigözlerinetakarb
ucamperdearkasındaadetagörünmezoluroradanpiyasanındu
rgunluğunuhayatınağırlığınıdevletmemuriyetindebellibirg
elirleçalışanlarınsaadetinianlatırmemurluğubırakıpdaelkasi
bühabibullahhadisineuyduğuiçinevetsırfbununiçin
…

Figure 2: An Example of Pre-Processed Book

We used character n-gram for authors' stylometric features and extracted bigram, trigram, quadrigram words from each author's training book. Figure 3 is shown how the text is split into n-gram in the work.

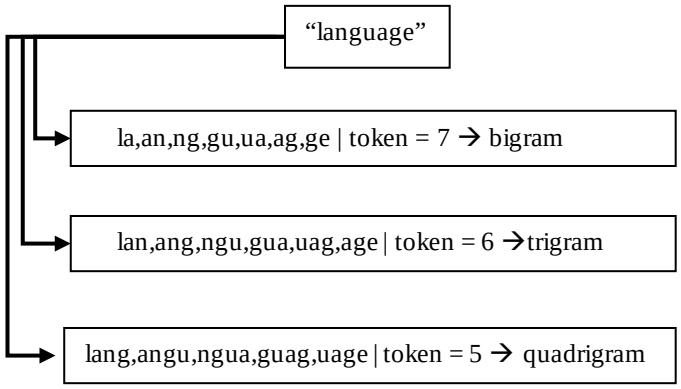


Figure 3:Bigram, Trigram, Quadrigram Words

After that, we calculated their how many times has it occurred in the text and sort by frequency. We took the most frequent 200 n-gram as an author stylometric feature. Figure 4 shown that exemplifies the bigram feature of an author. The trigram and quadrigram are similar to this example.

an:9897|ar:8683|la:8678|in:8441|en:8105:|er:7734:|de:737
0:|le:6841|bi:6549|ma:6353|di:6342|ın:6151|nd:6104|da:5
939|ir:5778|ya:5742|ak:5638|ka:5458|am:5141|ni:4684|m
e:4612 |…

Figure 4: An Author's Bigram Frequency Vector

These processes are done for each book. Then comparison is made with the Naïve Bayes formula and the classification process is performed.

B. Classification

Classification: the process of categorizing objects according to pre-specified classes. Our problem is that which one of the registered writers is already trained in the system written a book newly introduced into the system. The classification process can be done statistically or machine learning based [26]. We used Naïve Bayesian approach to classify the book.

The Naïve Bayes (NB) classifier is widely used to in decision mechanisms. NB is based on conditionally independent variables and has ability large number of features to take into account [25]. Equation 1 and Equation 2 view that Naïve Bayes formula.

$$P(C_i|X) = \frac{P(X|C_i)\,P(C_i)}{P(X)} \quad (1)$$

$$P(C_i|X) = P(X|C_i)\,P(C_i) \quad (2)$$

In our case, all authors are a class and we calculate using feature vector of an author and extracted feature vector of a book NB result. A book entered into the system is calculated separately for each author, and the book that belongs to the biggest of calculates values. This method named maximum-likelihood estimates.

$$C \leftarrow \arg max_{C_i} P(C_i) \prod_{j=1}^{n} P(X_j = x_j|c_i) \quad (3)$$

The mathematical classification formula applied in the study is shown in Equation 23. According to the formula, the probability of each gram word contained in the stylometry feature vector of the book in the trained author stylometry vector space is calculated and then and then all the possibilities (likelihood) and prior value are multiplying. This calculated value gives the probability that the book belongs to the author. The class of the book is determined as the biggest of the probability values calculated for all authors.

C. Evaluation

We already know who written 100 books we put on our system. We wise up to if it's true that estimating of the system between author and book matching. Therefore only the process of evaluating the performance rates of bigram, trigram, quadrigram remains.

Accuracy, Error rate, precision, recall, f-score values are our evaluation metric in the work.

Accuracy value is calculated by how many book are correctly estimated. Equation 4 shown that accuracy formula where A_T is Total Accuracy of all books, while T_E is True Estimated and D_T is total books inserted in the system. Error rate is obtained by subtracting the accuracy from one. Equation 5 view that error rate formula where E_R is Total Error Rate and A_T is Total Accuracy.

$$A_T = \frac{T_E}{D_T} \quad (4)$$

$$E_R = 1 - A_T \quad (5)$$

Precision and recall values for each book are calculated separately in the study and we get the average via dividing by count of the author introduced to the system. And then, we calculated the f-score value with these value.

$$P = \frac{\sum_{i=1}^{n} \frac{T_{Pi}}{T_{Pi} + F_{Pi}}}{A_C} \quad (6)$$

$$R = \frac{\sum_{i=1}^{n} \frac{T_{Pi}}{T_{Pi} + F_{Ni}}}{A_C} \quad (7)$$

$$F - Score = \frac{2\ x\ P\ x\ R}{P + R} \quad (8)$$

We made this transactions for the bigram and others separately. The outcomes was compared according to their performance comparatively.

D. Experimental Results

The following operations show the results of calculating the accuracy and error rate of the bigram, trigram and quadrigram vector spaces.

$$A_{T_{bigram}} = \frac{12}{100} = 0.12 \quad (9)$$

$$E_{R_{bigram}} = 1 - 0.12 = 0.88 \quad (10)$$

$$A_{T_{trigram}} = \frac{71}{100} = 0.71 \quad (11)$$

$$E_{R_{trigram}} = 1 - 0.71 = 0.29 \quad (12)$$

$$A_{T_{quadrigram}} = \frac{82}{100} = 0.82 \quad (13)$$

$$E_{R_{quadrigram}} = 1 - 0.82 = 0.18 \quad (14)$$

The operations below show the results of calculating the precision, recall and f-score of each vector spaces.

$$P_{bigram} = 0.108068 \quad (15)$$

$$R_{bigram} = 0.08813 \quad (16)$$

$$F - Score_{bigram} = 0.09279211 \quad (17)$$

$$P_{trigram} = 0.748215 \quad (18)$$

$$R_{trigram} = 0.6321 \quad (19)$$

$$F - Score_{trigram} = 0.68527358 \quad (20)$$

$$P_{quadrigram} = 0.767605 \quad (21)$$

$$R_{quadrigram} = 0.7056 \quad (22)$$

$$F - Score_{quadrigram} = 0.73529765 \quad (23)$$

All calculations are shown above. It was not necessary to clarify how the calculations were done due to the explanation in previous chapters.

The sum of entire the work is demonstrated on Table 1 and Figure 5 show values on the graph. Explanation and conclusions about the table will be presented in the next section.

Table 1: The Brief of All Calculations

N-Gram	A_T	E_R	P	R	$F - Score$
Bigram	0.12	0.88	0.108068	0.0881	0.09279211
TriGram	0.71	0.29	0.748215	0.6321	0.68527358
QuadriGram	0.82	0.18	0.767605	0.7056	0.73529765

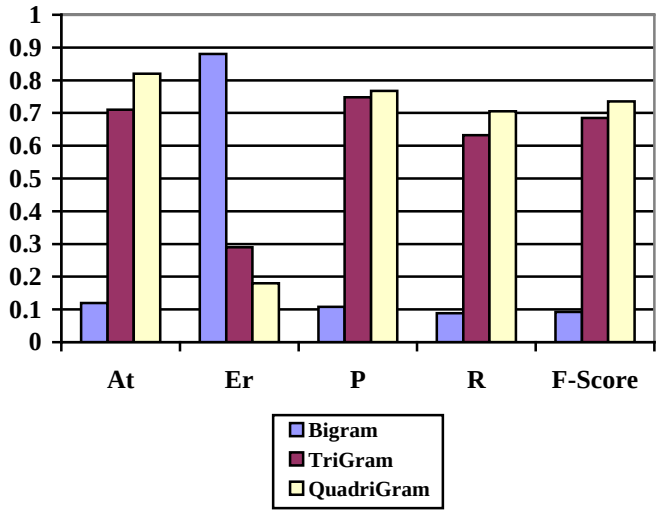


Figure 5: Results for N-Grams

V. CONCLUSION

We observed the achievements of the n-gram vector spaces about author detection. As a result of our observations, bigram vector space is failed. Besides, trigram and quadrigram gave good accuracy result. It should be noted the best performance belongs to the quadrigram.

We can say that with n-gram technique, the performance is being more successful as the number of n increases in the author recognition process by looking at the output. This is because, as long as the pattern size is longer, the likelihood of writing stylometry becomes more specific. As the size of n decreases, commonly used character string are observed on all stylometry vector. This reduces the classification performance of the naïve Bayes classifier.

It is achieved 82% accuracy with the quadrigram giving the best output in the study. However, the performance can be improved by varying author stylometry features. In other words, by adding the stylometric properties mentioned in section three, the author's different stylometric properties can be added to the vector space.

In another study, it is possible to study how long the n-gram pattern increases the performance and at what length it begins to fall negatively. We will look at the results of the naïve Bayes classification of n-gram vector spaces in 100 book and draw the performance curve in this case.

How n-gram properties will react to different classifiers on another curiosity. We can examine the output of our work on different classifiers. Machine learning and statistics based some classifiers like neural network, K-NN, perception will be tested.

REFERENCES

[1] C. Qian, T. He, and R. Zhang, "Deep Learning based Authorship Identification," pp. 1–9, 2016.

[2] W. Daelemans, "Explanation in computational stylometry," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7817 LNCS, no. PART 2, pp. 451–462, 2013.

[3] De Morgan Sophia Elizabeth, *Memoir of Augustus De Morgan*. London :Longmans, Green,.

[4] T. C. Mendenhall, "The characteristic curves of composition.," *Science*, vol. 9, no. 216, p. 297, 1887.

[5] C. Mascol, "Curves of Pauline and Pseudo-Pauline style I," *Unit. Rev.*, no. 30, pp. 452–460, 1888.

[6] C. Mascol, "Curves of Pauline and Pseudo-Pauline style II," *Unit. Rev.*, no. 30, pp. 539–546, 1888.

[7] G. U. YULE, "On sentence-length as a statistical characteristic of style in prose: With application to two cases of disputed authorship," *Biometrika*, vol. 30, no. 3–4, pp. 363–390, 1939.

[8] G. Lynch, "Computational Stylometry and Analysis of Style : A Study of Characterization in Playwrights," no. June 2009, 2015.

[9] S. T. Piantadosi, "Zipf's word frequency law in natural language: A critical review and future directions," *Psychon. Bull. Rev.*, vol. 21, no. 5, pp. 1112–1130, 2014.

[10] , Zipf and G. K, *Human behavior and the principle of least effort: An introduction to human ecology*, vol. 47. 1949.

[11] F. Mosteller and D. L. Wallace, "Inference in an Authorship Problem," *Journal of the American Statistical Association*, vol. 58, no. 302. pp. 275–309, 1963.

[12] F. Mosteller and D. L. Wallace, *Inference and disputed authorship: The Federalist*. 1964.

[13] D. I. Holmes, "Authorship attribution," *Comput. Hum.*, vol. 28, no. 2, pp. 87–106, 1994.

[14] J. Rudman, "The State of Authorship Attribution Studies: Some Problems and Solutions," *Comput. Hum.*, vol. 31, pp. 351–365, 1998.

[15] E. Stamatatos, "A survey of modern authorship attribution methods," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 3, pp. 538–556, 2009.

[16] F. Hassan, "N-Gram Based Text Author Verification," *2012 Int. Conf. Innov. Inf. Manag. (ICIIM 2012)*, vol. 36, no. Iciim, pp. 74–78, 2012.

[17] V. Kešelj, F. Peng, N. Cercone, and C. Thomas, "N-gram-based author profiles for authorship attribution.," *Pacific Assoc. Comput. Linguist.*, pp. 255–264, 2003.

[18] G. U. Yule, *The Statistical Study of Literary Vocabulary*. Cambridge University Press, 1944.

[19] F. J. Tweedie and R. H. Baayen, "How Variable May a Constant be? Measures of Lexical Richness in Perspective," *Comput. Hum.*, vol. 32, no. 5, pp. 323–352, Sep. 1998.

[20] A. A. Raza, A. Athar, and S. Nadeem, "N-Gram Based Authorship Attribution in Urdu Poetry," *Proc. Conf. Lang. Technol.*, no. January 2009, pp. 88–93, 2009.

[21] H. Baayen, H. Van Halteren, and F. Tweedie, "Outside the cave of shadows: using syntactic annotation to enhance authorship attribution," *Lit. Linguist. Comput.*, vol. 11, no. 3, pp. 121–132, 1996.

[22] N. Ali and N. Ali, "ThinkIR : The University of Louisville â€™ s Institutional Repository Text stylometry for chat bot identification and intelligence estimation . By," 2014.

[23] O. de Vel, A. Anderson, M. Corney, and G. Mohay, "Mining e-mail content for author identification forensics," *ACM SIGMOD Rec.*, vol. 30, no. 4, p. 55, 2001.

[24] A. Abbasi and H. Chen, "Writeprints: A Stylometric Approach to Identity-Level Identification and Similarity Detection in Cyberspace," *ACM Trans. Inf. Syst.*, vol. 26, no. 2, pp. 1–29, 2008.

[25] C. D. Manning and H. Schutze, *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: MIT Press, 1999.

[26] M. Koppel, J. Schler, and S. Argamon, "Computational Methods in Authorship Attribution," *J. Am. Soc. Inf. Sci. Technol.*, vol. 60, no. 1, pp. 9–26, 2008.

The Effect of Over-sampling and Under-sampling Techniques in Medical Datasets

Mehmet HACIBEYOGLU[1] and Mohammed Hussein IBRAHIM[1]

[1] Necmettin Erbakan University, Konya/Turkey, hacibeyoglu@konya.edu.tr
[1] Necmettin Erbakan University, Konya /Turkey, mibrahim@ konya.edu.tr

Abstract **- A well balanced dataset is crucial for the performance of the data mining classification algorithms. In medical datasets, the percentage of normal labeled classes is higher than the percentage of abnormal labeled ones, which is called as class imbalance problem in data mining. If training dataset is imbalanced, the classification algorithm generally predicts the labels of the majority class instances correctly and the minority class instances incorrectly which leads to a major problem for artificial intelligence based medical diagnosis systems. To overcome this problem, many researchers proposed over-sampling and under-sampling techniques in the literature. Over-sampling techniques increase the number of minority class instances, where the randomly chosen instances from minority class is duplicated and added to the new training dataset or synthetic instances are generated from the minority class. Under-sampling techniques decrease the number of majority class, where the randomly chosen subset of majority class is combined with the minority class instances as the new training dataset. In this study, the effect of over-sampling and under-sampling techniques in medical datasets is examined. For the experimental study, several medical benchmark datasets and well-known classification algorithms are used.**

Keywords **- Under-sampling, over-sampling, classification, medical datasets.**

I. INTRODUCTION

In many real-time applications, the distribution of the samples in the dataset is not uniform. If the number of samples of a class is higher than the other classes, this dataset is said to be quite skewed [1]. For the imbalance data set, majority class and minority class have two important terms. Majority class is said to be a class with more number of samples and minority class is said to be a class with a relatively small number of samples [2]. This class uniformity in the data set can lead to risky results in important classification problems such as medical diagnosis prediction [4], detecting fraud in online banking [5] and intrusions detection [6]. In this case the classes with more samples are well learned in the training process of the classification algorithm and the classes with fewer samples are learned worse. In addition, it is also possible that the trained classification system predicts everything as a majority class and ignores the minority class. Several sampling techniques have been proposed by researchers to solving this problem and these methods enter the data preprocessing category in the data mining. The simplest sampling techniques are random over-sampling and random under-sampling. The random under-sampling selects random samples to delete from the majority class and the random over-sampling randomly duplicated samples of a minority class. It is useful to write this, there may be loss of information in the under-sampling, and oversampling can lead to overfitting in the learning phase [3] and more than one technique has been proposed to solve this problem due to these disadvantages. Nitesh V. Chawla et al, has developed an over-sampling technique called Synthetic Minority Over-sampling Technique (SMOTE), and the experiment of this technique was performed using C4.5, Ripper and Naive Bayes classifier algorithms and the SMOTE technique performed well. Because of this success, the algorithm has been used in many areas of data mining [7]. The minority class in the datasets affects to the classification accuracy of classification algorithms. A clustering-based under-sampling technique was developed by Yen et al. to enhance the classification accuracy for minority class. Experimental results show that clustering-based under-sampling technique is outperformed other under-sampling techniques [8]. Classification methods developed by researchers are used in many important areas. Li et al. used over-sampling and under-sampling techniques to increase the classification accuracy of medical data sets [9]. Liu et al. tested that the over-sampling and under-sampling techniques on the imbalanced text dataset affected the performance and classification accuracy [10].

II. IMBALANCED DATASET PROBLEM SOLUTIONS

The imbalanced datasets is a problem for data mining classification algorithms. This problem can be tackled using two different solution approaches. First approach solve the imbalanced problem in the data level. This approach consists of over-sampling, under-sampling and hybrid techniques. Under-sampling techniques generate a subset of original datasets by removing some of instances of the majority class [11]. Tomek links [12] and Neighborhood Cleaning Rule [13] are the under-sampling techniques in the literature. Over-sampling techniques generate a super-set of original dataset by replication some of the instances in the minority class or generating new ones form minority class. Synthetic Minority Oversampling Technique (SMOTE) [14] is the well-known over sampling technique in the literature. Hybrid techniques combine over-sampling and under-sampling techniques. SMOTE—Tomek links [15], SMOTE—ENN [15], and Borderline-SMOTE1 [16] are the hybrid techniques in the literature.

III. EXPERIMENTAL STUDY

In order to evaluate the effect of over-sampling and under-sampling techniques in medical datasets, four well-known datasets from UCI is used for experiments [17]. The characteristics of the used datasets are shown in Table 1.

Table 1: Datasets characteristics.

Dataset	Instances	Features
Pima	768	9
Breast Cancer	699	11
Bupa	345	7
Hepatitis	155	20

All datasets have 2 classes and the imbalance ratio between these two classes is shown in Table 2.

Table 2: Imbalance ratio of the datasets.

Dataset	Majority Class	Minority Class	Imbalance Ratio
Pima	500	268	0.536
Breast Cancer	458	241	0.526
Bupa	200	145	0.725
Hepatitis	123	32	0.26

Oversampling and under sampling were performed on the dataset using Weka data mining tool prior the classification. In the over-sampling process, the number of instances in the minority class was increased to the number of instances in the majority class by using the Synthetic Minority Oversampling Technique (SMOTE). In the under-sampling process, the number of instances in the majority class was decreased to the number of instances in the minority class by using spread subsample technique. The number of instances in each class is shown in Table 3 after over-sampling and under-sampling processes.

Table 3: Number of instances in each class after over-sampling and under-sampling processes

Dataset	Over-sampling	Under-sampling	Imbalance Ratio
Pima	500	268	1.0
Breast Cancer	458	241	1.0
Bupa	200	145	1.0
Hepatitis	123	32	1.0

A. Classification Methods

In the classification process, we used three well-known classification algorithm K-nearest neighbor (K-nn) [18], naïve bayes (NB) [19] and support vector machine (SVM) [20].

The K-nn algorithm is one of the simplest and most widely used classification algorithm in the machine learning. This algorithm performs classification according to the class values of the k nearest neighbors where k is given by the user. Firstly, the distance between the unseen instance and each sample in the training set is calculated with distance function like Euclidean, Chi square, and Minkowsky. Secondly, k nearest instances are selected. Finally, the class of the unseen instance is assigned to the maximum class value in the selected instances.

NB classifier is a simple probabilistic classification method based on the Bayesian theorem. In this classifier, attributes are regarded as independent from each other, all the instances are equally important and an attribute does not contain information about another one. In this approach, the probability of belonging to each class of the unseen instance is calculated. The class of unseen instances is set to the class with the highest probability.

SVM uses statistical learning theory which was developed by Vapnik. SVM is very useful classifier because it has a simple structure and high performance in practical applications. Besides, the number of samples to be used in SVM is not important. This approach based on the decision planes which separates the training instances into two subsets such that in each subset all elements are similar. The class label of the unseen instance is assigned to the class value of the subset which it belongs to.

B. Evaluation Measures

The confusion matrix is used for comparing the performance of the classifiers. The confusion matrix is a specific table where each row represents the actual class values of the instances and each column represents the predicted class values of the instances, as in Table 4.

Table 4: Imbalance ratio of the datasets.

	Predicted: Yes	Predicted: No
Actual: Yes	True Positive (TP)	False Negative (FN)
Actual: No	False Positive (FP)	True Negative (TN)

Where, TP are the positive instances that classified correctly, TN are the negative instances that classified correctly, FP are the negative instances that classified as positive, and FN are the positive instances that classified as negative.

Different evaluation measures can be calculated using the confusion matrix. In this study, we used accuracy, sensitivity, specificity, and precision measures.

$$Accuracy\ (ACC) = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Accuracy is the ratio of samples correctly classified by a classifier to the number of all instances.

$$Sensitivity\ (SEN) = \frac{TP}{TP + FN} \quad (2)$$

Sensitivity is the calculation of the number of instances correctly classified as positive divided by all positive instances.

$$Specifity\ (SPEC) = \frac{TN}{TN + FP} \quad (3)$$

Specificity is the ratio of the number of samples correctly classified as a negative to the number of all samples whose actual value is negative.

$$Precision\ (PPV) = \frac{TP}{TP + FP} \quad (4)$$

Precision is the calculation of the number of instances correctly classified as a positive divided by all instances classified as positive.

C. Results and Discussion

In this section, we are going to show the experimental results of classification algorithms with original, over-sampled and under-sampled datasets. The three well-known classification algorithms NB, SVM, and K-NN were used in the classification process with tenfold cross-validation, which is the most used technique in the literature and estimates the mean of the errors obtained on ten different testing subsets. The experimental results of NB, SVM and K-NN are shown in Table 5, Table 6, and Table 7, respectively.

Table 5: Experimental results of K-NN.

Dataset	Accuracy			Sensitivity			Specificity			Precision		
	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread Subsample	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample
Pima	0.727	**0.789**	0.705	0.727	**0.789**	0.705	0.646	**0.789**	0.705	0.721	**0.791**	0.705
Breast Cancer	0.969	**0.977**	0.963	0.969	**0.977**	0.963	0.964	**0.977**	0.963	0.969	**0.977**	0.963
Bupa	0.617	**0.675**	0.628	0.617	**0.675**	0.628	0.605	**0.675**	0.628	0.621	**0.677**	0.628
Hepatitis	0.813	**0.874**	0.719	0.813	**0.874**	0.719	0.512	**0.874**	0.719	0.796	**0.874**	0.733

Table 6: Experimental results of NB.

Dataset	Accuracy			Sensitivity			Specificity			Precision		
	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample
Pima	**0.763**	74.4	71.828	**0.763**	0.744	0.718	0.693	**0.744**	0.718	**0.759**	0.745	0.719
Breast Cancer	0.96	**0.963**	96.266	0.96	**0.963**	**0.963**	**0.967**	0.963	0.963	0.962	**0.963**	**0.963**
Bupa	0.554	**0.565**	0.548	0.554	**0.565**	0.548	**0.612**	0.565	0.548	**0.609**	0.59	0.57
Hepatitis	0.845	**0.915**	0.813	0.845	**0.915**	0.813	0.729	**0.915**	0.812	0.853	**0.915**	0.813

Table 7: Experimental results of SVM.

Dataset	Accuracy			Sensitivity			Specificity			Precision		
	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample	Imbalanced data	SMOTE	Spread subsample
Pima	0.773	**0.758**	0.737	**0.773**	0.758	0.737	0.666	**0.758**	0.737	**0.769**	0.758	0.738
Breast Cancer	0.97	**0.971**	**0.971**	0.97	**0.972**	0.971	0.966	**0.972**	0.971	0.97	**0.972**	0.971
Bupa	0.582	**0.63**	0.586	0.583	**0.63**	0.586	0.424	**0.63**	0.586	**0.757**	0.639	0.597
Hepatitis	0.851	**0.915**	0.734	0.852	**0.915**	0.734	0.661	**0.915**	0.734	0.847	**0.915**	0.736

When the experimental results are analyzed, it is seen that over-sampled datasets with SMOTE technique outperformed original datasets and under-sampled datasets with spread subsample technique. The classifiers are better learned and the classification accuracy is improved by increasing the number of instances belonging to minority class. The main problem in the medical data cluster is that the instance numbers of the patients are less than the instance numbers of the non-patients.

Therefore, classifiers cannot correctly predict the patient instances while can correctly predict the non-patient instances. According to the experimental results, we can say that, over-sampling techniques can enhanced the performance of the classification algorithms in medical datasets.

IV. CONCLUSION

In this paper, the effect of over-sampling and under-sampling techniques in Medical Datasets is examined. Medical datasets are often imbalanced datasets which may reduce the performance of data mining classifiers. In the literature, over-sampling and under-sampling techniques are frequently used to balance imbalanced datasets. So, we used over-sampling SMOTE technique and under-sampling spread subsample technique to balance four medical datasets. Experimental results show that over-sampled datasets can learn more efficiently and predict patient instances more successfully.

REFERENCES

[1] Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering, 2(4), 42-47.

[2] Chawla, N. V. (2009). Data mining for imbalanced datasets: An overview. In Data mining and knowledge discovery handbook(pp. 875-886). Springer, Boston, MA.

[3] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008, December). A comparative study of data sampling and cost sensitive learning. In Data Mining Workshops, 2008. ICDMW'08. IEEE International Conference on (pp. 46-52). IEEE.

[4] Amato, F., López, A., Peña-Méndez, E. M., Vaňhara, P., Hampl, A., & Havel, J. (2013). Artificial neural networks in medical diagnosis.

[5] Wei, W., Li, J., Cao, L., Ou, Y., & Chen, J. (2013). Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web, 16(4), 449-475.

[6] Denning, D. E. (1987). An intrusion-detection model. IEEE Transactions on software engineering, (2), 222-232.

[7] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

[8] Yen, S. J., & Lee, Y. S. (2009). Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications, 36(3), 5718-5727.

[9] Li, D. C., Liu, C. W., & Hu, S. C. (2010). A learning method for the class imbalance problem with medical data sets. Computers in biology and medicine, 40(5), 509-518.

[10] Liu, A. C. (2004). The effect of oversampling and undersampling on classifying imbalanced text datasets. The University of Texas at Austin.

[11] Ramentol, E., Caballero, Y., Bello, R., & Herrera, F. (2012). SMOTE-RSB*: a hybrid preprocessing approach based on oversampling and undersampling for high imbalanced data-sets using SMOTE and rough sets theory. Knowledge and information systems, 33(2), 245-265.

[12] Tomek I (1976) Two modifications of CNN. IEEE Trans Syst Man Commun 6:769–772

[13] Wilson DL (1972) Asymptotic properties of nearest neighbor rules using edited data. IEEE Trans Syst Man Commun 2(3):408–421

[14] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: Synthetic minority over-sampling technique. J Artif Intell Res 16:321–357

[15] Batista GEAPA, Prati RC, Monard MC (2004) A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explor 6(1):20–29

[16] Han H, Wang WY, Mao BH (2005) Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. International conference on intelligent computing (ICIC05) LNCS 3644. Springer, pp 878–887

[17] Blake, C. (1998). UCI repository of machine learning databases. http://www. ics. uci. edu/~ mlearn/MLRepository. html

[18] Fix, E., & Hodges Jr, J. L. (1951). Discriminatory analysis-nonparametric discrimination: consistency properties. California Univ Berkeley.

[19] Lewis, D. D. (1998, April). Naive (Bayes) at forty: The independence assumption in information retrieval. In European conference on machine learning (pp. 4-15). Springer, Berlin, Heidelberg.

[20] Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142). Springer, Berlin, Heidelberg.

Analysis of the Co-authorship Network of Turkish Engineering Research Society

İ.TÜRKER[1], R.DURGUT[2] and O.FINDIK[3]

[1] Karabuk University, Karabuk/Turkey, iturker@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, rafetdurgut@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

***Abstract* – Co-authorship networks provide a broad view to the connectivity properties of scholars, together with patterns of knowledge diffusion in scientific society. Network science provides a substantial framework for discovering the dynamics of these interactions those are defined by co-authoring a paper together. We constructed a complex network consisting of co-authorship links between authors, using the data retrieved from Web of Science Core Collection. Date retrieved is limited to 67248 publications addressed from Turkey in engineering field, including the timespan between 1975 and 2018. Analysis performed through this massive dataset resulted a complex network of 78883 nodes (authors) and 194232 edges (co-authorship links). Authors exhibit an average degree (neighbor) of 4.925, which increases to 6.687 in weighted analysis. Network exhibits an invincible clustering coefficient of ~0.8, while the average path length is close to 18. Together with the power-law consistent degree distribution that labels the network as scale-free, we also presented top "most central" authors of this network with respect to betweenness, closeness and eigenvector centrality measures, each defining the "importance" of an author in different aspects.**

***Keywords* - Co-authorship networks, scientific collaboration networks, complex networks, centrality, bibliographic analysis.**

I. INTRODUCTION

COMPLEX interconnected structures are everywhere in nature, from biologic to social networks. They exhibit some universal properties independent from the scope of the network, like being scale-free or having low path length and high clustering [1]. Social networks, consisting of connections between individuals are good resources for exploring structural properties of real networks [2]. In the other hand, network science provides a matured framework for discovering node and edge based characteristics of social networks.

Co-authorship networks, defined by interactions between authors those co-author the same paper, are good resources of social networks, since they also have time resolution captured by the publication years of scientific papers [3-5]. Data repositories those are publicly or academically available provide detailed metadata about the papers including the author names and affiliations together with publication years etc. Therefore, they facilitate investigating the evolution of co-authorship networks in time, which is denominated as a concept named evolving complex networks [6].

In this study, we aimed to investigate the connectivity properties of co-authorship network of Turkish engineering research society. Data is retrieved from Web of Science Core Collection, a global repository for bibliographic data concerning high impact indexing services like Science Citation Index, Social Sciences Citation Index, Arts & Humanities Citation Index etc. The study is detailed in Data and Results section outlining the prominent properties of data and results of complex network analysis. The most central authors together with measures of network analysis are also presented in this section.

II. DATA AND RESULTS

Data retrieved from Web of Science Core Collection is limited to publications addressed from Turkey in engineering field, between years 1975 and 2018. This search resulted a massive dataset consisting of 67248 publications those are authored by 78883 distinct authors. These node set defined by authors are wired with 194232 edges, each edge defined by a co-authorship performed for a single paper. As an example, a paper co-authored by authors A, B and C result undirected edges between authors A-B, A-C and B-C.

A. Average Degree

Degree of an author is simply the number of co-authors during his/her academic life. This parameter may be studied both unweighted (single link to a co-author) or weighted (numerous links for a repeating co-authorship). The network with mentioned properties yield an unweighted average degree of 4.925, which increases to 6.687 for weighted analysis. These results indicate that each author approximately has 5 co-authors for this network.

B. Clustering Coefficient

Clustering is the tendency of forming clusters between nodes in a network. It is measured with the clustering coefficient, as a ratio of links between the neighbors of a single node, to the number of all possible links. For a node i with degree k_i the clustering coefficient is defined as:

$$C_i = \frac{2L_i}{k_i(k_i-1)} \quad (1)$$

Network exhibits an invincible clustering coefficient of 0.803, indicating the high tendency of forming cliques between author groups.

C. Average Separation

Another indicator if connectivity, average path length (also known as average separation) is the average of the shortest paths between all pairs of nodes. A path is a route that runs along the links of the network [1]. Real networks yield low separation, which is explained with the small-world-ness phenomenon indicating that there exists short paths between most node pairs in the network [7]. Average separation for our network is 17.803, a moderate value compared to similar real-world studies [8].

D. Degree Distribution

The degree distribution, p_k, demonstrates the probability of a randomly selected node to have degree k. It is a measure assumed to be in the core of network based calculations [1]. The form of degree distribution is an indicator of how information, viruses, rumors spread in a network, and how robust is a network against attacks resulting node and edge removal [9].

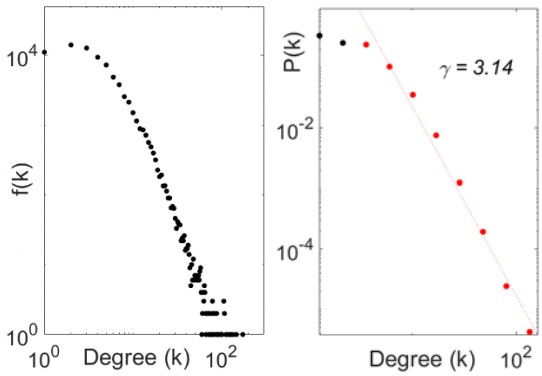


Figure 1: Degree frequencies (left) and normalized distribution (right) plots for the co-authorship network. Least-squared (LS) fitting is also presented for the tail of distribution in the right panel, resulting a power-law coefficient of γ=3.14.

The degree distribution for the network is presented in Fig.1 with scattered view of degree frequencies in the left panel and normalized and power-law fitted view in the right panel. The tail of degree distribution is power-law consistent with a coefficient of γ=3.14, after least squared fitting performed.

E. Centrality

Centrality measures, providing information on how a node is central by the means of network structure, make implicit assumptions about the manner of information flow through a network. Numerous measures have been developed, including degree centrality, closeness, betweenness, eigenvector centrality etc. [10]. The most common centrality measures are described below [11]:

i) Closeness centrality, C^C, is a measure defining how close a node i is to the all other nodes following the shortest paths. It is defined with equation:

$$C_i^C = \frac{N-1}{\sum_{j\in G; j\neq i} d_{ij}} \quad (1)$$

where N is the node count and d_{ij} is the shortest path length between nodes i and j in a network [12, 13].

ii) Betweenness Centrality, C^B is a measure of how a node takes role on the flow of information between other node pairs. It is defined as:

$$C_i^B = \frac{1}{(N-1)(N-2)} \sum_{j,k\in G, j\neq k\neq i} n_{jk}(i)/n_{jk} \quad (2)$$

where n_{jk} is the number of shortest paths between nodes j and k, and $n_{jk}(i)$ is the number of shortest paths between nodes j and k which contain node i [12, 13].

iii) Eigenvector Centrality, C^E characterizes the global prominence of a node in a graph. Thus, it takes into account the entire connectivity pattern in the network. Let A be the adjacency matrix for this graph; aij =1 if vertices i and j are connected by an edge and 0 if not.

$$Ax = \lambda x, \qquad \lambda x_i = \sum_{j=1}^{n} a_{ij} x_j, \qquad i = 1, \dots, n \quad (3)$$

The centrality of a vertex is proportional to the sum of the centralities of the vertices to which it is connected. λ is the largest eigenvalue of A and n is the number of vertices. [14]

We present top 20 authors in our network for these three centrality measures in Tables 1-3, together with the similar list presented in Table 4 with respect to node degrees. Records of some authors include only capital letters for names in Web of Science database, therefore they are listed in this way. We also identified the matching authors in more than one table and marked them in bold text in Tables 1-4.

Table 1: Top 20 authors according to normalized closeness centrality.

Surname, Name	Closeness Cent.
Alpat, Behcet	1.0
Akgun, Ugur	1.0
Albayrak, Elif A.	1.0
Aydin, Gural	1.0
Bilki, Burak	1.0
Cankocak, Kerem	1.0
Duru, Firdevs	1.0
Onel, Yasar	1.0
Ozok, Ferhat	1.0
Sonmez, Nasuf	1.0
Yetkin, Taylan	1.0
Mengi, Anil	1.0
Aytekin, Burcak Gundogdu	1.0
Kurt, Fatih	1.0
Suvaci, Ender	1.0
Bilgen, Mehmet	1.0
Akgiray, Ahmed	1.0
Dikmen, U.	1.0
Cayir, Akin	1.0
Eskiyerli, Murat Hayri	1.0

Table 2: Top 20 authors according to norm. betweenness centrality.

Surname, Name	Betweenness Cent.
Halici, U	0,486
Ulusoy, I	0,4856
Ozdemir, I.	0,346
Aytekin, O.	0,3453
Celik, Erdal	0,2053
Duzgun, H. S. B.	0,1407
Sahin, S.	0,1403
Yesiloglu-Gultekin, N.	0,1099
Ebeoglugil, M. Faruk	0,0951
Kilic, E	0,0942
Yilmaz, Mustafa	**0,0922**
Karahan, O	0,0847
Beksac, S	0,073
Colak, I	0,0638
Gokceoglu, C.	0,0626
Atalay, V	0,0625
Elmas, C	0,0621
Cetin, A. Enis	**0,0556**
Kose, A.	0,0544
Guney, Kerim	0,0539

Table 3: Top 20 authors according to norm. eigenvector centrality.

Surname, Name	Eigenvector Cent.
Undar, Akif	1
Akcevin, Atif	0,9628
Demir, Hilmi Volkan	**0,9205**
Bakir, Ihsan	0,7964
Yilmaz, Mustafa	**0,7534**
Pekkan, Kerem	0,7301
Orhon, Derin	**0,7276**
Alkan-Bozkaya, Tijen	0,7228
Cetin, A. Enis	**0,6984**
Haydin, Sertac	0,6899
Piskin, E	**0,6654**
Uysal, Murat	**0,6286**
Koyuncu, Ismail	**0,6203**
Denizli, Adil	**0,5999**
Dizge, Nadir	0,5869
Urey, Hakan	**0,5814**
Ceyran, Hakan	0,5736
Sasmazel, Ahmet	0,5661
Alatan, A. Aydin	**0,5625**
Ozbay, Ekmel	**0,5433**

Table 4: Top 20 authors according to node degree.

Surname, Name	Degree
Yilmaz, Mustafa	**147**
Demir, Hilmi Volkan	**146**
Uysal, Murat	**134**
Orhon, Derin	**128**
Cetin, A. Enis	125
Denizli, Adil	**124**
Ozbay, Ekmel	**117**
Urey, Hakan	**113**
Kaynak, Okyay	112
Piskin, E	**108**
Imeci, Taha	108
Koyuncu, Ismail	**106**
Soylak, Mustafa	106
Kisi, Ozgur	106
Yakuphanoglu, F.	101
Alatan, A. Aydin	**92**
Kurt, Gunes Karabulut	90
Yuksel, Mithat	89
Dundar, Gunhan	88
Cetin, Mujdat	88

As seen from the tables, the matching patterns are mostly evident between eigenvector centrality and degree tables (Tables 3 and 4). This indicates that these two parameters are in good correlation for our network.

III. CONCLUSION

As a complex interconnected structure, co-authorship network of Turkish engineering research society exhibits similar characteristics with recent studies concerning scientific collaboration networks. It has a power-law consistent degree distribution, with a degree coefficient just below 3. Each author has ~5 co-authors in average, while this parameter increases towards ~7 for weighted analysis. Average path, calculated close to 18, does not mimic the small-world property that requires lower values around 3-6 for real networks. We conclude this is mostly governed by the time dependent characteristics of this network, avoiding co-authorship links between authors active in different time windows.

Most central authors are presented as a result of this study, together with authors having greatest degrees. We outline that among these lists, two tables concerning eigenvector centrality and node degrees are in better agreement, having more common authors. Inspiring from these matching characteristics, we can conclude that investigating the correlations between all centrality measures and degree sequences is worth being the subject of another study.

REFERENCES

[1] A. L. Barabási, *Network Science*. Cambridge: Cambridge University Press, 2016.

[2] T. T. Yuen and T. A. Pickering, "Investigation of Social Networks and Discussions in STEM Education Communities on Twitter," in *Learning and Teaching in Computing and Engineering (LaTiCE), 2015 International Conference on*, 2015, pp. 128-133.

[3] A. Cavusoglu and I. Turker, "Scientific collaboration network of Turkey," *Chaos Solitons & Fractals*, vol. 57, pp. 9-18, Dec 2013.

[4] A. Ferligoj, L. Kronegger, F. Mali, T. A. B. Snijders, and P. Doreian, "Scientific collaboration dynamics in a national scientific system," *Scientometrics*, vol. 104, pp. 985-1012, Sep 2015.

[5] M. Perc, "Growth and structure of Slovenia's scientific collaboration network," *Journal of Informetrics*, vol. 4, pp. 475-482, Oct 2010.

[6] M. H. Li, J. S. Wu, D. H. Wang, T. Zhou, Z. R. Di, and Y. Fan, "Evolving model of weighted networks inspired by scientific collaboration networks," *Physica a-Statistical Mechanics and Its Applications*, vol. 375, pp. 355-364, Feb 2007.

[7] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47-97, Jan 2002.

[8] İ. Türker, "Generating clustered scale-free networks using Poisson based localization of edges," *Physica A: Statistical Mechanics and its Applications*, vol. 497, pp. 72-85, 2018/05/01/ 2018.

[9] W. Wang, M. Tang, H.-F. Zhang, H. Gao, Y. Do, and Z.-H. Liu, "Epidemic spreading on complex networks with general degree and weight distributions," *Physical Review E*, vol. 90, p. 042803, 10/06/ 2014.

[10] S. P. Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, pp. 55-71, 2005/01/01/ 2005.

[11] İ. Türker, "Evaluation of the Turkish Highway Network Analysis With Traffic Data," *Balkan Journal of Electrical and Computer Engineering*, vol. 6, pp. 13-19, 2018/02/01/ 2018.

[12] S. Wasserman and K. Faust, *Social network analysis: Methods and applications* vol. 8: Cambridge university press, 1994.

[13] P. Crucitti, V. Latora, and S. Porta, "Centrality measures in spatial networks of urban streets," *Physical Review E*, vol. 73, p. 036125, 2006.

[14] P. Bonacich, "Some unique properties of eigenvector centrality," *Social networks*, vol. 29, pp. 555-564, 2007.

The Effects of Ransom Software on IoT (Internet of Things) Systems

S. GENCAY[1], Y. CELIK[1]

[1]Karabük University, Computer Engineering Department, Karabük, semihgencay@ogrenci.karabuk.edu.tr

[1]Karabük University, Computer Engineering Department, Karabük, yukselcelik@karabuk.edu.tr

Abstract: ***In recent years, ransomware have become one of the most important threats facing both individuals and organizations. Attackers use strong encryption methods to create dangerous and comprehensive malware, making their ransom software flawless. Surveys indicate an increase in the number of assailants in parallel with the increase in the number of victims and the increase in illegal income. In the early years, especially at the risk of individual ransom viruses, more complex attacks are now beginning to appear that lead to the encryption of multiple machines targeting companies and every device connected to the Internet. In this study, ransom virus attacks against Internet of Things (IOT) network which is formed by connecting multiple devices with each other are examined and a study is presented about the measures to be taken.***

Keywords: ***Ransomware, Internet of Things, encryption, decrypting passwords, preventive measures, threats, security.***

I. INTRODUCTION

Ransom viruses are malicious software that has been shown to destroy personal, industrial or government network-related resources by encrypting them [1]. The cybercriminal smartphone, which continuously develops its methods and software, aims to use all the smart devices connected to the network after the networked computers and tablets as a ransom collection tool [2]. However, the current trends in technology, such as the proliferation of internete connected devices, the perception of personal data of users in social media, and the spread of cryptographic currency units, allow hackers to easily infiltrate devices.

Attacks on systems using ransom software are increasing day by day and getting more complicated. Individuals or groups using such harmfulness are developing methods and measures against harms that they meet or meet daily.

As more and more objects are connected to the Internet, security concerns with individual, institutional and public institutions are increasing. For example, even today, some automobiles that use advanced technology can be affected by attacks. These virtual attacks can lead to loss of life, as well as to the theft of automobiles. Unlike traditional virus software threats, ransom attacks on IOT can be more devastating than just security services that can lead to significant information breaches, such as confidentiality, integrity and availability, as well as causing financial losses [3].

Individual fraud in IoT devices is difficult because attackers need to determine the right owners of the devices they are requesting for the ransom money. In addition, most IOT devices are controlled / managed by other devices; Therefore, the attacker needs to find the device that controls the system to leak and encrypt the data. Successful attacks are only possible if the attacker knows the topological settings of the IOT network.

A typical IOT setup includes heterogeneous devices with embedded sensors connected together over a network, as shown in Figure 1. IOT devices often have low power, small memory and limited processing capacity. Gateways are used to connect IoT devices to the outside world and to provide data and services to IoT users remotely [4].

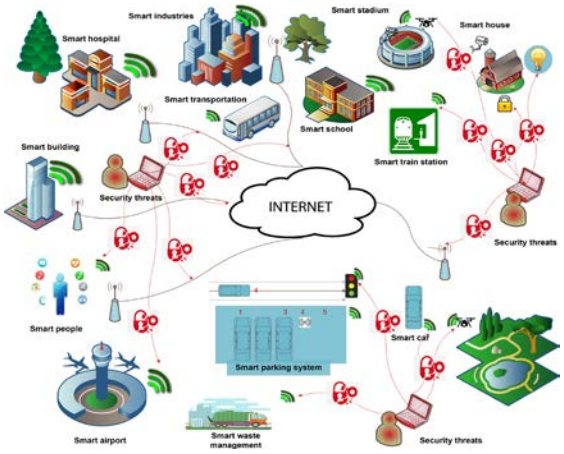


Figure 1. A Sample IOT and Security Release Zones

A.Types of Ransom Software

There are two basic types of modern ransom software, the hybrid randy software that combines locker and crypto together.

The contamination of both types of malware is similar. A user may infect various types of ransom viruses with his machine. With downloads from unsafe sites, spam, or different forms of social engineering, users are exposed to infection. However, the damage after infection is different for both types. Locker software blocks access to a user on an infected machine, but usually leaves the underlying system and the file untouched. They can change the functionality of devices in IoT systems, restrict user access, disable user interfaces, disable sensors in the system, and cause Denial of Service (DoS) attacks to reduce device performance. Locker ransom attacks are usually initiated on front-end IoT devices to prevent user access.

On the other hand, crypto ransom software is a type of data encryption that prevents people from accessing vital files or data (e.g., documents, images, videos, etc.) using encryption methods. Therefore, the files that have been attacked will not work until a ransom is paid and the decryption key is obtained. Crypto ransom software targets IOT systems where mainstream data are stored rather than front-end devices.

Hybrid ransom attacks, combining encryption and locking mechanisms, are more dangerous than device data and functionality can be compromised. A hybrid ransom attack can be more dangerous than it can target front-end and back-end IOT devices and systems. Technically, it is difficult to launch hybrid ransom attacks due to device heterogeneity, ownership and multi-tiered distribution of IOT systems, but such attacks can easily paralyze an entire IOT network, including front-end devices and back-end application servers.

B.Transmission Methods of Ransom Viruses to Iot Systems

C.Content distribution network (CDN) and malicious ads

If malware is buried in multimedia and internet traffic, large amounts of ransom can be distributed [5]. Attackers can block CDN (Content Distribution Network) traffic on back-end edge networks and front-end IOT devices. Ransom software can keep CDN traffic using back-end cache servers and internal memory of IoT devices. Attackers can also capture IoT device users through malicious advertisements; where the material posted via the CDN appears legitimate, and malware is installed on the device when users click on the ad.

D.Botnets and Downloaders

Ransom software can also use botnets that roam silently over IoT networks. Attackers can use phish-e-mails, asking users to download the upcoming files or click on certain links. When the botnet is enabled as a result of the actions of users or devices in response to a phishing e-mail, the entire IoT network is seized. A botnet can also be a tool for self-propagating ransom software that can cause DoS attacks on an IOT network [6]. When a device / network security is compromised, the device / related information can be sold to other ransom software and botnet operators. Therefore, data and devices are constantly threatened even after you remove the ransom software from the network.

E.Social Engineering

Ransom software uses social engineering tactics to allow attackers to portray themselves as legal authorities. It is an easy tool to enter user systems and collect user information. However, IOT devices often do not provide direct interaction with external users. In this case, ransom software can be launched by external users, presenting themselves as legitimate users / devices on the IOT network.

F.Fancy Software Services

Given that IoT devices are largely dependent on application services and cloud data centers, attackers can block device cloud traffic and inject ransom software. At the end, the device may see the ransom software as a subscribed service. However, when IOT uses virus services, the entire IOT network is also threatened.

G.Latest Technology Surveys on IOT Security

Cisco Inc. IOTs will have up to 50 billion connected devices by 2020 [7]. With the emergence of smart homes, smart cities and smart devices, the Internet of Objects (IOT) is emerging as a potential growth area. The storage and networking capacity of IoT devices is limited and is therefore more vulnerable to attack than other endpoint devices such as smartphones, tablets or computers. For this reason, in addition to the protocols used for networking, communication, and management in network security issues, IOT layered architects have been reviewed and categorized [8].

It is also a major problem to investigate the security and confidentiality risks and produce solutions in intelligent home systems where many systems operate as integrated. Survey research aimed at minimizing these risks and raising the awareness of users has led to the need to pay particular attention to the safety of wireless sensor networks [9]. At the

same time, a new multi-layered cloud architecture model has been developed to solve the security risks and communication problems arising from the fact that different devices used in smart home systems are manufactured by different manufacturers and using different communication methods [10]. In this model, ontology based security service framework is designed to solve heterogeneity problems better. Another work focuses on user-centered IOT systems used in smart homes. Foreseeing cyber threats for these systems and suggesting a focal point for digital researchers. The designed system is more heterogeneous than existing systems and it also includes the ability to decide "networking" [11].

Another technology that has recently been developed is mobile cloud computing technology. The Internet of objects is also a new technology that is growing rapidly in the field of telecommunication. The main objective of the interaction and cooperation between the information transmitted through the wireless networks and the objects is to fulfill the goal to be shown as a unified entity to them. In particular, we have proposed a system that combines the above two technologies to explore common features and explore the benefits of integration and security risks, and the contribution of the mobile cloud system to IOT has been explored [12].

Another method of securing IOTs is access control. This method includes integrity, authenticity, irresolution, identification, verification and authorization. Management control approaches are divided into two categories: symmetric and asymmetric protocols [13], which are examined by considering evaluation criteria such as authentication, extensibility, flexibility, scalability and privacy freedom.

A four-layer security architecture has been developed to detect and retrieve information from devices connected to IOT systems. The network layer supporting the connection between these architectural wireless or wired devices consists of a service layer that provides and manages the services required by users or applications, an application interface layer consisting of interaction methods between users and applications, and a detection layer integrated with the IOT end-nodes [14].

In networking technologies, the view that hardware-centric approach is a software-based model transformation in the recent period is accepted by a large segment. In the same way, software-based approaches have gained importance in IOT systems where network security is very important. In this study, which describes this approach based on security architecture concept and software defined networks (SDN), the work of the so-called SDN-Domain architecture is described and SDN is aimed to achieve more efficient and flexible network security [15].

The development of IOT systems has also become widespread in the healthcare sector. However, the safety of the devices used in the health system is of great importance as it is directly related to the safety of human life. Attacks in this network can cause disruptions that can put many patients' lives at risk. This is why it is tempting for ransom software. This field can lead them to this field with the thought that they can earn high income with a chaos they can bring to the fair. E-health systems and their applications, which are defined as wireless networks connecting healthcare devices in the work on the security and safety of IOT-based health services, are covered. Security issues specific to e-health applications are covered, including personal health information and the specific sensitivity of privacy. It focuses on possible types of attacks that can occur at the network and transport floor of the OSI model, including denial-of-service (DoS) attacks that can slow down the system, unauthorized attacks, Sybil attacks, and out-of-sync attacks. Strategies such as the use of cryptographic codes, prevention of attacks, random numbers, sequence numbers and time stamps have been proposed and their performance has been analyzed [16].

IOT systems have been expanding in recent years, with advances in technology and devices. This expansion has caused many different applications in medical, civil, maritime, military and local areas. Each of these areas has different requirements and challenges. But there is a common problem in all of them is data security. Data security is an important element for any IOT network, but simple data security may not be enough for modern IOT systems. There is a need for a secure, end-to-end IOT system solution that allows wireless sensors to connect to any computer in the world while ensuring data and network security. The system proposed in response to this need can protect an IOT solution against various attacks such as data breaches, Denial of Service (DoS) and unauthorized access [17].

IOT provides interactive, common, connected, and intelligent nodes in an autonomous fashion, while providing all kinds of services. The wide distribution of IOT objects, the clarity and relatively high processing power make them an ideal target for cyber attacks. Moreover, many IOT nodes collect and process special information. This makes the systems a data mine for malware. For this reason, the ability to detect security and especially the attacked nodes arises in the successful installation of IOT networks, with the collection and protection of evidence of attack or malicious activity [18].

Security issues in the messaging system, which allows the devices in the system to negotiate with each other, still pose a major threat. To solve this problem, a secure Message Queue Telemetry Transport (MQTT) mechanism called AUPS (Autenticated Publish Subscribe) has been developed [19]. The mechanism has developed MQTT, a popular communications protocol in the IOT system, by providing a secure broadcast / subscription system within the protocol. The proposed mechanism introduces a key management framework and introduces new policies, thus providing flexible control of the flow of information in IOT systems with MQTT support. The proposed system is released as an open source under an Apache v.2 license. The system needs to be tested in a larger and more complex environment where problems with synchronizing policies between hosts can arise. Another study has focused on providing secure communication between IOT devices [20]. As a result, these devices have become potential targets for conventional Internet attacks. In order to deal with these problems, this work presents an architecture that allows IoT devices to use DTLS with mutual authentication mechanism.

In existing health systems, dissemination of IOT practices offers improved and useful health services for doctors and patients; because they are useful for a large number of medical fields with the help of Body Sensor Network (BSN) technology. However, the lack of security and privacy solutions is hampering the acceptance of the BSN. A secure IoT based health method, called BSN-Care, has been proposed via the BSN [21]. The proposed scheme is primarily designed to remove mutual authentication and forgery attacks from the middle.

It seems difficult to provide full safety of IOT systems. However, they can be secured with fast, efficient and fully automated software that recognizes known and unknown ransom software. This has already been suggested, and an approach called HELDROID has been proposed. This approach is based on the identification of "building blocks" that are necessary to prevent ransom software implementation [22].

1. Conclusion

Smart improvements in smart technologies have opened a new system called IoT. This study addresses ransom attacks and security concerns in IOT systems. First, it discusses the rise of ransom attacks and outlines the outlines. Secondly, the types of ransom viruses are mentioned. Thirdly, the methods of transmission of ransom viruses to IoT systems have been examined. Fourth, several reliable case studies have been presented to warn people about the weakness of IOT devices against threats. Fifthly, the latest research on IoT safety has been searched by classifying the literature. Finally, although IOT can facilitate different aspects of human life, most IOT devices are vulnerable to ransom attacks. For this reason, the importance of strengthening IOT security and reducing ransom attacks must be emphasized and IOT must be trusted by the user.

REFERENCES

[1] K. Murnane, "The malware bytes report: The 2016 malware threatlandscape", 2017

[2] https://www.eset.com/tr/about/newsroom/basin-bultenleri/araciniz-sizden-fidye-isterse/

[3] E. Bertino, N. Islam, "Botnets and internet of things security", Computer (2017) 76–79

[4] I. Yaqoob, E. Ahmed, M. H. ur Rehman, A. I. A. Ahmed, M. A. Al-garadi, M. Imran, M. Guizani, "The Rise of Ransomware and Emerging Security Challenges in the Internet of Things", 2017, Pages 444-458

[5] K. Cabaj, M. Gregorczyk, W. Mazurczyk, "Software-Defined Networking-based Crypto Ransomware Detection Using HTTP Traffic Characteristics ", 2016

[6] S. M. Cheng, P. Y. Chen, C.-C. Lin, H.-C. Hsiao, "Traffic-aware Patching for Cyber Security in Mobile IoT", 2017

[7] https://www.cisco.com/c/tr_tr/index.html

[8] A. K. Minhaj, S. Khaled, "IoT security: Review, blockchain solutions, and open challenges", 2017

[9] G. Cerullo, G. Mazzeo, G. Papale, B. Ragucci, L. Sgaglione, "IoT and Sensor Networks Security", 2016, Pages 77-101

[10] M. Tao, J. Zuo, Z. Liu, A. Castiglione, F. Palmieri, "Multi-layer cloud architectural model and ontology-based security service framework for IoT-based smart homes", 2017, Pages 1040-1051

[11] N. Akatyev, J. I. James, "Evidence identification in IoT networks based on threat assessment", 2017

[12] C. Stergiou, E. Kostas, Psannis, B. Kim, B. Gupta, "Secure integration of IoT and Cloud Computing" 2016, Pages 964-975

[13] I. Romdhani, "Securing the Internet of Things", 2017, Pages 119-130

[14] S. Li, "Security Requirements in IoT Architecture", 2017, Pages 97-108

[15] F. Olivier, G. Carlos, N. Florent, "New Security Architecture for IoT Network", 2015, Pages 1028-1033

[16] I. Romdhani, "Securing the Internet of Things", 2017, Pages 133–139

[17] A. Mathur, T. Newea, W. Elgenaidi, M. Raoa, G. Doolya, D. Toal, "Sensors and Actuators A: Physical", 2017, Pages 291-299

[18] M. Conti, A. Dehghantanha, K. Franke, S. Watson, "Future Generation Computer Systems", 2017, Pages 544-546

[19] A. Rizzardi, S. Sicari, D. Miorandi, A. C. Porisini, "Information Systems", 2016, Pages 29-41

[20] G. L. Santos, G. C. Rodrigues, L. Z. Granville, L. M. R. Tarouco, "A dtls-based Security Architecture for the İnternet of Things", 2015, Pages 809-815

[21] P. Gope, T. Hwang, "A Secure IoT-Based Modern Healthcare System Using Body Sensor Network", 2016, Pages 1368-1376

[22] A. Kharraz, W. Robertson, D. Balzarotti, L. Bilge, E. Kirda, "A Look Under the Hood of Ransomware Attacks", 2015, Pages 3-15

A Fast and Simple Computer Aided Lighting Simulator

B.AKGÜL[1] , A.BİLİCİ[2], H. KUTUCU[3]

[1]Bartin University, Bartin/Turkey, bayramakgul@bartin.edu.tr
[2]Bartin University, Bartin/Turkey, abilici@bartin.edu.tr
[3]Karabuk University, Karabuk/Turkey, hakankutucu@karabuk.edu.tr

***Abstract* - In any working environment, to perform visual tasks efficiently and accurately, adequate and appropriate lighting should be provided. The quantity and quality of illumination in any work place control by the type and duration of activity. The Illumination can be provided by daylight and artificial light together or separately. To calculate the illumination intensity at point C on the R plane from a light source at point A where the light distribution in space is known; besides the direct rays from point A to point C, the rays reflected from other planes must also be considered. By using this information, for inner work area a computer-aided modeling application has been developed by developing an algorithm that recommends the number and location of the selected luminaire considering the minimum illumination level recommendation specified in EN 12464-1 standard (which is European Standard that specifies lighting requirements for indoor work places). In this study, it was aimed to provide visual comfort in closed areas by focusing on only artificial light illumination.**

***Keywords* - Illumination, light intensity, eulumdat, reflection, computer aided modeling**

I. INTRODUCTION

Computers certainly have become indispensable tools in many fields because of the conveniences they provide to users. Computer-aided modeling has played an important role in testing and monitoring to prevent possible damages. In order to provide the desired level of illumination in the indoor workspaces, an application was developed in a simulator environment instead of physically mounting and testing lighting luminaries. There are commercial applications in the industry such as Dialux [1] and Relux [2], but these applications do not recommend luminaires which are appropriate for the indoor working areas.

The International Energy Agency (IEA) reported in 2015 that more than 3000 TWh of the global electricity production was used for lighting purposes [3]. In case of necessity of comfort, many simulation studies have been done on lighting in terms of energy saving. MATLAB, Simulink, MATLAB / GUI, EnergyPlus, DOE, JAVA, LabVIEW, DaySim and other simulation tools are the most popular simulation tools for indoor illumination [4].

In this study, a simulation was performed using OpenGL in .Net Framework. In the simulation, the type of the workspace can be selected. The location and the number of the suitable luminaires for the selected workspace are provided according to the minimum illumination level specified in EN 12464-1 standard, then luminaires are placed in the appropriate places in the workspace. The provided illumination levels according to the performed operations are reported to the user.

In this study, for any luminaires with known light distribution in space, the lighting intensity is evaluated on a desired plane in a three-dimensional closed area. The study consists mainly of two parts: 1) location/luminaire database and determining the type of luminaires for the location 2) determining the number of luminaires and lighting intensity.

II. INDOOR AREAS & LUMINAIRES DATABASE

The database consists of two parts. In the first part of the database, there are many information about the lighting solutions such as the quantity and the quality of illumination for most indoor work places and their related areas. In addition, some recommendations are given for good lighting practice. This data is taken from EN 12464-1 [5].

The second part of the database covers "eulumdat" information about luminaires. EULUMDAT is a data file format which is used to determine photometric light distribution data from light sources such as lamps and luminaires. In the "eulumdat" file, with the light distribution values of luminaire, there are manufacturer information, dimensions, luminous size, power, lumen, color temperature, lamp type and number of lamps are given [6].

Producers create the "eulumdat" file of each luminaire in their laboratories. The data in an EULUMDAT file is usually measured using a goniophotometer. A goniophotometer is a device to measure the light emitted from an object at different angles [7]. Visualization of light distribution values of a luminaire in an eulumdat file in 2D is shown in Figure 1 and in 3D is shown in Figure 2. Also the dimensions of luminaire are visualized in Figure 1. For our simulation, addition to the "eulumdat" data, a luminaire image is also added to the database as shown in Figure 1.

A look-up table was created to determine the proper type of luminaire for the indoor work area. In this table, there is information about which luminaire should be used in which indoor work area. In the simulation environment, the recommendation of luminaire is made according to this table. After the work area is determined, the first proposed luminaire is automatically selected to use. if necessary, the user can choose another luminaire.

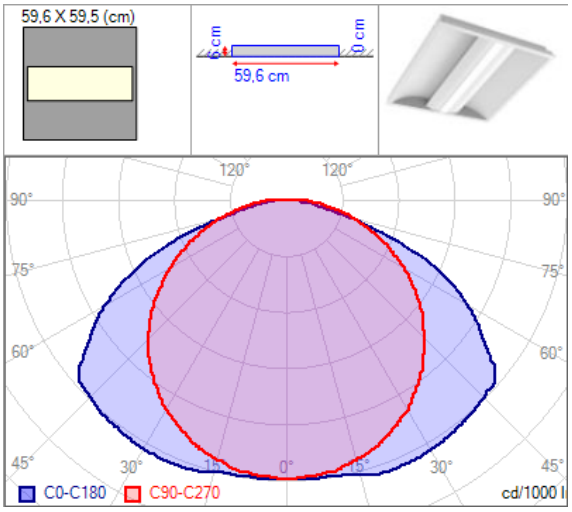


Figure 1: Light distribution values in 2D & dimensions and image of a luminaire.

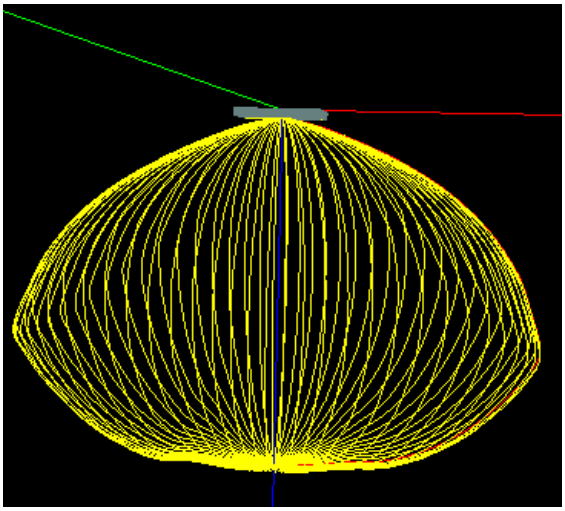


Figure 2: Light distribution values in 3D.

III. LIGHTNING CALCULATION

In order to calculate the illumination intensity at a point *C* on the *R* plane from a light source at a point *A*, we use equation (1).

$$E = \frac{I \cos\theta}{r^2}, \tag{1}$$

where I refers to the intensity of light, r is the Euclidean distance between the points A and C and θ represents the angle between light ray and normal of the R plane. For each luminaire, the light intensities $I_{\alpha,\theta}$ at any α and υ according to the polar coordinate system are given in "eulumdat" file. The angles α and υ are shown in Figure 3. Visualization of all I values is shown in Figure 2. The figure is used as the light distribution of luminaire in the space.

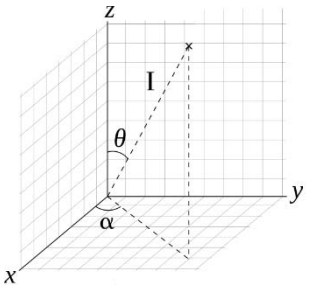


Figure 3: The light intensity $I_{\alpha,\theta}$ at angles α and θ.

The Euclidean distance r between points A and C is calculated as in equation (2).

$$r = \left\|\overrightarrow{AC}\right\| = \sqrt{(C_x - A_x)^2 + (C_y - A_y)^2 + (C_z - A_z)^2} \tag{2}$$

Equations (3) and (4) are used to find the angles α and υ, respectively.

$$\alpha = \arctan\left[\frac{(C_y - A_y)^2}{(C_x - A_x)^2}\right] \tag{3}$$

$$\theta = \arccos\left[\frac{C_z - A_z}{r}\right] \tag{4}$$

After finding α and υ, the $I_{\alpha,\theta}$ is obtained from "eulumdat" file after some interpolation operations. In this case, the lighting intensity calculation at a point C is carried out using equation (5).

$$E = \frac{I_{\alpha,\theta} \cos\theta}{r^2} \tag{5}$$

This calculation is related to direct illumination with no reflections. Lighting calculations at outside areas (road or park) can be done using only this equation [8].

But at indoor areas, reflections also should be used for illumination. In this case, as shown in Figure 4, it is necessary to find angles α and υ for a ray exiting from a point A, reflected at a point B on a plane P and reach to a point C at a plane R. Besides, the ray's total path $\sum r = r_{\overline{AB}} + r_{\overline{BC}}$ should be calculated.

As it can be seen in Figure 4, a ray which is exiting from a point $A(A_x, A_y, A_z)$ reflect at a point $B(B_x, B_y, B_z)$ which is on a plane P (corner points P_1, P_2, P_3 and P_4) and reach to a point $C(C_x, C_y, C_z)$ which is on a plane R (corner points R_1, R_2, R_3 and R_4). The angle between ray and the normal vector of the plane P is ϕ.

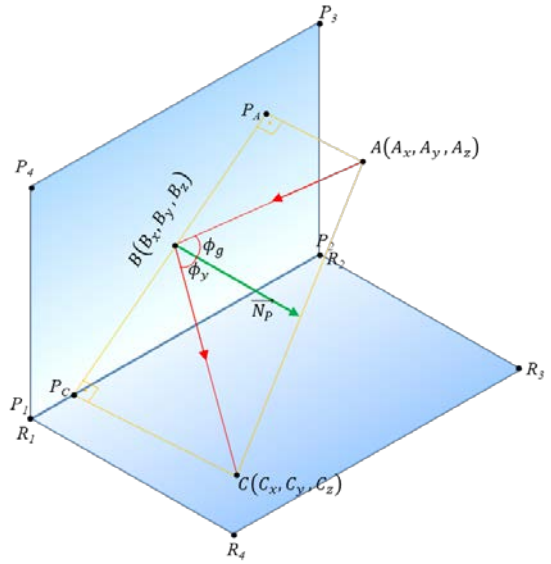


Figure 4: The path of a ray exiting from a point A, reflected at a point B and reaching to a point C.

To find point $B(B_x, B_y, B_z)$, firstly we must find P_A and P_C which are projection points of A and C on plane P. The scalar multiplication of any two vectors lying on two perpendicular planes is equal to 0. Therefore, we can write triple equations (6) to find P_A and triple equations (7) to find P_C according to vector lying on right trapezoid P_AP_CCA. ($\overrightarrow{N_P}$ is the normal vector of plane P)

$$\begin{aligned} \overrightarrow{P_AA}.\overrightarrow{P_1P_2} &= 0 \\ \overrightarrow{P_AA}.\overrightarrow{P_2P_3} &= 0 \\ \overrightarrow{P_AP_1}.\overrightarrow{N_P} &= 0 \end{aligned} \tag{6}$$

$$\begin{aligned} \overrightarrow{P_CC}.\overrightarrow{P_1P_2} &= 0 \\ \overrightarrow{P_CC}.\overrightarrow{P_2P_3} &= 0 \\ \overrightarrow{P_CP_1}.\overrightarrow{N_P} &= 0 \end{aligned} \tag{7}$$

Now, because of incoming and reflected ray angles (ϕ_g, ϕ_y), as seen in Figure 4, are equal to each other, by using $(\overset{\Delta}{P_C}CB)$ and $(\overset{\Delta}{P_A}AB)$ triangle similarities, we can write equation (8) to find point $B(B_x, B_y, B_z)$.

$$\frac{|P_AA|}{|CP_C|} = \frac{B_X - P_{AX}}{P_{CX} - B_X} = \frac{B_Y - P_{AY}}{P_{CY} - B_Y} = \frac{B_Z - P_{AZ}}{P_{CZ} - B_Z} \tag{8}$$

To find the ray's total path $\sum r = r_{\overline{AB}} + r_{\overline{BC}}$, we can use the equations (9) and (10).

$$r_{\overline{AB}} = \left\|\overrightarrow{AB}\right\| = \sqrt{(B_x - A_x)^2 + (B_y - A_y)^2 + (B_z - A_z)^2} \tag{9}$$

$$r_{\overline{BC}} = \left\|\overrightarrow{BC}\right\| = \sqrt{(C_x - B_x)^2 + (C_y - B_y)^2 + (C_z - B_z)^2} \tag{10}$$

Now, we can find angles α and υ by using equation (11) and (12).

$$\alpha = \alpha_{\overline{AB}} = \arctan\left[\frac{(B_y - A_y)^2}{(B_x - A_x)^2}\right] \tag{11}$$

$$\theta = \theta_{\overline{AB}} = \arccos\left[\frac{B_z - A_z}{r_{\overline{AB}}}\right] \tag{12}$$

The intensity of the light will decrease in a certain proportion (ρ) while reflecting. After the reflection, the illumination at C will be as in equation (13).

$$E_C = \frac{(\rho I_{\alpha,\theta})\cos\theta}{(\sum r)^2} \tag{13}$$

The reflection coefficient (ρ) differs for each colors and materials [9,10]. Reflection coefficient in default calculations; 0.8 for ceilings (light white), 0.5 for the walls (gypsum plaster) and 0.3 for the floors(dark laminate parquet).

The following steps are performed in order to make a lighting calculation in a working area:

1. *Identify the working area (type and dimensions)*
2. *Select luminaire (using look-up table)*
3. *Create "suggestion list" (quantity of luminaire)*
4. *Number of lighting points = 100 (for suggestion)*
5. *Determine points for lighting calculation ($E_{i,j}$)*
6. *Select a suggestion from "suggestion list" (using binary search)*
7. *Calculate the illumination according to equation 4 at each point (with no reflection)*
8. *If the desired lighting is not provided, go to step 6.*
9. *Number of lighting dots = 1000 // this value can be increased for more precision*
10. *Calculate the reflection from each plane for each luminaire according to equation 13 at the determined points*
11. *Calculate average, minimum, maximum lighting levels*

A. Determining points for lighting calculations

While the lighting calculations is made, the area is divided by the number of points previously determined by the user. while suggestion searching for number of luminaires 100 points is selected. After the number and locations of luminaire is determined, approximately 1000 points uses for actual calculation is performed.

We assume z = 0 at the floor and z = room height (generally 270 cm) at the ceiling. In working area, the lighting height is generally selected as z=85 cm (or table height). The z value of luminaire point is determined according to the use type of luminaire. For example, if luminaire is a "built-in" type, then z=room height. If luminaire is surface-mounted, then z = (room height)-(luminaire height). When determined the lighting calculation points, the z value of all points are set to working height (generally 85 cm).

The following algorithm is used to select the coordinates of the points which will be made lighting calculations.

Finding lighting calculation point-coordinates ($C_{i,j}$)

$L_x = x_{max} - x_{min}$ // x-len of room
$L_y = y_{max} - y_{min}$ // y-len of room
$A = L_x * L_y$ // area of room
$d = (A/np)^{1/2}$ // number of points (by user)
$n_x = (L_x/d)$ // number of points on x axis
$n_y = (L_y/d)$ // number of points on y axis
$d_x = L_x/n_x$ // distance between points on x axis
$d_y = L_y/n_y$ // distance between points on y axis
$x = x_{min} + d_x/2$
$y = y_{min} + d_y/2$
$i=0, j=0$
while($x < x_{max}$)
  while($y < y_{max}$)
    $C_{i,j}$ = point(x,y)
    $y = y + d_y$
    $j = j + 1$
  end
  $x = x + d_x$
  $i = i + 1$
end

B. Construction Suggestion List and Selection

Along with the type of luminaire selected, the room dimensions and the color of the walls (because of reflections) are directly related to how many should be selected for the desired minimum light intensity. In this part, a list is created for the recommendations of how many luminaires can be used. This list is called "suggestion list". To create "suggestion list", it is first calculated how many luminaires can fit horizontally and vertically relative to their sizes. Then, the "suggestion list" is created according to the following algorithm.

Create "Suggestion list"
List = {} // empty list
MaxNumX = L_x/W_{lum} // W_{lum} is weight of luminaire
MaxNumY = L_y/H_{lum} // H_{lum} is height of luminaire
for i=1 to MaxNumX
  r1 = L_x/i // distance between lums. at x axis
  for j=1 to MaxNumY
    r2 = L_y/j // distance between lums. at y axis
    r = max(r1,r2)/min(r1,r2) //ratio between dist on //x and y axis
    if r <= ratio then
      List.Add (Suggestion(i,j,r))
    end-if
  end-for
end-for

The "ratio" is a variable that is defined as "the ratio between horizontal and vertical distance of luminaries" and used to prevent unsuitable luminaire layouts. The "ratio" was chosen as 2.5 according to our experiences. One of the all permutations of 6 luminaries which are 1x6, 2x3, 3x2 and 6x1 may be added into the "suggestion list". The list for such situations needs to be simplified. To simplify the list, the list is sorted by the number of luminaires. For equal cases, the one having minimum distance ratio (r) between the luminaries remains in the list but the others are deleted from list.

Now, we have a "suggestion list" which is sorted in increasing order by the number of luminaries. A suggestion is selected from the list based on the binary search from the middle of the "suggestion list". According to selected suggestion, lighting is calculated at 100 points in a certain tolerance with no reflections. If the calculated illumination intensity is more than desired, then a new suggestion is selected in the middle of the first half, otherwise a new suggestion is selected in the middle of the second half of the list. Repeating this process recursively, a "suggestion" is selected within the determined tolerances. For example, if (2x2) suggestion is selected, then the luminaires are distributed properly as shown in Figure 5. The tolerance is typically between 20% and 25% below the desired level of illumination. It is experienced that 20% to 25% of illumination has been done with reflections.

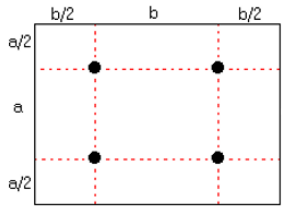


Figure 5: Regular distribution of luminaires.

After a "suggestion" is selected and the luminaries are distributed for appropriate locations, at each determined points the illumination is calculated including reflections for each plane. Finally, to calculate the average illumination intensity, all intensity values at points are added to summation and then divide to the number of points.

IV. Conclusion

In this study, we have developed a simulator environment instead of physically mounting and testing lighting luminaries. In the simulator environment, it is easy and quickly to calculate how many luminaires will be required according to proposed or selected luminaire type for the desired minimum illumination level which specified in EN 12464-1. On the other hand, it is also easy to calculate how much lighting will be provided in the area, according to the number of luminaires set for a room and even the different luminaries selected and placed on the specified locations. In the simulating area, we can visualize the illumination levels in 2D and 3D, then we can print out a report about process. Integration of the EN 12464-1 standard into the simulation environment, it can be considered as an advantage of our application. The interface of the proposed application is shown in Figure 6.

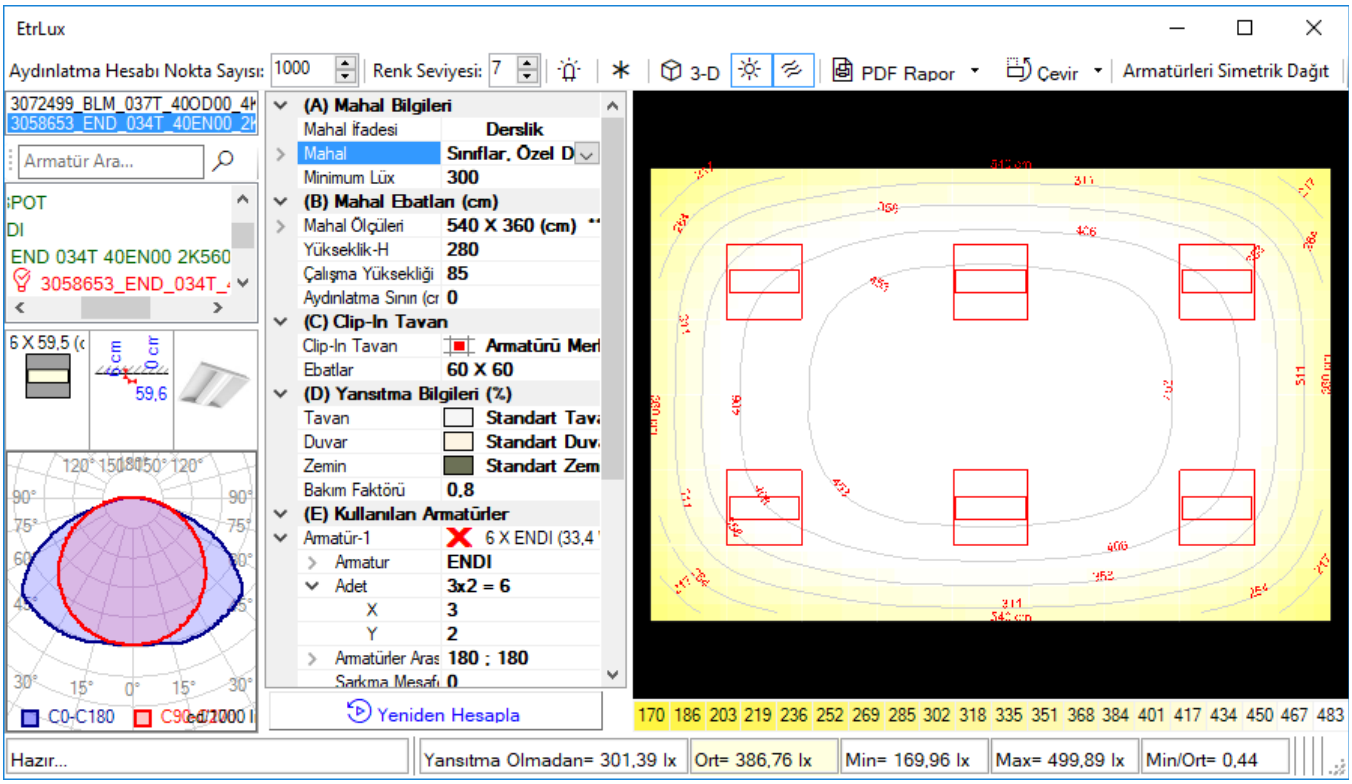


Figure 6: The interface of the lighting similator application.

REFERENCES

[1] Dialux. (2018, April). Available: https://www.dial.de/en/dialux/.

[2] Relux. (2018, April). Available: https://reluxnet.relux.com/en/.

[3] Dubois, M.-C. et al., "Retrofitting the Electric Lighting and Daylighting Systems to Reduce Energy Use in Buildings: A Literature Review", *Energy Research Journal*, 6, 25–41 (2015).

[4] Baloch, A. A. et al., "Simulation tools application for artificial lighting in buildings", *Renewable and Sustainable Energy Reviews*, vol. 82. pp. 3007–3026, 2018.

[5] "EN 12464-1, Light of work places Part1: Indoor work places," 2002.

[6] "EULUMDAT File Format Specification.", [Online]. Available: http://www.helios32.com/Eulumdat.htm. [Accessed: 01-Apr-2018].

[7] Marx, P., "New goniophotometers for lighting engineering laboratories", *Light & Engineering.*, vol. 5, no. 4, pp. 32–36, 1997.

[8] Zhou, H., Pirinccioglu, F. & Hsu, P., "A new roadway lighting measurement system", *Transp. Res. Part C Emerg. Technol.*, vol. 17, no. 3, pp. 274–284, 2009.

[9] Şahin, M., Oğuz, Y. & Büyüktümtürk, F., "İç Mekân Aydınlatmasında Renk Seçiminin Aydınlatma Ekonomisi ve Görselliğe Etkisi", *Electronic Journal of Machine Technologies*, vol. 2013, no. 10, pp. 15–26, 2013.

[10] Dincer, K., Köse, H., Dede, O., Serdar, P. & Tosun, M., "Duvar Renginin Aydınlatmada Güç Performansına Etkisinin İncelenmesi ve Örnek Bir Uygulama", *J. Selcuk.*, vol. 12, no. 2, pp. 25–38, 2013.

A Survey of Uncertainties in MAPE-K Control Loop

S. OUARETH[1], S. BOULEHOUACHE[2], and S. MAZOUZI[3]

[1] 20 Août 1955 University, Skikda /Algeria, s.ouareth@univ-skikda.dz
[2] 20 Août 1955 University, Skikda / Algeria, s.boulehouache@univ-skikda.dz
[3] 20 Août 1955 University, Skikda / Algeria, s.mazouzi@univ-skikda.dz

***Abstract* – Self-Adaptive Systems (SASs) are systems that monitor and adapt their behavior autonomously in response to dynamic state and environmental conditions. A typical architecture of SASs is constituted of a Manager (Autonomic) Sub-System that controls a Managed Sub-System. A well known architecture of the Autonomic Sub-System is the MAPE-K model. It is constituted of the Monitor, the Analysis, the Plan, and the Execution stages and the Knowledge Base. The major challenge of SAS is that all the stages are subject to uncertainty. Consequently, it has a significant impact on the adaptation quality. Currently, uncertainty is considered as a first-class concern in constructing Self-Adaptive Systems. However, few detailed works have been done about uncertainty in MAPE-K Control Loop. This paper intends to survey the most recent research on uncertainty in the MAPE-K using FRAMESELF architecture which is a detailed MAPE-K loop. Precisely, we present the sources of uncertainty in each process of the FRAMESELF model. In addition, we focus on missed sources of that we believe the community should consider.**

***Keywords* – Self-Adaptive Systems, MAPE-K Control Loop, Uncertainty.**

I. INTRODUCTION

Self-adaptive systems [1] (SASs) have the capability to autonomously change its behavior in response to changes in either the system's internal and environment states. Typically, a SAS [2] consists of a Managed Sub-System that implements the system functionalities and a Manager (Autonomic) Sub-System that implements the adaptation logic. Generally, the Manager Sub-System is implemented using the well established MAPE-K model. This last is constituted of a Monitor, an Analyzer, a Planner, and an Executer. The four stages are supported by a Knowledge base that allows the inter-components information exchange [3].

Uncertainty in SAS is considered as a first-class concern in the design, implementation, and deployment of this type of systems [4]. It has a significant impact on a SAS's ability to adapt its behavior and structure. During the last few years, the SAS's research community has made an important effort to describe the sources of uncertainty [5]. They are explored different theories (e.g., probability [6], fuzzy sets and possibility [7]) to mitigate these various types of uncertainty that affect SAS.

According to [5], uncertainty in SAS is due to complexity and loose coupling between the components of such systems. Consequently, the uncertainty in MAPE-K Control Loop is due to complexity and loose coupling between the components of the MAPE-K Control Loop, as well as between the components of this loop and the components of the system.

Actually, the MAPE-K Control Loop is the most popular model to build self-adaptation into software-intensive system [8]. However, few works have been done to study uncertainty in MAPE-K. Moreover, there is no comprehensive work that exhaustively compiles the different sources that affect a MAPE-K Control Loop. The goal of this paper is to survey the sources of uncertainty of the different stages of the MAPE-K Control Loop.

The remainder of this paper is organized as following: Section II introduces a brief overview of the SASs. Section III presents the MAPE-K Control Loop, then the FRAMESELF architecture and the details of its components. Section IV describes the different sources of uncertainty in MAPE stages (Monitor, Analyze, Plan, and Execute). Here, we use the FRAMESELF architecture as a vehicle to precisely illustrate the sources of uncertainty. Section V presents a discussion of this survey. Finally, the paper concludes in section VI with a summary.

II. SELF-ADAPTIVE SYSTEMS

Self-Adaptive Systems (SASs) [1] are able to autonomously modify their behavior at run-time in response to change or unexpected environmental conditions. SAS must continuously provide acceptable behavior during execution. As software is used for more pervasive and critical applications, support for self-adaptation is increasingly seen as essential to avoid costly disruptions in system repair, maintenance and evolution [9]. This type of systems is used in automotive systems, telecommunication systems, and environmental monitoring.

Self-adaptation [10] is a mechanism of a behavioral and/or structural adjustment. It is triggered in response to an environmental change or a modification in the system itself. The self character of the adaptation ensures a minimal level human oversight. Moreover, it is necessary for applications that

must run continuously, even under adverse conditions and changing requirements.

Typically, SAS [2] consists of a Managed Sub-System and a Manager (Autonomic) Sub-System. The first one deals with the domain functionality. The second deals with the adaptations of the Managed System to achieve particular quality goals. It implements the self-adaptation by implementing the well established feedback control loop. In the IBM's Autonomic vision [3], the autonomic behavior is achieved by implementing the MAPE-K reference model.

III. MAPE-K Reference Model

MAPE-K is proposed by IBM and it represents the most popular model. The MAPE-K Control Loop [2] is the reference model for designing SAS in the context of Autonomic Computing. Figure 1 shows an Autonomic Element where the Manager (Autonomic) Sub-System implements the adaptation logic by means of the well established MAPE-K model, and the Managed System implements the system functionalities. It consists of sensors and effectors as well as four components that realize the feedback loop: Monitor-Analyze-Plan-Execute, supported by a Knowledge base that allows the transfer of information between the other components [3]. The Sensors, attached to the Managed Sub-System, monitor the Managed Sub-System to collect data reflecting the state of the system, while Effectors apply the changes to the Managed Sub-System [2].

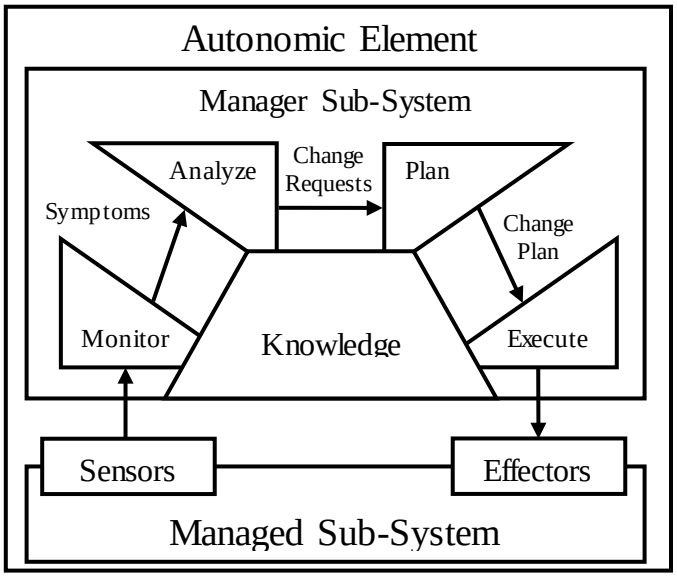


Figure 1 : MAPE-K Control Loop [2].

The MAPE-K Control Loop highlights four essential aspects of self-adaptation:

The monitor component is responsible of the collection of data captured by available sensors. The accumulated data is then cleaned, filtered, and pruned. Finally, it is stored for future reference to portray an accurate model of past and current states [3, 11].

The analyze component consists in exploiting the data contained in the symptoms provided by the Monitor. It treats them according to high level management policies and strategies. Then, it decides which sequences must be interpreted and which sequences must be discarded [3].

The plan component is responsible for planning adaptation actions, usually according to some guiding strategies [3].

The execute component provides the mechanisms to control the execution of the plan actions over the Managed Sub-System by means of effectors. Note that, the consistency of the system must be ensured during the execution of these actions [3].

There is a variety of models based on the MAPE-K Control Loop. In our case, we choose FRAMESELF architecture [12] to illustrate the sources of uncertainty. The reason for choosing FRAMESELF as the basic model is its detailed and separated description of each part of the control loop. Consequently, we can precisely identify the sources and the types of uncertainty in the MAPE-K.

The FRAMESELF architecture [12] is based on the IBM autonomic architecture reference (MAPE-K Control Loop). In this model, the monitor, analyze, plan, and execute operate as expert systems to emulate the decision making ability of human experts. The UML component diagram is used to describe the FRAMESELF modules and show how they are connected together.

A. Monitoring Component Diagram

Figure 2 describes the UML components diagram of the monitoring process. It is composed of sub-processes which can be summarized as follows [12]:

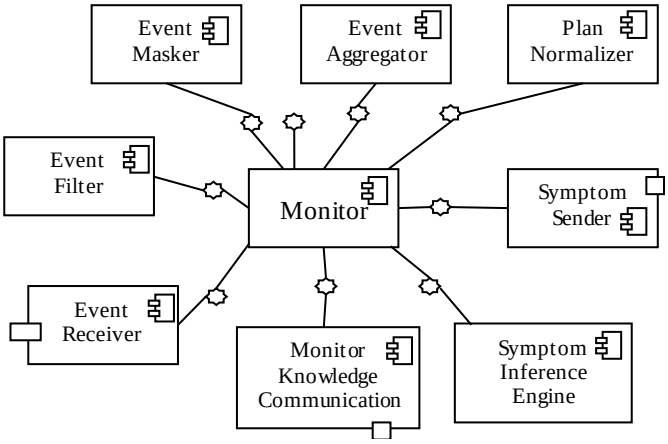


Figure 2 : Monitor components diagram [12].

- *Event Receiver* is responsible for collecting events from different sources.

- *Event Normalizer* makes it possible to render the captured event in a standardized format to make them consistent for processing.
- *Event Filter* determines the events that are relevant to the management platform.
- *Event Aggregation* is responsible for ignoring duplicates of the same event (caused by network instability).
- *Event Masking* is responsible for ignoring events pertaining to systems that are downstream of a failed resource.

The difference between *Filter* and *Masking* process is that: the masking allows to dynamically hiding unfiltered events depending on contextual changes.

- *Monitor Knowledge Communication* enables the monitor to read/write information from/to knowledge base.
- *Symptom Inference Engine* generates symptoms based on knowledge model and received events.
- *Symptom Sender* sends symptoms to the analyzer component.

B. Analyzing component diagram

Figure 3 describes the UML components diagram of the analyzing process. It is composed of sub-processes which can be summarized as follows [12]:

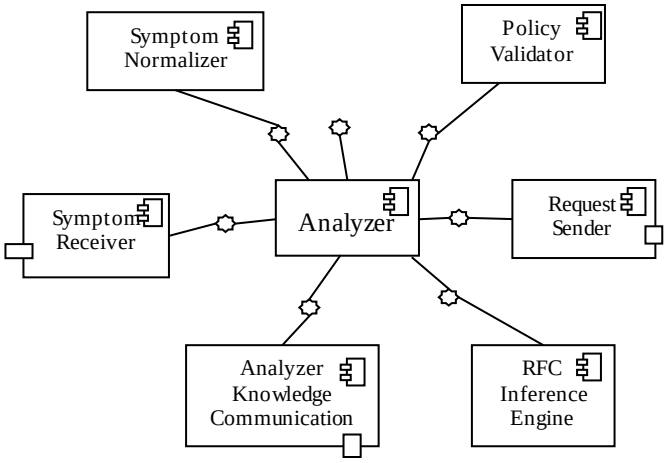


Figure 3: Analyzer components diagram [12].

- *Policy Validator* enables to validate a policy defined by the manager (located on the knowledge base) in order to guide the control loop operations.
- *Symptom Receiver* is responsible for receiving symptoms from Monitor source using.
- *Symptom Normalizer* makes it possible to render the received symptoms in a standardized format to make them consistent for analyzing.
- *Analyzer Knowledge Communication* enables to read rules and policies from the knowledge base.
- *Request Inference Engine* receives symptoms as input, checks knowledge model, then generates Change Requests as a response.
- *Request Sender* sends Change Requests to the planner component.

C. Planning component diagram

Figure 4 describes the UML components diagram of the planning process. It is composed of the following sub-processes [12]:

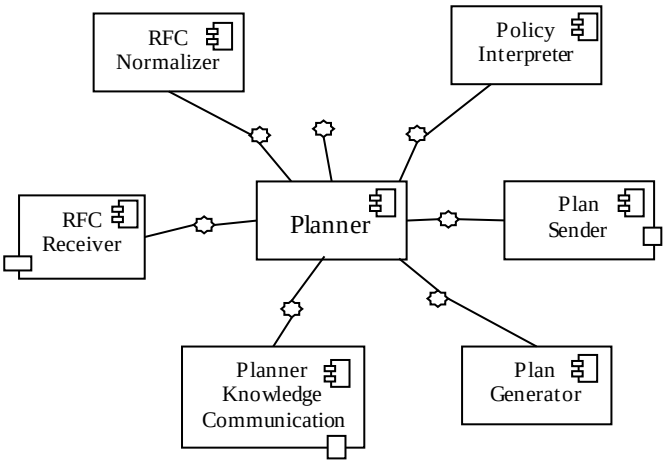


Figure 4 : Planner components diagram [12].

- *Policy Interpreter enables* to processes a policy (located on the knowledge base) in order to apply it to guide the decisions that affect the autonomic manager behavior.
- *Change-Request Receiver receives* Change Requests from the analyzer component.
- *Change-Request Normalizer makes* it possible to render the received Change Requests according to a predefined template in a standardized format to make them ready for planning.
- *Planner Knowledge Communication enables* to handle the deployment graph model and read policies from the knowledge base.
- *Plan Generator* receives Change *Requests* as input, checks knowledge model, then generates Change plan as response.
- *Plan Sender* sends Change Plan to the Execute component.

D. Executing component diagram

Figure 5 describes the UML components diagram of the executing process. It is composed of the following sub-processes [12]:

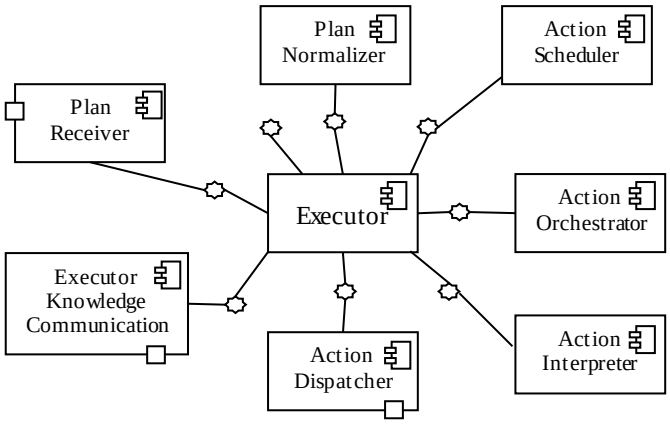


Figure 5 : Executor components diagram [12].

- *Plan Receiver* receives Change Plan from the planner component.
- *Plan Normalizer* makes it possible to render the received plan in a standardized format to make them easy to parse by the Action interpreter.
- *Executor Knowledge Communication* enables to handle the deployment result graph model from the knowledge base.
- *Action Interpreter* focuses on processing received actions and checking whether the executor has the rights and means to execute them.
- *Action Orchestrator* describes the automated arrangement and coordination of actions in time.
- *Action scheduler* determines, for simultaneous actions, which action to be executed next to load balance the system.

IV. UNCERTAINTY IN MAPE-K CONTROL LOOP

In this section, firstly, we briefly present the literature survey on sources of uncertainty in the MAPE-K, then describe these sources using FRAMESELF architecture (more details in section III).

A. Literature Survey

The literature in the area of self-adaption is extensive. During the past decade, the uncertainty in the SAS has seen a great interest. However, A few works have been conducted to study uncertainties in MAPE-K Control Loop, we focus our discussion here to those works that are of utmost relevance to present the uncertainties in MAPE-K as follows.

Cheng and Garlan [13] described three sources of uncertainty in control loop (problem-state identification, strategy selection, and strategy outcome), which used Rainbow approach [14] to mitigate them. The first challenging source of uncertainty is problem-state identification, which is related to the monitoring and analysis parts of the MAPE-K Control Loop. Once a problem is identified, a strategy is selected to resolve this problem; it is the second source of uncertainty which is related to the planning part of this loop. Finally, the uncertainty in strategy outcome is related to execution part of the loop, by specifying if a given strategy has succeeded or failed in execution.

Following the same approach, Cámara et al. [8] proposed other version of different sources of uncertainty that may affect different parts of MAPE-K and based on the works of [13] and [15]. In their work, they also identify different sources of uncertainty in parts of the MAPE-K loop as follows: in monitoring component, event detection can be affected by different uncertainties related to the inherent imperfections of sensors, which can be summarized as follows: (i) Sensing latency: information processing can take more time by the monitoring infrastructure. By the time it is arrived to analysis component, it may be outdated. (ii) Sensing inaccuracy: inaccuracies in sensor readings can also introduce uncertainties because the sensor cannot distinguish deviations between the ideal reading and detected reading. (iii) Sensing reliability: Sensors may fail, it may start providing values that are far from the real state. Analysis and Planning can be also affected by uncertainties, which can be summarized as follows: (i) Analysis/Planning latency: the uncertainty concerned the reasoning mechanisms themselves, such as the time requires to complete analysis/planning, particularly important when sophisticated planning for solution synthesis takes place. (ii) Model uncertainty: Analysis and planning can be affected by uncertainty, related to models employed to reason about the need to adapt, and how to adapt. Finally, there are different uncertainties that affect execution as following: (i) Execution Latency (or execution time): the time between the moment when the adaptation execution is triggered, and the moment when it is actually executed will not be deterministic in the general case. (ii) Effectors reliability: The effectors will not always successfully execute its required action.

B. Sources of Uncertainty in MAPE-K Control Loop

Before describing the sources of uncertainty in the Monitoring, Analyzing, Planning, and Executing, we present the uncertainty caused by the sensors and effectors, based on the literature survey of uncertainty in MAPE-K.

1. Uncertainty in Sensors/Effectors

1.1. Uncertainty caused by sensors:

There is a probability that the sensor will not capture data accurately, i.e. it will capture data that are far from the real state, or no values at all [8]. For example: a dramatic drop in the temperature sensor reading.

1.2. Uncertainty caused by Effectors:

There is always a possibility that the effectors will not successful execute its required action [8]. For example: effectors fail.

2. Uncertainties in monitoring process

There are two sources of uncertainty in monitoring process that have to be considered, which are:

2.1. Event *Normalizer, Event Filter, Event Aggregation, and Event Masking*

Pre-processing of collected events consists of normalizing, filtering, masking, and aggregating these events. As a result, they can take time to be processed by these sub-processes. By the time it is analyzed by Analyzing process it may be outdated [8].

2.2. *Event* Aggregation*:*

Aggregation in the FRAMESELF architecture is used to ignore the duplication of events. If we take aggregation in another meaning: Aggregation takes a set of data as input and produces a single piece of data as a result. Consequently, aggregation is considered a source of uncertainty, the incompleteness of the input data can provide data that is far from the real state (especially in complex system). As an example: Massive Open Online Courses (MOOCs) provide rich data sets about learner interactions, collaborations ...etc. This data must be aggregated to determine the path learning for each learner. Incompleteness of data may lead the learner to an inappropriate path.

3. Uncertainties in analyzing process

The source of uncertainty in analyzing process is:

3.1. Request Inference Engine

Within this process, and before generated Change Requests, it must determine which symptom should be adapted and which symptom should be discarded. By the time it is analyzed, the system's adaptation goals may be changed, so that the identification of symptoms-state is uncertain. For example: modifying adaptation goals to meet needs that occur over time to preserve the required security level in a changing system.

4. Uncertainties in planning process

There are two sources of uncertainty in planning process that have to be considered, which are:

4.1. *Policy Interpreter*

When actions are executed, several consequences are possible, i.e. the availability of resources can be changed. At this time, the Policy Interpreter sub-process may be interpreted the policies, therefore, the determination of the decisions to adapt the system is uncertain. As an example: the format of the policy rule is: When event If condition Do action.

```
When changed (sys://storage/memory@free)
If (sys://storage/memory@free >= 10*1024)
Do {action1 (); }
```

Suppose that, the Policy Interpreter (sub-process of planning) validates this condition, i.e. the "free" parameter is greater than or equal to 10*1024. On the other hand, within the executing process, an action is already programmed to be execute, where the outcome of this action being to decrease the value of the "free" parameter.

4.2. *Plan Generator*

The Plan Generator sub-process is responsible to generate a Change Plan according to adaptation strategy selection. After he chooses his adaptation strategy, he may change his mind based on changes in system state (Uncertainty in Decision-Making) [15].

5. Uncertainties in executing process

There are three sources of uncertainty in executing process that have to be considered, which are:

5.1. Action Interpreter

The dynamics of the system can lead the Action Interpreter sub-process to determine an incomplete, inaccurate, or unreliable decision.

5.2. Action Orchestrator

There is a temporal uncertainty because the time between the instant in which Action Orchestrator sub-process determined the order and coordination of the actions in time, and the instant in which the effectors is actually executed the actions, will not be deterministic in the general case [8].

5.3. Action scheduler

The misjudge of the outcome of running a workflow or the delay in determining of the execution of the action in time, can induce problems in the system balance. For example, in the fire monitoring system, the delay in reporting a fire can be regarded as a failure of the monitoring system and can have dramatic consequences.

V. DISCUSSION

Uncertainty has a significant impact on the ability of SASs to achieve its quality objectives. In particular, this survey presented the uncertainty in the MAPE-K Control Loop. In fact, the authors [8] [13] were able to establish the main uncertainties in MAPE-K. They are based on the MAPE-K loop proposed by IBM. However, there is no comprehensive work that exhaustively and precisely compiles the different sources that affect this loop.

The paper presented a comprehensive view for the sources of uncertainty in MAPE-K. In addition, we based on the FRAMESELF architecture (a detailed MAPE-K) to describe them. As consequently, it is an important study towards an effective treatment of uncertainty.

Finally, the survey paper gives the scientific community the opportunity to improve the adaptive ability of the system, in order to satisfy its objectives by addressing the uncertainties described in this work.

VI. CONCLUSION

This survey paper is focusing on the uncertainties in the MAPE-K Control Loop. We presented the most recent progresses and researches raised in this field.

At first, we introduced an overview of SASs. Then, we presented the MAPE-K Control Loop and FRAMESELF architecture. Next, we described the sources of uncertainty in each process of the loop, used the FRAMESELF architecture which is based on the MAPE-K, in order to precisely describe them, and to enhance the self-adaptive capabilities of the MAPE-K Control Loop. Before that, we presented the main works that are of utmost relevance to discover the uncertainties in the MAPE-K.

In our future efforts, we plan to manage uncertainties in the MAPE-K loop by using one of the existing stochastic approaches. In the other hand, we plan to focus on the Monitor part of the MAPE-K feedback loop, and in particular on the aggregation of data collected using Data Mining approach.

REFERENCES

[1] Salehie, M., & Tahvildari, L. (2009). Self-adaptive software: Landscape and research challenges. ACM transactions on autonomous and adaptive systems (TAAS), 4(2), 14.

[2] Kephart, J. O., & Chess, D. M. (2003). The vision of autonomic computing. Computer, 36(1), 41-50.

[3] Computing, A. (2006). An architectural blueprint for autonomic computing. IBM White Paper, 31, 1-6.

[4] Garlan, D., Cheng, S. W., Huang, A. C., Schmerl, B., & Steenkiste, P. (2004). Rainbow: Architecture-based self-adaptation with reusable infrastructure. Computer, 37(10), 46-54.

[5] Esfahani, N. (2011, November). A framework for managing uncertainty in self-adaptive software systems. In Automated Software Engineering (ASE), 2011 26th IEEE/ACM International Conference on (pp. 646-650). IEEE.

[6] Kolmogorov, A. N. (1956). Foundations of the theory of probability Chelsea New York USA.

[7] Zadeh, L. A. (1965). Fuzzy sets, Information and Control8 (3): 338–353. Google Scholar.

[8] Cámara, J., Garlan, D., Kang, W. G., Peng, W., & Schmerl, B. (2017). Uncertainty in Self-Adaptive Systems Categories, Management, and Perspectives. CMU-ISR-17-110.

[9] Vassev, E., Hinchey, M., Balasubramaniam, D., & Dobson, S. (2011, June). An ASSL approach to handling uncertainty in self-adaptive systems. In Software Engineering Workshop (SEW), 2011 34th IEEE (pp. 11-18). IEEE.

[10] Oreizy, P., Gorlick, M. M., Taylor, R. N., Heimhigner, D., Johnson, G., Medvidovic, N., Quilici, A., Rosenblum, D., & Wolf, A. L. (1999). An architecture-based approach to self-adaptive software. IEEE Intelligent Systems and Their Applications, 14(3), 54-62.

[11] Brun, Y., Serugendo, G. D. M., Gacek, C., Giese, H., Kienle, H., Litoiu, M., Muller, H., Pezzè, M., & Shaw, M. (2009). Engineering self-adaptive systems through feedback loops. In Software engineering for self-adaptive systems (pp. 48-70). Springer, Berlin, Heidelberg.

[12] Alaya, M. B., & Monteil, T. (2012, June). FRAMESELF: A generic autonomic framework for self-management of distributed systems-Application on the self-configuration of M2M architecture using semantic and ontology. In International Conference on Collaboration Technologies and Infrastructures (IEEE WETICE 2012),.

[13] Cheng, S. W., & Garlan, D. (2007). Handling uncertainty in autonomic systems. Int'l Wrkshp. on Living with Uncertainty.

[14] Garlan, D., Cheng, S. W., Huang, A. C., Schmerl, B., & Steenkiste, P. (2004). Rainbow: Architecture-based self-adaptation with reusable infrastructure. Computer, 37(10), 46-54.

[15] Esfahani, N., & Malek, S. (2013). Uncertainty in self-adaptive software systems. In Software Engineering for Self-Adaptive Systems II (pp. 214-238). Springer, Berlin, Heidelberg.

Design and Control of Excavator Type Work Machine Simulator

O. ÇEVİK[1] and İ. ÇAYIROĞLU[1]

[1] Karabuk University, Karabuk/Turkey, oguzcevik@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, icayiroglu@karabuk.edu.tr

Abstract **- This study aims to simulate the actual working environment of the excavator type work machine and to train the operator the driving movements on a robot excavator. In this study, the design, manufacture and control of training simulator was realized for the purpose of obtaining driving ability of excavator on simulator before actual machine use was started. In our country, the training of the work machine operators is carried out on real machines.**

For use in work, small size excavator robot which has two palettes and composed of boom, stick and bucket and able to control these organs with hydraulic flow such as real machine was manufactured. The simulation of hydraulic components to real machine movements is provided and with the designed drive circuit, the movement of the pallets and the return of the upper structure were performed with wireless control over the RF channel. Control functions of the excavator robot resemble the use of real machine driving. A camera was placed in the driver's cabin on the robot and the camera images were transferred to the video glasses system of the instructor outside the robot working area. The purpose of using video glasses is to simulate actual use. The trainer has the impression that he is using the excavator completely inside the robot independent of the outside space.

Three different experimental studies were carried out using the excavator robot. Experimental environment has been prepared so that the operator who will receive the training can perform these applications. These are: 1) Canal digging, 2) Ramp up and lowering from ramp, 3) Using the excavator as a crane.

Keywords **- Hydraulic, Mechanical, Excavator robot, Remote control, Operator Training**

I. INTRODUCTION

Training of operators who use excavator type work machines which frequently used in the field of construction and building sector is done on actual machines. This situation has drawbacks in terms of both trainers and machines. Passing the operator's simulator training before using the machine will help prevent many losses.

Most of the simulators in the market are based on providing the virtual reality with the software and representing the movements with the movable seat system. With the pneumatic system placed under the armchair, the educator better understands the dynamic movements of excavator. Some simulator types appear below. These simulators which are sold commercially attract attention with their high prices.

Figure 1: Working in a virtual environment with a moving seat

Figure 2: Mechanically simulated real-world simulator model

Figure 3: Working on the stationary seat with the joysticks

The robot is designed to simulate six different movements same as the actual machine control functions by the remote controller. Thus actual machine movements are achieved by wireless communication. In order for the educator to perceive the visual vision within the real machine, the camera images in the cabin are transferred to the video glasses system of the educator.

When the studies done in the literature are examined, there is not much research on the use of robots as simulators. These studies mostly focus on the application and development of two and three dimensional simulations in the virtual environment.

Weir has designed a general purpose driving simulator and created a human-machine interface (HMI) that offers the possibility to test with different vehicles [1].

Sasaki et al. have developed a pneumatic robotic system in order to perform remote control of the construction machine in their work [2].

Sasaki et al. have shown that the portable 40 kg robot arm module is carried on two diggers with a volume of 0,025 m3 and 0.28 m3 buckets and have successfully performed movements with camera images [3].

Engel et al. presented real-time operation of excavators with joystick control equipment in a simple excavator simulator program in their work [4].

Kwon et al. used a small-sized hydraulic machine for the excavator simulator in their work. Pressure sensors are placed on the excavator and the angular positions of the moving components relative to the pressure during simulation are shown graphically [5].

Liu et al. have proposed a general architecture of the system for remotely controllable excavator applications and have thoroughly examined each construction in terms of its functionality [6].

DiMaio et al. have studied a method that can calculate real mechanical reaction movements from a virtual excavation environment for performance calculations and controller design [7].

Haga et al. performed a control system design for the excavation work of excavators in their work. During excavation, the height of the boom and stick should be controlled so that the bucket position should move horizontally. With this control system, the depth of excavation was realized with the same feature as manual operator using, and the repetitive movements at the time of excavation were reduced. As a result, business productivity has been increased [8].

Wang et al. have proposed a parallel hybrid model with fuzzy logic control in a hydraulic excavator in order to provide better energy efficiency in their work and presented a mathematical model by analyzing the system [9].

In this study, the design and control of a training simulator was carried out in order to acquire the driving ability of the excavator type work machine. For use in working; from aluminum material, a 1/15 scale excavator robot which consists of boom, stick and bucket and controls these organs by hydraulic flow like the truth was manufactured. The robot has two pallets and can move independently with two DC motors, and can drive the upper structure by a DC motor independently from pallets. The proportional size of boom, stick and bucket and the speed of hydraulic system movement are designed in accordance with the truth. Produced robot appears below.

Figure 4: Manufactured excavator robot.

The hydraulic components, pallets and upper structure movements are simulated to the actual machine movements, and the remote control movements are likened to the operator usage of the actual machine. Six different movements are realized by wireless communication. It is aimed to simulate the actual use by placing a camera in the driver's cabin on the robot, and camera images are transferred to the video eyewear system of the trainer who is outside the robot working area.

Arm travel speeds are not fixed but are proportional to the position of the joysticks in the remote control device.. Also, the rotation speed of upper structure is not constant but is proportional to the joystick. The weight that the bucket can lift when it is at the farthest distance from the ground level is measured 2 kg in the present system. It was measured that the bucket could lift a weight of 4 kg at the ground level closest to the main body.

The designed system is based on the control of the robot by transmitting the commands from the remote controller to the driver circuit through the wireless communication receiver and by processing these commands in the driver circuit. Control diagram is shown below.

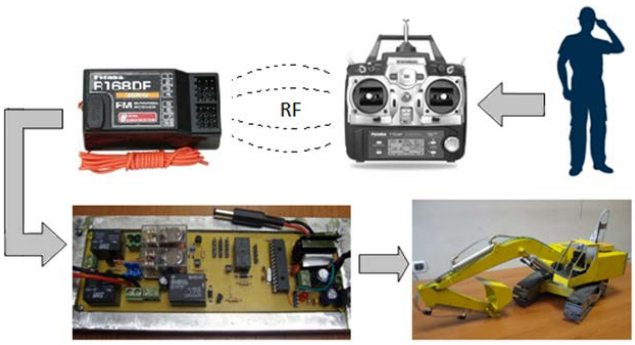


Figure 5: System integration

II. MATERIAL AND METHODS

In this study, a robot with 1/15 scale of a real excavator machine was designed and 2 and 3 dimensional drawings were made in Autocad program. Later, hydraulic calculations of the boom, stick and bucket that will move with the hydraulic system was done. After the design, a prototype machine was produced from aluminum material Which has 32cm pallet length, 22cm width, 21cm height and arm open total length

90cm.

Hydraulic working system of excavator is carried out by directing the flow of the pressurized hydraulic oil produced by the hydraulic pump to the target cylinder by the valve. Important variables in the hydraulic system are cylinder inner area, cylinder maximum length, hydraulic flow volume and hydraulic pressure. While the cylinder inner area and pressure determine the maximum load that can be lifted, the hydraulic flow rate determines the cylinder travel speed.

The force calculation in cylinder is calculated by multiplying each cylinder internal area by hydraulic pressure.

$$F = P * A \quad (1)$$

The maximum force in each cylinder is found by comparing the distance from the center of gravity to moment point when maximum arm opening and the bucket is full. Accordingly, the greatest force is seen in cylinders that lift boom arms.

In the case of the maximum arm opening, the arm weight in the full state of the bucket is 2.5 kg and the length of the center of gravity to moment point is predicted as 45 cm in length. The point at which the boom cylinder is connected is 5.5 cm from the moment point. The required hydraulic pressure calculation is shown in the following equation.

$$F_{(arm)} * L_{(arm)} = F_{(boom)} * L_{(boom)}$$

(2)

$$2{,}5 * 9{,}81 * 0{,}45 = F_{(boom)} * 0{,}055$$

$$200{,}7\ \text{N} = F_{(boom)} = P * A = P * \pi * r^2 \qquad (r = 0{,}008\text{m})$$

$$200{,}7/201{,}1\ \text{N/m}^2 * 10^6 = P \cong 1\ \text{Mpascal}$$

The approximate pressure value from this calculation is 10 bar. The pressurized oil from the hydraulic pump directs cylinder movement with the 5/2 proportional control valve. Three movements of the arm mechanism require three separate valves.

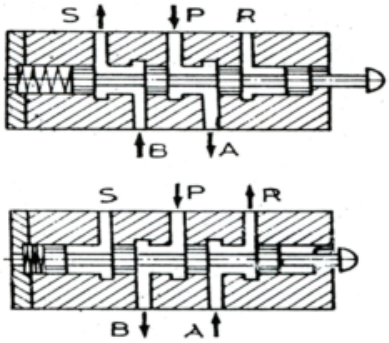


Figure 6: 5/2 proportional directional control valve

The driver circuit which read the information from the related channels of the receiver and control the three motors with hydraulic pump on the robot is seen below. This circuit is designed with 5 relays, a 16F876 PIC controller and a L298 motor driver.

Figure 7: Improved control circuit

Futaba 7CAP 7 channel controller is used for remote control application. A 1.2 GHz wireless AV image transmitter that is different frequency from the remote control is used to transmit the camera images. 3D video glasses system is used for image monitoring. The images taken from the mini camera placed in the cabinet are taken from the AV receiver and presented to the educator.

III. Results and Discussion

When the pallet motors and pump are controlled by the relay because of the high current draw, the motor that rotates the upper structure is controlled by the L298 drive integration. For this reason, as the pallets rotate at a constant speed, the upper structure moves at variable speed. In the main circuit, the digital information that the microcontroller reads from the receiver is compared with the reference values to control the motors and the pump. The magnetic field of the motors at the moment of operation and the magnetic field on the outside affect the receiver and lead to deterioration of the control of the system. To prevent this, the receiver is housed in the aluminum cage, known as the faraday cage.

The hydraulic system parts are connected completely and the speed of the arms and the lifting force have been tried according to the pressure given by the pump and the similarity to the movement of real machine has attracted attention.

Main circuit can control robot motion functions smoothly according to commands. Experimental set was prepared to perform experimental work using excavator robot. The experimental set designed for canal digging, ramp on and use as a crane is shown below.

Figure 8: Experimental environment

In canal excavation application, the operator takes the soil

from the cannal and fills the truck. After a certain period of time, it is measured how much soil the truck filled with, how smooth the digging process is, whether it hit the truck.

The excavator is carried on the trailer while moving from one place to another. A ramp is used to lift the excavator onto the trailer. In our second application excavator loading practices to trailer were done by using ramp. The operator is instructed to load and download from the ramp without lowering the excavator. The number of times this process is performed within a certain period of time, the more points he receives.

In some environments, excavators can be used as a crane to save time when no crane is available. The operator must have an opinion inverse relationship between load and moment point for not overturning the excavator. The operator must have an idea and experience of how far the bucket is located when it is full. In the test environment, it is required to lift the masses of 1.75, 2.0, 2.5, 3.0 kg and turn the upper structure of the excavator by 90 degrees so as to leave the mass as far as possible. The operator's performance score is created if he left the weights without overturning the excavator in a certain period of time.

It is desirable that the inverse ratio between the load and the momentum of the operator is dominant in order for the excavator to overturn.

IV. CONCLUSION

The projected dimensions in the design phase ensure the similarity the robot appearance to the actual machine. It is seen that the pressure requirement of 10 bar determined in the calculations can move the arm organs smoothly while carrying the load. The amount of liquid pumped by the pump can move the arms at a speed that is in line with reality. With the use of proportional valve it is possible to move the arms very slowly.

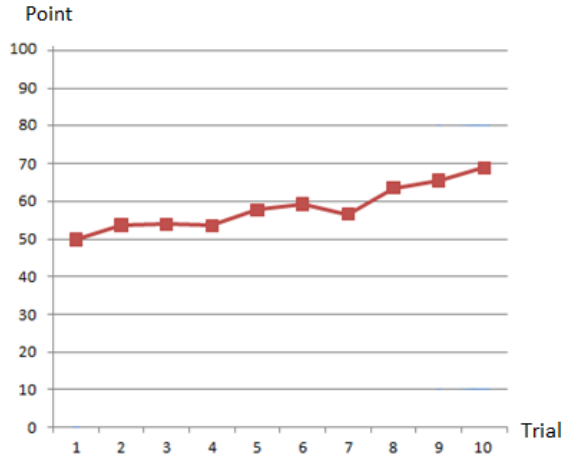


Figure 9: Channel digging application performance graph

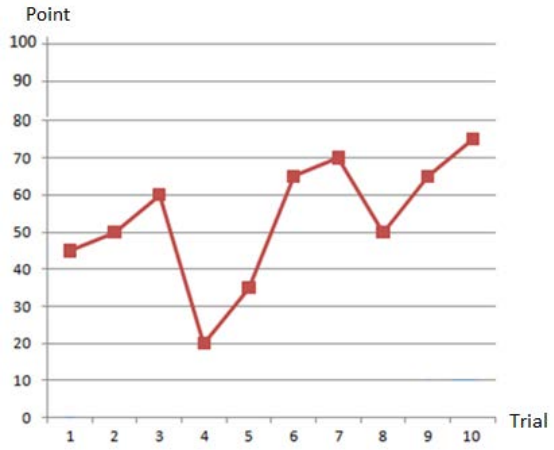


Figure 10: Ramp loading performance graph.

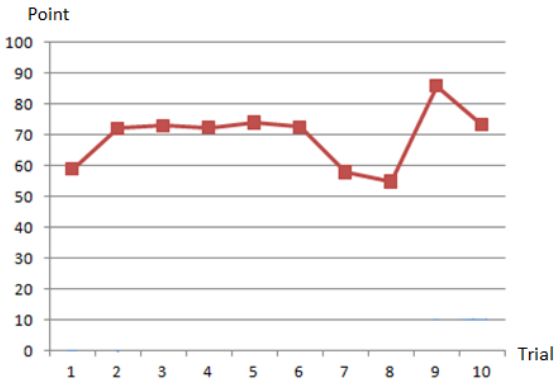


Figure 11: Performance graph of application usage as crane.

REFERENCES

[1] Weir, D. H., "Application of a driving simulator to the development of in-vehicle human–machine-interfaces", *Journal on IATSS Research*, (34): 16-21, 2010.

[2] Sasaki, T. and Miyata, T., "Development of remote control system of construction machinery using pneumatic robot arm", *IEEE International Conference on Intelligent Robots and Systems*, Japan, 748-753, 2004.

[3] Sasaki, T. and Kawashima, K., "Remote control of backhoe at construction site with a pneumatic robot system", *Journal on Automation in Construction*, (17): 907-914, 2008.

[4] Engel, S., Alda, W. and Boryczko, K., "Real-time computer simulator of hydraulic excavator", *AGH University of Science and Technology*, Poland, 1-9, 2010.

[5] Kwon, S. K., Kim, J., Jung, Y. M., Jung, C. S., Lee, C. D. and Yang, S. Y., "A hydraulic simulator for an excavator", University of Ulsan, *JFPS International Symposium on Fluid Power*, Korea, 2-6, 2008.

[6] Liu, Y., Hasan, M. S. and Yu, H. N., "Modelling and remote control of an excavator", Staffordshire University, *International Journal of Automation and Computing*, 349-358, 2010.

[7] DiMaio, S. P., Salcudean, S. E., Reboulett, C., Tafazoli, S. and Hashtrudi-Zaad, K., "A virtual excavator development and for controller evaluation", University of British Columbia, *International Conference on Robotics & Automation*, Belgium, 52-58, 1998.

[8] Haga M., Hiroshi W. and Fujishima K., "Digging control system for hydraulic excavator", *Mechatronics*, (11): 665-676, 2001.

[9] Wang D., Lin X. and Zhang Y., "Research on the energy regeneration systems for hybrid hydraulic excavators", *Automation in Construction*, (20): 581-587, 2011.

Low Cost and Practical Data Acquisition System Using Labview: An Application

H. BAKIR[1], M. S. BAŞARSLAN[2] and Ü. AĞBULUT[1]

[1] Duzce University, Düzce/Turkey, hsynbakr@gmail.com
[2] Doğuş University, İstanbul/Turkey, mbasarslan@dogus.edu.tr
[1] Duzce University, Düzce/Turkey, umitagbulut@duzce.edu.tr

Abstract-In almost all areas of the life-cycle, measurement, contolling and even achieving in stable/desired values of temperature have big importance. The man-kind's daily activities are influenced by various control systems in almost every direction. Also, control systems are widely used in all sectors of the industry. This study mainly focused on measuring and controlling the temperature values. The purpose is to keep stable the analog temperature data at the desired temperature by performing the necessary control procedures. In designed temperature control system, analog temperature values were measured by K type thermocouple. Since the temperature measured by the thermocouple is in the mV level, this data must be raised to 0-5 V to be supplied to the Arduino analog input (A0). The AD620 instrumentation amplifier was used for the upgrade. The digital output data from the Arduino PWM3 block is transmitted to the SSR (Solid State Relay) using the 74HC244 buffer. Temperature control was performed by the PID control software preparing in the LabVIEW program. In this study, temperature values are successfully obtained with a ±%1.5 accuracy via Labview.

Keywords **- Data Acquisition, LabVIEW, Temperature control.**

I. INTRODUCTION

CONTROL of a system or a process using computerized automation technologies has become very common in recent times. Thanks to automation applications, many processes in the industry are realized with very few mistakes.

Temperature is one of the most commonly measured and controlled events in the industry. In the industry many of the devices exhibit unstable behavior due to deterioration in their characteristics over a certain temperature [1]. These instabilities can cause financial loss in industrial applications, causing the operation and systems to deteriorate. Moreover, loss of temperature control in most applications in the industry can cause irreversible damage. For this reason, temperature measurement and control are important to the industry.

The applied techniques and equipment vary according to the application area (temperature level, physical and chemical structure of the environment, sensitivity, reading accuracy and speed and type of control output.) in temperature control. Thermocouples, RTDs, thermistors, semi-conductor based temperature sensors and pyrometer elements are frequently used today to convert temperature information into electrical signals [2-3]. After the electrical information is obtained, the desired control method is applied with digital or analog control circuits.

Controls and visual results in the designed system were performed by LabVIEW graphical control software [4-5]. In the design, the temperature of a resistor is fixed at a certain temperature using a closed-loop feedback control model.

II. TEMPERATURE MEASUREMENT AND CONTROL

A. Temperature Measurement

Temperature sensors are simply referred to as devices used to measure the ambient temperature. Increasing the importance of temperature measurement in industrial environments has led to the emergence of temperature measurement techniques that have distinct characteristics in different environments. It is necessary to select the most suitable temperature sensors according to usage environment, temperature measurement range and process conditions [6].

- Things to note when choosing a temperature sensor are as follows.
- Temperature reading accuracy.
- Temperature measurement range.
- Reaction rate versus temperature change and detection accuracy.
- The level of environmental constraints.
- Cost.

The temperature sensing elements usually operate by contacting the surface where the temperature is to be measured. In addition, non-contact temperature sensors are available. Known temperature sensors are thermocouples, RTD, thermistors, integrated circuit temperature sensors and surface contactless temperature sensors [7]. In this work, K type thermocouple was used for temperature measurement.

B. Temperature control

Temperature control can be done in different models (On-Off, PI, PD, PID) according to the area to be used in the industry. In this study, PID model is preferred for temperature control. For temperature control, the closed-loop feedback control method commonly used in the industry was preferred [8]. The designed system controls the temperature of a heated resistor with the software created in LabVIEW.

In closed loop control systems, also referred to as feedback control systems, an output value is measured by the measuring

element and the measured value is fed back to the input. This value is then compared to a reference value. A comparison result is obtained as an error signal. A control signal is generated in accordance with the structure of the error signal and the output variable. Figure 1 shows the block diagram of the closed-loop control system.

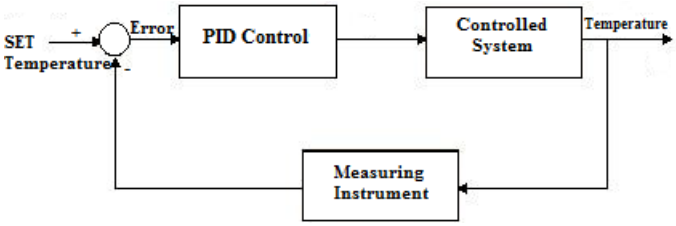


Figure 1: Block diagram of closed loop control system.

1) PID Control

PID (Proportional Integral Derivative) control is one of the most common control methods. PID controllers are used to calculate the output correction ratio; the error, the integral of the error and the derivative of the error. The PID control algorithm is calculated according to Equation (1) [9], [10].

$$u(t) = K_p e(t) + \frac{K_p}{T_i}\int_0^t e(t)\,dt + K_p T_d \frac{de(t)}{dt} \quad (1)$$

In equation (1), $e(t)$ is the error signal, $u(t)$ is the control input, K_p is the proportional gain, T_i is the integral time constant and T_d is the derivative time constant [11].

The expression in the Laplace frequency domain of the PID control algorithm is as in Equation (2).

$$u(s) = K_p\left[1 + \frac{1}{T_i s} + T_D s\right] E(s) \quad (2)$$

A PID control is consist of 3 parameters. These are K_p, T_i and T_d. These parameters, which are dependent on each other for a successful PID control, must be selected at appropriate values.

To determine the PID control parameters of the heater, the unit step response of the open loop control system is drawn in the computer and the transfer function of the heater is created. The PID parameters are determined by the Ziegler-Nichols method using the numerical data on the open loop unit step response.

The transfer function of a heater is written over the open-loop unit step response as shown in Equation (3) [12].

$$G(s) = \frac{K e^{-sT_d}}{1 + sT_1} \quad (3)$$

In Fig. 2, the ideal unit step response of the open loop control system is shown [13].

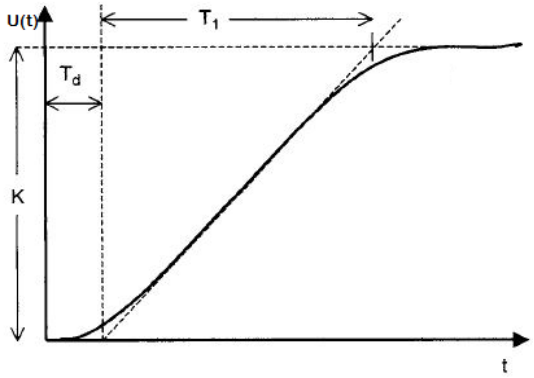


Figure 1: Ideal unit step response of open loop control system.

Table 1 shows the approximate values of the control methods according to the data obtained from the transfer function of the heater according to the Ziegler-Nichols parameter table.

Table 1. Ziegler-Nichols PID parameters.

Controller	K_P	T_i	T_d
P	$\frac{T_1}{KT_d}$	∞	0
PI	$\frac{0.9T_1}{KT_d}$	$3.3\,T_d$	0
PID	$\frac{1.2T_1}{KT_d}$	$2T_d$	$0.5T_d$

2) Determination of Heater Control Parameters

The temperature values obtained with the thermocouple are recorded for the heater's open-loop unit step response. The arduino's PWM duty-cycle value varies between 0-255. The PWM duty cycle value was started from 0 and the temperature read from the heater is set to 0 degree.

After this step the PWM duty cycle values are incremented one by one and the voltage is recorded every second until the temperature reaches 500 ° C. This process continued until the 255 PWM duty cycle value. The values obtained after the experiment are as follows:

$T_d = 75$ s.
$T_1 = 1200$ s.
$K = \frac{500}{255} = 1.96\ mV.$

The transfer function of the heater G (s) is as shown in Equation (4).

$$G(s) = \frac{1.96\,e^{-75s}}{1 + 1200s} \quad (4)$$

The PID parameters obtained using the Ziegler-Nichols PID parameter table was shown below.

$K_p = \frac{1.2 * 1200}{1.96 * 75} = 9.79$
$T_i = 2 * 75 = 150$ s.
$T_d = 0.5 * 75 = 37.5$ s.

$$\frac{U(s)}{E(s)} = 9.79\left[1 + \frac{1}{150s} + 37.5s\right] \quad (5)$$

The block diagram of the PID closed loop control system was shown in Fig. 3.

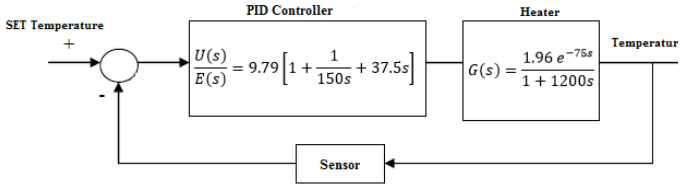


Figure 2: Block diagram of PID closed loop control system.

III. MEASURING AND CONTROLLING SYSTEM' HARDWARE BY LABVIEW.

DAQ (Data Acquisition) cards supplied by National Instruments are used to process time-dependent events such as temperature in this software. These equipment are comprehensive devices for such applications and their costs are also very high. In this study, data collection was carried out by the system that was designed to get rid of this necessity.

In designed temperature control system, analog temperature values were measured using K type thermocouple. The schematic view using in this study is given in Fig. 4.

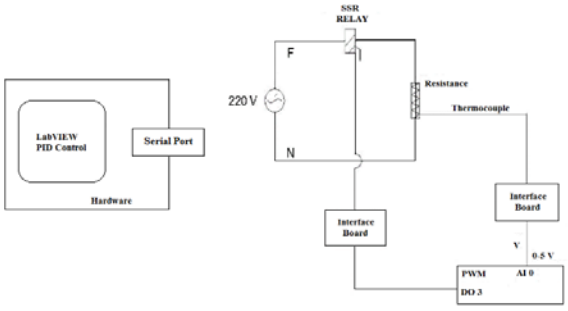


Figure 3: Block diagram of controlling and measuring temperature system.

IV. TEMPERATURE MEASUREMENT AND CONTROL SYSTEM SOFTWARE.

LabVIEW software, developed with measurement and instrumentation focus, has all the abilities of structural and object programming languages. Ready functions make the software easier. The PID control implemented in this application is easily created thanks to the LabVIEW PID blocks. The program front panel and block diagram views were shown in Fig. 5 and Fig. 8 respectively.

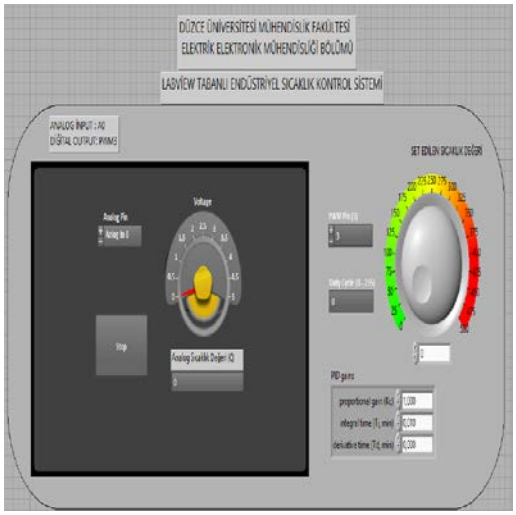


Figure 4: Users control panel.

The user front panel, prepared in LabVIEW, was tried to be compared to the temperature controllers in the market. When the user interface is examined, the user can select the analog input and digital output pins.

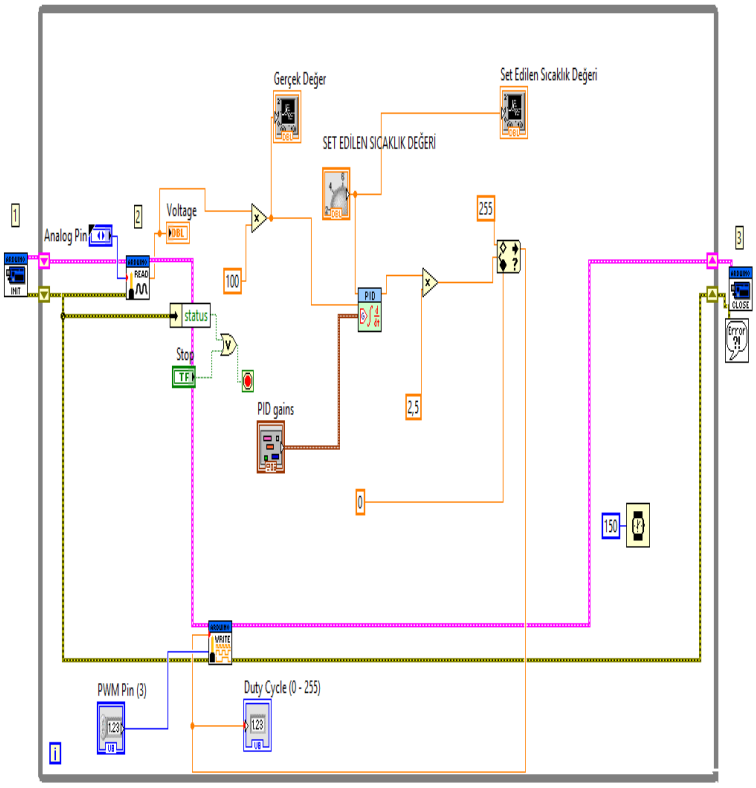


Figure 5: LabVIEW block diagram of the system.

V. RESULTS AND DISCUSSION

An application has been studied on temperature measurement. In the LabVIEW program, the temperature control system was operated at different temperature values and the following results were obtained.

Correct detection of the coefficients in PID controllers is very important for successful control. PID coefficients were

determined by the Ziegler-Nichols method from the transfer function which is the result of the open loop system response tests. One of the priorities of this study is that the PID control process can be performed in PC environment by LabVIEW and at the same time, the real time monitoring of the system response in LabVIEW has made it a great convenience and success in detecting real coefficients.

The control system designed as shown in Fig. 7. Here the user set the temperature to 200 °C. Thanks to the implemented PID control system software, the system temperature is fixed at approximately 202.37 °C. The temperature data were obtained with ±%1.5 accuracy.

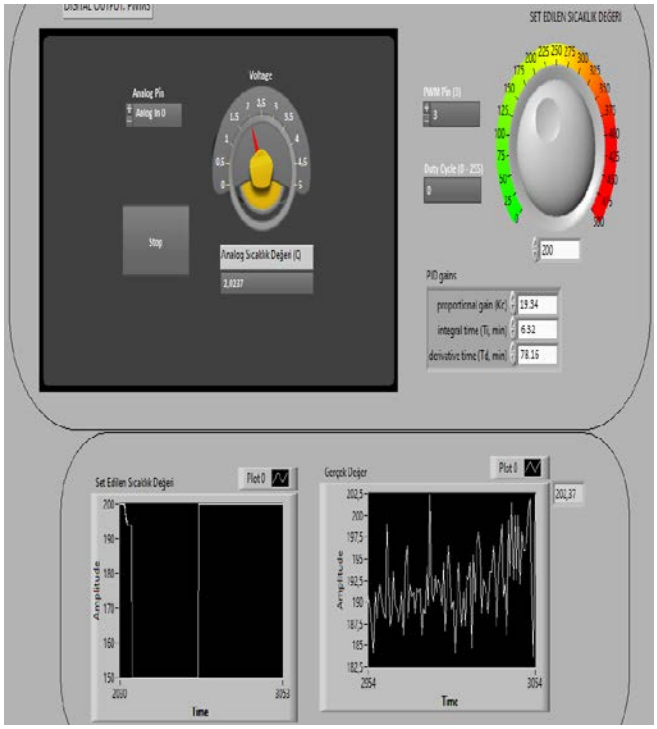


Figure 6: Testing system I.

The control system designed as shown in Fig. 8 has been tested. Here the temperature is set at 150 ° C. Thanks to the preferred PID control system software, the system temperature is fixed at approximately 150.37 °C. When the obtained values were compared, it was determined that the control process was successful.

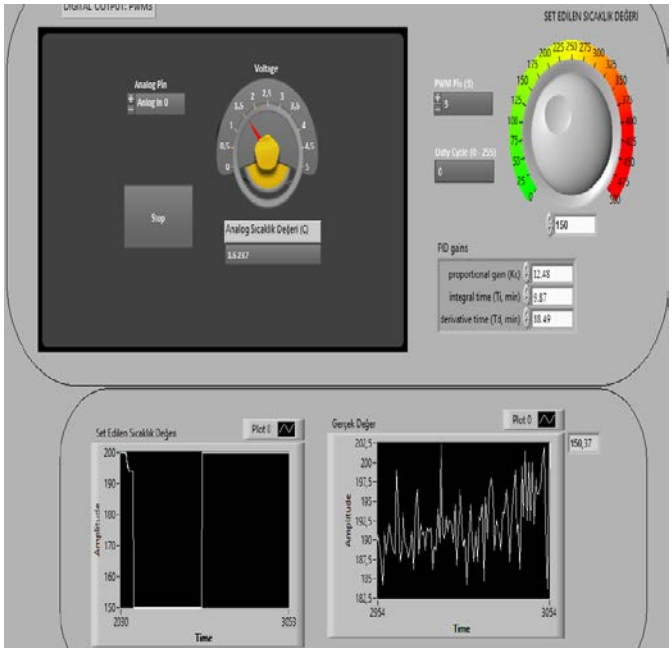


Figure 7: Testing system II.

Performing PID control in the LabVIEW environment allows easy adjustment of values such as PID gain coefficients, PID output ranges. In microcontrollers, these operations are either done with additional control hardware or by configuring and changing the software each time the data is changed. In addition, obtaining more sensitivity results should put a cooler to the system in future studies.

REFERENCES

[1] A. Pérez-González, O. Begovich-Mendoza, and J. Ruiz-León, "Modeling of a greenhouse prototype using PSO and differential evolution algorithms based on a real-time LabView™ application," *Applied Soft Computing*, vol. 62, pp. 86-100, 2018.

[2] M. Nascimento, M. S. Ferreira, and J. L. Pinto, "Real time thermal monitoring of lithium batteries with fiber sensors and thermocouples: A comparative study," *Measurement*, vol. 111, pp. 260-263, 2017.

[3] A. Murmu, B. Bhattacharyya, and S. Munshi, "A synergy of voltage-to-frequency converter and continued-fraction algorithm for processing thermocouple signals," *Measurement*, vol. 116, pp. 514-522, 2018.

[4] Z. Mingle, Y. Jintian, J. Guoguang, and L. Gang, L. "System on temperature control of hollow fiber spinningmachine based on LabVIEW," *Procedia Engineering*, vol. 29, pp. 558-562, 2012.

[5] R. L. Golda, M. D. Golda, T. D. Peterson, and J. A. Needoba, "Graphical coding data and operational guidance for implementation or modification of a LabVIEW®-based pHstat system for the cultivation of microalgae," *Data in brief*, vol. 12, pp. 463-470, 2017.

[6] A. Ü. Keskin, "Design of a PID controller circuit employing CDBAs," *International journal of electrical engineering education*, vol. 43, no. 1, pp. 48-56, 2006.

[7] D. Ibrahim, "Microcontroller-based temperature monitoring and control," *Newnes*, 2002.

[8] R. Solea, M. Găiceanu, C. Eni, and B. Codreș, "Performance Comparison of Sliding Mode Controller and Conventional Pid Controller for DC Motor," 2013.

[9] M. J. Grimble, "Integral minimum variance control and benchmarking," *Journal of Process Control*, vol. 14, no. 2, pp. 177-191, 2004.

[10] M. Oono, Q. Yuan, N. Kurita, and T. Ishikawa, "Analysis of a new linear actuator made of Soft Magnetic Composite material with spring," *In Electrical Machines and Systems, ICEMS 2009. International Conference on IEEE*, pp. 1-6.

Implementation of NURBS curves on the LCD touch screen using FPGA

Y. OZTURK[1] and C.OZCAN[2]

[1] Department of Computer Engineering, Karabuk University, Karabuk, Turkey, yasin.ozturk5@hotmail.com
[2] School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, ozcan@purdue.edu

***Abstract* - Computer-aided curve and surface design models are used to design surfaces that do not have a particular shape. Since these surfaces cannot be expressed by known mathematical functions, curve and surface modeling methods such as Bezier, B-Spline and Non-Uniform Rational B-Spline (NURBS) have been developed. In this study, an application is developed that allows the NURBS curve algorithm, which is used as a method in 3D modeling, to be computed using Field Programmable Gate Array (FPGA) with parallel processing capability and displayed on the LCD touch screen. The B-Spline basis function which NURBS based on are sampled mathematically. The parametric values used in the display of the NURBS curves and surfaces are presented on the LCD with user interaction. Parametric values such as control points, knot vector, and weight vector are separately sampled for curves and mathematical solutions are given. The NURBS curve for user-defined control points can be displayed in real time on the screen. In addition, an implementation of the NURBS algorithm in the Visual Studio platform was developed and obtained results were compared.**

***Keywords* - NURBS curves, 3D modeling, FPGA, parallel processing, LCD**

I. INTRODUCTION

In recent years, the importance given to the design of products development and accordingly the design development methods are improving rapidly due to the new technology and the facilities that we have found more in our lives. For this reason, computer aided design/production (CAD/CAM) has become essential in industrial areas. There are many different design methods used in computer graphics. Nonuniform Rational B-Spline (NURBS) surfaces are used almost every field because of the flexible nature of these methods and their local variability. The NURBS algorithm is a method for computer aided curve and surface designs and it offers great flexibility and precision [1].

This method has emerged with the development of the Bezier method which is presented by Pierre Bezier for use in automobile design [2]. The NURBS algorithm generates smooth curves with smooth transitions and provides the possibility to model the desired object with the surfaces created between these curves. NURBS, which is used in CAD/CAM systems, is a commonly used method especially in the modeling of parts to be processed in CNC machines. In the literature there are different studies on B-Spline curves in many fields [3-6]. In the studies carried out, it was observed that the performance of CNC machines is very high in designs modeled with NURBS [7,8]. A real-time NURBS surface interpolation has been developed for high precision three-dimensional CNC applications [8]. Preprocessing methodology for CL points in CNC machines for high speed machining is proposed using NURBS curve fitting technique [9]. A Field-Programmable Gate Array (FPGA) based hardware computation methodology is presented for Bezier, B-Spline and NURBS parametric trajectory interpolations commonly used in CNC machines [10].

NURBS is accepted as a mathematical subject beyond many professional fields of study. In some studies, the underlying mathematical theory of the NURBS method has been discussed and various solution methods, algorithms and equations have been developed. A solution method has been developed with NURBS for continuous Euler equations [11]. A real-time interpolation algorithm for NURBS curves has been developed and its performance on servo motors has been analyzed [12]. Geometry-based triangulation of trimmed NURBS surfaces was investigated [13,14]. A geometric condition is presented on the control polygons ensuring that the NURBS curves are injective (one to one) at the positive weight points [15]. In another study, necessary and sufficient injective conditions were provided by examining the geometric conditions of the non-intersecting NURBS surfaces [16].

Computer aided curve and surface design models are used to design free-form surfaces. It is quite easy to design these curved surfaces in computer environment thanks to some design programs today. However, the operation time of the system has gained importance due to the importance given to the productivity and speed in CAD/CAM systems in industrial areas. For this reason, the NURBS algorithm is implemented on different microcontrollers outside the computer, and the performance of the NURBS algorithm is examined. In some studies, NURBS surface interpolation developed for CNC applications was realized in real time on DSP and the execution times are presented [7,8,17]. In these studies, it was observed that the interpolation time on DSP is more than 1ms. Because DSP processors handle serial commands, the processing times are long.

However, in FPGA chips, interpolation times are shorter due to the ability to perform parallel processing [18]. Zhao et al. designed the NURBS interpolation on a motion control board

with DSP and FPGA on it. Interpolation calculations are performed on the FPGA as hardware and motion control of the servos is provided by DSP [19]. In similar studies, real-time NURBS interpolation for high-speed motion control on FPGA was developed and algorithm times were obtained [18,20]. When the studies are examined, the result of the parametric interpolation calculations on the FPGA with parallel processing capability is completed in a shorter time.

In this study, we designed an application that allows the NURBS curve algorithm to be displayed on the touch screen in real time on an FPGA with parallel processing capability. The results were compared with the measurements obtained in the computer environment.

II. NURBS INTERPOLATION ALGORITHM

The NURBS technique includes Bezier and B-Spline curves as content. It is especially the improved form of the B-Spline method. While it is more useful than Bezier curves, B-Spline curves also have disadvantages. For example, a change in the degree of the curve also affects the knot vector. This means to do all the operations from the beginning. However, the NURBS technique can be modified locally on the curve by the weights of the control points without making any changes to the curve or knot vector. For this reason, the NURBS technique is easier and more convenient for us.

The most important feature distinguishing the NURBS technique from the B-Spline method and makes it more useful is that the control points have a weight. With this weight it is easier to give the desired shape to the curve. We can control and modify the structure of the curve locally by changing only the weight of the control point without changing any other properties of the curve. NURBS curve from the $p.$ order can be expressed as [1]:

$$R_{i,p}(u)=\frac{N_{i,p}(u)w_i}{\sum_{j=0}^{n}N_{j,p}(u)w_j} \tag{1}$$

and

$$C(u)=\sum_{i=0}^{n}R_{i,p}(u)P_i \tag{2}$$

where $\{w_i\}$ is the weights of control points and $\{R_{i,p}(u)\}$ is the rational basic functions of the NURBS curve. These functions are piecewise functions in the interval $u\in[0,1]$. Some of the features of the $R_{i,p}(u)$ basic functions are as follows:

1. For $\forall i,p\ \ R_{i,p}(u)\geq 0$ on interval $u\in[0,1]$.
2. For $\forall u\in[0,1],\ \sum_{i=0}^{n}R_{i,p}(u)=1.$
3. $R_{0,p}(0)=R_{n,p}(1)=1$
4. For $p>0$ all $R_{i,p}(u)$ reaches a maximum value in the range $u\in[0,1]$.
5. For $u\in[u_i,u_{i+p+1}),\ R_{i,p}(u)=0$.

A. Knot Vector

One of the main factors in achieving regional control in B-Spline curves is knot vector. The B-Spline basis functions $\{N_{i,p}(u)\}$ that form the curve are shaped according to the knot vector. For this reason we can say that the knot vector can determine the sensitivity of the curve. Because the basis functions are constructed according to the knot vector ranges. In other words, when the curve is formed, which basis function is used in which region is determined by the knot vector.

We need $m+1$ knot values for $n+1$ control point unit on the $p.$ order B-Spline curve defined as $m=n+p+1$. For example, for 5 control points $(n=4)$ on the third-order B-Spline curve $(p=3)$, we need knot values as $\underbrace{(n+p+1)}_{m}+1=(4+3+1)+1=9$. The knot vector has different forms of use. The knot vector which we will use in this study is as follows:

$$T=[\underbrace{a,...,a}_{p+1},t_{p+1},...,t_{m-p-1},\underbrace{b,....,b}_{p+1}] \tag{3}$$

In this study, we will assume $a=0$ and $b=1$ unless otherwise stated. In this case, the generated knot vector will be as follows,

$$T=[\underbrace{0,...,0}_{p+1},t_{p+1},...,t_{m-p-1},\underbrace{1,...,1}_{p+1}] \tag{4}$$

In this case, the knot vector for a 2nd and 3rd order curve with 4 control points will be,

$$T=\begin{bmatrix}0 & 0 & 0 & 0 & \frac{1}{2} & 1 & 1 & 1 & 1\end{bmatrix}\quad p=3$$

$$T=\begin{bmatrix}0 & 0 & 0 & \frac{1}{3} & \frac{2}{3} & 1 & 1 & 1\end{bmatrix}\quad p=2$$

Fig. 1 shows the B-spline curves with different degrees and knot vectors on the same control points.

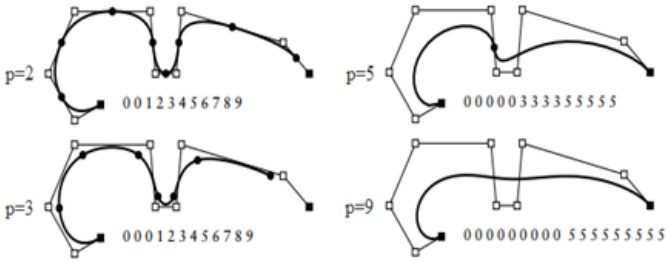


Figure 1: B-Spline curves with different degrees and knot vectors on the same control points [21].

B. *B-Spline Basis Functions*

We can describe a B-Spline basis function in the p. order as follows:

$$N_{i,0}(u)=\begin{cases}1 & t_i \le u < t_{i+1}\\ 0 & diğer\end{cases}$$

$$N_{i,p}(u)=\frac{u-t_i}{t_{i+p}-t_i}N_{i,p-1}(u)+\frac{t_{i+p+1}-u}{t_{i+p+1}-t_{i+1}}N_{i+1,p-1}(u)$$

where the range $u\in[t_i,t_{i+1})$ for functions $N_{i,0}(u)$ is consecutive knot points. Fig. 2 presents a diagram showing the basis functions that must be included in the knot vector range.

The regional control which is the most important feature of the B-Spline curves is provided by basis functions. Because the effect of basis functions on the curve is limited to a certain region. In other words, the B-Spline basis functions are equal to zero except for their regions. For this reason, the shape of the curve depends only on a few adjacent control points. Knot vectors for different grade curves and corresponding basis functions are shown in Fig. 3.

$[t_0,t_1)$ $N_{0,0}$

$N_{0,1}$

$[t_0,t_1)$ $N_{1,0}$ $N_{0,2}$

$N_{1,1}$ $N_{0,3}$

$[t_0,t_1)$ $N_{2,0}$ $N_{1,2}$ $N_{0,4}$

$N_{2,1}$ $N_{1,3}$ ⋮

$[t_0,t_1)$ $N_{3,0}$ $N_{2,2}$ ⋮

$N_{3,1}$ ⋮

$[t_0,t_1)$ $N_{4,0}$ ⋮

⋮ ⋮

Figure 2: Diagram for B-spline basis functions.

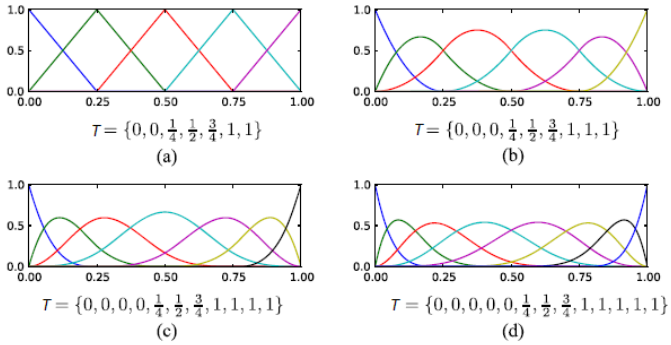


Figure 3: Knot vectors for different grade curves and corresponding basis functions [22].

III. Implementation of NURBS Algorithm with Computer Application

We developed an application in Visual Studio environment to observe the operation of the NURBS modeling method and to verify the algorithm. In application, the control points of the curve are determined by the user. With the setting of the second control point, the application begins to model the NURBS curve. Each subsequent point is regarded as the last control point and the application curve is re-modeled according to the end point. A NURBS curve with 6 control points drawn by application is given in Fig. 4.

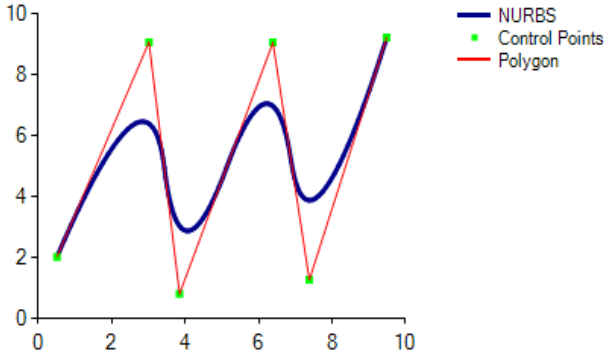


Figure 4: A NURBS curve with 6 control points.

Another thing to note here is the change in proportional basis functions when there is an increase in the number of control points. The rational basis functions of the curve in Fig. 4 are given in Fig. 5.

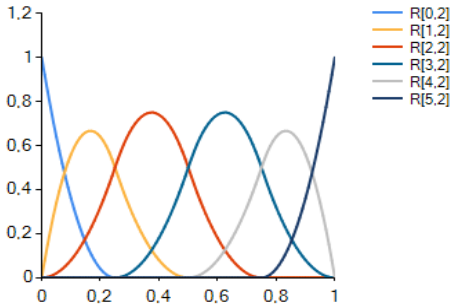


Figure 5: Rational basis functions.

The flow diagram of the algorithm steps of the implementation is given in Fig. 6.

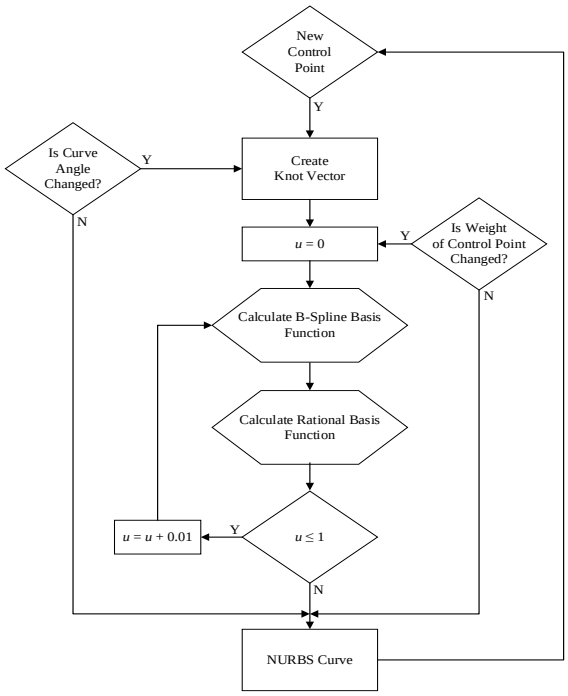


Figure 6: Flow diagram of NURBS algorithm.

IV. FPGA HARDWARE DESIGN

The designed control board provides serial channel and Ethernet communication interfaces and LCD control interfaces. The touch screen is communicated through the Inter Integrated Communication (I2C) interface. The LCD screen is controlled via Low Voltage Differential Signal (LVDS) discrete lines. Designs of control and communication blocks were made in VHDL language.

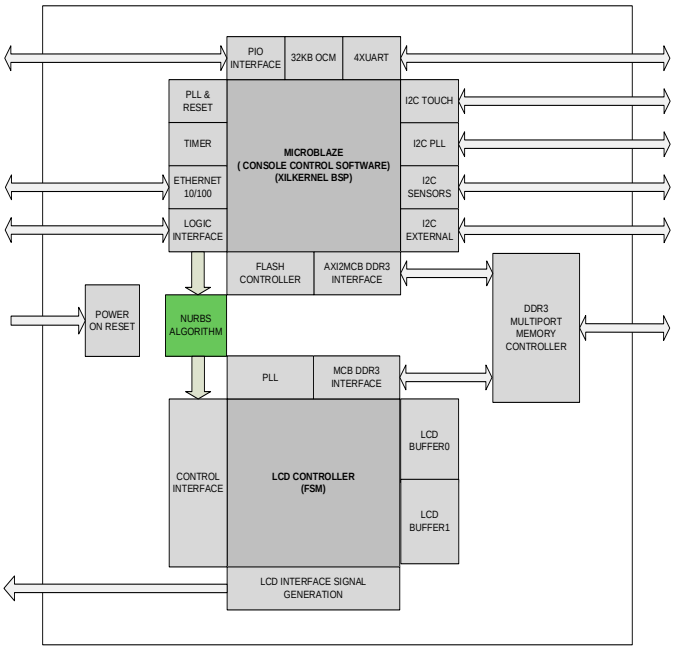


Figure 7: FPGA design.

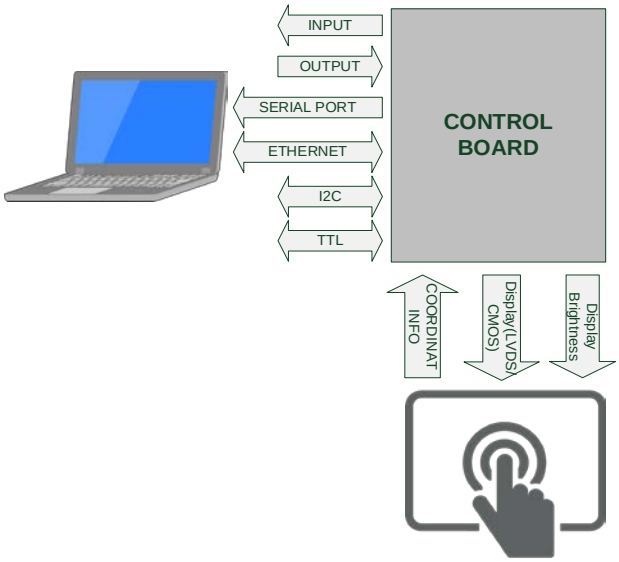


Figure 8: Hardware system design.

The NURBS algorithm was implemented in VHDL on the FPGA. In the implementation of NURBS algorithm, floating point processing elements in FPGA are used. In this developed system, the XC6SLX45 FPGA belonging to Spartan6 family of XILINX is used on the control board. The block diagram of the FPGA design is shown in Fig. 7.

V. NURBS CURVE MODELING SYSTEM ON FPGA BASED TOUCH SCREEN

NURBS curves are developed in real time on the touch screen with user interaction. The hardware design of the system is given in Fig. 8. The control points required for the modeling of the curve are determined on the touch screen. The NURBS algorithm given in Equation (2) starts to run in real time with the selected control points. Algorithm steps are performed on the FPGA in the control board. Each new point specified on the screen is considered as the last control point and the curve is re-modeled according to this point. The modeled curve is displayed on the screen in real time. A sample of a NURBS curve modeled on the system is given in Fig. 9.

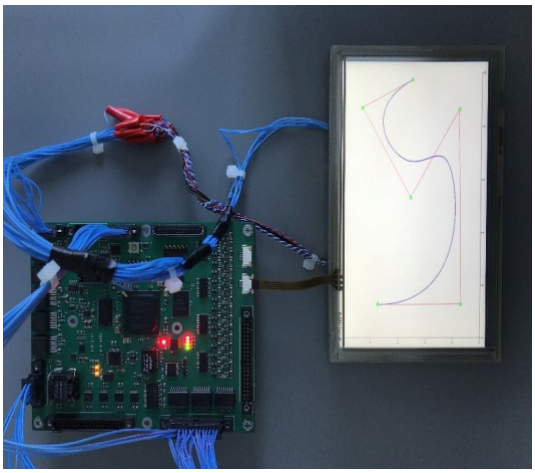


Figure 9: NURBS curve drawn in real time.

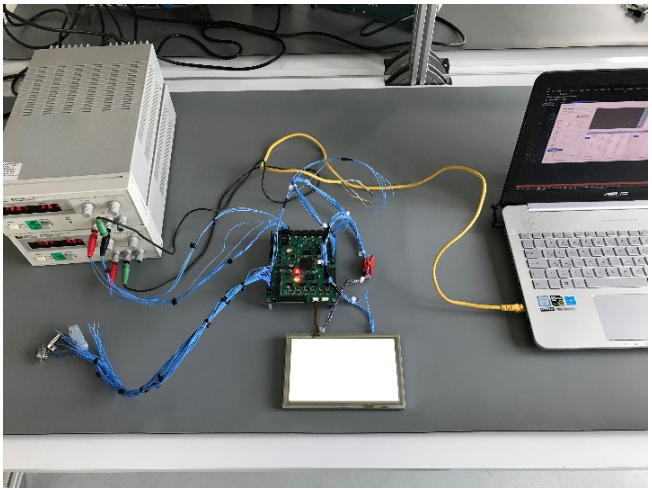


Figure 10: NURBS modeling system.

In the NURBS modeling method, it is very difficult to model the curve without using a computer or a microprocessor. Because even for a NURBS curve with 4 control points, there are approximately 25 transaction lines for $\forall u \in [0,1]$ value. In this case, if the *u* parameter is given sensitivity with a percent precision, the process line for modeling the curve is approximately 2500. This calculation is much more for a surface model with thousands of control points, so even the capacity of the microcontroller to be used is important. For this reason, it may not be possible to realize the NURBS algorithm in any microprocessor.

In the system we have implemented, the control board displays the curve modeled on the screen by performing the NURBS algorithm with the coordinate information received using the touch screen. At the same time, it sends the results by communicating with the computer via Ethernet interface. The general structure of the system is given in Fig. 10. Since all these operations are performed at the same time, the processor capacity on the board must match this. Microcontrollers such as DSP may be insufficient in algorithms that require a lot of cyclic operation because they perform serial processing. Thus it is easier to perform these operations on FPGAs with parallel processing capability. Because the FPGA can perform an operation while at the same time it can receive the second operation's data and with all this, it can show the results on the screen. For these reasons, FPGA is used to realize the NURBS algorithm in this study.

VI. Experimental Studies

In the developed system, the coordinates of the control points required for the algorithm are taken from the user through the screen and the NURBS curve is generated by realizing the NURBS algorithm on the FPGA is displayed on the screen. The control points and the control polygon of the points along with the curve are marked on the screen. At the same time, the NURBS curve was modeled with an application developed on the Visual Studio platform on the computer and the results were compared. The results of the execution times for NURBS modeling systemare given in Table 1.

Table 1: Execution times of NURBS modeling system.

Number of Control Points	Computer (μs)	FPGA (μs)
4	17000	85
8	19000	96
16	26000	130
25	30000	150
50	44000	220
80	58000	290
100	74000	370
120	90000	450

When obtained results are compared, the implementation speed of the FPGA is approximately 200 times faster than the computer. The computer used in the study has a 2.6GHz Intel i7 processor and 16GB RAM while the FPGA has a clock speed of 100MHz. As a result, it has been determined that the use of FPGA in modeling systems with many processing lines developed in industrial areas in CAD/CAM systems significantly affects the system speed. The performance of the systemwill be even better when a faster FPGA is used.

VII. Results

In this study, we developed a real time NURBS modeling algorithm with user interaction on the touch screen. Similar studies have been investigated by searching the literature for the hardware to be used in the system. From these studies, FPGA is preferred because of the parallel processing ability of the algorithm for the speed performance of the system. Mathematical solutions are presented for the NURBS algorithm. The algorithm is also implemented in a computer environment and the obtained results are compared. According to the results, acceleration was achieved about 200 times. Future work will be based on the CAD/CAM system performance by applying it to the mechanical design of the modeling system.

References

[1] L. Piegl and W. Tiller, "The NURBS Book," *Springer-Verlag, New York*, 1997.

[2] D.F., Rogers, "An Introduction to NURBS:With Historical Perspective," *Morgan Kaufmann*, 2001.

[3] L. Piegl and W. Tiller, "Curve and surface constructions using rational B-splines," *Computer-Aided Design*, 19(9):485-498, 1987.

[4] H. Shim and E. Suh, "Contact treatment algorithm for the trimmed NURBS surface," *Journal of Materials Processing Technology*, 104(3):200-206, 2000.

[5] E. Ulker and A. Arslan, "Automatic knot adjustment using an artificial immune system for B-spline curve approximation", *Information Sciences*, 179(10):1483-1494, 2009.

[6] E. Ulker and A. Arslan, "The calculation of parametric NURBS surface interval values using neural networks," *ICCS:2006*, pp. 247-254, 2006.

[7] M.C. Cheng, M.C. Tsai and J.C. Kuo, "Real-time NURBS command generators for CNC servo controllers," *International Journal of Machine Tools & Manufacture*, 42(7):801-813, 2002.

[8] M.C. Tsai, C.W. Cheng and M.Y. Cheng, "A real-time NURBS surface interpolator for precision three-axis CNC machining," *International Journal of Machine Tools & Manufacture*, 43(12):1217-1227, 2003.

[9] J. Wu, H. Zhou, X. Tang and J. Chen, "Implementation of CL points preprocessing methodology with NURBS curve fitting technique for high-speed machining," *Computers & Industrial Engineering*, vol.81, pp.58-64, 2015.

[10] J.J. Perez, R.A. Rios, R.J. Troncoso and L. Velezquez, "FPGA-based hardware CNC interpolator of Bezier, splines, B-splines and NURBS curves for industrial applications," *Computers & Industrial Engineering*, 66(4):925-932, 2013.

[11] X. Meng, G. Hu, "A NURBS-enhanced finite volume solver for steady Euler equations," *Computers & Industrial Engineering*, vol.359, pp.77-92, 2018.

[12] X. Zhiming, C. Jincheng and F. Zhengjin, "Performance Evaluation of a Real-Time Interpolation Algorithm for NURBS Curves," *Int J Adv Manuf Technology*, 20(4):270-276, 2002.

[13] G.V.V. Kumar, P. Srinivasan, K.G. Shastry and B.G. Prakash, "Geometry based triangulation of multiple trimmed NURBS surfaces," *Computer-Aided Design*, 33(6):439-454, 2001.

[14] L.Piegl and W.Tiller, "Geometry-based triangulation of trimmed NURBS surfaces," *Computer-Aided Design*, 30(1):11-18, 1998.

[15] X. Yi Zhao, C.G. Zhu, "Injectivity of NURBS curves," *Journal of Computational and Applied Mathematics*, vol.302, pp.129-138, 2016.

[16] X.Y. Zhao, C.G. Zhu and H. Wang, "Geometric conditions of non-self-intersecting NURBS surfaces," *Applied Mathematics and Computation*, vol.310, pp.89-96, 2017.

[17] M.C. Tsai, C.W. Cheng, "Real-time variable feed rate NURBS curve interpolator for CNC machining", The International Journal of Advanced Manufacturing Technology, 23(11-12):865-873, 2004.

[18] Z. Huan, Z. Limin, X. Zhenhua and D. Han, "Development of FPGA Based NURBS Interpolator and Motion Controller with Multiprocessor Technique," *Chinese Journal of Mechanical Engineering*, 26(5):940-947, 2013.

[19] Z. Huan, Z. Limin, X. Zhenhua and D. Han, "Design of a FPGA-based NURBS interpolator," *Conference [Online]*. Available: https://www.researchgate.net/publication/221105501.

[20] H.T. Yau, M.T. Lin, and M.S. Tsai, "Real-time NURBS interpolation using FPGA for high speed motion control," *Computer-Aided Design*, 38(10):1123-1133, 2006.

[21] D. Breen, W. Regli and M. Peysakhov, "B-Splines and NURBS," *Geometric and Intelligent Computing Laboratory Department of Computer Science Drexel University*, [Lecture Notes], pp. 22.

[22] C. Kadapa, W.G. Dettmer and D. Peric, "NURBS based least-squares finite element methods for fluid and solid mechanics", *International Journal for Numerical Methods in Engineering*, 101(7):2015.

Intrusion Detection with Probabilistic Neural Network: Comparative Analysis

İ.ATAY[1]

[1] Okan University, Istanbul/Turkey, contact@ibrahimatay.com

Abstract **- The use of machine learning techniques has significantly increased recently. The classification of normal or abnormal situations in network traffic is successfully applied with machine learning techniques. It is possible to encounter False Positive situations during the classification process. With Probabilistic Neural Network (PNN) model, it is aimed to explore the intrusion and its types within network traffic with probabilistic distribution. Knowledge Discovery Dataset (KDD99) will be used in this study.**

Keywords **- Probabilistic Neural Network, Intrusion Detection, KDD99.**

I. INTRODUCTION

The use of machine learning techniques has significantly increased recently. Especially, intrusion detection applications have increased with machine learning techniques. Intrusion detection systems monitor the information packages within the network and detect the normal and abnormal situations within the scope of the given rules

There are various data sets available in the literature to experience the success of intrusion detection systems. DARPA Intrusion Detection Evaluation (DARPA) and Knowledge Discovery Dataset (KDD99)[1] data sets are the most commonly used ones in the literature. Said data sets contain the traffic status of normal and abnormal situations realized in the network traffic.

Various machine learning techniques have been used to detect the intrusion within network traffic. Decision Tree, Support Vector Machines and Bayes are the most commonly used techniques in the studies [2-3]. However; it is an important factor to reach acceptable False Positive values in the techniques applied.

There are two approaches in intrusion detection systems being abuse and abnormality detection [4]. Detection of abuse involves the intrusion situations which are well-known and applied in the past. It is possible to detect such situations with the defined rules within the system. Detection of abnormality occurs with encountering unexpected situations within the network traffic. It is required to know the network traffic history in order to be able to detect the unexpected situations.

In this article, it is aimed to detect the normal and abnormal situations on the network traffic in the light of statistical values, using Probabilistic Neural Network (PNN)[5] model. The application has been developed using Visual C#. KDD99[6, 7] data set has been used in the study.

The article includes information about the data set in 2nd section, literature search about the study in 3rd section, problem to be examined in 4th section, Probabilistic Neural Network(PNN) in 5th section, the application realized in 6th section, and the conclusions of the study in 7th section.

II. DATASET

Knowledge Discovery Dataset (KDD99) is the data set used for performance comparison in intrusion detection systems. Data set includes unprocessed TCP data which was monitored for 9 weeks. Training data set contains 7-week and 5 million connections. Test data set contains 2-week and 2 million connections.

Test data set includes additional intrusion situations which are not included in training data set in order to experience the success of the intrusion situations. Training data includes 39 intrusion situations, and test includes 22 intrusion situations different from training data [1, 6]. The intrusions in data set are classified in 4 main categories. These categories are Denial of Service (DOS), Probing (Probe), Remote to Local (R2L) and User to Root (U2R).

There are 494,021 records in KDD99 data set. There are 97,277 (19.69%) normal, 391,458 (79.24%) DOS, 4,107 (0.83%) Probe, 1,126 (0.23%) R2L and 52 (0.01%) U2R intrusion connections in the records [7]. There are 41 attributes in data set and each has a label assigned as intrusion type or normal. Table 1 shows the labels and number of samples that appears within the data set.

Table 1: Class labels and the number of samples that appears [7].

Attack	Number of Samples	Minus Repeated Number of Samples	Class
Black	2,203	994	DOS
Land	21	19	DOS
Neptune	107,201	51,820	DOS
Pod	264	206	DOS
Smurf	280,790	641	DOS
Teardrop	979	918	DOS
Satan	1,589	908	PROBE
Ipsweep	1,247	651	PROBE
Nmap	231	158	PROBE
Portsweep	1,040	416	PROBE
Normal	97,277	87,831	NORMAL

Guess_passwd	53	53	R2L
ftp_write	8	8	R2L
Imap	12	12	R2L
Phf	4	4	R2L
Multihop	7	7	R2L
Warezmaster	20	20	R2L
Warezclient	1,020	1,020	R2L
Spy	2	2	R2L
Buffer_overflow	30	30	U2R
Loadmodule	9	9	U2R
Perl	3	3	U2R
Rootkit	10	10	U2R

III. RELATED WORKS

First intrusion detection system was designed by James P. Anderson[12] in 1980. Anderson designed the system so as to monitor, inspect and control all incidents occurring in computer systems or network traffic. If there is any security problem encountered within the process, a warning will be sent to the relevant personnel and unit and the measures will be taken for possible risks. Figure 1 shows the design of Anderson.

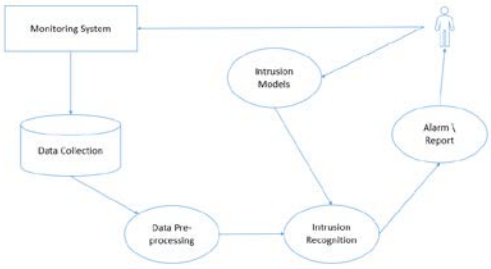


Figure 1: James P. Anderson Intrusion Detection Model.

The types of intrusion which can be applied within network traffic are described below.

- **Denial of Service (DOS):** Intrusion is made by sending connection request to a server more than it can carry, exploiting the gaps of TCP/IP protocol. The use of service by the actual users is prevented as a result of the intrusion.
- **Probing (Probe):** It is made by monitoring the active server or systems to learn IP address, open port or operating system.
- **Remote to Local (R2L)** : These are the situations in which the unauthorized access by users is encountered.
- **User to Root (U2R):** It includes the situations in which the user can make transactions beyond his/her authorization.

Various machine learning techniques have been used in the study to detect intrusion in network traffic. Decision Tree, Support Vector Machines and Bayes are the techniques which are most commonly used in the studies.

IV. PROBLEM DEFINITION

There are many traffic patterns within network traffic. Many intrusion patterns are recorded in databases thanks to past experience. However, developing computer systems brought with them the formation of different intrusion patterns.

In the intrusion model designed by James P. Anderson[8], there are network traffic monitoring, inspecting and controlling functions. However, the intrusion techniques differ with the development of computer systems. Therefore, machine learning models which can form patterns regarding new intrusion situations should be used within network traffic.

It is required to learn based on past data and to form new patterns based on new data in the process of forming new patterns in intrusion detection applications. This will cause encountering False Positive situations. However, the success of the process has increased with the use of acceptable threshold value in the study.

V. PROBABILISTIC NEURAL NETWORK (PNN)

Probabilistic Neural Network(PNN) [5] model is also known as Bayes-Parzen in the literature. PNN has the same advantages as statistical classification model. Additionally; Back-Propagation Neural Network model has advantages with its calculation power.

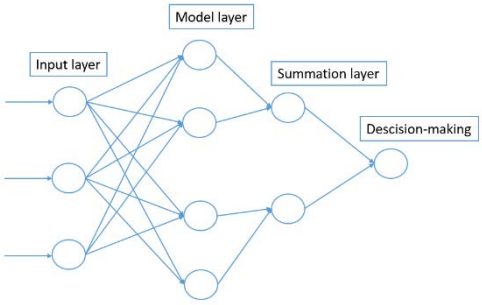


Figure 2: Layer PNN Model.

PNN can be designed with 4 layers as shown in figure 2. In the first layer of the design, the attributes are input and the weights are calculated (W). In the second layer, the connection between the input attributes and weight values should be established. Using the transfer function, vector at the size of 1xN is added to the connections established. Transfer function is shown in figure 3.

$$S_{ij} = exp\,(-\frac{||\,X_j - X_i||}{\sigma^2})$$

Figure 3: PNN Transfer Function [7]

Transfer function range is 0 and 1. As the weight distances between the transferred vector and model vector increase, output values are approached. In the third layer, close neural templates are formed. Number of neural and number of model is equal at this layer. Layer function is shown in Figure 4.

$$f_k(x_i) = \frac{1}{M_k} \sum_{\forall x_i \in y_k} S_{ij}$$

Figure 4: 3 Function of formation of layer template values [7]

In the fourth section, vectors are classified depending on Bayesian decision rules. Layer function is shown in figure 5.

$$P(y_i = 1\,|X_i) > \frac{L_{1,0}(X_i) - L_{0,0}(X_i)}{L_{0,1}(X_i) - L_{1,1}(X_i) + L_{1,0}(X_i) - L_{0,0}(X_i)}$$

Figure 5: Bayesian function [7]

When erroneous classification is assumed in the transaction, situation input is made on the vector. In the other case, vector distribution is applied. Therefore, correct classification cost is always higher. In the function shown in Figure 5, classification score range is defined as [1].

VI. APPLICATION

In the network intrusion model designed by James P. Anderson [12], there are monitoring, inspecting and controlling stages. As KDD99 data set is used in the study, only controlling stage will be realized. The study has been performed using Visual C# language. The application has been developed asynchronously due to the process it realizes.

PNN model has been realized with the application prepared. PNN model consists of two parts being training and test. In training part, Bayesian model is used on data, and the classification model as shown in Figure 6 is formed. In Test, the classification model as shown in Figure 7 is applied with the experience obtained in training part.

Begin

Initialization

J = 0;

Do

J ← j + 1

Normalization: $x_{jk} \leftarrow \frac{x_{jk}}{\sqrt{\sum_i^d x_{ji}^2}}$;

Learning process: $w_{jk} \leftarrow x_{jk}$;

If x ∈ w_i then $a_{ic} \leftarrow 1$;

Until j = n;

End

Figure 6: Training Algorithm

Begin

Initialization

k = 0

x = test sample

Do

$k \leftarrow k + 1; z_k \leftarrow w_k^t x$

If $a_{kc} = 1$ **then** $g_c \leftarrow g_c + e^{\frac{x_k - 1}{\sigma^2}}$;

Until k = n;

Return class ← $arg\,max(g_i(x))$;

End

Figure 7: Test Algorithm

A series of transactions has been applied before KDD99 data set. As shown in Table 1, repeated transaction records have been cleared.

Table 2: Performance Values of Machine Learning Models related to KDD99 Data Set Classification.

	DOS	**PROBE**	**R2L**	**U2R**
Bayes	%99,62[9]	%100 [9]	%99,35[14]	%99,47[14]
Support Vector Machine	%100[10]	%100[12]	%99,42[15]	%100[10]
Decision Tree	%99,98[11]	%99,66[13]	%99,70[16]	%92,5[16]
Probabilistic Neural Network(PNN)	**%99,78**	**%99,30**	**%99,48**	**%99,85**

KDD99 data set classification study was performed with the application developed. The study is shown in Table 2 with performance comparison in literature study performed with KDD99 data set.

VII. CONCLUSION

This study has been prepared in order to model a system exploring the intrusion situations within the network traffic. Within the framework of the rules determined within network traffic, the most successful intrusion detection model is seen as Decision Tree. However, the intrusion situations realized outside the defined rules are not included in performance value. This situation may cause various loss in different intrusion situations.

PNN which is discussed in the study makes classification using Neural Network and Bayes statistical model. The success of the applied classification model will increase with the growth of training set. Additionally, memorization problem of Neural Network structures has been eliminated with the use of Bayes statistical model within the model. In Table 2, PNN has been compared to the other model applied on KDD99 data set in the literature.

REFERENCES

[1]P. Aggarwal and S. Kumar, "Analysis of KDD Dataset Attributes - Class wise For Intrusion Detection," *Procedia - Procedia Comput. Sci.*, vol. 57, pp. 842–851, 2015.

[2]J. Zhao, J. Zhao, and J. Li, "Intrusion detection based on clustering genetic algorithm," *2005 Int. Conf. Mach. Learn. Cybern.*, no. August, p. 3911–3914 Vol. 6, 2005.

[3] S. Sheen and R. Rajesh, "Network intrusion detection using feature selection and Decision tree classifier," *TENCON 2008 - 2008 IEEE Reg. 10 Conf.*, pp. 1–4, 2008.

[4]D. E. Denning, "An Intrusion-Detection Model," *IEEE Trans. Softw. Eng.*, vol. SE-13, no. 2, pp. 222–232, 1987.

[5]X. Huafeng, "Probabilistic Neural Network and Its Application," *2010 Int. Conf. Comput. Des. Appliations (ICCDA 2010)*, vol. 2, no. Iccda, pp. 0–3, 2010.

[6]"KDD Cup 1999 DataSet," 1999. .

[7]A. A. Olusola, A. S. Oladele, and D. O. Abosede, "Analysis of KDD &apos ; 99 Intrusion Detection Dataset for Selection of Relevance Features Analysis of KDD ' 99 Intrusion Detection Dataset for Selection of Relevance Features," vol. I, no. January, pp. 16–23, 2016.

[8]James P. Anderson, "Computer Security Threat Monitoring and Surveillance." 1980.

[9]F. Jemili and M. Zaghdoud, "Intrusion Detection based on ' Hybrid ' Propagation in Bayesian Networks," *Intell. Secur.*, 2009.

[10] Q. Mu, Y. Chen, and Y. Zhang, "Incremental SVM algorithm to intrusion detection base on boundary areas," *2012 Int. Conf. Syst. Informatics, ICSAI 2012*, no. Icsai, pp. 198–201, 2012.

[11] V. Sharma and A. Nema, "Innovative Genetic Approach for Intrusion Detection by Using Decision Tree," *2013 Int. Conf. Commun. Syst. Netw. Technol.*, pp. 418–422, 2013.

[12] J. Wang, T. Li, and R. Ren, "Real time IDSs based on artificial bee colony-support vector machine algorithm," *3rd Int. Work. Adv. Comput. Intell. IWACI 2010*, pp. 91–96, 2010.
[13] M. Bahrololum, E. Salahi, and M. Khaleghi, "Machine Learning Techniques for Feature Reduction in Intrusion Detection Systems: A Comparison," *Comput. Sci. Converg. Inf. Technol. 2009. ICCIT '09. Fourth Int. Conf.*, pp. 1091–1095, 2009.
[14] D. M. Farid and M. Z. Rahman, "Anomaly network intrusion detection based on improved self-adaptive Bayesian algorithm," *J. Comput.*, vol. 5, no. 1, pp. 23–31, 2010.
[15] Y. Zhang, "Application of Improved Support Vector Machines in Intrusion Detection," no. 3, pp. 0–3, 2010.
[16] A. Alazab, M. Hobbs, J. Abawajy, and M. Alazab, "Using Feature selection for intrusion detection system," pp. 296–301, 2012.

How to manage Software Architecture documentation in Scrum Framework

M. SAVASCI, F. CETIN, C.CAKIR and O. FINDIK

Siemens A.S., Istanbul/Turkey, mustafa.savasci@siemens.com
Siemens A.S., Istanbul/Turkey, fatihcetin@siemens.com
Siemens A.S., Istanbul/Turkey, caglar.cakir@siemens.com
Karabuk University, Karabuk/Turkey, oguzfce@gmail.com

Abstract* - In software development processes, architectural documents can be prepared by the beginning of the project like Big Design Up front (BDUF) if software projects requirements are prepared with details and reviewed by the team. However, software projects with a lot of uncertainties or ambiguous requirements do not have any process to prepare and record architectural decisions during the software development. In recent years, agile software development has become very popular in the software industry. Therefore software development teams try to adopt their software processes against changing requirements and dynamic market conditions with using agile methodologies. Because of the underlying philosophy of agile, agile teams started paying more attention to working product over comprehensive documentation and big design up front. However agile software development also contains architectural and design decisions during the development. One of the agile principles also points that the best architectures, requirements, and designs emerge from the agile team during the development. Nevertheless, agile methodologies do not offer any processing cycle for architectural documentation in their process. In this article, we will propose how the architectural decisions will be documented in agile frameworks, which one is the most popular Scrum framework.

Keywords – Scrum, Software Architecture, Architectural Documentation.

I. INTRODUCTION

IN traditional software development processes like "The Waterfall Model" [1], architectural decisions and documents are prepared at the beginning of the project like Big Design Up front (BDUF). However, in the early phase of the software project, lots of different variables and uncertainties exist like details of the product, project plan, undetailed or unambiguous requirements [2]. Uncertainties and complexities in software project cause the wrong architectural decisions which are decided by software architects or software development team. In last decade, agile software development has become very popular in software development, especially software companies who'd like to adopt software processes against changing requirements, complex problems, and dynamic market conditions. Because of the underlying philosophy of agile, agile teams started paying more attention to working product over comprehensive documentation [3] and big design up front. However agile software development also contains architectural and design decisions during the development. [4, 5] One of the agile principles [3] also points that the best architectures, requirements, and designs emerge from a self-organized team during the development. Nevertheless, agile methodologies (like scrum framework) do not offer any process for architecture documentation and also any role for a software architect. In this paper, we will propose a process for how to handle software architecture documentation in the scope of Scrum framework without violating any scrum values and agile principles.

The paper is structured as follows: the next sections describe Scrum Framework, agile architecture (emergent architecture) and agile architectural documentation, how to combine Scrum and software engineering best practices; and then finally we proposed an architectural documentation review meeting for Scrum practitioners.

II. SCRUM FRAMEWORK

Scrum word was first used by Takeuchi and Nonaka in their article "The New New Product Development Game" [6] at 1986. Their article was shown that great performance is achieved when teams are small and self-organized and when such teams are fed with objectives and not managed via command and control way. Scrum framework inherits same philosophy and also name from this article for developing complex products [7, 8]. Scrum framework is defined in Scrum Guide as: "A framework within which people can address complex adaptive problems, while productively and creatively delivering products of the highest possible value" [8].

Scrum is simple and lightweight agile framework, straightforward, and easy to implement in different business areas. Due to these advantages, scrum is easily understood and enacted by the community. Recent agility report also shows that Scrum is the most common agile approach used by the software organizations [9].

The Scrum framework defines roles, events, artifacts, and rules by default. These scrum components are mentioned in the scrum guide as follows: "Each component within the framework serves a specific purpose and is essential to Scrum's success and usage" [8]. Any change in these scrum components cause a lost value in the scrum. Instead of traditional software development roles, scrum has only three

roles: Product Owner, Development Team, and Scrum Master. When we check the scrum guide, architecture or architect word is mentioned only two place: first sentence is "Scrum recognizes no sub-teams in the Development Team, regardless of domains that need to be addressed like testing, *architecture*, operations, or business analysis;" and the second sentence is "They collaborate and interoperate through sophisticated development *architectures* and target release environments."[8]. Role definitions of Scrum and also these sentences prove that scrum does not have an architect role. A development team is responsible for software architecture and architectural related work in Scrum.

III. AGILE ARCHITECTURE

Kruchten defines agile architecture as "The architecture should emerge gradually sprint after sprint, as a result of succession of small refactoring" [5]. "Emerge" and "Refactor" words are critical words for understanding the agile architecture. Emergence is defined by Goldstein: "[Emergence is] the arising of novel and coherent structures, patterns, and properties during the process of self-organization in complex systems" [10]. A refactoring always purposes to produce the simplest and smallest architecture that works and fulfills existing functionality in software. Because the simplest and smallest software architecture is the best architecture, can be maintained and changed easily by a development team [11, 28].

Many agile practitioners perceive software architecture and its related work in the context of the plan-driven development paradigm [12]. But on the other hand, research of Falessi shows that %95 percent of IBM software developers perceived software architecture as important for agile values and complex software development [13].

A good question arises in here "how can align scrum and agile architecture". While the scrum framework generally focuses to deliver a potentially releasable increment with functional requirements at most 30 days, software architectural approaches focus on developing software non-functional requirements like security and scalability. Scrum team delivers functional requirements for validating the non-functional requirements. Thus, the architectural works which are in the increment can be delivered along with the new functions of the software. While there are more architectural works in the first sprints, more improvements will be made with the progressive sprints that provide new features [14]. However, the Scrum team should be aware of business priorities that are trying to prevent good architecture [15]. If Product Owner does not prioritize architectural work, the development team should guide the product owner on the importance of architectural work.

The scrum framework also provides rapid feedback in the architectural solutions on the working software. This feedback loop provides great chance for all stakeholders to find better architectural solutions later that they couldn't have understood sooner [16]. Incremental architecture evolution keeps the software synchronization with changing market conditions.

IV. ARCHITECTURAL DOCUMENTATION

Architecture documentation is one of the key documentation for software documentation due to its critical content. But in agile software development processes, software documentation is often considered as an item that should not be given importance or even should not be applied. Nevertheless, Agile manifesto mentioned, "Working software *over comprehensive* documentation" [3]. In the truth of fact, agile processes mean that comprehensive documentation is less valuable than working software due to unnecessary and comprehensive maintain costs, and these documentation works have not any value without working software [17]. As well as out-of-date architectural documents can mislead software developers or cause them to mismanage the software. From understanding the agile documentation, do not document everything like waterfall documentation, document important information in "just enough" [18].

As indicated in the agile manifesto "The most efficient and effective method of conveying information to and within a development team is face-to-face conversation". The most efficient method for communicating is two persons talking face to face in front of whiteboard [11]. Documentation, which is a one-way communication style, can never be replaced by face-to-face communication, which is a two-way communication style. According to Coplien and Bjornvig, "all documentation (including architectural) is written for two reasons: to remember things, and to communicate them." The main purpose to create architectural documentation is to save the design and its logic as a kind of collective memory [19].

Instead of using architectural documents, some scrum teams use "design wall" to get an overview of the software architecture. The "design wall" is just a big whiteboard containing the most important design scrawls and printouts of the most important design documentation [20]. Unfortunately, these "design walls" do not provide any solution for a history of the transition of architecture and not suitable for the not co-located software developers.

The most of the architectural documentation templates like "arc42" usually have more than 30-50 pages and contain lots of complex information [21]. As we have explained above, architectural documentation should serve in a lean and simple way. Traditional architectural document templates can be used, but these template documents must be reviewed by the development team and unnecessary parts in the document should be removed by the development team. Lean architecture documentation reduces documentation effort and eases to review, update and communicate [22].

V. SCRUMAND

Scrum is an agile mindset, a process framework and focuses on to develop a product in the complex and chaotic environment. Therefore, a body of the knowledge is defined explicitly in "The Scrum Guide" [23]. Due to Scrum does not fully describe the whole process as other software methodologies, every scrum team should develop own specific Scrum in an empirical way without violating any Scrum rules and values. As each scrum team develops its

own scrum, the scrum can deviate from the good or the bad way. Deviation of the scrum components in a bad way is called "ScrumBut" and deviation of scrum components in a good way is called "ScrumAnd" [24].

"Scrum And" is defined by Development teams, who understand scrum values and definitions and implement it correctly, can add best practices of software engineering to their software development activities and processes without violating scrum values. ScrumAnd example can be given as "We use Scrum and we are continuously building, testing and deploying our increments every Sprint" [25].

One of the ScrumAnd implementations will be given in the next section for implementation of architecture documentation review meeting.

VI. SURVEY ABOUT SOFTWARE ARCHITECTURE DOCUMENTATION

An internal simple survey was conducted to see roughly software architecture processes and software architecture document processes in our company's agile projects. 26 agile practitioners participated in this survey and all participants' projects have a software architecture document.

65 percent of participants' software architecture document is updated by software architects and 35 percent of participants' software architecture documentation is updated by the development team. When we ask a question about the update status of software architecture documentation, the result shows that only 11 percent of software architecture document is up-to-date. 84 percent of participants said that software documentation update effort is a medium or little effort. 31 percent of software architecture documentation contains between 0 and 20 pages.

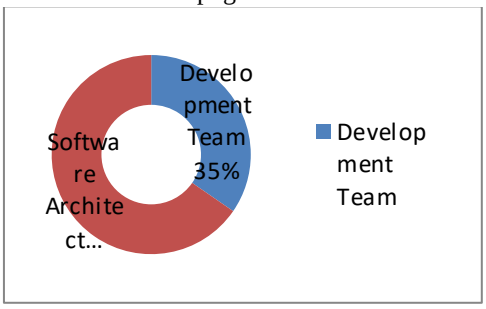


Figure 1: Responsibility of Software Architecture

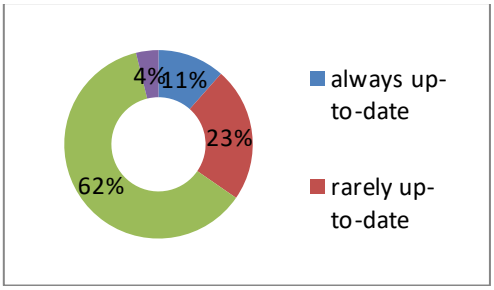


Figure 2: Update status of Software Architecture document

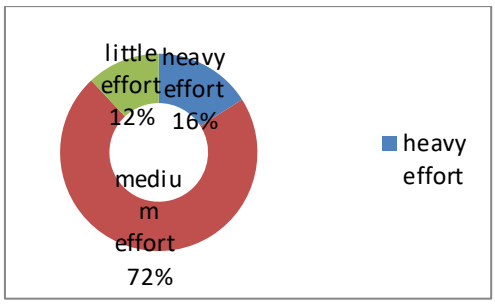


Figure 3: How much effort is needed for Software Architecture?

According to the survey respondents, only 11 percent of the documents appear to be always up-to-date, although they do not need too heavy effort to update the software architecture document and there is a responsible role for software architecture. Survey result shows that a defined process is needed for the software architecture in agile projects.

VII. ARCHITECTURE DOCUMENTATION REVIEW MEETING IN SCRUM

Scrum is founded on empiricism [8]. Scrum framework attempts to minimize the risks in an iterative way, using the knowledge and experience that emerges as the result of empirical experimentation. In the scrum, every single sprint and increment is actually a single project. Scrum limits the failure of a product to a sprint, preventing the product from going to general project failure. After every Sprint Review, done increment can be released by Product Owner. Like the product, software architecture is also a living artifact and evolves with respect to time.

Evaluation of the software architecture and documentation in each sprint by the scrum team and stakeholders ensures that the failure of the software architecture is limited to a sprint. Therefore, architectural documentation should be up-to-date and ready for the review meeting at the end of every sprint.

We propose a meeting where the software architecture is evaluated by Scrum Team and stakeholders. Architecture review meeting's scope and details should be decided by Scrum Team with respect to scrum values and agile principles. This architecture review meeting is a "ScrumAnd". Scrum master should monitor the architecture review meeting. If something does not fit the definition and values of the scrum, the scrum master should intervene in the meeting and ensure that the meeting serves the product properly. Architecture review meeting is at the most three-hour meeting for one-month Sprints like Sprint Retrospective. Scrum Team can decrease meeting duration but cannot increase the meeting duration with respect to the time-boxed rule of Scrum.

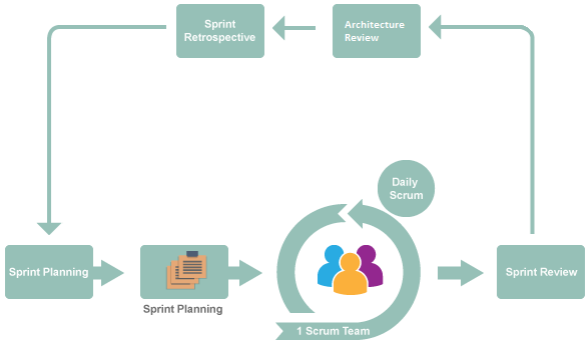


Figure 4: Scrum Event with Architecture Review Meeting

Architecture review meeting inputs are updated in the agile software documentation and the latest "done" increment. Architectural documentation provides high-level information to Stakeholders for the evaluation of the software architecture. Architecture documentation should not contain so much technical and technological details [26]. Technological and technical details cause duplication in architecture document and source code and also harden to communicate with stakeholders.

The architectural review meeting should be between Sprint Review and Sprint Planning. Done user stories from Sprint or the architectural decisions taken by the development team are updated with stakeholders and scrum team at this meeting. This meeting's outcomes provides for good feedbacks and also inputs for Sprint Planning meeting for better estimates of related user stories.

Architectural related developments together with non-functional requirements gain weight in order to achieve minimum viable product in the first sprints [27]. After reaching the minimum viable product, the balance between architectural related work and a functional requirement changes to the functional side. The first sprints architectural review meetings are much longer than the upcoming sprint meetings due to the above reasons. After a certain number of sprints, architectural change may not be necessary. In this case, the scrum team has to decide if a meeting is needed and if not, the scrum team has the right to cancel the meeting.

VIII. CONCLUSION

Instead of choosing between agility and software architectural related work, the development teams must integrate software architecture into their project as needed. Since the Scrum framework is particularly suitable for working in complex product development areas, it needs software architecture documentation to verify building right architecture with stakeholders.

The proposed architecture review meeting can be adapted easily in the Scrum teams. The scope of the architecture document and architecture review meeting can be adjusted by the Scrum Team respect to Scrum values. Unnecessary architectural documentation sections should be removed and inadequate parts strengthened.

IX. FUTURE WORK

In future studies, applying the proposed architectural documentation review meeting to the several scrum teams and gathering the implementation results to see and improve how useful the proposed work is to the scrum teams.

REFERENCES

[1] W. W. Royce, "Managing the Development of Large Software systems," Proceedings of IEEE WESCON 26, pp. 1-9, Aug. 1970.
[2] S. McConnell, Software Estimation: Demystifying the Black Art; Microsoft Press: Redmond, WA, USA, 2006
[3] K. Beck et al., "Agile Manifesto," [Online] Available: http://agilemanifesto.org/, 2001.
[4] P. Abrahamsson, M. A. Babar, and P. Kruchten, "Agility and architecture: Can they coexist?" IEEE Software, vol. 27, no. 2, pp. 16–22, Mar.–Apr. 2010.
[5] P. Kruchten, "Software Architecture and Agile Software Development-A Clash of Two Cultures?", proceeding of 32nd ACM/IEEE International Conference on Software Engineering, 2010.
[6] H. Takeuchi, I. Nonaka, "The new product development game", Harvard Business Review (64) pp. 137–146. 1986
[7] K. Schwaber, "Scrum Development Process," in OOPSLA Business Object Design and Implementation Workshop, J. Sutherland, et al., Eds., London: Springer, 1997.
[8] J. Sutherland, K. Schwaber. "The Scrum Guide." [Online] Available: http://scrumguides.org/scrum-guide.html, 2016.
[9] VersionOne Inc. (2017, April). The 11th annual State of Agile Report. [Online] pp. 10. Available: http://www.agile247.pl/wp-content/uploads/2017/04/versionone-11th-annual-state-of-agile-report.pdf
[10] J. Goldstein. Emergence as a construct: history and issues. Emergence: complexity and organization (1.1) 49–72. [Online] Available: http://www.anecdote.com.au/papers/EmergenceAsAConsutructIssue1_1_3.pdf; 1999.
[11] M.A. Babar, A.W. Brown, I. Mistrik, "Agile Software Architecture" Elsevier, pp. 192, 2014.
[12] R.L. Nord, and, J.E. Tomayko, "Software Architecture-Centric Methods and Agile Development", IEEE Software, 23(2): pp. 47-53. 2006.
[13] D. Falessi, G. Cantone, S.A. Sarcia, G. Calavaro, P. Subiaco, C. D'Amore. Peaceful coexistence: agile developer perspectives on software architecture. IEEE Software, 27, pp. 23-25, 2010.
[14] H. Doshi, "Scrum Insights for Practitioners", pp.100. 2016.
[15] J. Madison, Agile-Architecture Interactions. IEEE Software, 27(2), pp. 41-47, 2010.
[16] K. Schwaber, M. Beedle, Agile Software Development with Scrum, Prentice Hall, pp. 23–30, 2002.
[17] S. Sherman, I. Hadar, "Identifying the Need for a Sustainable Architecture Maintenance Process", 5th Int. Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), ICSE, June 2012.
[18] R. Hoda, J. Noble, S. Marshall, "Documentation strategies on agile software development projects." International Journal of Agile and Extreme Software Development, 1(1), pp. 23-37, 2012.
[19] J. Coplien, G. Bjornvig, "Lean architecture for agile software development". Chichester: Wiley; 2010.
[20] H. Kniberg. "Scrum and XP from the Trenches". InfoQ Enterprise Software Development Series 2nd edition- Director's cut, pp.68, 2015.
[21] Dr G. Starke. [Online] Available: http://arc42.org/examples
[22] I. Hadar, S. Sherman, E. Hadar & J.J. Harrison, 2013. Less is more: Architecture documentation for agile development. San Francisco, CA, IEEE, pp. 121-124, 2013
[23] K. Schwaber, [Online] "Scrum is simple, just use it as is!!" Available: https://kenschwaber.wordpress.com/2017/11/11/scrum-is-simple-just-use-it-as-is/ 2017.
[24] K. Schwaber, "Scrum But Replaced by Scrum And" https://kenschwaber.wordpress.com/2012/04/05/scrum-but-replaced-by-scrum-and/ 2012.
[25] V.A. Krishna, A.B. Basu, "Scrum: Is it Scrum But or Scrum And. En: Proceedings – 2011 Annual IEEE India Conference: Engineering Sustainable Solutions, INDICON-2011, 2011

[26] B. Selic. "Agile Documentation Anyone?". IEEE Software, Volume 26 Issue 6, IEEE Computer Society Press Los Alamitos, CA, USA, Nov., pp.11-12, 2009
[27] H. Doshi, "Scrum Insights for Practitioners", p 100 2016.
[28] H.P. Breivold, D. Sundmark, P. Wallin, and S. Larsson, "What does research say about agile and architecture?" Proc. 15th International Conference on Software Engineering Advances, August 2010, pp. 32–37, 2010.

Software Architecture Documentation in Agile

C.CAKIR, F. CETIN, M. SAVASCI and O. FINDIK

Siemens A.S., Istanbul/Turkey, caglar.cakir@siemens.com
Siemens A.S., Istanbul/Turkey, fatihcetin@siemens.com
Siemens A.S., Istanbul/Turkey, mustafa.savasci@siemens.com
Karabuk University, Karabuk/Turkey, oguzfce@gmail.com

***Abstract* - Over the last decade, agile practices become very popular amongst software development. According to one of the phrases of the Agile Manifesto "Working software over comprehensive documentation" sometimes could lead practitioners to the misunderstanding of "documentation is not valuable" or "not needed at all". Because of this understanding from software community, agile practitioners do not give enough attention to architectural related documents. However, documentation is also a communication way between people and this communication should be also simple and lean considering agile principles. Commonly used traditional architectural documentation is very comprehensive and detailed. Creation and maintenance of this architectural documentation take too much effort for agile teams. Therefore, existing architectural template documents cannot serve agile teams in the best way. Rather than using existing architectural documents, this article presents a new lightweight architectural documentation template that can be used maintained easily in agile projects.**

***Keywords* – Agile, Software Architecture, Architectural Documentation.**

I. INTRODUCTION

Nowadays more and more software-driven businesses feel the pressure of responding dynamic market requirements and satisfying customers' needs. These challenges push software industry to embrace agile software methodologies rapidly [1].

Since the publication of *Agile Manifesto* over a decade ago the use agile practices in the software industry has increased significantly as a mechanism for increasing adaptation to changing market requirements. In today's software business, usage of agile methodologies comes to the point that number of surveys shows that agile is now the dominant approach amongst software companies [1]. However, there are still questions and concerns regarding software architecture and documentation practices.

Software architecture describes the key components and their interactions with each other or different parts of the system; therefore it provides the skeleton of the system to all stakeholders [1, 2]. Regardless of the software development methodology, architectural information should be kept under documentation in order to build common understanding between stakeholders. According to Clements et al. documentation is important for building, maintaining and reusing software architecture [3].

According to the "*Working software over comprehensive documentation*" principle of agile, documentation effort mostly seen by practitioners as it is not required and time-consuming. In fact the underlying meaning of this principle is that comprehensive documentation is less important due to high cost of maintenance effort [4]. Considering the importance of the software architecture and agile approaches, finding the middle ground between these concepts could be an ideal way for the software industry and that's the aim of this article.

In the next sections, we will detail agile approaches and architectural related processes, and the tension between them. After that, we will discuss documentation effort of these two cultures. Finally, we will introduce a lightweight architecture documentation template that can be used in agile practices.

II. AGILE VS ARCHITECTURE

According to many agile practitioners, software architecture practices seem to be in the context of plan-driven development paradigm [5]. They feel big upfront design and maintenance of software architecture requires too much work and have little benefit to the customer, additionally all these efforts contains high ceremony processes [6]. Thapparambil [7] writes that "no agile methods discuss Architecture at any length." Architectural design is thought to be unimportant activity amongst agile approaches; therefore there is not enough detailed information about architectural related activities [8]. Thapparambil [7] also sees refactoring as the main activity to replace architecture in the agile world.

On another side, companies working with well-established architectural practices often tend to see agile methodologies as amateurish and more suitable for relatively small applications [9]. According to them, architectural related software development processes cannot be followed using the agile methodologies [6]. Satashi Basaki [9] criticizes agile practitioners with these words "It seems that many agile method users misunderstand what agile methods are, just ignore architecture, and jump onto refactoring". He believes threating "refactoring" as the only remedy for the architectural challenges is a big problem.

Considering both sides opinions it's obvious that the tension between two cultures lies under adaptation versus anticipation. Agile practitioners want to decide architectural decisions at the "last responsible moment". Because of the

changing requirements and uncertainties, they have right to think so [9]. On the other hand, some projects need to be planned carefully when the software quality expectations are in at a high level and changes in requirements do not expected to happen very often. Rick Kazman [1] noted this issue as: "No one would want to fly in an aircraft with flight control software that had not been rigorously planned and thoroughly analyzed. Similarly, no one would want to spend 18 months planning an e-commerce website for their latest cell phone model, or video game, or women's shoe style (all of which are guaranteed to be badly out of fashion in 18 months)"

However, paying more attention to architectural disciplines increased significantly amongst agile approaches [5, 10, 11]. Architects have also seemed happy with agile practices. Kati Vilkki surveyed more than 2,400 developers, testers, architects, and managers at Nokia Siemens Networks. The result of the survey showed that more than 70 percent of the architects were satisfied with agile development approaches in their work [12].

Most of the development teams frequently find themselves into this dilemma "should I do agile or architecture" but the real question should be "how much architecture should I need up front or how long should I delay this process until requirements become solid enough" [1]. A short answer to this question is "it depends". Rick Kazman details this answer with the following criteria:

- If you are working on a large complex system with well-defined requirements doing a large amount of architectural work at the beginning is beneficial.
- Be ready to change and maintain architecture when quality attribute requirements emerge and solidify.
- On smaller projects with uncertain requirements, try to find optimal architectural concepts with major functionalities are included. Do not spend lots of time with upfront work.

Both of the two cultures carry their own good values inside. Except for some specific domains, it should not become a choice to pick one of them. Mixing and choosing the right time and effort to combine these cultures might be the key to the success in the software industry.

III. ARCHITECTURAL DOCUMENTATION

Because of the "*Working software over comprehensive documentation"* [13] principle, documentation gets a bad reputation amongst agile approaches. In fact, the underlying meaning of this principle is, eliminate creation and maintenance of the unnecessary and comprehensive documentation due to its high cost [4]. But many agile practitioners forget the fact that documentation has two valuable assets "communication" and "collective memory" [1, 14]. Architectural documentation considered as one of the most important document due to containing the key information about the product which every stakeholder might need it for different reasons. Another principle of agile development "*the most efficient and effective method of conveying information to and within a development team is face-to-face conversation*" describes the way of knowledge sharing, but Clements at al [15] refuses this idea and points the importance of the architectural documentation with these words: "…think of a maintainer, who has inherited the system years after all the original developers have left, trying to understand where to begin. On a project involving hundreds of developers, do you really want your architects to spend all of their time answering the same questions over and over? Or would you rather let documentation serve that purpose, while also making sure the developers get the same answer every time?"

Considering the philosophy of the agile the amount of architectural documentation should be "just enough" [16]. If no one is going to read it, there is no point to generate it but if there is anyone, the document should be lean and simple enough. According to Cockburn, the correct amount of architectural documentation is exactly as needed to accomplish the next step. Any other effort to exceed this limit is a waste of time and money [17]. Finding the middle ground between agile and architecture documentation process brings the following benefits to the product:

- Achievement of critical quality attributes.
- Building a communication channel for the future generations.
- Enabling general system analysis.

IV. PROPOSAL

The main challenges related to architectural documentation could be listed as:

- Understandability of the document by each stakeholder (architects, reviewer, analyst, developer etc.), the complexity of the document.
- Setting the correct information to the right place or finding required information easily [18, 19].
- Keeping the document up to date. Stakeholders eventually will lose the confidence of quality for outdated document [19, 20].

The purpose of the lightweight architectural documentation template is to help software development teams to eliminate or minimize the challenges stated above; another goal of this proposal is, decreasing the creation and the maintenance effort of the architectural document.

The contents of proposed template inspired from one of the most common architectural document templates of "arch42" [21]. As a natural behavior of agile, the contents of this template can be changed from team to team. The main motivation when editing the document should be "Describe what you can't get from code". Here are the contents of the proposed lightweight architecture documentation template:

A. Context

A short set of explanations, context diagram (or both) that answer these questions:

- What is the system all about?
- Who are the users of this product?
- What is the main motivation to build this product?

B. Functional Overview

Summarization for the key functionalities of the system (use cases, user stories)

C. Quality Attributes/Tree

A small set of the most critical quality attributes like performance, security, flexibility etc. Short explanations could be filled for each attribute.

D. Constraints

A small set of the most important constraints that need to be aware of.

E. Software Architecture

This section is the big picture of the system and can contain different types of view like logical, development, physical views. It is preferred to put only diagrams into this section.

F. Decision Log

A short paragraph or set of architectural decision records which explains the most important decisions.

V. CONCLUSION

Agile and architectural processes seem like totally different concepts but both of the cultures carry great artifacts for the software industry. Although agile philosophy advice less documentation, creating a simple and lean software architecture documentation brings great benefits to working product. Documentation creates an alternative communication channel and builds historical memory. These benefits would likely to pay off after the team member changes or project size grows up after few years.

REFERENCES

[1] M.A. Babar, A.W. Brown, I. Mistrik, "Agile Software Architecture" Elsevier, pp. 192, 2014.

[2] R.C. de Boer and H. van Vliet, "On the similarity between requirements and architecture," Mar. 2009.

[3] P. Clements, D. Garlan, R Little, R. Nord, and J. Stafford. (2003). Documenting Software Architectures: Views and Beyond. Proceedings of the 25th International Conference on Software Engineering, pp. 740 741.

[4] S. Sherman and I. Hadar, "Identifying the Need for a Sustainable Architecture Maintenance Process", 5th Int. Workshop on Cooperative and Human Aspects of Software Engineering (CHASE), ICSE, June 2012.

[5] R. Nord,.L, and J.E Tomayko, Software Architecture-Centric Methods and Agile Development, IEEE Software,2006. 23(2): pp. 47-53.

[6] M.A. Babar, 2009. An exploratory study of architectural practices and challenges in using agile software development approaches. Proceeding of the 2009 Joint Working IEEE/IFIP Conference on Software Architecture and European Conference on Software Architecture, WICSA/ECSA 2009, Cambridge, pp: 81-90

[7] P. Thapparambil, Agile architecture: pattern or oxymoron?, Agile Times, 2005. 6(1): pp. 43-48.

[8] C. Hofmeister, et al., A general model of software architecture design derived from five industrial approaches, Journal of System and Software, 2007. 80(1): pp. 106-126.

[9] P. Abrahamsson, M. A. Babar, and P. Kruchten, "Agility and architecture: Can they coexist?" IEEE Software, vol. 27, no. 2, pp. 16–22, Mar.–Apr. 2010.

[10] T. Ihme, and P. Abrahamsson, Agile Architecting: The Use of Architectural Patterns in Mobile Java Applications, International Journal of Agile Manufacturing, 2005. 8(2): pp. 1-16.

[11] R. Parsons, Architecture and Agile Methodologies - How to Get Along, in WICSA. 2008.

[12] K. Vilkki, "Impact of Agile Transformation," Flexi Newsletter, vol. 2, no. 1, 2008, pp. 5–6.

[13] K. Beck et al., "Agile Manifesto," [Online] Available: http://agilemanifesto.org/, 2001.

[14] A. Cockburn, Agile software development: The cooperative game, 2nd ed. Reading, MA: Addison-Wesley 2007.

[15] P. Clements, J. Ivers, R. Little, R. Nord, and J. Stafford, (2003), "Documenting Software Architectures in an Agile World", Carnegie Mellon University, Software Engineering Institute, CMU/SEI2003-TN-023, Pittsburgh, PA.

[16] R.Hoda, J. Noble & S. Marshall, "Documentation strategies on agile software development projects." International Journal of Agile and Extreme Software Development, 1(1), pp. 23-37, 2012.

[17] A. Cockburn, Agile Software Development. Boston, MA: Addison-Wesley, 2002.

[18] R. C. de Boer and H. van Vliet, "Architectural Knowledge Discovery With Latent Semantic Analysis: Constructing a Reading Guide for Software Product Audits", Journal of Systems and Software, vol. 81, no. 9, pp. 1456-1469, 2008.

[19] S. Sherman, I. Hadar, and M. Levy, "Enhancing Software Architecture Review Process via Knowledge Management", The Sixteenth Americas Conference on Information Systems, AMCIS, Lima, Peru, August 2010.

[20] A. Jansen, P. Avgeriou, and J.S. van der Ven, "Enriching Software Architecture Documentation," The Journal of System and Software. vol. 82, pp. 1232-1248, 2009.

[21] Dr. G. Starke, [Online] Available: http://arc42.org/

Tracking the Architectural Quality: "W-Model of Software Architecture"

O. TAVILOGLU and F. CETIN

Siemens A.S., Istanbul/Turkey, onur.taviloglu@siemens.com
Siemens A.S., Istanbul/Turkey, fatihcetin@siemens.com

Abstract **- Keeping the quality goals in software is not a matter of chance. It needs proper planning and continuous attention of the development team. The main responsible for achieving the quality goals is the software architect. The architect needs to contribute in: clarifying and defining the goals, make a design that maps the required quality. However, that's not enough. The architect needs to govern all development and test activities to move towards the goals in harmony as a team. In the end, software quality depends on the design decisions, implementation and the also definition of the goals itself. A process for tracking the architectural quality is presented in this paper with an analogy to the development processes: "V Model" and "W Model".**

Keywords **Software Architecture, Software Quality, Architecture Review, Non-functional Requirements, Quality Attribute Workshop**

INTRODUCTION

SOFTWARE development processes present activities, roles, artifacts, meetings and lots of relations to come up with predefined steps for developing good quality software. Although being criticized by the agile practitioners, the V Model of software development [1] is still a good resource for providing an overview of software development activities that are both horizontally and vertically traceable. The W Model is introduced by Paul Herzlich [2] after the V Model. It can be seen in figure 1. In "W Model" planning the test activities and designing the software starts in parallel.

The proposed method in this paper adapts the W Model to handle the architectural activities while keeping the essence same. A set of techniques for tracking architectural quality will also be presented.

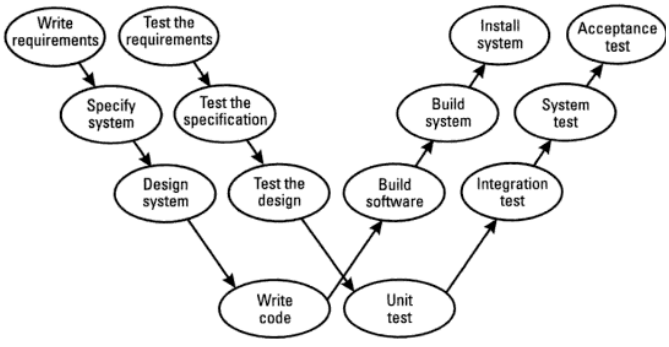


Figure 1: The original W Model [2]

Software architecture might be named as the infrastructure of the software. Software without an infrastructure might address the required functionality but not probably in a way that it also addresses performance, reliability, usability, and scalability requirements -so-called quality attributes- in a sustainable way. The software infrastructure (or software architecture) should address the quality attributes (QA). The quality attributes should be identified, designed, implemented, and tested in a proper way. The core functionality is also considered architectural. It has an effect on the architectural design as suggested in Attribute Driven Design [3]. However, only the quality perspective of software architecture is in the focus of this paper.

I. ADAPTING THE W MODEL FOR SOFTWARE ARCHITECTURE

In the original presentation of the V-Model, time is an axis. But here in the W Model of architecture, the time is not clearly defined. Only, the relations between activities are shown. The relations can be read as Architecture Design needs Quality Attributes properly defined. Architecture Review is needed to verify that the Architecture Design is correctly made. The original W Model also uses the Risk-Based Testing Methodology. Here the architecture review methodologies get integrated with Risk-Based Testing to improve the original W model.

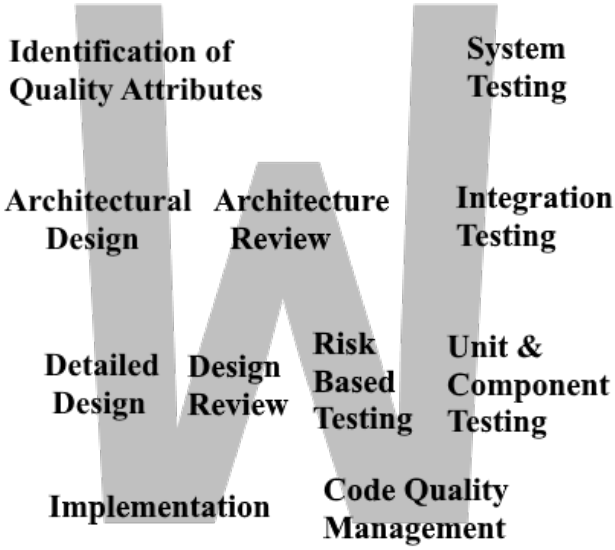


Figure 2: W Model of software architecture

II. ELEMENTS IN THE MODEL

The elements in the model are activities that needed to track the architecture. They are described in the next section in detail while making suggestions on methodologies and techniques that can be used within the activity.

A. Identification of Quality Attributes

Quality attributes are the key input to the architecture design. They define how the software system should behave in particular environments. The most common way of representing a quality attribute is to use quality attribute scenarios (QAS). They can be identified by quality attribute workshop [4] as suggested by SEI. However, it requires a wide range of contributors from marketing to the project management and test team. It might also require for them to spend a couple of days on the topic. Finding a time slot for all stakeholders might not be possible all the time. Therefore it is also common to use Quality Attribute Utility Tree and Mini Quality Attribute Workshop [5]. The Quality Attribute Utility Tree only presents the architect's perspective. In that case, this perspective has to be agreed on by the product manager. The Mini Quality Attribute Workshops simulate the regular "quality attribute workshop" with fewer people by assigning all roles of stakeholders to the limited number of contributors. Whichever method is used, the first step in the architecture process should be to define the Quality Attributes in a testable way.

B. Architectural Design

Architecture design is similar to the generic software design. It has similar steps like, identification of requirements, defining the scope, creation of software objects/components and structuring them. Attribute Driven Design focuses on the structuring part. It provides well-defined steps to structure the software objects according to quality attributes.

The main difference of architecture design from the generic software design is that it has to address the quality attributes. ADD suggests using design tactics to achieve them. Architectural blueprints, domain-specific patterns can be used as well.

The architect should also consider the process that is used by the project. If an agile methodology is used than the architecture should also grow with each increment. The architecture should be stable enough to bring a product like quality and also be flexible enough to embrace change [6, 7].

C. Detailed Design

Architecture design identifies systems, subsystems, and components. The entities that are in smaller granularity like subcomponents, units, modules, classes need to be designed as well. Design at this level is not the direct responsibility of the architect. However, the architect has to communicate the architecture and guide the development team in a certain direction so that they can prepare the detailed designs in harmony with the architecture design. The architect has to own the role of a coach and/or a sparring partner in that case. The development team might have expertise on technical constraints. The architect has knowledge of required quality attributes. They need to come up with a negotiation. The architectural design might be updated or refactored after the detailed design.

D. Implementation

The implementation should be done in compliance with the detailed design and hence with the architectural design. The code reviews or pair programming can be used to check the compliance.

E. Design Review

The architect reviews the detailed designs made by the developers/experts to see that they are not causing any architectural drift. The reviews shouldn't be necessarily done with official documentation or tools. It can be done with meetings. The design and code review can also be combined with a pair or mob programming or with the application of Test Driven Development.

F. Architecture Review

Architecture reviews are used to verify that the architectural decisions that are made in the Architecture Design phase are correctly addressing the architecturally significant requirements. It can also include steps to correctly identify the Quality Attributes. ATAM [8] methodology basically focuses on the Quality Attributes and the corresponding design decisions are analyzed to identify the risks, nonrisks, tradeoffs and sensitivity points as a result. There are different review methods for architecture reviews. The selection of the methodology depends on multiple factors. The project phase is one of the factors. However, the focus of this paper is on tracking the architecture quality continuously. From this perspective, the lightweight version of ATAM is a good candidate to be used as a part of the W Model. Lightweight version basically uses the same approach as the classical version. However, it does not require the high number of stakeholders to attend. The architects might drive the architecture review internally as well. Lightweight version only focuses on the Quality Attribute Utility Tree. On the contrary, the classic version has the brainstorming of the Quality Attribute Scenarios and the architectural analysis of the generated quality attributes scenarios included. By not including those parts in the lightweight version, some ceremonies of the classic version can be skipped. The lightweight version of ATAM can be used iteratively to make it as effective as the classic version [9].

Different scenario-based review methodologies all share the same essence:

1) Identify the scope of the review in terms of quality attributes.
2) Identify the architectural decision/patterns/tactics that address quality attributes.
3) Find the outcomes of using the identified patterns especially in terms of risks.

The steps above can be modified to use different sets of tools and techniques, also steps like presenting the business drivers

and the presentation of the architecture steps from ATAM can be added according to the following factors:

1) Development phase: In the first phase of the development it would also beneficial to include the review of the quality attributes in the review. By doing so, it can be made sure that QAs are correctly identified.
2) Reviewer Team: It wouldn't make sense to include the review of the QAs in the later phases of the development unless the reviewer team has some project external members. The project internals would be familiar with QAs in the later phases. However, it is very important to know the business factors that affect the design decisions when project externals are reviewing the project with a scenario-based methodology.

G. Risk-Based Testing

Risks and sensitivity points generated from the Architecture Review becomes an input for the Risk-Based Testing (RBT) Methodology [2]. RBT itself does not only focus on architectural decisions and quality attributes but on all risks. RBT can be used as the main testing strategy of any project. Even if it's not used as the main testing strategy it can be used to plan testing activities to test the risks regarding QAs. According to the nature of the risk, a decision is made to address it within a level of testing that is described after the Code Quality Management section.

H. Code Quality Management

SQALE methodology [11] is a good example of how static Analysis tools can be used to track architectural quality in terms of tracking the Technical Debt and more. Doing code reviews, using coding guidelines would also help to keep up with the desired quality level. However, CQM is not enough by itself to guarantee a certain level of quality.

I. Unit & Component Testing of Quality Attributes

After making risk analysis on QAs some of the risks might be addressed in unit and/or component testing level. Getting profiling information, stressing the limits of performance requirements, monitoring the memory management system regarding leaks or fragmentation can be done with the sense of unit and/or component testing.

J. Integration Testing

Most of the architectural decisions would be tested within the integration testing level. Integration testing normally uses gray box testing approach where specifically injected testing interfaces are used while the design internals is taken into consideration. The risks that are in the scope of multiple components would be addressed by integration testing. For example, if the reliability of the system is handled by an active redundant design, an integration test set would be a perfect candidate to test the systems redundancy behaviors.

K. System/Acceptance Testing

The system testing will always have some tests specifically for QAs no matter how good the testability is within the software as it focuses the system as a whole. Even if some of the risks are covered within the integration test phase, they can also be a part of the system test phase from the black box perspective.

III. Using the Model In An Agile Way

The steps defined in the previous section can be adapted to the used software development process. They all can be used in different phases and incrementally.

Identification of the quality attributes should be done at the beginning of the project and shouldn't be left in the later phases. The QAs should be ranked according to the importance and the expected cost. After defining them the architect can start designing a system that addresses the low hanging fruits from the business perspective. It would also a good choice for the architect to address the risky quality attributes in the first phases, at least with a prototyping approach.

The architect should design a system that grows with each increment as suggested in the architecture design section above. The communication and the runtime infrastructure should be the first topic that the architect should focus. The design might be evolved with each increment. Hence the architect should assist the project management in the project planning phase.

Detailed designs might be handled within each increment. It is not feasible to make an upfront detailed design in an agile project. It's the responsibility of the architect to guide the development team with their designs in each increment.

Design reviews might be a part of the detailed design process itself. The implementation part could also be integrated with the design and the review phases. Such as applying TDD and pair programming and making active design reviews between the architect and the development team in parallel.

Scenario-Based architecture reviews can be adapted or get integrated into the development cycles as described in the previous section. The risks identified by the architecture reviews will be an input to the Risk-Based Testing Workshops.

Risk-Based Testing workshops are an important activity within each increment. They should be a part of the increment planning meetings and the detailed design phase in each increment. All design decisions, technology selections; requirements might have an implicit effect on the risks. Also, the risks that are identified within the architecture reviews become an input for RBT workshops. For each QA, there should at least be one relevant test or activity. The relevance and the addressing of the QA are handled within the RBT workshops. In the end, the architect should have a traceable relation between the test and the QA. The risks might also be handled by additional architecture or design reviews. However, these will probably expert based reviews.

Static analysis should become a part of the check-in policy. The code that cannot fulfill the static analysis requirements shouldn't be merged to the major branch. A continuous integration tool might be used together with the quality tracking tools.

Unit/Components tests that are identified by the RBT should be a part of the definition of done. So that, the work of the development team cannot end without guaranteeing that the implementation and the design are free of identified risks.

In the integration test phase, the QAs should be tested from the regression testing perspective. Once an integration test is

identified in the RBT workshops, it becomes a candidate for regression test set as all development activities might affect QAs in a negative way even if there's no direct risk identified.

System/Acceptance testing should also be done iteratively. The agile philosophy suggests having a usable software part added to the system in each cycle. Hence the customer using the software in each increment becomes a tester. However, the customer might not have the necessary skills or the know-how to test the software in terms of the QAs. In these cases, the development team should use set of tests from the black box testing [12] perspective to stress the system as a whole in terms of QAs.

IV. CONCLUSION

The adaptation of W model for tracking the quality of software architecture is defined within the paper. The methodologies and techniques to be used within each element of the W Model of architecture are described in detail. The key approach of the model is getting the Risk-Based Testing and the architecture review methodologies to be used together in harmony on a quality attribute driven architecture design process. In the last section, it is exemplified how the model can be tailored for agile software development processes.

V. REFERENCES

[1] P. Rook, "Controlling software projects". *Software Engineering Journal*, pp.7-16, 1986.

[2] Gerrard, Paul, and Neil Thompson. "Risk-based e-business testing", *Artech House*, pp.56-59, 2002.

[3] L. Bass, M. Klein and F. Bachmann, "Quality attribute design primitives and the attribute driven design method", *International Workshop on Software Product-Family Engineering*, pp. 169-186. Springer, Berlin, Heidelberg, 2001.

[4] M.R. Barbacci, R.J. Ellison, A. Lattanze, J. Stafford, C. B. Weinstock, and W. Wood. "Quality attribute workshops", 2002.

[5] T. de Gooijer. "Discover Quality Requirements with the Mini-QAW", *Software Architecture Workshops (ICSAW), 2017 IEEE International Conference*, pp. 196-198. IEEE, 2017.

[6] S. Bellomo, P. Kruchten, R. L. Nord, and I. Ozkaya. "How to Agilely Architect an Agile Architecture.", *Cutter IT Journal* 27, no. 2, pp. 12-17, 2014

[7] R. L. Nord, I. Ozkaya, and P. Kruchten. "Agile in distress: Architecture to the rescue." *In International Conference on Agile Software Development*, pp. 43-57. Springer, Cham, 2014.

[8] R. Kazman, M. Klein, M. Barbacci, T. Longstaff, H. Lipson, and J Carriere. "The architecture tradeoff analysis method." *In Engineering of Complex Computer Systems, 1998. ICECCS'98. Proceedings. Fourth IEEE International Conference on*, pp. 68-78. IEEE, 1998.

[9] L. Bass, P. Clements, R. Kazman. "Software Architecture in Practice, Third Edition". *Addison-Wesley Professional*, pp. 415-417, 2012.

[10] F. Bachmann "Design Peer Reviews the ATAM Style." *CrossTalk November/December*. pp. 8-10, 2011.

[11] J.-L. Letouzey, "The SQALE method for evaluating technical debt." *In Proceedings of the Third International Workshop on Managing Technical Debt*, IEEE Press, pp. 31-36, 2012.

[12] L. Copeland, "A practitioner's guide to software test design", *Artech House*, pp. 9-12, 2004.

An Interpretation System from Turkish to Turkish Sign Language

M. F. KARACA[1] and Ş. BAYIR[2]

[1] Gaziosmanpasa University, Tokat/Turkey, mfkaraca@gmail.com
[2] Karabuk University, Karabuk/Turkey, safakbayir@karabuk.edu.tr

***Abstract* – Turkish Sign Language (TSL), which is a mother tongue of hearing-impaired individuals, is a natural language. Expression patterns used in this language are carried out within the settings of the rules of the language.**

In this study, a textual interpretation system from Turkish to TSL was developed. Within this context, a corpus consists of 230 sentences was composed. Beside the sentences included in the corpus, also the interpretation of the sentences entered by the user interactively can be made. The rules of both languages were taken into consideration in the interpretation of the sentences.

Firstly, Turkish sentences were parsed to words and then, morphological analyses of the sentences were performed with Zemberek. As a result of the morphological analyses, root/stem of the words and the affixes attached to the words were determined. Some affixes are not considered as necessary in TSL. Therefore, the affixes which are not considered as necessary should be determined and ignored within the interpretation. Moreover, various rules were constructed for the transformation by considering the results of the morphological analysis rules and usage samples.

Descriptions related to 489 signs and 6 non-manual signs were made in order to express the sentences included in the corpus in TSL. The number of the signs out of 489 were as follows: 81 static, 408 dynamic, 334 single, 155 repetitive, 6 sign union and 2 word combination. 6 non-manual signs were as follows; baş önde (head ahead) "bö", baş yukarıda (head up) "by", kaş yükseltme (eyebrow raising) "ky", kaş indirme (eyebrow lowering) "ki", past aspect "di" (past tense suffix) and continuous aspect "yor" (present continuous tense suffix). Usage numbers of the non-manual signs were as follows: 66 "bö", 12 "by", 38 "ky", 40 "ki", 94 "di" and 31 "yor".

***Keywords* - Turkish Sign Language, Turkish, grammar, interpretation system, natural language processing.**

I. INTRODUCTION

TURKISH Sign Language is accepted as one of the oldest sign language in the world [1]. TSL which is a natural language has its own grammar rules. Processes from derivation of signs to the use of affixes are carried out within the settings of these grammar rules. Furthermore, it is not possible to say that these two languages, Turkish and TSL, are identical although they display various similar features as in terms of syntax [2].

In this study, an interpretation system in which textual transformations are carried out from Turkish to TSL was developed. Grammar rules of both language are taken into consideration within the processes. An application was developed with C# programming language in Visual Studio environment for the interpretation system.

II. DATASET

Firstly, a corpus was consisted of 130 Turkish sentences related to 9 themes included in TSL course and 100 sentences related to 4 themes included in Turkish-1 course provided in Hearing-Impaired Primary Schools. The numbers of the sentence types obtained in the corpus are as follows: 160 declarative sentences and 26 yes/no type questions, 40 sentences including question words begin with "ne" type questions and 4 sentences including direct quotations.

III. NATURAL LANGUAGE PROCESSING

In this study, Zemberek [3], an open source Natural Language Processing Library for Turkic languages [4], was utilized for the morphological analysis of the words forming Turkish sentences. As a result of the morphological analyses, root/stem of the words and the affixes attached to the words were determined. Some of the affixes which exist in Turkish but are neglected in TSL were omitted within the transformation.

Zemberek can provide more than one analysis in the morphological analyses of the words. Generally, the first analysis is the accurate one but it is seen that the accurate analysis is not in the first order in the analyses of some words. For example, the accurate analysis of the word *kimsin* was provided in the 2nd order. These kinds of situations were detected and the accurate one among the order of the analyses which is going to be utilized was determined.

The sign + was used in TSL in order to display the affixes. Any word including plural suffixes of *-ler* or *-lar* in Turkish is expressed as SIGN+ÇOK combination in TSL: Ev *ler*→EV+ÇOK. The derivational suffixes of *-sız* and *-siz* which are used to make a noun from a noun in Turkish is expressed as SIGN+HİÇ combination in TSL: Ses *siz*→SES+HİÇ. Various rules were generated for the special cases like these.

IV. INTERPRETATION PROCESS

In this section, descriptions related to non-manual signs, sign unions and sign-word combinations were made. It was observed that non-manual signs were used according to the sentence type (for example, interrogative sentence) and the affixes used. The

procedure of expressing non-manual signs in TSL sentences are carried out within this process.

The beginning point of a non-manual sign was denoted with ↑ and the ending point was denoted with ↓; ↑di•GELMEK~↓di•.

Non-manual signs of "by" and "ky" were added to the beginning and ending of the sign of DEĞİL which is a negative expression; ↑by,ky•DEĞİL↓by,ky•

More than one non-manual sign can be seen in a sentence. When the two non-manual signs were occurred sequentially, these were represented as combined, not as separated; ↓di•↓bö,ki•→↓di,bö,ki•

Signs created with sign unions were denoted with ^; Hastane→HASTA^YER.

Signs created with word combinations were denoted with #; Ne zaman→#NE ZAMAN#.

V. Experimental Results and Conclusion

In this paper, since the ways of expressing Turkish sentences in TSL were explained with samples, researchers willing to carry out studies on TSL will be informed on the issues which should be taken into consideration in the interpretation process.

Interpretation samples from Turkish to TSL realized within the settings of this study were given in Table 1. It is seen that there are sentences including components of affirmative, negative, question, sign unions and sign-word combinations and also non-manual signs.

Table 1: Interpretation samples.

Turkish Sentence	TSL Sentence
Bu bisiklet onun.	BU BİSİKLET~ ONUN~.
Antalya, güzel bir ildir.	ANTALYA~, GÜZEL~ BİR İL.
Babam gelmedi.	BENİM+BABA ↑di•GELMEK~↓di•+↑by,ky•DEĞİL↓by,ky•.
Matematik zor değil.	MATEMATİK~ ZOR ↑by,ky•DEĞİL↓by,ky•.
Uçak, erken geldi.	UÇAK, ERKEN~ ↑di•GELMEK~↓di•.
Hastanede kan verdim.	HASTA^YER KAN ↑di•VERMEK↓di•+BEN.
Okula gidiyorum.	OKUL~ ↑yor•GİTMEK~↓yor•+BEN.
Annem yemek yapıyor.	BENİM+ANNE YEMEK ↑yor•YAPMAK~↓yor•.
Bayram ne zaman?	↑bö,ki•BAYRAM #NE ZAMAN#↓bö,ki•?
Kim geldi?	↑bö,ki•KİM ↑di•GELMEK~↓di,bö,ki•?
Çayı kim içti?	↑bö,ki•ÇAY~ KİM ↑di•İÇMEK↓di,bö,ki•?
Adın ne?	↑bö,ki•SENİN~+AD NE↓bö,ki•?
Engelli misin?	↑bö,ky•ENGELLİ~ SEN+[SPAR]↓bö,ky•?
Ödevini bitirdin mi?	↑bö,ky•SENİN~+ÖDEV ↑di•BİTİRMEK↓di•+SEN [SPAR]↓bö,ky•?

Descriptions related to 489 signs and 6 non-manual signs were made for the application. It was seen that the signs were expressed as in the following numbers: 81 static, 408 dynamic, 334 single, 155 repetitive, 6 sign union and 2 word combination. The positions / situations where the non-manual signs were used are given as follows: "bö", "ky" and "ki" were used in interrogative sentences, "by" with "ky" together were used in negative sentences, "di" was used as a past tense suffix and "yor" was used as a present continuous tense suffix. It was seen that the number of the non-manual signs used within the sentences are as follows: 66 "bö", 12 "by", 38 "ky", 40 "ki", 94 "di" and 31 "yor".

It should be taken into consideration that the hearing-impaired individuals prefer to use simpler and shorter sentences instead of complex and long sentences in communication and they understand these types of sentences easily. It would be possible to transform the interpretation process into more meaningful process if this situation and also writing Turkish sentences (such as punctuation marks) are taken into consideration.

References

[1] U. Zeshan, "Aspects of Türk İşaret Dili (Turkish Sign Language)", *Sign Language and Linguistics*, 6(1):43-75, 2003.

[2] A. Z. Oral, *Türk İşaret Dili çevirisi*, Ankara: Siyasal Yayınevi, 79, 2016.

[3] Zemberek, "Açık kaynak kodlu Türkçe doğal dil işleme kütüphanesi", http://code.google.com/p/nzemberek, 2015.

[4] M. D. Akın ve A. A. Akın, "Türk dilleri için açık kaynaklı doğal dil işleme kütüphanesi: Zemberek", *Elektrik Mühendisliği*, 431:38-44, 2007.

Performance Evaluation of kNN, Support Vector Machines and Artificial Neural Network on Optical Character Recognition

M. F. KARACA[1] and Ş. BAYIR[2]

[1] Gaziosmanpasa University, Tokat/Turkey, mfkaraca@gmail.com
[2] Karabuk University, Karabuk/Turkey, safakbayir@karabuk.edu.tr

***Abstract* – Optical Character Recognition (OCR) is the extraction of letters, numbers, punctuation marks, shortly the texts, from the images in the digital settings. As a result of this process, electronic images are converted into editable texts.**

In this study, the performances of character recognition techniques on upper-case letters, lower-case letters and numbers were evaluated in terms of various features. At first, training images with 10 typefaces were generated. Then, testing images were obtained by adding extra 10 typefaces into the system. The training images were formed in 36pt font size. However, the testing images were also formed in 12, 36 and 60pt font sizes in order to see the effects on OCR.

Firstly, images were parsed into the characters, then, these images were stretched or squeezed into 100x80, 50x40, 25x20 and 5x4 pixels. Then, thresholding was applied to image files and each character was expressed as vectors having pixel value either 1 or 0.

For the OCR process, 3 algorithms were used; k-Nearest Neighbors (kNN), Support Vector Machines (SVM) and Artificial Neural Network (ANN). Euclidean Distance, Inner Product and Cosine Similarity were used for the measurement of similarity in kNN.

The following results were obtained when the results are evaluated in terms of average means; the best classification was realized with ANN and the least with Inner Product (k=3) in upper-case letters, the best classification was realized with ANN and the least with Inner Product (k=10) in lower-case letters and the best classification was realized with ANN and the least with Inner Product (k=10) in numbers. The best classification was realized with 36pt and the least with 12pt in upper-case and lower-case letters; and the best classification was realized with 60pt and the least with 12pt in numbers.

***Keywords* - Optical character recognition, kNN, Support Vector Machines, Artificial Neural Network.**

I. INTRODUCTION

As a result of technological developments, enormous increase is seen in the amount of data in digital media. Image files including textual content are also one of these digital data. Characters must be recognized in order to convert the texts included in the files into editable forms. This situation increases the importance of optical character recognition.

With the increase of data in digital settings, it is seen that OCR is used in various fields; recognition of the scrolling texts in news as video processing [1], license plate recognition systems [2].

In this study, recognition of upper-case letters, lower-case letters and numbers were realized with various techniques.

II. DATASET

Training images with 10 typefaces were generated. Then, testing images were obtained by adding extra 10 typefaces into the system. While the training images were formed in 36pt font size, the testing images were formed in 12, 36 and 60pt font sizes.

III. OPTICAL CHARACTER RECOGNITION PROCESS

Images were parsed into the characters, then, these images were stretched or squeezed into 100x80, 50x40, 25x20 and 5x4 pixels. Thresholding was applied to both training and testing images to obtain binary form as black which is 1 (pixel values which are equal to 200 and above) and white which is 0 (pixel values which are below 200).

5 feature detection methods used in recognition were as follows; with its boundaries (method 1), with up/down boundaries (method 2), with only vertical histogram (method 3), with only horizontal histogram (method 4), with both horizontal and vertical histograms (method 5). In the study, 29 classes were determined for the upper-case letters (A, B, C, …, Z) and 29 classes for lower-case letters (a, b, c, …, z) separately and 10 classes for the numbers (0, 1, 2, …, 9).

Recognition process was implemented with kNN (k-Nearest Neighbors with Euclidean Distance, Inner Product and Cosine Similarity), Support Vector Machines (SVM) and Artificial Neural Network (ANN). The value of k was determined as 3, 5, 7 and 10 for kNN.

At first, the similarities between testing images and training images are calculated in kNN. Then, these similarity values are sorted and the class which has the highest frequency among the first k value is assigned as the class of the testing images [3]. It is aimed to obtain the maximum possible margin and optimal hyper plane in order to realize the best successful classification in SVM [4]. Multilayer perception was used in ANN. At first, the network was initialized with the initial values and then the

weights between neurons were updated by training the network [5].

IV. EXPERIMENTAL RESULTS AND CONCLUSION

Results were given as upper-case letters, lower-case letters and numbers separately. Results related to recognition of upper-case letters are as follows:

- ✓ The best recognition results were obtained with SVM and the worst results with kNN (Inner Product); 96.21% and 0.00%.
- ✓ Successful recognitions over 95% were realized via SVM with 2 times and via ANN with 4 times.
- ✓ In terms of average results;
  - The most successful was ANN and the least successful was kNN,
  - The best recognition results were obtained with images having 100x80 pixel numbers and the worst recognition results were obtained with images having 5x4 pixel numbers.
  - The most successful method was the method 1 and the least one was the method 4 in terms of feature detection.

Results related to recognition of lower-case letters are as follows:

- ✓ The best recognition results were obtained with ANN and the worst results with kNN (Inner Product); 94.66% and 0.00%.
- ✓ Recognitions could not be achieved over 95%.
- ✓ In terms of average results;
  - The most successful was ANN and the least successful was kNN,
  - The best recognition results were obtained with images having 100x80 pixel numbers and the worst recognition results were obtained with images having 5x4 pixel numbers.
  - The most successful method was the method 1 and the least one was the method 3 in terms of feature detection.

Results related to recognition of numbers are as follows:

- ✓ Both the best and the worst recognition results were obtained with SVM; 99.50% and 10.00%.
- ✓ Successful recognitions over 95% were realized via SVM with 12 times and via ANN with 18 times.
- ✓ In terms of average results;
  - The most successful was ANN and the least successful was kNN,
  - The best recognition results were obtained with images having 100x80 pixel numbers and the worst recognition results were obtained with images having 5x4 pixel numbers.
  - The most successful method was the method 1 and the least successful method was the method 3 in terms of feature detection.

REFERENCES

[1] E. Dikici and M. Saraclar, "Sliding Text Recognition in Broadcast News," *IEEE 16th Signal Processing and Communications Applications Conference (SIU)*, Aydın, 1-4, 2008.

[2] E. Tamer and B. Çizmeci, "A Different Approach for License Plate Recognition System," *IEEE 17th Signal Processing and Communications Applications Conference (SIU)*, Antalya, 357-360, 2008.

[3] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufmann Publishers, 348-350, 2006.

[4] A. K. Uysal and S. Gunal, "A novel probabilistic feature selection method for text classification," *Knowledge-Based Systems*, 36:226-235, 2012.

[5] T. M. Mitchell, *Machine Learning*, New York: McCraw Hill, 97-100, 1997.

Comparison of Classification Algorithms on NFC-Based Public Transport Data

U. DEMİR ALAN[1] and D. BİRANT[2]

[1] Graduate School of Natural and Applied Sciences, Dokuz Eylul University, Izmir/Turkey, ufuk.demir@live.com

[2] Department of Computer Engineering, Dokuz Eylul University, Izmir/Turkey, derya@cs.deu.edu.tr

***Abstract* – In recent years, with the growing popularity of smartphone, near field communication (NFC) based mobile applications commenced to be used in public transportation. This development provides an opportunity to collect additional and province independent data about passengers and so it allows development of better data mining applications. The present study is conducted to compare classification algorithms on public transport data collected by NFC-based mobile phone ticketing application for the first time. In this paper, five popular classification algorithms have been considered to investigate various target attributes in terms of accuracy rates: Naive Bayes, C4.5 Decision Tree, Random Forest, Support Vector Machines, and k-Nearest Neighbor. The study presented in this paper can be useful to provide decision support for public transportation.**

***Keywords* - Classification algorithms, data mining, public transport, near field communication, mobile application.**

I. INTRODUCTION

TECHNOLOGY in the mobile industry has been moving towards the integration of electronic ticketing in public transportation. Currently, it is possible to use smartphones as a contactless paper ticket or as a smart card. There are two main technologies that have been widely used for mobile ticketing: near field communication (NFC) and quick response (QR) code. In fact, QR code generation and usage on the field is easier but if the subject is the public transportation, finding QR reader terminal is difficult. After Android 4.4 KitKat operating system development, the NFC technology in mobile phones became to support ISO 7816-4 standard communication protocol on contactless interface. This technological development opens the NFC usage on transportation because all over the world the readers at the transportation vehicles have generally capability to communicate in ISO standards.

When compared with the traditional systems, smartphone usage in public transportation has numerous advantages: ease of use, less likely to be lost, greener solution, province independence, ease of system integration, interoperability, easy upgrade and passenger-specific loyalty support. Considering this motivation, an NFC-based mobile ticketing application developed[1] and commenced to be used in some cities in Turkey. In this study, the real-world data collected from smartphone users was analyzed to provide decision support. To the best of our knowledge, this is the first study that compares classification algorithms on public transport data collected by NFC enabled mobile phone application.

The novelty and main contributions of this paper are as follows. First, this paper briefly explains the previous studies that apply classification algorithms on public transport data. Second, it gives information about NFC-based mobile ticketing application developed for public transportation. Third, it compares five different classification algorithms on real-world data in terms of accuracy: Naive Bayes, C4.5, random forest, support vector machines, and k-nearest neighbors. The study presented in this paper can be useful for different purposes in public transportation i.e. personalization, outlier detection, scoring, and recommendation.

This paper is structured as follows: In the following section, related literature and previous studies on the subject is summarized briefly. Section 3 explains NFC-based mobile application developed for public transportation. In section 4, materials and methods used in this study are described. Section 5 explains the application of the classification algorithms on the datasets and presents the obtained results with discussions. Finally, Section 6 gives some concluding remarks and future directions.

II. RELATED WORK

In this section, technical studies that use classification algorithms on public transport data are discussed briefly.

Since the passengers are the key constituents of the public transportation, the public/private operators give the highest priority to meeting and satisfying passengers' demands. For example, passenger information services gain importance from day to day. Answering the following problems have begun to take an important place for the operators to provide high quality services to passengers.

- when the vehicle will arrive at the station [1],
- which transport mode is preferred [2] and
- how to reach a destination (multi-model journey planning) and how long it takes [3]

[1] This project was supported by the Scientific and Technological Research Council of Turkey under Grant TEYDEB 1501-3150065 and Kent Kart Ege Elektronik San. Tic. A.Ş. Company from Turkey.

In recent years, the studies are centralized around the arrival time estimation [1], transport mode prediction [4], and quality of service classification [5][6] in public transportation. When addressing these problems, data mining algorithms have been widely used to make correct classifications and estimations.

To address the first problem (arrival time estimation of buses), Kee et al. [1] compared several classification algorithms: decision tree, random forest, Naive Bayes, AdaBoost and neural network. If a bus will arrive in the first 15 minutes of the hour, it is classified as Q1; if a bus will arrive in the 16th to 30th minutes, it is classified as Q2; and the same applies for Q3 and Q4.

To deal with the second issue (transport mode detection), Guvensan et al. [4] developed a mobile application to predict eight classes including stationary, walking, car, bus, tram, train, metro, and ferry. When the performances of the classifiers are compared, the results indicate that random forest has higher performance (80.62%) in terms of classification accuracy than C4.5, k-nearest neighbor, and Naive Bayes.

To handle the third issue (quality of service – QoS), several classification studies [5][6] have been done. Tsami and Nathanail [5] applied C4.5 decision tree algorithm on questionnaire data related to QoS collected from public transport passengers in Greece. In their study, service quality in public transport was classified into five categories: very bad, bad, medium, good and very good. Stottelaar [6] classified Twitter messages about public transport by sentiment (positive, neutral, negative), content and type of transport (bus, train) to detect problems, using C4.5 decision tree algorithm.

Besides to above researches, some studies [7][8] have been performed on the classification of driving styles of the bus drivers. This problem is very important when the subject is public transportation in terms of fuel consumption and passengers' safety. Karginova et al. [7] compared several classifiers (including neural networks, decision trees, random forests and k-nearest neighbor) to classify the driving styles into two categories: normal and hard.

Many attempts have been made to derive useful information about the behavior of the passengers traveling, using classification algorithms. For example, Ahlawat et al. [9] tried to estimate the trip purpose of the passengers which can be thought of a key attribute in the context of the behavior of the travelers. They applied Naive Bayes algorithm on the trip survey data to classify different trip purposes: commuting to work, commuting to school, leisure, business and returning home.

Data mining studies conducted in public transport shows that classification task plays an important role when addressing different types of public transport problems. Differently from the existing studies, our study focuses the classification of passengers based on their characteristics using real-world transport data collected through NFC-based mobile phone ticketing application.

III. NFC-BASED MOBILE TICKETING APPLICATION

The mobile application that developed for NFC-enabled phones is the point of interaction with the passengers. This application actually uses an open-loop system, but it behaves like a closed-loop contactless transport ticket, which means that it is the emulation of the travelling card. All sensitive data (i.e. personal data, balance, social rights etc.) is not stored at the mobile site. Instead of this, completely account-based and host-centric system was developed [10]. Figure 1 shows the general structure of the system.

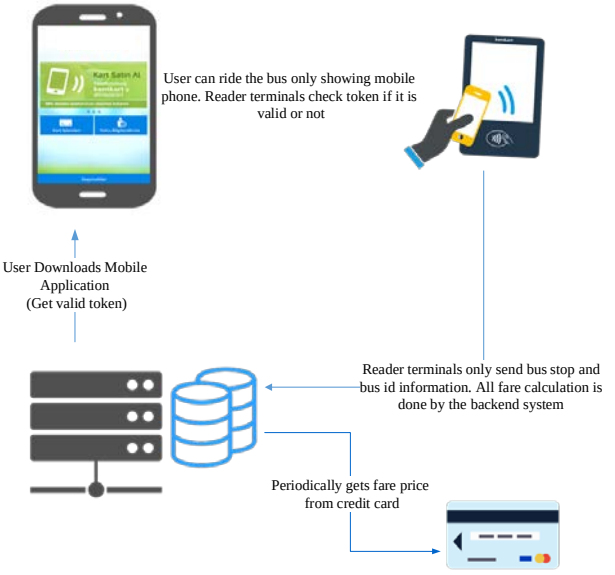


Figure 1: General structure of the system

Figure 2 shows the use case diagram drawn for the usage of the NFC-enabled phone at the vehicle. When the NFC-enabled phone is touched to the validator, in contrast to the traditional transportation systems, the validator doesn't calculate the fare and deduct the money. Only token is approved by the validator.

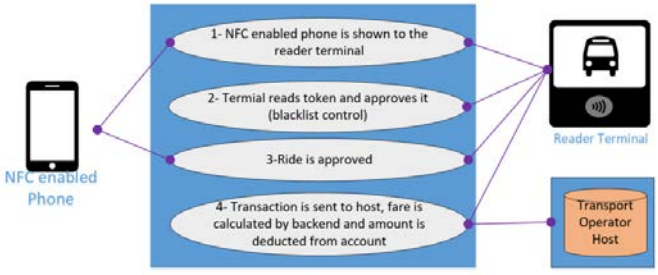


Figure 2. Use case diagram of NFC usage on the public transport vehicle

The user needs to register to the system before starting to use the application. Users can provide their personal data according to the voluntary principle. The information about these passengers was used in the study presented in this paper. The basic steps from the user perspective are listed as follows:

- The user sends a request to purchase mobile card by opening the application.
- The user is routed to the account creation screen.

- The password for the account is generated.
- One-time password (OTP) is sent to the user.
- The user enters the OTP for the account approval.

The NFC mobile application was developed to enable several operations like virtual card creation, purchase, and usage. There are multiple ways to develop Android apps: Hybrid and Native. Since we are going to use NFC technology and keep performance on the front, we preferred to develop Native applications with Java. At the mobile application side, all services that communicate to the server were implemented as RESTful web services. The data interchange format was defined in JavaScript Object Notation (JSON).

The libraries used in the development of the application can be grouped into three main categories.

- *UI Libraries*: These are libraries that affect the user-visible parts of the application. They are related to the appearance of the application directly.
- *Database Libraries*: They provide database operations required for the application.
- *Backend Libraries*: They are backplane libraries that are included in the functional parts of the application. So, they are used for operations in which the user does not see directly.

The most important and difficult part of the mobile application development is to construct the file structure of the virtual card. Android is an open operating system, for this reason, it is more vulnerable to attacks. Therefore, proving the security of the virtual card usage is the essential part of this study. Details about NFC procedures between validators and smartphone, as well as provided hardware and software security techniques can be found in [11].

IV. Materials and Methods

Classification is the prediction of the class of an object on the basis of its attributes [12]. There are many different classification algorithms used for data mining problems. The most popular ones are C4.5 decision tree, random forest, Naive Bayes, support vector machine and k-nearest neighbor.

Decision Tree (DT): Decision tree learning is a method widely used in data mining to create a classifier that predicts the value of a target parameter based on several input parameters. As one of the decision tree algorithms, the C4.5 algorithm constructs a decision tree that has a flowchart-like structure in which each internal node splits the instance space into two or more sub-spaces based on a "test" on an attribute, each branch represents the outcome of the test, and each leaf node contains a class label [13]. The C4.5 algorithm uses entropy and normalized information gain to construct a decision tree. Entropy is a formula to calculate the homogeneity of a sample data and is the measure of uncertainty. Information gain is calculated by the taking difference between the target attribute entropy and the entropy of each branch. Both the discrete and continues attribute values, also missing values in the dataset can be handled by the C4.5 algorithm.

Random Forest (RF): The random forest algorithm builds more than one decision trees using different subsets of the data and evaluating different subsets of features at each node. It is an ensemble learning method and uses bootstrap samples to train multiple classifiers [14]. Each bootstrap sample creates a new tree that is different from the original tree because, at each tree generation, a random number of attributes are tested. The outputs of the predictors are combined and the result is evaluated by a voting mechanism. The random forest algorithm generally produces highly accurate results because it uses many different characteristics to make a prediction.

Naive Bayes (NB): It is another classification algorithm that utilizes Bayes' theorem to attempt to predict class labels. It is a probabilistic and simple model that uses maximum-likelihood estimation. Naive Bayes classifier is based on assumption that each attribute in a class is independent of the presence of any other attribute.

Support Vector Machine (SVM): It is a supervised machine learning algorithm that tries to find the optimal separating hyperplane as a decision surface in such a way that maximizes the margin of the training data [15]. SVMs learn a linear binary classification boundary. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification by mapping their inputs into high-dimensional spaces by using the kernel trick.

K-Nearest Neighbor (kNN): K-Nearest Neighbor algorithm is a technique in which an object is classified by a majority vote among the classes of its *k* neighbors in the training sample that are closest to this instance [16]. It uses a distance measure (i.e. Euclidean distance) to compute the *k* nearest neighbors of a given point.

V. Experimental Studies

A. Dataset Description

A server-based fare collection and passenger management system has been used in several cities in Turkey. Approximately for four months, live transportation data was collected from the field.

The dataset used in classification task in this study contains 24,485 NFC usage records collecting from public transport vehicles located at five different cities in Turkey. There are eight attributes in the dataset: gender, age, marital status, passenger type, region, day, time, and city.

We combined raw data obtained from two different sources:

- Passenger information getting via mobile application:
  The usage of a mobile phone in payment allows us to identify the passengers and obtain demographic data about them. Users can provide their personal data according to the voluntary principle. The basic parameters about a passenger are as follows:
  - *Gender*: The gender of the passenger.
  - *Marital Status*: The marital status of the passenger.
  - *Age*: The age of the passenger.
- Reader terminals data (NFC transactions sending via services): The basic parameters about boarding with NFC are as follows:
  - *Usage Date Time*: Boarding date and time information.
  - *Bus Stop Information*: The bus-stop id and bus-stop name that riding was realized.

- *Route Information*: The route id and route name where public transport vehicle went.
- *City*: The city name that the transaction has occurred.
- *Passenger Type*: The type of the passenger that has four distinct values: adult-card, student-card, teacher-card, and retired-card. This info is used when calculating the fare of each ride at the server side.

Such variety of data allows us to develop better data mining applications. So, classification algorithms can be used for different purposes i.e. personalization, outlier detection, scoring, and recommendation.

B. Dataset Preparation

Data preparation is the process of collecting, integrating, cleaning, structuring, transforming and organizing data, which is taken before data analysis is carried out.

In this study, the following data preparation steps are performed on the dataset before it is used for classification.

Data collection: Approximately for four months, NFC usage records were collected from public transport vehicles located in five different cities in Turkey.

Data integration: The data was obtained from two sources and merged to form a single dataset: (i) one source that includes personal user information and (ii) another source that contains automatic fare collection data.

Data preprocessing: Especially real-world data is often incomplete and have many errors. Data preprocessing is a proven method of resolving such issues. In this study, the noisy data was cleaned from the original data.

Data transformation: It is possible to map and convert data from one format to another. In this study, the following transformation operations were carried out:

- Within the scope of this study, boarding date attribute was formatted into seven days of the week, from Monday to Sunday.
- Similarly, boarding time attribute was converted into four different categories: morning, afternoon, evening and night.
- The raw dataset contains global positioning system (GPS) coordinates. For better classification, the GPS coordinates were converted into region concept. The logic behind this conversion is to divide each city into four regions according to the GPS coordinates. Two GPS points are taken from city borders and the region is divided into 4 regions. There are five cities and twenty regions but only fifteen of them were used actively in the system.

Data discretization: Data discretization is the process of converting continuous data attribute values into a finite set of intervals. Discretization process can be classified into different categories from different perspectives such as supervised or unsupervised, top-down or bottom-up, static or dynamic, local or global, nominal or ordinal, univariate or multivariate, direct or iterative. The discretization technique applied in this study is the type of unsupervised, top-down, static, global, nominal, univariate, and direct. The age attribute was discretized into four segments: young, middle-aged, adult and old.

Table 1 shows the values of the data set for each attribute.

Table 1: Attributes in the dataset

Attribute Name	Attribute Values
Day	*from* Monday *to* Sunday
Time	Morning 5am - 12pm Afternoon 12pm - 6pm Evening 6pm - 12am Night 12am - 5am
City	five cities in Turkey
Region	15 regions from five cities
Passenger type	Adult-card Student-card Teacher-card Retired-card
Marital status	Single Married Divorced
Age	Young $age \leq 20$ Middle-aged $20 < age \leq 35$ Adult $35 < age \leq 60$ Old $age > 60$
Gender	F (Female) M (Male)

Table 2 shows a sample data used in this research.

Table 2: Sample data

Day	Time	City	Region	Pas. Type	Marital Status	Age	Gender
Friday	Morning	A	A1	Adult	Married	Adult	F
Sunday	Afternoon	E	E2	Student	Single	Middle-aged	M
Friday	Afternoon	B	B1	Teacher	Married	Adult	M
Monday	Morning	A	A4	Adult	Single	Young	F
Tuesday	Evening	C	C4	Student	Single	Young	F
Friday	Night	B	B3	Student	Single	Young	F

C. Comparison of Classification Algorithms

The application was developed by using Waikato Environment for Knowledge Analysis (WEKA) open source data mining library.

In this study, five classification algorithms (DT, RF, NB, SVM, and kNN) were applied to transportation data collected through NFC-based mobile phone ticketing application. The algorithms were executed five times and a different attribute was selected as a target column in each try. The number of neighbors, *k* for kNN classifier was selected as 5. The input parameters of other classifiers were left as default Weka parameters. The results were compared according to their correct estimation ratios (accuracy rates). Classification accuracies of the applied algorithms were obtained using 10-fold cross-validation. In addition, the most important parameters were determined by selecting the attributes at the high levels in the construct decision tree, because a decision tree algorithm recursively chooses "most significant" attributes as a

root of tree/subtree.

The comparative results are given in Table 3. According to the results, Naive Bayes algorithm has lower performance when compared to other algorithms. The random forest algorithm has a little higher accuracy rate for some attributes than the C4.5 algorithm. Decision tree-based algorithms (C4.5 and random forest) have the highest performance in all cases. The experimental result also indicates that it is possible to predict the regions and marital statuses of the passengers with high accuracy rates. According to the results, the most important features in NFC-based transportation data are generally age and passenger type.

Table 3: Comparison of classification algorithms on NFC-based data

Target Attribute	Classification Accuracy (%)					Significant attributes
	NB	RF	DT	SVM	KNN	
Region	95.18	**95.58**	95.50	95.36	94.74	City, day, age
Marital Status	82.09	86.77	**87.07**	86.38	86.17	Age, gender, pas type
Gender	76.83	85.65	**85.77**	81.25	85.36	Age, pas type, marital status
Age	75.65	**77.67**	77.56	75.45	77.35	Marital status, pas type, gender
Passenger Type	65.13	**76.96**	76.84	72.71	76.67	Age, gender, region

Figure 3 shows the classifiers' average ranks. Classifiers are rated according to their accuracies. It is done by assigning rank 1 to the classifier with the best accuracy and increasing the rank for each evaluated classifier until giving rank 5 to the worst one. According to the results, when the random forest algorithm is applied, it obtains the lowest rank value which means that it outperforms the others in terms of accuracy.

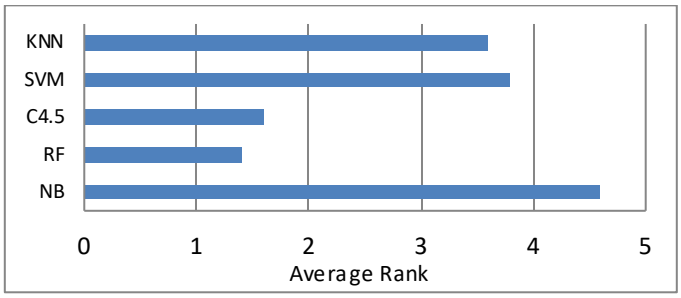


Figure 3: The average rank of classifiers

VI. Conclusion And Future Work

The goal of this study is to devise an accurate classification model and to determine the important features over real-world NFC-based transportation data. We collected data through NFC-based mobile ticketing application developed as a part of the server-based automatic fare collection and management system. Demographic features of the passengers and their mobile ticketing usage transactions were analyzed empirically by using five classification algorithms: Naive Bayes, C4.5 decision tree, random forest, support vector machines, and k-nearest neighbor. We observed that decision tree-based algorithms achieved high classification accuracies; especially random forest algorithm is the best one when compared with the others.

As further work, passengers can be grouped according to their characteristics using clustering algorithms. In addition, passengers can be scored by providing a scoring mechanism on the basis of the mobile ticketing data in future studies.

Acknowledgment

We thank Kent Kart Ege Elektronik San. Tic. A.Ş Company for believing in this project and supporting the implementation of the proposed system on their terminals and servers.

References

[1] C. Y. Kee, L. Wong, A. T. Khader, and F.H. Hassan, "Multi-label classification of estimated time of arrival with ensemble neural networks in bus transportation network," in *Conf. Rec. 2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, pp. 150-154.

[2] M. A. Shafique, and E. Hato, "Classification of travel data with multiple sensor information using random forest," *Transportation Research Procedia*, vol. 22, pp. 144-153, 2017.

[3] D. E. Kiss, and C. Csiszar, "Evaluation of multimodal journey planners and definition of service levels," *International Journal of Intelligent Transportation Systems Research*, vol. 13(3), pp. 154-165, June 2014.

[4] M. A. Guvensan, B. Dusun, B. Can, and H. I. Turkmen, "A novel segment-based approach for improving classification performance of transport mode detection," *Sensors for Transportation (MDPI)*, vol. 18, pp. 1-19, December 2017.

[5] M. Tsami, and E. Nathanail, "Guidance provision for increasing quality of service of public transport," *Procedia Engineering*, vol. 178, pp. 551-557, 2017.

[6] B. Stottelaar, "Detection of public transport problems using location extended community based sensing", in *Conf. Rec. 17th Twente Student Conference on IT*, June 25, 2012.

[7] N. Karginova, S. Byttner, and M. Svensson, "Data-driven methods for classification of driving styles in buses," *SAE Technical Papers*, ISSN 0148-7191, No: 2012-01-0744.

[8] C. M. Martinez, M. Heucke, and F. Y. Wang, "Driving style recognition for intelligent vehicle control and advanced driver assistance: a survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19(3), pp. 666-676, March 2018.

[9] G. Ahlawat, A. Gupta, and A. K. Vatsa, "Predicting behavior of passengers using data collected through smart cards," in S. Kohli, A.V.S. Kumar, J.M. Easton, and C. Roberts (Eds.), *Innovative Applications of Big Data in the Railway Industry*, chapter 6, pp. 127-156 November 2017.

[10] U.D. Alan, and D. Birant, "Server-based intelligent public transportation system with NFC," *IEEE Intelligent Transportation Systems Magazine*, vol. 10(1), pp. 30-46, January 2018.

[11] S. Arslan, V. Demirel, and İ. Kuru, "A public transport fare collection system with smart phone based NFC interface," *International Journal of Electronics and Electrical Engineering*, vol. 4(3), pp. 258-262, June 2016.

[12] C. C. Aggarwal, *Data Mining: The Textbook*. Switzerland: Springer, 2015.

[13] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Francisco: Morgan Kauffmann, 1993.

[14] C. Zhang, and Y. Ma, *Ensemble Machine Learning: Methods and Applications*. New York: Springer, 2012.

[15] Y. Ma, and G. Guo, *Support Vector Machines Applications*. Switzerland: Springer, 2014.

[16] B. Steele, J. Chandler, and S. Reddy, *Algorithms for Data Science*. Switzerland: Springer, 2016.

An Efficient Human Action Recognition Framework with Pose-based Spatiotemporal Features

S.AGAHIAN[1], F. NEGIN[2], and C. KÖSE[1]

[1] Karadeniz Technical University, Trabzon/Turkey, saeid@ktu.edu.tr
[1] Karadeniz Technical University, Trabzon/Turkey, ckose@ktu.edu.tr
[2] INRIA, Sophia Antipolis/France, farhood.negin@inria.fr

***Abstract* – In the past two decades, human action recognition has been among the most challenging tasks in the field of computer vision. Recently, extracting accurate and cost-efficient skeleton information became available thanks to the cutting edge deep learning algorithms and low-cost depth sensors. In this paper, we propose a novel framework to recognize human actions using 3D skeleton information. The main components of the framework are pose representation and encoding. Assuming that human skeleton can be represented by spatiotemporal poses, we define a pose descriptor consists of three elements. The first element contains the normalized coordinates of the raw skeleton joints information. The second element contains the temporal displacement information relative to a predefined temporal offset and the third element keeps the displacement information pertinent to the previous timestamp in the temporal resolution. The final descriptor of the skeleton sequences is the concatenation of frame-wise descriptors. To avoid the problems regarding high dimensionality, PCA is applied on the descriptors. The resulted descriptors are encoded with Fisher Vector (FV) representation before they get trained with an Extreme Learning Machine (ELM).**

The performance of the proposed framework is evaluated by three public benchmark datasets. The proposed method achieved competitive results compared to the other methods in the literature.

***Keywords* - Skeleton-based, 3D Action Recognition, Extreme Learning Machines, and RGB-D.**

I. INTRODUCTION

Non-intrusive human action recognition has gained great attention among computer vision researchers owing to its extensive domain of applicability. It can be employed as a solution for the problems arising in video surveillance, human-computer interaction as well as in robotics and rehabilitation etc. [1-3].

Regardless of the enormous effort has put in researching this domain, many challenges in action recognition have remained unresolved. To have a general purpose solution to those challenges, an action recognition framework should be capable to cope with viewpoint variation, occlusion, body size variation of subjects, inter and intra-class variation and etc.

Conventionally, an action recognition framework consists of two main parts. In the first part, the action descriptors got extracted from the input modalities (e.g. RGB or depth map) followed by the second step where a classifier (e.g. SVM) is trained using a representation of the extracted descriptors (e.g. bag-of-visual-words or Fisher vectors).

To obtain the action descriptors, many works have tried pose features as a high-level description of actions. However, in early days of research in pose estimation topic, body part detection was not reliable and computationally expensive [4]. Moreover, using low and mid-level descriptors has a major disadvantage that such descriptors do not represent semantic information [5]. With the introduction of depth sensors, extraction of body part information from depth images became economical. Since then the depth sensors have become ubiquitous providing high-level semantic features as a remedy. Recently, the revival of deep learning methods helped to recover this information only from RGB images with high accuracy and reliability without even requiring the depth information [6, 7].

Other than describing body shape and pose, the temporal evolution of these configurations is important in recognition of a performed action. To preserve temporal data and represent action models embedded with this information, temporal pyramid (TP) approaches [8] were most welcomed techniques in the past years. However, after great achievements of Recurrent neural networks (RNN) and long short-term memory (LSTM) networks in speech and text recognition, researchers became motivated to use them in modeling the temporal evolution of sequences such as skeleton pose sequence appearing in successive frames of a video [9]. Although these methods achieve better results compared to the previous methods such as HMMs and TPs, their computational complexity is restraining for real-time tasks.

The second part of a hypothetical recognition framework is its feature encoding and classification components. Two broad categories of classification methods have been used for this purpose: Generative and Discriminative methods. Hidden Markov Models (HMM) is the most popular method in the first category which is used in action classification context. Despite their success in various applications, HMMs cannot benefit

from the discriminative power of kernel machines or metric learning. Discriminative methods such as Support Vector Machines (such as SVM or random forests) estimates the conditional probability of a labeled sequence given the observations. Although these methods achieved good results on different action recognition tasks [10, 11], they are more computationally complex and given the observations, they cannot make explicit estimations. Hybrid models such as Fisher Kernels [12] take the advantages of the both method's positive aspects and tackle their drawbacks. Fisher vector encoding [13] is the improved version of the Fisher kernel that is employed to encode image features and carries all the advantages of the hybrid representation models.

In this paper, we introduce an action recognition framework considering all the above-mentioned ideas. Figure 1 illustrates the data flow of the proposed method. We describe a pose in an action using a simple descriptor reserving temporal information. Embedding temporal information in the descriptors helps to avoid the limitation of bag-of-visual-words methods. Principal Component Analysis (PCA) method is used to overcome the problem of the high dimensionality of the features. In this way, we implicitly reduce the size of the calculated Fisher vectors which results in improved efficiency of the framework. In the final step, we use Extreme Learning Machine (ELM) [14] for classification.

We evaluated our framework by three public benchmark datasets and the reported results show that the proposed method is capable of producing competitive results when it is compared with state-of-the-art methods in the literature.

II. RELATE WORK

In standard pose-based action recognition pipeline, human poses at each frame are described from the extracted features. Then a calculated feature vector for the whole action is used for classification or reasoning.

The pose-based features are mainly categorized into four groups: displacement features [15], features based on orientation [16] or raw joint position information [17] and multimodal features [4, 11, 18, 19]. The extracted pose descriptors should be represented by a vector of a fixed size before they fed as input to the classifier. A traditional way of doing this kind of representation is to learn a codebook of the extracted features and use it as a visual dictionary to encode the features based on their distribution regarding the learned dictionary [20, 21]. These bag-of-visual-words approaches suffer from missing temporal information. A more recent approach is to use Fisher vector for encoding the features which result in better recognition performance [22, 23].

The final stage of the pipeline is classification. Most of the conventional frameworks use SVM in their final step [22, 24, 25]. In this paper, unlike these approaches we use ELM classifiers as it has shown efficiency in learning and reliability in classification [23, 26, 27].

Due to the recent breakthrough of deep learning methods in image processing researchers are encouraged to use them on videos and recognize dynamic content in the image sequences.

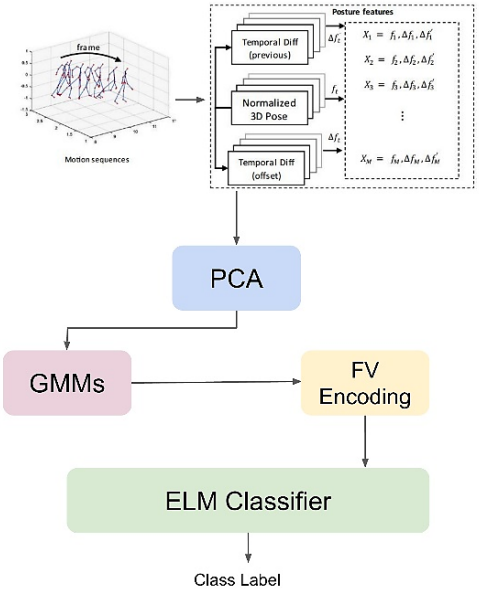


Figure 1: The proposed pipeline. Training of the GMMs and the ELM model only run in the training phase. The other steps are shared in training and testing.

In this regard, some works try to map the skeleton pose information into color or texture space and use the generated images as an input of a deep network [28]. Some others employ RNN or LSTM networks to find patterns in the evolution of pose sequences, hence, recognize the performed actions [29-32]. Although we do not use deep representation learning in this work, the proposed method is compared with those methods. The results show that our framework achieves comparable and competitive performance.

III. PROPOSED METHOD

As input, our framework accepts a high-dimensional sequence of vectors representing the actions. Given an action with T frames and J skeleton joints, we define:

$$S = \{P_t \mid \forall\, t \in (1 \dots T)\} \qquad (1)$$

Where P_t is a set of skeleton joints at tth frame of the sequence and is defined as $P_t = \{p_t^i \mid \forall\, i \in (1 \dots J)\}$ and p_t^i is the 3D coordinates of the ith joint of skeleton p at tth frame. The origin of the coordinate system for the position information is relative to the center of the recording sensor. In order to make the skeletons invariant to the location of the subjects, the origin of the coordinate system is transformed to the location of *hip center* joint of the skeleton. Moreover, to make the skeletons scale invariant, similar to [33], we choose a random reference skeleton and we rescale the limbs of the remaining skeletons to the same size of the reference skeleton, preserving the original angles between the joints. Additionally, to make the poses rotation invariant, a rotation operation is applied on the joints relative to the camera view angle. All these preprocessing steps makes our skeletons normalized and suitable for defining a robust descriptor.

Our descriptor contains geometrical information of the

skeletons, however, at the same time, the temporal information is implicitly embedded in it. So, we define a feature vector f_t describing the features of *t*th frame. f_t is simply the concatenation of the normalized coordinates of the joints. To embed the temporal information another vector Δf_t is defined that represents temporal relations of the poses in the sequence. First, a random offset frame is selected (t'). If the current pose (t) occurs before the offset, the feature vector will contain the original joint features, otherwise, it will also store the distance between the current pose and all the poses in the range of the temporal offset:

$$\Delta f_t = \begin{cases} f_t & 1 \le t < t' \\ \frac{f_t - f_{t-t'+1}}{\|f_t - f_{t-t'+1}\|} & t' \le t \le T \end{cases} \quad (2)$$

This component of the descriptor models the temporal information in the global range of the sequence. In addition to this component, we define another one f_t' to store the local temporal dependency by keeping the displacement between the adjacent elements of the pose sequence:

$$\Delta f_t' = \begin{cases} f_t & t = 1 \\ f_t - f_{t-1} & 2 \le t \le T \end{cases} \quad (3)$$

The final descriptor is the concatenation of the three components: $F = [f_t, \Delta f_t, \Delta f_t']$. The rest of the pipeline consists of learning the coefficients of PCA, parameters of the Gaussian mixture models, and the parameters of the ELM classifier.

A. Fisher Vector Encoding

Fisher vector is an image representation which is used for encoding visual vocabularies. Different from the classical bag-of-visual-words representation which uses first-order statistics, FV encoding employs GMM to build the visual dictionary, hence, it utilizes first and second order statistic altogether. To calculate the FV, first, parameters of a GMM distribution fitting the distribution of the given descriptors are detected. The calculated GMM can associate a feature vector x_i to component *k* of the mixture with a strength provided by its posterior probability. So, if the mean and deviation of the *k* components of the GMM is calculated as:

$$u_k = \frac{1}{N\sqrt{\omega_k}} \sum_{i=1}^{n} q_{ki} \frac{(x_i - \mu_k)}{\sigma^k},$$

$$v_k = \frac{1}{N\sqrt{2}\,\omega_k} \sum_{i=1}^{n} q_{ki} \left[\left(\frac{(x_i - \mu_k)}{\sigma^k} \right)^2 - 1 \right] \quad (4)$$

Where q_{ki} is the posterior probability of vector x_i with respect to component *k* and the gradient with respect to ω is discarded. The Fisher vector encoding of the observed feature vector is the stacking of the u_k and v_k vectors of each *k* components of the Gaussian mixtures. Prior to building the vocabulary, we apply PCA to the feature vectors to avoid the problem pertinent to the high dimensionality of the data. We set the size of *k* by evaluation of a range of parameters (Table 1).

B. Extreme Learning Machine

Instead of using popular classifiers such as SVM, ANN etc. for action classification, we use ELM. ELM is a single-layered feed-forward classification and regression method that can be used as a fast and reliable alternative for the mentioned methods. ELM is successfully applied in different domains and despite its short training time, it achieved viable accuracy even when it applied to large-scale video classification problems [23]. ELM is fast, efficient and economical when compared to the training time of the large-scale deep convolutional networks making it an ideal choice for real-time applications. It is recently used for classification of the pose-based methods [21, 26, 27]. In ELM there is only one hidden layer and unlike conventional neural networks that learn the mapping of the input and the hidden layers by the back-propagation algorithm, it is assumed that the hidden layer is known and does not need any tuning. Since the mapping between the input and the hidden layer is known, the weights of the mapping hidden nodes to the output node can be calculated analytically. Therefore, in a C-class ELM classifier, the aim is to minimize the training error and norm of the output weights at the same time. ELM allows for multi-class classification. We use the calculated Fisher vectors as input to train the ELM classifier.

IV. EXPERIMENTS

A. Datasets

We evaluate our framework with three public benchmark dataset. We assume that only one person is present in the scene performing the assigned action.

UTKinect Action Dataset: this dataset [16] is collected by Kinect v1 and consists of 10 actions performed by 10 subjects which are recorded at 30 fps. In total, there are 200 sequences available in the dataset including RGB, depth map and skeleton information of the actions. This dataset is challenging due to occlusion, variation of subjects' position and orientation towards the camera and significant difference between the speed and duration of the actions.

Florence 3D Action Dataset: this dataset [34] is recorded from 10 subjects in different environments performing 9activities. In total, there are 215 action sequences in this dataset. The dataset includes only RGB and skeleton information. Some challenges in this dataset include laterality and similarity of the actions.

UTD-MHAD Dataset: this dataset [35] is recorded by Kinect v2, provides RGB, depth and skeleton information. The dataset includes 27 actions performed by 8 subjects. The actions in the dataset are divided into four categories: *sports actions*, *hand gestures*, *daily activities*, and *exercise*.

B. Results and Discussion

As explained, we need to tune the parameters of the framework in order to achieve the best performance. In total, we need to individually tune four parameters.

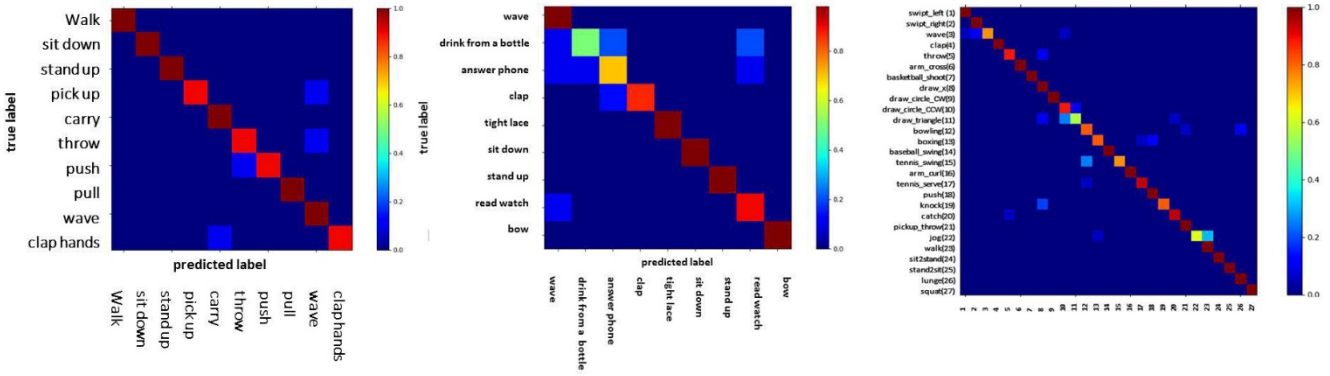


Figure 2: *Confusion matrices for: Left UTKinect; Center Florence 3D; Right: UTD-MHAD datasets*

Table 1. The optimal values of the parameters found for each dataset.

Setting	Dataset		
	UTKinect	Florence	UTD-MHAD
offset	7	6	9
GMM num	8	12	12
PCA cof num	100	60	40
Neuron num	3100	3100	3100

The offset parameter is the selected temporal offset that is used to calculate the temporal difference (Δf_t) while constructing the descriptors. Second parameter is the optimal number of the Gaussian mixture components used for calculation of the feature vector distribution. Moreover, different values for the PCA coefficients need to be investigated to find the best one. Finally, we need to determine the optimal neuron number for the ELM classifier. Table 1 shows the best parameter detected for each one of the datasets. We empirically investigated a range of different values searching for the optimal one. In the investigated range, the detected parameters ensure the optimal accuracy.

For the UTKinect dataset, we follow a leave-one-subject-out protocol in the experiments where we use subjects 1, 3, 5, 7, and 9 for training and subjects 2, 4, 6, 8, and 10 for testing. The overall result of our framework and comparison with the state-of-the-art methods applied in this dataset is reported in Table 2.

Table 2. Results and comparison with state-of-the-art on the UTKinect dataset.

Method	Accuracy (%)
HOJ3D [16] (LOSeqO)	90.9
Lie Group [18]	97.0
Spatiotemporal SHs [27]	93.0
PAIRWISE JOINTS [36]	94.4
Our method	**97.0**
LMNN [37] (LOOCV)	98.0
Multilayer LSTM [38]	95.9
ST-LSTM [30]	95.0
TS-LSTM [29]	96.9

Our framework achieves competitive results compared to the top methods in handcrafted and deep learning methods. We outperform all of the handcrafted methods by obtaining the second best accuracy among all of the reported methods. As it is showed in Figure 2-Left our method succeeds to recognize 7 out 10 activities with 100% accuracy. The similarity of motion patterns on the other three actions is the main reason those actions to get confused with each other (such as in ''throw'' and ''wave hand'' actions).

Table 3. Results and comparison with state-of-the-art on the Florence 3D dataset. *Notice that in these methods both skeleton and depth information is used and the evaluated protocol was easier.

Method	Accuracy (%)
Multi-Part Bag-of-Poses [34]*	82.0
Motion Trajectories [39]*	87.0
Elastic functional coding[40]	89.6
Our method	**89.6**
Li group [18]	90.8

Table 3 shows the results of applying our method to the Florence 3D dataset. The proposed framework obtain second-best performance among the state-of-the-art methods. The method proposed in [18] achieved a better accuracy with a small margin compared to ours. Again similarity of the action dynamics causes the confusion. According to Figure 2-Center, the highest confusion rate belongs to the ''Drink from bottle'' action which its motion pattern is very similar to ''Answer phone''. The other problem that makes this dataset a challenging one is the problem of laterality which makes the dataset to have a high intra-class variation. In some of the action classes, different hands are used to perform the same action. The same action is performed using left hand in some sequences and right hand in some other.

The common protocol for experiments on the UTD-MHAD dataset which is suggested by the providers is to take half of the subject with odd ids for training (subjects 1, 3, 5, and 7) and half of them with even ids for testing (2, 4, 6, and 8). We use the same protocol in our experiments.

Table 4 shows the overall accuracy of our method on this dataset and its comparison with the other approach in the literature. It can be noticed that our method achieves state-of-the-art accuracy in this challenging dataset outperforming both handcrafted and deep learning approaches. It is interesting to see that the proposed method beats all of the deep neural-network-based methods.

Table 4. Results and comparison with state-of-the-art on the UTD-MHAD dataset.

Method	Accuracy (%)
Kinect & Inertial [35]	79.1
Kinect & Inertial fusion [41]	91.5
ELC-KSVD [42]	76.1
Cov3DJ [43]	85.5
Our method	**92.6**
SOS_based CNN [28]	86.9
JTM_CNN [44]	85.8

The closest method to ours is [50]. However, this method uses information from the inertial sensors fused with joint information. By analyzing the confusion matrix of this dataset (Figure 2-Right) it can be concluded that the actions that share common poses usually get confused with each other. For example ''Jog'' action is classified with *62%* accuracy, while in *38%* of samples it is misclassified as ''Walk''. These actions are very similar to each other and share similar skeleton poses, hence, the calculated descriptors will be similar and will result in confusion of the classifier. Nevertheless, 15 out of 27 actions in the dataset is classified with 100% accuracy. Even though this dataset includes larger action classes, our framework achieves high accuracy that also ensures the scalability of the proposed method.

The training and classification of the proposed framework are efficient thanks to the reduced dimensionality of the feature vectors by the PCA method and the cost-effective calculations of the ELM algorithm. Based on the obtained results the framework is highly accurate in action recognition task owing to the discriminative descriptors encoded by the Fisher vector representation.

V. CONCLUSION

In this paper, we proposed a pose-based human action recognition framework empowered by spatiotemporal action descriptors. Instead of generative or discriminative methods, we used a hybrid Fisher vector method for encoding the action descriptors of the video sequences. The suggested framework is validated with three publicly available benchmark 3D action datasets and produced competitive results on those. As a future study, we will focus on improved modeling of the temporal dependencies of the action poses and calculation of discriminative poses to enable the framework to distinguish between the similar poses.

REFERENCES

[1] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," ACM Computing Surveys (CSUR), vol. 43, p. 16, 2011.

[2] R. Poppe, "A survey on vision-based human action recognition," Image and vision computing, vol. 28, pp. 976-990, 2010.

[3] S. Vishwakarma and A. Agrawal, "A survey on activity recognition and behavior understanding in video surveillance," The Visual Computer, vol. 29, pp. 983-1009, 2013.

[4] A. Eweiwi, M. S. Cheema, C. Bauckhage, and J. Gall, "Efficient pose-based action recognition," in Asian Conference on Computer Vision, 2014, pp. 428-443.

[5] S. Sadanand and J. J. Corso, "Action bank: A high-level representation of activity in video," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 1234-1241.

[6] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," arXiv preprint arXiv:1611.08050, 2016.

[7] G. Chéron, I. Laptev, and C. Schmid, "P-cnn: Pose-based cnn features for action recognition," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3218-3226.

[8] L. Zelnik-Manor and M. Irani, "Event-based analysis of video," in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, 2001, pp. II-II.

[9] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436-444, 2015.

[10] S. Chernbumroong, S. Cang, A. Atkins, and H. Yu, "Elderly activities recognition and classification for applications in assisted living," Expert Systems with Applications, vol. 40, pp. 1662-1674, 2013.

[11] F. Negin, F. Özdemir, C. B. Akgül, K. A. Yüksel, and A. Erçil, "A decision forest based feature selection framework for action recognition from rgb-depth cameras," in International Conference Image Analysis and Recognition, 2013, pp. 648-657.

[12] T. Jaakkola and D. Haussler, "Exploiting generative models in discriminative classifiers," in Advances in neural information processing systems, 1999, pp. 487-493.

[13] S. Ma, J. Zhang, N. Ikizler-Cinbis, and S. Sclaroff, "Action recognition and localization by hierarchical space-time segments," in Computer Vision (ICCV), 2013 IEEE International Conference on, 2013, pp. 2744-2751.

[14] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," Neurocomputing, vol. 70, pp. 489-501, 2006.

[15] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 1290-1297.

[16] L. Xia, C.-C. Chen, and J. Aggarwal, "View invariant human action recognition using histograms of 3d joints," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, 2012, pp. 20-27.

[17] A. A. Chaaraoui, J. R. Padilla-López, P. Climent-Pérez, and F. Flórez-Revuelta, "Evolutionary joint selection to improve human action recognition with RGB-D devices," Expert systems with applications, vol. 41, pp. 786-794, 2014.

[18] R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 588-595.

[19] M. Zanfir, M. Leordeanu, and C. Sminchisescu, "The moving pose: An efficient 3d kinematics descriptor for low-latency action recognition and detection," in Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2752-2759.

[20] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, "Learning realistic human actions from movies," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

[21] S. Agahian, F. Negin, and C. Köse, "Improving bag-of-poses with semi-temporal pose descriptors for skeleton-based action recognition," The Visual Computer, February 21 2018.

[22] D. Oneata, J. Verbeek, and C. Schmid, "Action and event recognition with fisher vectors on a compact feature set," in Computer Vision (ICCV), 2013 IEEE International Conference on, 2013, pp. 1817-1824.

[23] G. Varol and A. A. Salah, "Efficient large-scale action recognition in videos using extreme learning machines," Expert Systems with Applications, vol. 42, pp. 8274-8282, 2015.

[24] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in European conference on computer vision, 2010, pp. 143-156.

[25] S. M. Yoon and A. Kuijper, "Human action recognition based on skeleton splitting," Expert systems with Applications, vol. 40, pp. 6848-6855, 2013.

[26] X. Chen and M. Koskela, "Skeleton-based action recognition with extreme learning machines," Neurocomputing, vol. 149, pp. 387-396, 2015.

[27] C. Youssef, "Spatiotemporal representation of 3D skeleton joints-based action recognition using modified spherical harmonics," Pattern Recognition Letters, vol. 83, pp. 32-41, 2016.

[28] Y. Hou, Z. Li, P. Wang, and W. Li, "Skeleton optical spectra based action recognition using convolutional neural networks," IEEE Transactions on Circuits and Systems for Video Technology, 2016.

[29] I. Lee, D. Kim, S. Kang, and S. Lee, "Ensemble Deep Learning for Skeleton-Based Action Recognition Using Temporal Sliding LSTM Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1012-1020.
[30] J. Liu, A. Shahroudy, D. Xu, A. K. Chichung, and G. Wang, "Skeleton-Based Action Recognition Using Spatio-Temporal LSTM Network with Trust Gates," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
[31] J. C. Núñez, R. Cabido, J. J. Pantrigo, A. S. Montemayor, and J. F. Vélez, "Convolutional Neural Networks and Long Short-Term Memory for skeleton-based human activity and hand gesture recognition," Pattern Recognition, vol. 76, pp. 80-94, 2018.
[32] B. Li, M. He, Y. Dai, X. Cheng, and Y. Chen, "3D skeleton based action recognition by video-domain translation-scale invariant mapping and multi-scale dilated CNN," Multimedia Tools and Applications, pp. 1-21, 2018.
[33] R. Vemulapalli, F. Arrate, and R. Chellappa, "R3DG features: Relative 3D geometry-based skeletal representations for human action recognition," Computer Vision and Image Understanding, vol. 152, pp. 155-166, 2016.
[34] L. Seidenari, V. Varano, S. Berretti, A. Del Bimbo, and P. Pala, "Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses," in Computer vision and pattern recognition workshops (CVPRW), 2013 IEEE conference on, 2013, pp. 479-485.
[35] C. Chen, R. Jafari, and N. Kehtarnavaz, "Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor," in Image Processing (ICIP), 2015 IEEE International Conference on, 2015, pp. 168-172.
[36] M. Liu, C. Chen, and H. Liu, "Learning informative pairwise joints with energy-based temporal pyramid for 3D action recognition," in Multimedia and Expo (ICME), 2017 IEEE International Conference on, 2017, pp. 901-906.
[37] D. C. Luvizon, H. Tabia, and D. Picard, "Learning features combination for human action recognition from skeleton sequences," Pattern Recognition Letters, 2017.
[38] S. Zhang, X. Liu, and J. Xiao, "On geometric features for skeleton-based action recognition using multilayer LSTM networks," in Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on, 2017, pp. 148-157.
[39] M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo, "3-d human action recognition by shape analysis of motion trajectories on riemannian manifold," IEEE transactions on cybernetics, vol. 45, pp. 1340-1352, 2015.
[40] R. Anirudh, P. Turaga, J. Su, and A. Srivastava, "Elastic functional coding of human actions: From vector-fields to latent variables," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3147-3155.
[41] C. Chen, R. Jafari, and N. Kehtarnavaz, "A real-time human action recognition system using depth and inertial sensor fusion," IEEE Sensors Journal, vol. 16, pp. 773-781, 2016.
[42] L. Zhou, W. Li, Y. Zhang, P. Ogunbona, D. T. Nguyen, and H. Zhang, "Discriminative key pose extraction using extended lc-ksvd for action recognition," in Digital lmage Computing: Techniques and Applications (DlCTA), 2014 International Conference on, 2014, pp. 1-8.
[43] M. E. Hussein, M. Torki, M. A. Gowayyed, and M. El-Saban, "Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations," in Twenty-Third International Joint Conference on Artificial Intelligence, 2013.
[44] P. Wang, Z. Li, Y. Hou, and W. Li, "Action recognition based on joint trajectory maps using convolutional neural networks," in Proceedings of the 2016 ACM on Multimedia Conference, 2016, pp. 102-106.

Real Time Drowsiness Driver Detection and Low Cost Warning System

F.ÇAKMAK[1] and N.AYDIN ATASOY

[1] Karabuk University, Karabuk/Turkey, nesrinaydin@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, furkancakmak@yandex.com

Abstract **- The number of vehicles in our country is increasing day by day. Therefore, the most important reason for traffic accidents is human. These accidents are caused by reasons such as nuisance, insomnia, alcohol consumption, excessive speed, not obeying traffic rules.**

In this study, it is determined that the driver in the car is drowsy. If the driver closes own eyes and yawning, alert system gives a warning. In this real time application, eyes and mouth are detected using Viola-Jones algorithm. Eyes and mouth according to head movements are performed with Kanade-Lucas-Tomasi algorithm. A counter is used for no iris detection on the eyes and mouth is too more open state. These are using parameters in the developed system. The developing prototype system works successfully.

Keywords **– Low cost drowsiness driver detection, eyes and mouth detection, Viola-Jones algorithm, Kanade-Lucas-Tomasi algorithm.**

I. INTRODUCTION

THE number of vehicles in our country is increasing day by day thus increasing traffic accidents. The driver, pedestrian, vehicle and road can cause traffic accidents. The most important cause of the traffic accidents is the human. Causes of traffic accidents due to human factors are pessimism, obedience, insomnia for a long time, incorrect overtaking, excessive speed, using an alcoholic vehicle. Insomnia has a significant percentage of these reasons in traffic accidents [1]. There are many technological studies to remove this important effect. Generally, sleepy driver identification techniques are examined in 3 categories:

- examination of physical body properties,
- perception of driver's warning,
- perception of vehicle's reactions [2].

There has been great development in object recognition in recent years. Human face detection is an important research topic in the image processing application. Face features are usually used in the drowsiness driver detection applications. Examination of physical body properties methodologies includes opening area of mouth and eye, eye closure, eye blinking frequency and driver's head movements [3]. For this reason, face detection is detected first in these studies. With the rapid technological advancement of video cameras, monitoring technologies based on digital video analysis are developing. The most important goal is to remove restrictions on user behaviors.

The purpose of this work is to realize a low cost drowsiness drive detection system without restricting user movement. In the study, the head, eyes and mouth movements of the driver in the vehicle are followed and the state of being drowsy is determined. Real time images are obtained from the web camera. The driver's head movements are tracked using the Kanade-Lucas-Tomasi algorithm (KLT). The right eye, left eye, and mouth of the driver are detected using the Viola-Jones algorithm. Hough Circle Transform (HCT) is used to find iris on the eyes. The alert system is activated when the iris is not found and the driver's mouth is too more open. Thus, the alert tone rings at certain intervals. Opening the driver's mouth and blinking states are checked from the system. The low cost real time system that controls successfully eyes and mouth at the same time.

II. METHOD

Physiological based approach (pulse rate, heart rate and brain activity information) for drowsiness drive detection can usually give good results. But, these approaches must require physical connection such as connecting electrode to the driver body and disturb the driver at the moment driving [4]. Therefore, images are obtained using a web camera. Thus, it is realized at low cost.

In the study, face area is detected firstly. Object detection in the real time video is realized using Viola-Jones algorithm that is developed by Paul Viola and Michael Jones in 2001. Haar Cascade Classifier (HCC), which is successfully used in many object finder applications, is first proposed by Viola and Jones, developed by Lienhart [5]. The algorithm works in 3 basic steps:

- Creating an "integral image" for fast feature detection with rectangle features called Haar features;
- AdaBoost learning method;
- HCC for integrating many features efficiently.

Basic features are used for object detection are shown in Figure 2.

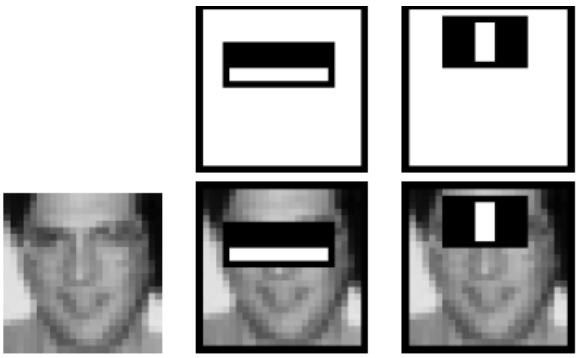


Figure 1: Using basic features in HCC.

Viola-Jones algortihm uses AdaBoost Machine Learning Method for Haar features and choosing threshold value. AdaBoost creates strong classifier by successively bringing weak classifiers. This method is suggested from Freund and Schapire in 1995.

According to method proposed by Freund and Schapire, a strong classifier is created from the linear combination of some weak classifiers trained for each attribute representing the sample space [6]. Figure 2 shows the working principle of the strong classifier created from sequential weak classifiers.

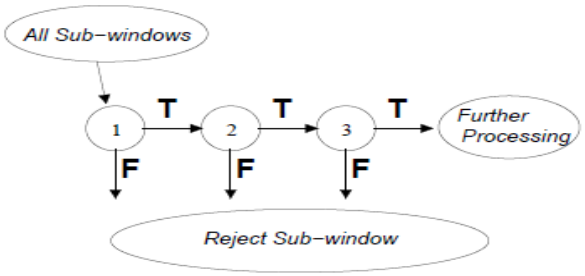


Figure 2: Schematic presentation of detection cascade [5].

In the study, eyes and mouth are detected after the face area is found using Viola-Jones algorithm. Head movements tracking is realized with KLT algorithm. KLT algorithm is highly preferred algorithm for feature tracking. KLT algorithm is used for detecting scattered feature points when motion of image is sufficiently small [7-10]. Thus, it is used in this study. KLT algorithm defines an alignment template *T(x)* and the tracker tries to end tracked points in *T(x)* according to the previous state, called initial points. The tracker will search different points to find alignment if the image and the template do not converge. The search process will stop when the best alignment is found. Thereafter, the new state will constitute the initial state for the next state [11, 12].

This system is low cost so images are capture from web camera. Iris on the eyes is detected using HCT that is shape based method. HCT is proposed from Paul Hough and is patent by IBM for detecting geometric shapes. HCT is a standard method for detecting shapes so is identified line, circle, and ellipse [13].

III. Implementation of Application

The application is developed on a machine with an Intel Core i7-4510U CPU, 2.00 GHz, 8 GB RAM, 64-bit operating system and using MATLAB R2017b. Algorithm steps of the application are shown in Figure 3. These steps are told below in orderly.

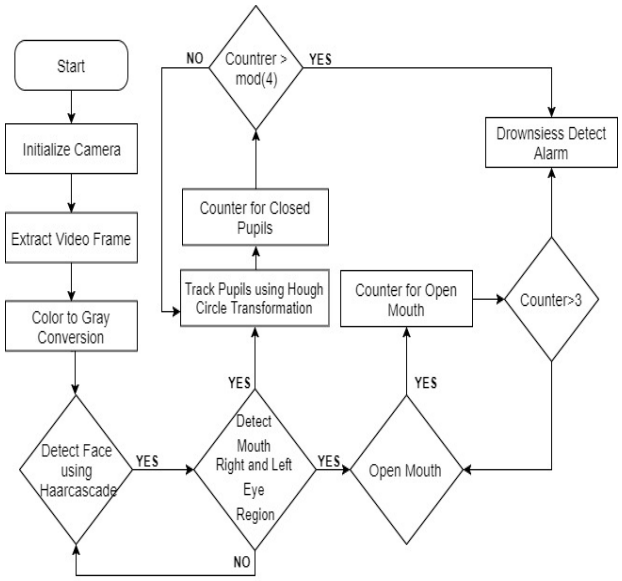


Figure 3: The general flow diagram of application.

Images are captured from web camera that are converted RGB to Binary according to threshold. The threshold value is determined according to the average of light values. Real time face region detection is realized using HCC. That is shown in Figure 4(a).

Figure 4: (a) Detecting face area.

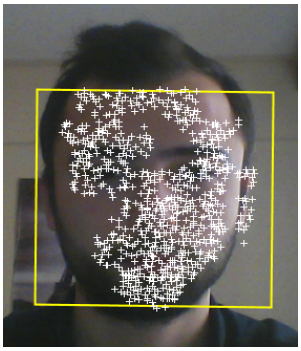


Figure 4: (b) Creating point trackers for tracking head movements.

Point trackers that are created using KLT algorithm for tracking head movements are shown in Figure 4(b). Probability of face area detection is increased using KLT algorithm.

Eyes and mouth on the face area are identified as shown Figure 5(a) and 5(b).

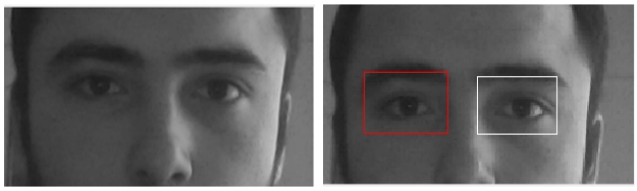


Figure 5: (a) Detecting right and left eye.

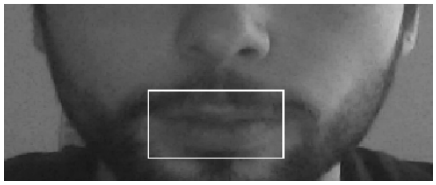


Figure 5: (b) Detecting mouth.

In the study, presence or absence of iris is a parameter for is determining the driver in the car is drowsy or not. Detecting iris is seen in Figure 7.

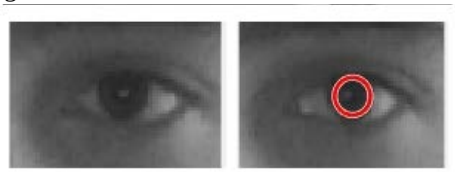


Figure 7: Detecting iris using HCT.

Right and left eye are detected separately. Because, if the driver looks at right or left side, open or close situation of each eye can be detected easily using web camera. Canny edge is applied on detection right or left eye image. Canny edge detection is a popular edge detection algorithm. After that, iris on the eye is detected using HCT.

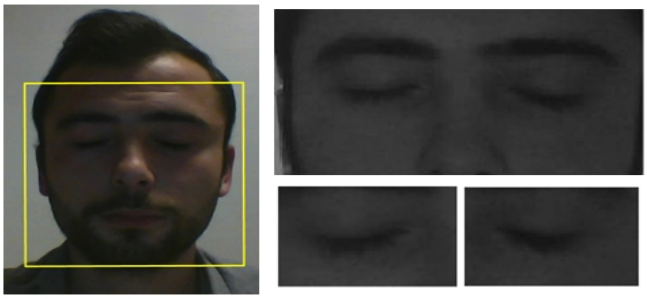


Figure 8: Iris isn't detected when eyes are closed.

Iris detection is realized for each frame on the real time image. A counter is created for short term of closed eye as blinking or not detecting iris. If iris situation isn't detected during 4 frame/4 sec as Figure 8, the counter is increased and alert system is activated.

Open and close mouth detection is realized using Viola-Jones algorithm. These are shown Figure 9(a) and (b).

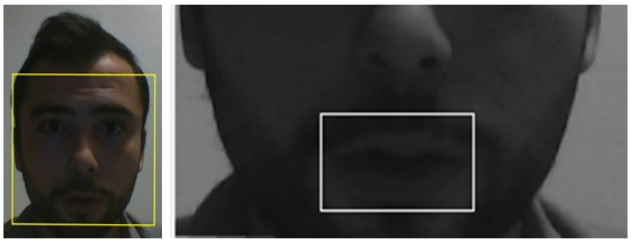


Figure 9: (a) Detecting closed mouth.

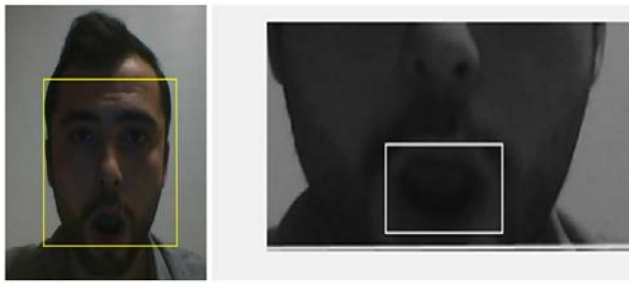


Figure 9: (b) Detecting open mouth.

It is controlled that the mouth is too open for detecting the driver's yawning. If the driver yawns 3 times consecutively, alert systemactives.

IV. CONCLUSION

In this study, the low cost is implemented a prototype system that detects yawns and eye closures to control whether the driver is drowsy or not. Input is captured by the web camera and the output is in the form of a buzzer that gives the driver an alert as and when drowsiness is detected. The system is identified successfully drowsiness driver. But, light condition is very important factor. And then, distance of between web camera and the driver is varied according to the driver length and seat situation. Optimum values are between 30 and 45 cm as shown in Table 1. Object detection cannot be realized for more distance. Skin color of the driver is not very important because of the chosen method.

Table 1: Determining distance and light.

Distance (cm)	**Light (brightness)**		
	Little	**Normal**	**Much**
15	Not detected face	Not detected face	Not detected face
30	Detected iris, mounth and eyes.	The most detected iris, mounth and eyes.	Detected iris, mounth and eyes.
45	Detected iris, mounth and eyes.	The most detected iris, mounth and eyes	Detected iris, mounth and eyes.

In future work, it is aimed to develop an embedded system by adding the steering movements of the driver so a more complex system will be developed. People will be able to use technology in expensive cars as low-cost and they will be able to install the system themselves.

REFERENCES

[1] Dr.İ. Hamit Hancı, Dr. Burcu Eşiyok, "Uyku ve Trafik Kazaları" Ege Üniversitesi Tıp Fak. Adli Tıp A.D. Trafik Gönüllüleri Derneği.

[2] Sezgin Hayırli, "Uykulu Sürücü Belirleme Sistemi", Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2005.

[3] Rajeshwari Sanjay Rawal, Sameer.S.Nagtilak, Drowsiness Detection Using RASPBERRY-PI Model Based On Image Processing, International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 10, Oct-2016.

[4] Tejasweeni Musale, Prof. B. H. Pansambal, Real Time Driver Drowsiness Detection system using Image processing, International Journal for Research in Engineering Application & Management (IJREAM) ISSN : 2494-9150 Vol-02, Issue 08, Nov 2016.

[5] P. Viola, M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[6] Freund Y, Schapire RE. Experiments with a new boosting algorithm. In: Machine Learning: Proceedings of the Thirteenth International Conference; July 1996; Bari, Italy. San Francisco, CA, USA: Morgan Kaufmann Publishers. pp. 148-156.

[7] Ritesh Boda and M. Jasmine Pemeena Priyadarsini, Face Detection And Tracking Using Klt And Viola Jones, ARPN Journal of Engineering and Applied Sciences, VOL. 11, NO. 23, December 2016.

[8] Sinha SN, Frahm JM, Pollefeys M, Genc Y. GPU-based video feature tracking and matching. In: Workshop on Edge Computing Using New Commodity Architectures; 2006.

[9] Tomasi C, Kanade T. Detection and Tracking of Point Features. Technical Report CMU-CS-91-132. Pittsburgh, PA, USA: Carnegie Mellon University, 1991.

[10] Fassold H, Rosner J. Realtime KLT Feature point tracking for high denition video. In: International Workshop on Computer Graphics, Computer Vision and Mathematics; 2009.

[11] Bagherpour P, Cheraghi SA, Mokji MbM. Upper body tracking using KLT and Kalman lter. Procedia Computer Science 2012; 13: 185-191.

[12] Nesrin AYDIN ATASOY, Abdullah Cavusoğlu, Ferhat ATASOY, Real-time motorized electrical hospital bed control with eye-gaze tracking, Turk J Elec Eng & Comp Sci (2016) 24: 5162-5172.

[13] L. Schwarz, H. R. Gamba, F. C. Pacheco, R. B. Ramos ve M. A.Sovierzoski, "Pupil And Iris Detection In Dynamic Pupillometry Using the OpenCV Library", 5th International Congress on Image and Signal Processing (CISP 2012), Curitiba, 2012.

Local Variance Switching Gaussian Filter

A. DEĞİRMENCİ[1] İ. ÇANKAYA[1] Ö. KARAL[1] and R. DEMİRCİ[2]

[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, adegirmenci@ybu.edu.tr
[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, icankaya@ybu.edu.tr
[1] Ankara Yıldırım Beyazıt University, Ankara/Turkey, karal@ybu.edu.tr
[2]Gazi University, Ankara/Turkey, rdemirci@gazi.edu.tr

Abstract **- Filtering can be used for different purposes in image processing. One of them is to reduce noise in the image. In this study, local variance based on switching filter is designed to remove Gaussian noise from gray scale images. During filtering process, local variances of each pixel is calculated and then pixels are classified as five clusters according to their local variance values by the k-means clustering method. Depending on the result of the clustering, variance of the Gaussian filter kernel is tuned. In the smooth regions, in which variances of the pixels are low, higher standard deviation Gaussian filter is applied. Higher variance pixels represent the edge pixels, therefore lower standard deviation Gaussian kernel is applied to preserve the edges. In clusters with medium variance pixels, the standard deviation value in the Gaussian filter is changed depending on the local variance values. Experimental results show that designed local variance based switching filter gives better performance to remove the Gaussian noise at various levels compared to the classical filters.**

Keywords **– Noise, Gaussian filter, Switching filter, Image processing.**

I. INTRODUCTION

NOISE affects the visual appearance in the picture, as well as distorting the information in the image. The deterioration of the information on images causes a reduction in the success of image processing techniques to be applied later, such as edge detection, segmentation. For instance, falsely detected edges may occur in the images (noisy pixels can be regarded as edges), and edges pixels cannot be detected. In another example, background pixels can be detected as object pixels and the objects cannot be segmented adequately. To overcome these problems, noise in the images needs to be reduced, and this is done by applying filters to the images. During the filtering process, the noise in the image is reduced, however this can cause loss of information. Due to this reason, different types of filters are developed to effectively reduce the noise [1,2].

Gaussian noise is one of the additive noise which can be originated from during acquisition process of the imaging. To reduce the Gaussian noise in the images, Gaussian and averaging filters are commonly used. Since, weights and mask size are constant in the classical filters, the performance of these filter are depending on the selected filter parameters. Therefore, adaptive or switching filters have been studied by researchers. For example, Lee performed a noise filtering by using local statistics of the image [3]. In this study, filtering methods were developed for additive noise, multiplicative noise, combination of both additive and multiplicative noise. Filtering algorithm was developed from local mean and local variance of the pixels. Rank et al. designed local feature based adaptive recursive filter [4]. This filter was composed of two phases which were feature detector and adaptive recursive low-pass filter. In the first phase, edges, flat regions and spots were detected. In the second phase, adaptive recursive lowpass filter was applied to the pixel based on the detected features which were determined in the first phase. To preserve the edges, weak low-pass filter or no filter was applied to the edges. Strong low pass filter was applied to the flat regions to increase the smoothness.

The size of the filter can be adjusted during the filtering process. For instance, Vijaykumar et al. proposed an adaptive window based filter for removing Gaussian noise in both gray scale and color images [5]. The size of the filter mask was changed by comparing the local variance value with the threshold value. It was obtained by multiplication of the smoothing factor and the noise variance of the image. To calculate the noise variance, flat region was founded from the noisy image and this region's variance was set as a noise variance. Later, processed pixel's variance value was calculated. If the local variance was smaller than the threshold value, processed pixel's intensity value was changed by the average value within the filter size. Otherwise, filter size was increased, and the process was applied again for the bigger mask size. Nguyen et al. performed an adaptive noise detection and filtering method to remove Gaussian noise [6]. In this method, local maximum, local weighted mean and local weighted activity was used for detection of the noise. Modified Gaussian filter was applied to the pixels where noise is detected. In the small local activity regions, smoothing becomes higher. However, smoothing is lowered in the high activity regions. Ilangi et al. proposed three different hybrid filtering methods to remove Gaussian noise from medical images [7]. The first method was called as hybrid cross median filter. In this method, filtered pixel intensity value was altered with the median value of the left transversal (LT) neighbors median value, right transversal (RT) neighbors median value and the processed pixel's intensity value. The other methods were named as hybrid min filter and hybrid max filter. In these methods, processed pixel's intensity value was equal to minimum or maximum value of the, LT neighbors median, RT neighbors median and the processed pixel's intensity value. Leavline et al. used fast

multiscale directional filter banks (FMDFB) to remove from the gray scale images [8]. To obtain decomposed coefficients, the FMDFB method was applied to the noisy image. After, threshold was determined from the median absolute deviation and variance of the noise. Then, soft thresholding is implemented to the transformed coefficients. Finally, noise-reduced image was obtained by applying inverse FMDFB to the thresholded coefficients. Murugan et al. designed a two-stage system to remove Gaussian noise from gray scale images [9]. In the first stage, the Winery filter was applied to the noisy image to reduce the gaussian noise at a certain level. In the second stage, local and global histograms were used for further noise reduction. The globally most probable gray level was founded for the image, and the local most probable gray level was calculated for each pixel. Absolute difference of the minimum of local and global most probable gray level was compared with the threshold value. If the difference was higher than threshold value, it was considered as noisy pixel and replaced with the minimum of global and local most probable gray level.

In this study, local variance based switching image filter was designed. For this purpose, local variance was calculated for every pixel in the noisy image. Local variance values were grouped into five clusters by k-means clustering methods. The switching mechanism of the designed filter was changed based on the cluster of the local variance values of the pixels. Designed filter is tested with the classical test images which are Cameraman, Forest, GantryCrane and Peppers. According to the results, switching filters has a higher performance compared to the classical filters.

II. METHODS

A. Gaussian Filter

The weights in the Gaussian filter mask is determined by using Gaussian function. 2-dimensional Gaussian function is defined as

$$F(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

where σ defines standard deviation. x and y correspond to the coordinates. In the application of Gaussian filter, two parameters are defined by the user. These are the mask size and the standard deviation value. Based on these two parameters, smoothness of the image increases or decreases. The first one is the mask size. When the mask size of the filter is increased, the smoothness of the processed pixel increases because more pixels are considered in computation of the processed pixel. Second parameter is the standard deviation. It affects the distribution of the Gaussian function. Therefore, it determines the distribution of the weights in the filter mask. The effect of the standard deviation in the range of 0.5 to 2.5 on the Gaussian filter mask is shown in Figure 1. As shown in Figure 1 (a), when the standard deviation equal to the 0.5 in the Gaussian filter mask, 61.93% of the processed element is determined by the center element of the mask and 33.52% is determined by the four diagonal neighbors. The increase in standard deviation causes the weights to approach each other in the filter mask. For example, as demonstrated in the Figure 1 (e), the weights of the Gaussian filter mask are nearly equal in the 3X3 mask size, when the standard deviation equals to 2.5.

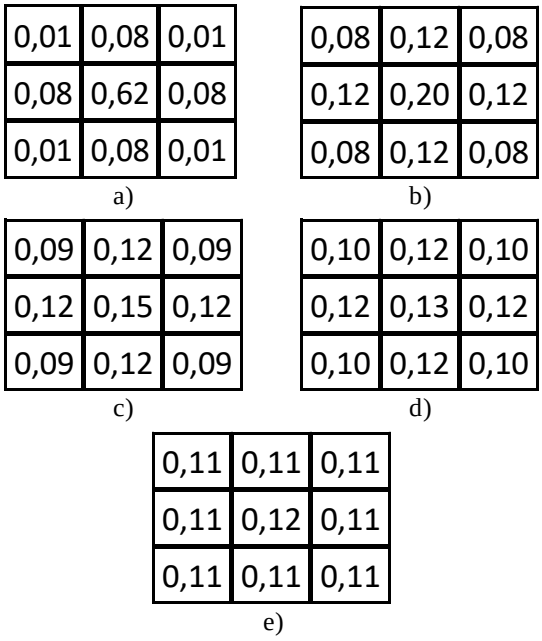


Figure 1: Gaussian filter masks with different standard deviation values a) 0.5, b) 1.0, c) 1.5, d) 2.0, e) 2.5

B. Local Variance and K-Means Clustering

Local variance in the images is defined by

$$\sigma^2=\frac{1}{KL}\sum_{i=-\frac{KL}{2}}^{\frac{KL}{2}}\sum_{j=-\frac{KL}{2}}^{\frac{KL}{2}}\left[f(x+i,y+j)-\mu\right]^2 \tag{2}$$

where f is the input image, K and L determines the local width and height and μ represents mean, which is calculated by

$$\mu=\frac{1}{KL}\sum_{i=-\frac{KL}{2}}^{\frac{KL}{2}}\sum_{j=-\frac{KL}{2}}^{\frac{KL}{2}}f(x+i,y+j) \tag{3}$$

The local variance value of the pixel gives the distribution between the local average intensity value and the neighborhood pixels. This information can be used to locate the edges of the image, smooth areas, and noisy pixels. Because the local variance value is expected to be low values in the smooth regions, high values in the edges and between the low and high values in the noisy pixels. For the most part, the local variance value is increasing depending on the noise intensity and can be grouped by clustering methods.

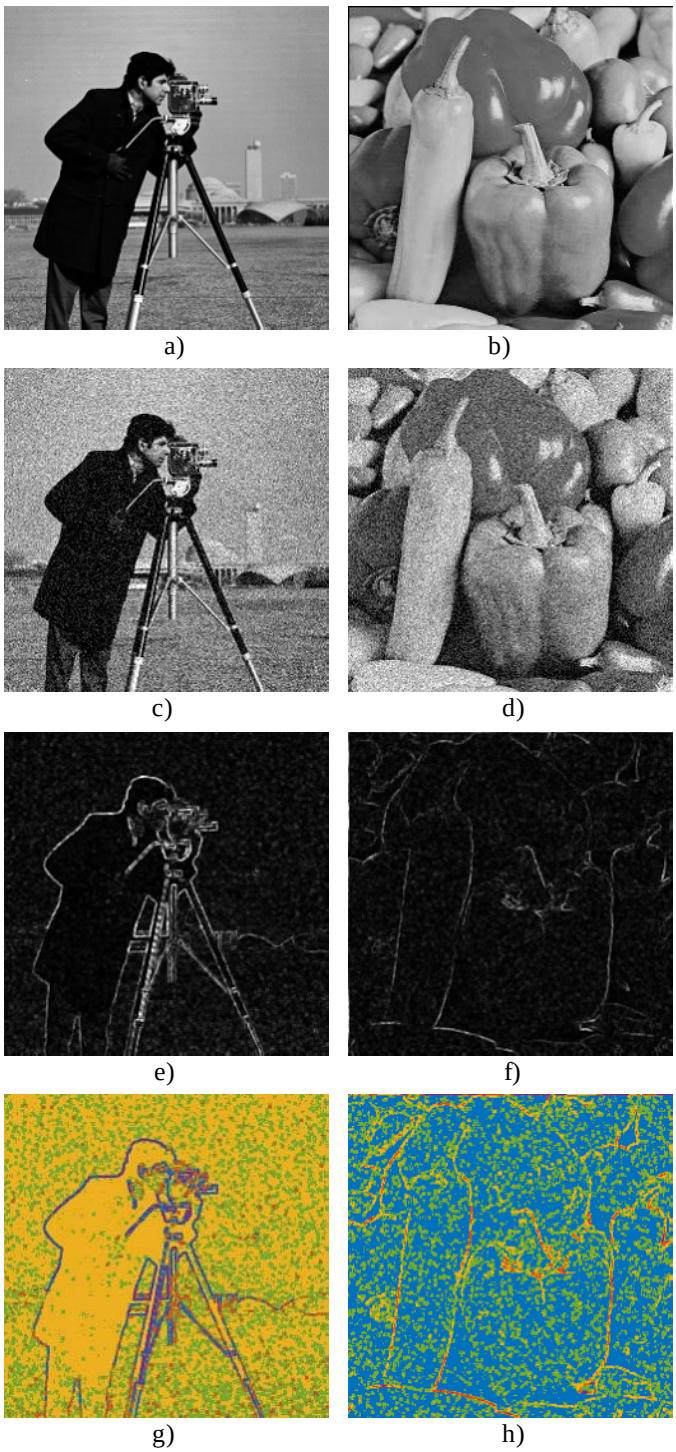


Figure 2: a) Original Cameraman image, b) Original Peppers image, c) Cameraman with $\sigma^2 = 0.01$ Gaussian noise, d) Peppers with $\sigma^2 = 0.01$ Gaussian noise, e) gray scale image of the local variance of (c), f) gray scale image of the local variance of (d), g) k-means clustering result of (e), h) k-means clustering result of (f)

After clustering, the data in the same cluster must be similar and the data in different clusters should be quite different from each other. There have been many methods to cluster the data. One of the widely used clustering algorithm is the k-means clustering method [10]. Therefore, this method is used in images obtained from different fields such as medical [11,12] and satellite [13,14].

In this method, the input data is separated into the number of k clusters. Additionally, algorithms that automatically determine the number of clusters are also developed. k-means clustering method is performed in 6 steps. In the first step, number of clusters are determined. In the second step, number of k points are determined randomly as a cluster centers. In the third step, distance between each point and the clusters centers are calculated. In the fourth step, each data is assigned to the cluster which gives the minimum distance. After that, new cluster centers are computed. The third and the forth step is repeated until cluster centers are not changing. The Cameraman and Pepper images are shown in Figure 2 (a) and (b), $\sigma^2 = 0.01$ Gauss noise added images are shown in Figure 2 (c) and (d). In Figure 2 (c) and (d), $\sigma^2 = 0.01$ Gaussian noise added Cameraman and Pepper images are presented. In Figure 2 (e) and (f), the local variance of each images are visualized. The result of k-means clustering is shown in Figure 2 (g) and (h). Clusters are demonstrated with different colors.

III. SWITCHING FILTER

The parameters of the filters are constant in the application of standard filters. Therefore, the same filter is applied to the entire image. This makes the choice of filter type and parameters to be applied more important. Because, depending on the selected parameters, edge preserving and noise reduction performance is affected. For example, when a Gaussian filter with a small standard deviation value is selected, the amount of blurring in the image becomes small. However, the reduction of the noise becomes low. In the other case, when the standard value in the Gaussian filter is set to higher values, noise reduction becomes higher, but amount of blurring is increased compared to the Gaussian filter with lower standard deviation. When these factors are considered, it is expected that the switched filters perform better than the standard filters. Because, the filter type and parameters can be adjusted during the filtering based on the features of the image.

In this study switching filter is designed for Gaussian noise reduction. The switching mechanism is based on the local statistics of the pixels. For that purpose, local variance of each pixel is calculated within the 3x3 window. Later, local variance of each pixel is separated into five groups by k-means clustering method. Following that, the switching mechanism of the filter is changed based on the cluster of the pixel. Clusters of the highest variance are corresponded to the location of the edges. For this reason, a σ=0.5 Gaussian filter is applied in these pixels. The regions, where the local variance is minimum, is corresponded to the homogenous regions. So, the σ of the Gaussian filter is selected as the maximum value which is 2.5. In this way, the homogeneity in these regions is increased and the noise is effectively suppressed. Clusters between maximum and minimum variance clusters are considered as noisy pixels. Generally, magnitude of the local variance and the noise intensity vary in direct proportion. Thus, the magnitude of the σ in the Gaussian filter are changed depending on the local variance cluster of the pixel. Gaussian filter with σ values of 2,

1.5 and 1 is applied to the three clusters between the maximu m and minimum set. Since Gaussian filter with small σ is applied to the low noise density regions, the information in these regions is preserved. In addition, by applying a Gaussian filter with a high σ value to the regions with high noise intensity, the noise is effectively reduced. The algorithm of the designed switching filter is given in Figure 3.

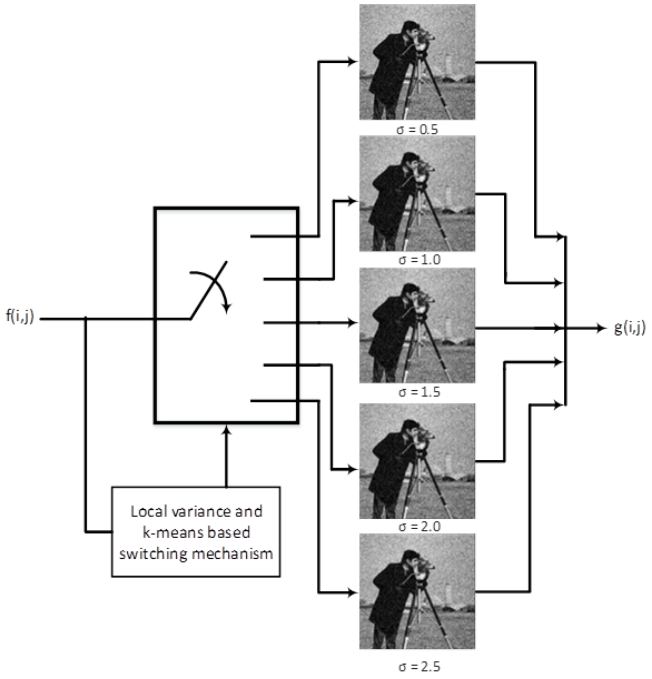


Figure 3: Switching mechanism of the designed filter

IV. EXPERIMENTAL RESULTS

The designed switching filter is tested with several grayscale images to evaluate the performance. For this purpose, zero mean Gaussian noise with various variance levels are added to the test images. To measure the performance, mean square error (MSE) and peak signal to noise ratio (PSNR) are calculated. MSE is described as

$$MSE = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(f(i,j)-g(i,j)\right)^{2} \quad (4)$$

where *M* and *N* are the size of the image. *f* represents the input image and *g* corresponds to the filtered image. PSNR, is calculated by

$$PSNR = 20\log_{10}\frac{I_{\max}}{\sqrt{MSE}} \quad (5)$$

where I_{max} equal to the 255 in 8-bit uniformly quantized grayscale image. A filter, which gives a lower MSE value and a higher PSNR value, is more successful in reducing noise compared to the other filters.

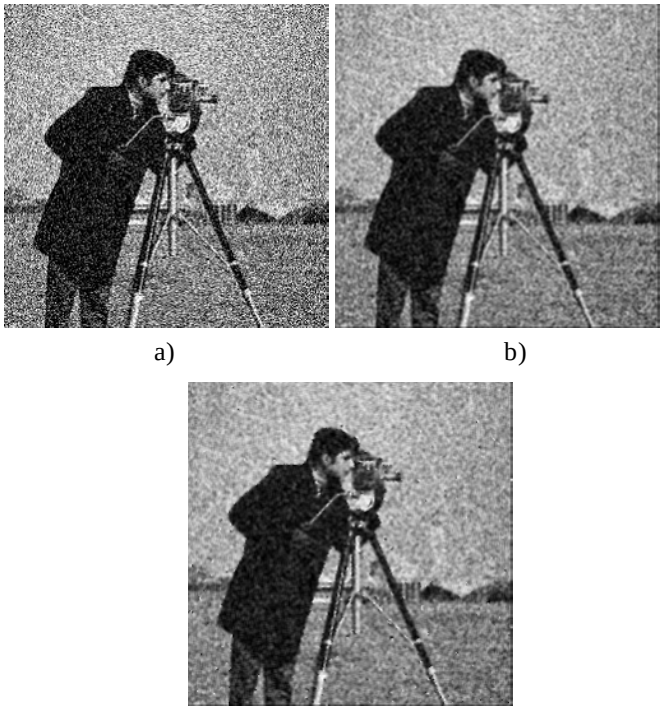
Table1: Experimental results: MSE

Images	Noise Level	Noisy	Gauss σ=1	Median	Average	Switching Filter
Cameraman	0,01	593	224	252	271	220
	0,02	1155	308	361	347	296
	0,03	1683	387	469	418	371
	0,04	2192	473	593	496	463
Forest	0,01	508	192	281	228	187
	0,02	932	267	373	297	257
	0,03	1329	349	456	374	339
	0,04	1707	429	541	448	424
Gantry Crane	0,01	633	206	249	244	203
	0,02	1228	287	367	315	273
	0,03	1802	365	485	384	347
	0,04	2325	435	590	447	427
Peppers	0,01	627	175	175	203	168
	0,02	1226	252	292	271	237
	0,03	1770	332	409	343	314
	0,04	2291	396	517	400	386

Table 2: Experimental results: PSNR

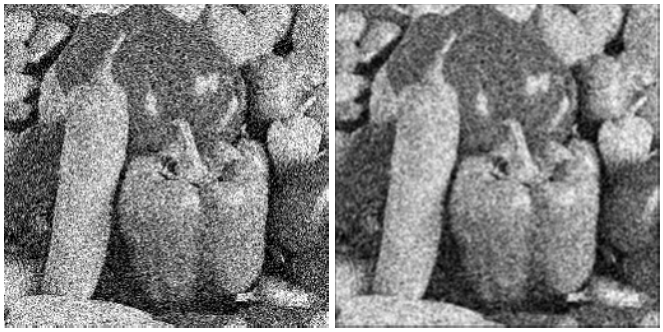
Images	Noise Level	Noisy	Gauss σ=1	Median	Average	Switching Filter
Cameraman	0,01	20,40	24,63	24,12	23,80	24,70
	0,02	17,50	23,25	22,56	22,73	23,42
	0,03	15,87	22,26	21,42	21,92	22,44
	0,04	14,72	21,38	20,40	21,17	21,48
Forest	0,01	21,07	25,30	23,65	24,55	25,40
	0,02	18,44	23,87	22,42	23,40	24,03
	0,03	16,90	22,70	21,54	22,40	22,83
	0,04	15,81	21,80	20,80	21,61	21,86
Gantry Crane	0,01	20,12	24,99	24,16	24,26	25,06
	0,02	17,24	23,55	22,48	23,14	23,77
	0,03	15,57	22,51	21,28	22,28	22,73
	0,04	14,47	21,74	20,42	21,63	21,82
Peppers	0,01	20,16	25,71	25,71	25,07	25,89
	0,02	17,24	24,12	23,48	23,80	24,39
	0,03	15,65	22,92	22,02	22,78	23,16
	0,04	14,53	22,15	21,00	22,12	22,26

Performance of the designed filter is compared with the averaging, median, σ = 1 Gaussian filter. Results of MSE and PSNR are given in the Table I and II, respectively. During the test, a 3X3 mask size is chosen in the filters. As seen from the Table I, MSE result of the Switching filter is lower than the compared filters. Additionally, PSNR result of the designed filter is higher than the compared filters. The closest results to the switched filter are in the Gaussian filter. In Figure 4 and 5, several test images and their filtering results are shown. As can be seen, the designed switching filter is able to preserve the edges while reducing Gaussian noise.



a) b) c)

Figure 4: a) Noisy image, b) Gaussian filtered image with σ=1, c) Switching filter



a) b) c)

Figure 5: a) Noisy image, b) Gaussian filtered image with σ=1, c) Switching filter

V. CONCLUSIONS

In this study, switching filter was designed for noise reduction from gray scale images. In the switching mechanism, local statistics and the clustering methods were used. The local variance of each pixel was calculated and according to the local variance values, the switching mechanism of the filter was automatically changed. The switching mechanism automatically adjusts according to the noise intensity of the pixel. Homogeneity was increased in smooth regions while blurring was decreased in the edges. Thus, the information on the edges was preserved. In the future, the type and size of the filter could be changed in the switching mechanism to improve the performance of the filter and the visual quality of the image.

REFERENCES

[1] R. C. Gonzalez, and R. E. Woods, Digital image processing, 2002.

[2] R. Szeliski, Computer vision: algorithms and applications. Springer Science & Business Media, 2010.

[3] J. S. Lee, "Digital image enhancement and noise filtering by use of local statistics," IEEE transactions on pattern analysis and machine intelligence, vol. 2, pp. 165-168, 1980.

[4] K. Rank, and R. Unbehauen, "An adaptive recursive 2-D filter for removal of Gaussian noise in images," IEEE Transactions on Image Processing, vol. 1, no. 3, pp. 431-436, 1992.

[5] V. R. Vijaykumar, P. T. Vanathi, and P. Kanagasapathy, "Adaptive window based efficient algorithm for removing gaussian noise in gray scale and color images," In Conference on Computational Intelligence and Multimedia Applications, vol. 3, pp. 319-323, December 2007.

[6] T. A. Nguyen, M. J. Kim, and M. C. Hong, "Fast and efficient Gaussian noise image restoration algorithm by spatially adaptive filtering," In Picture Coding Symposium (PCS), pp. 122-125, December 2010.

[7] G. Ilango, and R. Marudhachalam, "New hybrid filtering techniques for removal of Gaussian noise from medical images," ARPN Journal of Engineering and Applied Sciences, vol. 6, no. 2, pp. 8-12, 2011.

[8] E. J. Leavline, and S. Sutha, "Gaussian noise removal in gray scale images using fast Multiscale Directional Filter Banks," In Recent Trends in Information Technology (ICRTIT), pp. 884-889, June 2011.

[9] V. Murugan, and R. Balasubramanian, "An Efficient Gaussian Noise Removal Image Enhancement Technique for Gray Scale Images," World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering, vol. 9, no. 3, 2015

[10] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, vol. 1, pp. 281-297, 1967

[11] L.H. Juang, M.N. Wu, MRI brain lesion image detection based on color-converted K-means clustering segmentation, Measurement, vol. 43, no. 7, pp. 941-949, 2010.

[12] S. Khanmohammadi, N. Adibeig, S. Shanehbandy, "An improved overlapping k-means clustering method for medical applications", Expert Systems with Applications, vol. 67, pp. 12-18, 2017.

[13] A. Rekik, M. Zribi, M. Benjelloun and A. b. Hamida, "A k-Means Clustering Algorithm Initialization for Unsupervised Statistical Satellite Image Segmentation", 2006 1ST IEEE International Conference on E-Learning in Industrial Electronics, Hammamet, pp. 11-16, 2006.

[14] T. Celik, "Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and k-Means Clustering", in IEEE Geoscience and Remote Sensing Letters, vol. 6, no. 4, pp. 772-776, 2009.

Image Compression Using SVD Method

A. SUTCU[1] A. DEGIRMENCI[1] O. KARAL[1] and I. CANKAYA[1]

[1] Ankara Yildirim Beyazit University, Ankara/Turkey, aslihansutcu@gmail.com
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, adegirmenci@ybu.edu.tr
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, karal@ybu.edu.tr
[1] Ankara Yildirim Beyazit University, Ankara/Turkey, icankaya@ybu.edu.tr

Abstract **- As is known, it is often difficult to store and transmit images used in various computer applications. A possible solution to this problem is to use one of the known data compression techniques because they help to reconstruct the image with a lower number of measurements. In this study, a new Singular Value Decomposition (SVD)based technique is proposed to compress images. The advantage of using SVD is that it both has energy compression capability and is easily adaptable to local statistical variations of the image. Furthermore, the SVD can be implemented with non-square, non-reversible matrices of size m x n. However, how to determine the threshold value for image compression in the SVD technique is still one of the fundamental problems. In this study, the desired threshold value is calculated by dividing the sum of the differences between the obtained singular values by the rank of the matrix. Simulation results confirm the feasibility of our proposed method.**

Keywords **– Singular Value Decomposition, Lossy Image Compression, Data Compression**

I. INTRODUCTION

IMAGE processing is a widely used method for obtaining an enhanced image from digital images recorded on electronic devices or for extracting some useful information therefrom. For example, almost everyone today has a smartphone, and these phones are taking dozens of photos every day. Because people want to immortalize their photographs with their time. This causes the number of captured and stored images to increase. For this reason, large amounts of memory space are required to store such images in digital media. This problem can be overcome by compressing images containing large amounts of data at appropriate compression ratios. Many algorithms have recently been suggested for compressing and reconstructing digital images. These algorithms can be basically divided into two groups: lossless and lossy compression algorithms. In lossless algorithms, the original image can be completely reconstructed from the compressed image [1]. Lossless image compression algorithms are required for sensitive data applications such as medical images, legal business documents, and remote sensing [2-7]. In lossy algorithms, the original image cannot be reconstructed from the compressed image because some information is lost. However, lossy compression algorithms are more commonly used in image compression since they can achieve high compression ratios (CR) compared to lossless compression algorithms. Lossy compression algorithms are commonly used to compress multimedia applications such as video, audio, and image that are used in internet and media activities [8-15].

In recent years, many compression algorithms have been proposed [15-20]. Singular Value Decomposition is one of the well-known and commonly used lossy image compression techniques. In SVD-based algorithm, an image matrix is decomposed into three matrices such as U, S, and V in which U and V are orthogonal and S is a diagonal matrix containing the sorted singular values of the image matrix in descending order. Compression is done by removing some small singular values in the diagonal of S matrix. However, in the literature survey, there is no specific method for determining which singular values should be eliminated in SVD-based image compression. This study suggests a new thresholding technique based on differences between singular values to remove singular values.

The study is organized as follows. In Section 2, the singular value decomposition (SVD) is given in details. Section 3 discusses the implementation of proposed SVD model. The experimental results are given in Section 4 followed by conclusion in the last section.

II. SINGULAR VALUE DECOMPOSITION

Singular Value Decomposition is a method used to decompose the given matrix into three matrices such as U, S, and V in which U and V are orthogonal and S is a diagonal matrix. More specifically, suppose we are given a matrix A with *m* rows and *n* columns, with rank *k* with $k \le n \le m$. Then the SVD representation of the matrix A is given by

$$A = USV^T \quad (1)$$

where, $U \in R^{mxm}$ and $V \in R^{nxn}$ are orthogonal matrices and $S \in R^{mxn}$ is a diagonal matrix in which the entries along the diagonal of S are singular values of A. In matrix form of Equation (1) can be written as follows.

$$U = \begin{bmatrix} u_{11} & u_{12} & \cdots & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & \cdots & u_{2m} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ u_{m1} & u_{m2} & \cdots & \cdots & u_{mm} \end{bmatrix} \quad (2)$$

$$S = \begin{bmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & 0 & \sigma_{\min(m,n)} \end{bmatrix} \quad (3)$$

$$V = \begin{bmatrix} v_{11} & v_{12} & \cdots & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & \cdots & v_{2n} \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots \\ v_{n1} & v_{n2} & \cdots & \cdots & v_{nn} \end{bmatrix} \quad (4)$$

The singular values in the S matrix are arranged along the diagonal in descending order as $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min(m,n)}$. Each singular value in the diagonal is unique and equal to the square root of the eigenvalues of matrix A. The rank of the matrix A is equal to the number of nonnegative singular values.

III. Implementation of Proposed SVD Algorithm

In SVD theory, any *m* by *n* rectangular matrix can be decomposed into three matrices. An image can be defined by an *m* by *n* matrix where *m* denotes the pixel height of image and *n* denotes the pixel width of the image. When an image is stored in a computer environment, each pixel in the image is represented by a number denotes the intensity value that represents the relative darkness or brightness. In grayscale images, the range of the intensity is from 0 (black) to 1 (white). On the other hand, in colour images, each colour pixel is divided into three main components: red, green and blue. Therefore, the range of each single colour pixel is calculated as a grayscale image.

To illustrate the performance of the proposed SVD-based compression technique, seven set of sample images were used. By applying SVD method on every image, the image matrix is decomposed into three different matrices U, S and V. But simply applying SVD method alone does not compress the image. To compress the given image, some singular values should be ignored according to determined threshold value. Therefore, the selection of threshold value is important. Because, If the threshold value is set to large, the compression ratio will increase, but quality will decrease. In this paper, the proposed threshold value is determined from the following formula.

$$\text{Thr}_{\text{Value}} = \frac{\sum_{i=2}^{\text{rank of S}} |\sigma_i - \sigma_{i-1}|}{\text{rank of S}} \quad (5)$$

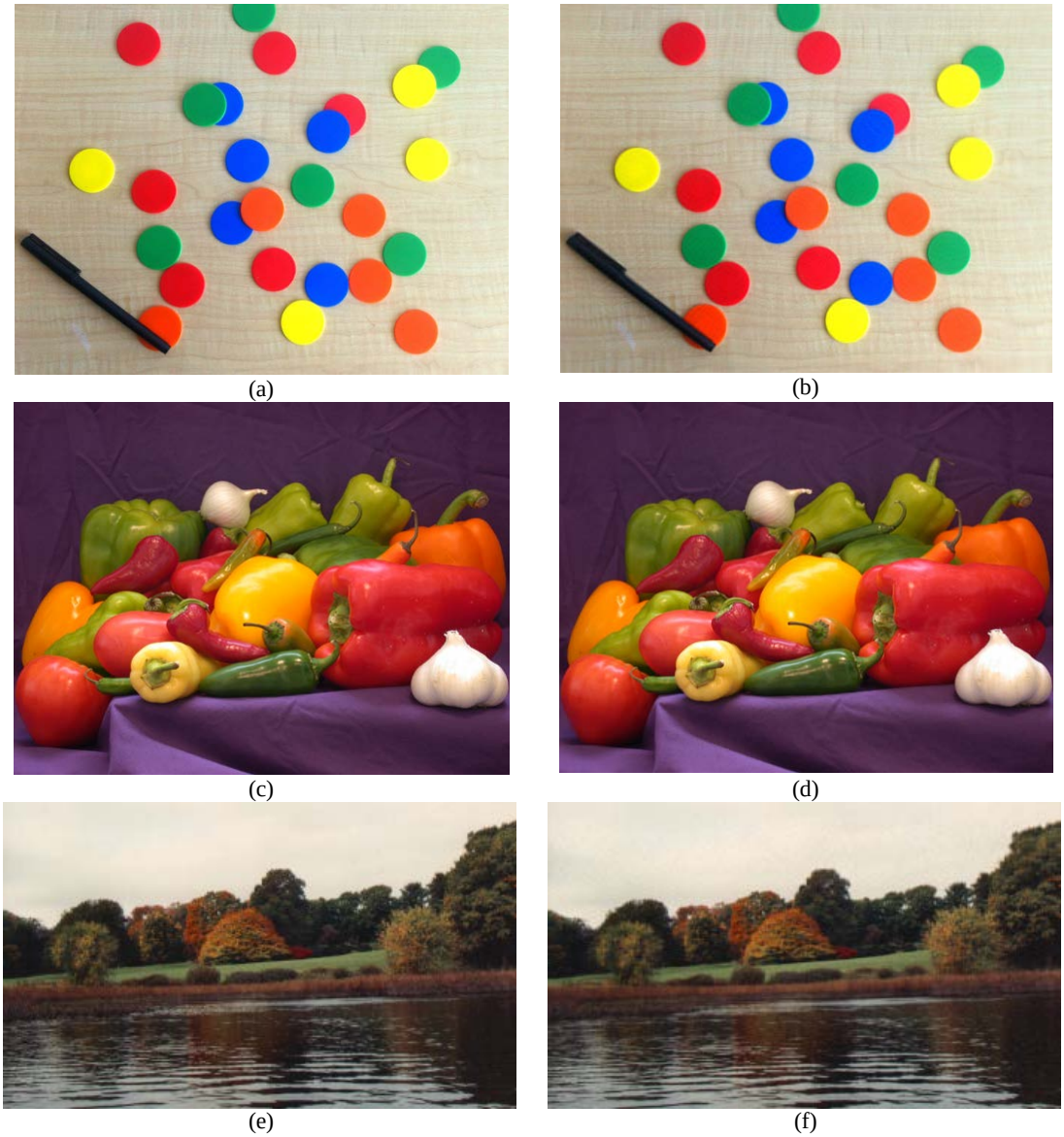
This threshold formula is applied to each R, G and B channels separately.

After applying this threshold value to the singular values in the S matrix obtained from the input image, singular values that are less than the threshold value are ignored. In this way, the reconstructed image has been compressed.

IV. Experimental Results

The experimental results are obtained through simulation by MATLAB R2015a installed on a personal computer with Intel® Core™ i7-4510U CPU processor 2.00 GHz. 8GB RAM and 64bit Windows 10 operating system. The performance of the proposed method is evaluated based on three parameters: Peak Signal to Noise Ratio (PSNR), Compression Ratio (CR) and Mean-Square Error (MSE). Their definitions are given as follows.

$$\text{PSNR} = 20\log 10\left(\frac{255}{\sqrt{MSE}}\right) \quad (6)$$



(a) (b) (c) (d) (e) (f)

Figure 1: (a), (c), and (e) original images and (b), (d), and (f) compressed images

Table 1: Experimental results

Names of images			peppers	autumn	tape	coloredChips	pears	onion	hestain
Performance Metrics	CR		0.79	0.64	0.81	0.43	0.67	0.73	0.64
	MSE	R	$7.46\text{x}10^{-5}$	$2.15\text{x}10^{-4}$	$3.34\text{x}10^{-5}$	$1.50\text{x}10^{-4}$	$1.14\text{x}10^{-4}$	$2.73\text{x}10^{-4}$	$2.16\text{x}10^{-4}$
		G	$2.44\text{x}10^{-5}$	$1.93\text{x}10^{-4}$	$2.22\text{x}10^{-5}$	$1.40\text{x}10^{-4}$	$1.02\text{x}10^{-4}$	$1.29\text{x}10^{-4}$	$1.07\text{x}10^{-4}$
		B	$2.08\text{x}10^{-5}$	$1.56\text{x}10^{-4}$	$2.32\text{x}10^{-5}$	$1.06\text{x}10^{-4}$	$4.06\text{x}10^{-5}$	$4.39\text{x}10^{-5}$	$3.50\text{x}10^{-4}$
	PSNR	R	89.40	84.79	92.89	86.38	87.55	83.77	84.79
		G	94.25	85.26	94.66	86.69	88.03	87.04	87.82
		B	94.95	86.17	94.47	87.90	92.04	91.70	82.69

$$CR=\frac{Compressed\_image\_size}{Original\_image\_size} \quad (7)$$

$$MSE=\sqrt{\frac{1}{mn}\sum_{x=1}^{m}\sum_{y=1}^{n}(I(x,y)-C(x,y))^2} \quad (8)$$

where, $I(x, y)$ and $C(x, y)$ denote the original and compressed image intensity values, respectively.

Figure 1 shows only three of the seven images used to compress. From Figure 1, the proposed SVD-based compression algorithm, such as the JPEG compression algorithm, can reduce the file size of the original image without any quality distortion. The numerical analysis for each image in terms of CR, MSE and PSNR is given in Table 1.

From the Table 1 it is seen that the proposed SVD algorithm for the autumn image (e and f in the Figure 1) has CR = 0.64. The PSNR values are 84.79, 85.26 and 86.17, for each colour channel (R, G, B) respectively. Moreover, the MSE values are $2.15x10^{-4}$, $1.93x10^{-4}$, and $1.56x10^{-4}$ for each R, G and B channel, respectively. Similarly, for the other images the CR, PSNR, and MSE values appear in the other columns of the Table 1.

V. CONCLUSION

SVD-based compression techniques are commonly used to compress grayscale and colour images. However, in SVD-based compression techniques, the compression ratio is determined by eliminating singular values less than the arbitrarily determined threshold value. As the number of singular values decreases, the value of the compression ratio increases, but the quality of the compressed image decreases. Therefore, it is necessary to select the appropriate value of threshold value to choose between compression ratio and image quality. In this study, appropriate threshold value is calculated by dividing the sum of the differences between the obtained singular values by the rank of the matrix. The proposed SVD based compression algorithm offers very less MSE and RMSE error, hence the characteristic features of the original image are preserved very well by the reconstructed image. For future applications, it may be interesting to use this technique for other one-dimensional data (biomedical signals).

REFERENCES

[1] A.M. Rufai, G. Anbarjafari, H. Demirel,'' Lossy medical image compression using Huffman coding and singular value decomposition,'' in Conf. 2013 IEEE Signal Processing and Communications Applications,pp. 1–4.

[2] L.F.R. Lucas, N.M.M. Rodrigues, L.A.D Cruz, S.M.M. de Faria "Lossless Compression of Medical Images Using 3-D Predictors,'' IEEE Trans. Medical Imaging, 36 (Issue 11), pp. 2250-2260,2017.

[3] C. Narmatha, P. Manimegalai, S. Manimurugan "A Lossless Compression Scheme for Grayscale Medical Images Using a P2-Bit Short Technique," Journal of Medical Imaging and Health Informatics, 7 (Issue 6), pp. 1196-1204, 2017.

[4] S. Haddas, G. Coatrieus, M. Cozic, S. Bouslimi "Joint Watermarking and Lossless JPEG-LS Compression for Medical Image Security," Innovation and Research in BioMedical engineering, 38 (Issue 4), pp. 198-206, 2017.

[5] Y. Deigant, V. Akshat, H. Raunak, P. Prinjal, J. Avi "A Proposed Method for Lossless Image Compression in Nano-satellite Systems" in Conf. 2017 IEEE Aerospace, MT, Big Sky.

[6] B. Rusyn, Y. Lysak, A. Lukenyuk, L. Pohreliuk "Lossless Image Compression in the Remote Sensing Applications"in Conf. IEEE First International on Data Stream Mining & Processing, Ukraine, pp. 23-27, 2016.

[7] D. V. Meegan, L. B. Stelmach, and W. J. Tam, "Unequal weighting of monocular inputs in binocular combination: Implications for the compression of stereoscopic imagery," Journal of Experimental Psychology: Applied, vol. 7(2), pp. 143-153, 2012.

[8] TC Lu, CY Chang," A survey of VQ codebook generation." J. Inf. Hiding Multimedia Signal. Process. 1(3), pp. 190–203, 2010.

[9] A.M.U. Ahamed, C. Eswaran and R. Kannan, "Lossy Image Compression based on Vector Quantization using Artificial Bee Colony and Genetic Algorithms," International Conference on Computational Science and Technology, Kota Kinabalu, Sabah, November 2016. (To be published in the journal: Advanced Science Letters)

[10] J.S. Walker, Y. Chen, T.M. Elgindi "Comparison of the JPEG2000 lossy image compression algorithm with WDR-based algorithms" in Conf. IEEE International on Image Processing, 2006.

[11] T. Ma, M. Hempel, D. Peng, H. Sharif "A survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems" Common Surveys Tutorials, IEEE, pp. 963-972,2013.

[12] F. Douak, R. Benzid and N. Benoudjit ,"Color image compression algorithm based on the DCT transform combined to an adaptive block scanning," AEU International Journal of Electronics and Communications, vol.65, Issue 1, pp. 16-26 , Jan 2011.

[13] A.M.Raid, W.M.Khedr, M. A. El-dosuky and W. Ahmed, "Jpeg Image Compression Using Discrete Cosine Transform - A Survey," International Journal of Computer Science & Engineering Survey (IJCSES) Vol.5, No.2, April 2014.

[14] Z. Eldin, H., Elhosseini, M. A., & Ali, H. A. "Image compression algorithms in wireless multimedia sensor networks" A survey. Ain Shams Engineering Journal, 6(2), 481-490, 2015.

[15] S.S. Umadi and G.D. Kulkarni "An implementation of efficient Low power VLSI Architecture for image compression system using DCT and IDC, T" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 4(8). 2015.

[16] G. Vijayvargiya, S. Silakari,and Pandey, R. (2013). A survey: various techniques of image compression. International Journal of Computer Science and Information Security, Vol. 11, No. 10, October 2013.

[17] A. P., Singh, A.Potnis and A Kumar, "A Review on Latest Techniques of Image Compression," International Research Journal of, 727-734, 2016.

[18] M. Hemalatha, and S. Nithya, "A Thorough Survey on Lossy Image Compression Techniques," International Journal of Applied Engineering Research, 11(5), 3326-3329, 2016.

[19] A. J. Hussain, A. Al-Fayadh, and N. Radi, "Image Compression Techniques: A Survey in Lossless and Lossy algorithm," Neurocomputing, 2018.

[20] P. Dhumal and S. S. Deshmukh, "Survey on Comparative Analysis of Various Image Compression Algorithms with Singular Value Decomposition," International Journal of Computer Applications (0975 – 8887) Volume 133 – No.6, January 2016.

Online Mine Detection Experiment with an Unmanned Underwater Vehicle

M.BERİK[1] and S. KARADENİZ KARTAL[1]

[1] Bulent Ecevıt University, Zonguldak/Turkey, merve-berik@yandex.com
[1]Bulent Ecevıt University, Zonguldak/Turkey, sedakaradeniz@gmail.com

***Abstract* – In this study, real-time online mine detection under water is performed using an unmanned underwater vehicle. According to the underwater position of the vehicle, the data from the vehicle camera is used to identify mine in a real-time video stream. Here is an approach for underwater mine detection based on the use of trained classifiers. Algorithm performance in real time video shooting of the vehicle is optimized to reduce false positive rate by aiming to identify a mine segment of each picture frame. According to the results obtained, it was ensured that mine parts are successfully detected under changing conditions with incorrect positive detection. Algorithms are developed using the Python program.**

***Keywords* – Mine detection, unmanned underwater vehicle, trained classifiers, haar cascade classifier.**

I. INTRODUCTION

UNMANNED underwater vehicles are used for many civilian and military application areas [1], [2]. The use of unmanned underwater vehicles, underwater image processing, underwater archaeological investigations, geological and geophysical applications, research of underwater resources, underwater cable and pipe laying and military applications are of great importance [3]. The use of unmanned underwater vehicles is a good practice in terms of safety and reduces cost Where it is not possible to use the diver for long periods in dark and deep water for the detection of mines placed under and besides to ships.

There are various classification methods for the identification of the desired object [4]-[6].In this study, the solution was reached with an approach based on the use of trained cousin Haar classifiers for mine detection. Classifiers represent a specific part of a mine image and the location of the desired view [7].In classifier training, a positive picture and undesired negative pictures that do not contain the object are added in order to be able to find any object desired to be detected under water [8].In order to minimize the risk of misclassification, an improvement is made to reduce false negative and false positive rates the least [9].

II. UNMANNED UNDERWATER VEHICLE

This is a remotely controlled underwater exploration and observation vehicle that is used in this study and is shown as figure 1, an underwater observation vehicle that underwater mine detection work is performed from the images taken with the camera on it. With an average of about 10 kg in the air and a depth of up to 200 meters, this vehicle can quickly perform tasks such as underwater observation, instant high-resolution video and photo shooting, data collection, and underwater mapping. Thanks to the thrusters located at right, left and vertical positions at specific distances from the center of gravity of the vehicle, the vehicle can make forward, depth motion and yaw and pitch angles in the water. The vehicle is controlled by the operator console and the data from the vehicle is observed. Data transmission and power transmission between the vehicle and the operator console are provided by cable. The camera, which has 128 degrees vertical, 96 degree horizontal viewing angle and 700 TVL (TV line) resolution, is placed in the waterproof zone at the front of the vehicle [10], [11].

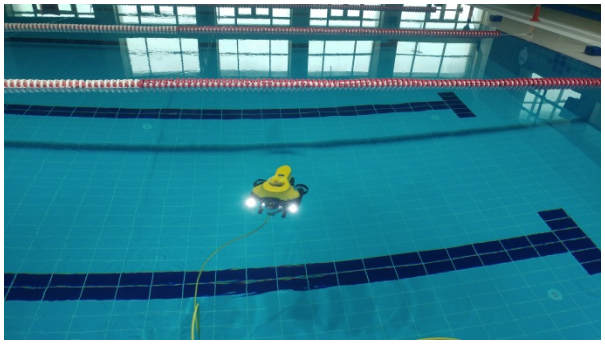


Figure 1: The unmanned underwater vehicle used in the experiment at the swimming pool of Bulent Ecevit University (BEU).

III. HAAR CASCADE CLASSIFIER

In the Haar cascade method, a specific algorithm is used to find the desired object. Various images and videos resembling this desired object are introduced to the computer by various images and videos which include the desired object in the discovery process. It tries to find that object by scanning the introduced images or video frames. In order to find any object, negative images are also called positive and undesirable objects. The images that are to be perceived to be able to recognize desired objects on the image coming from the camera and the objects that are not to be perceived are trained by classifier training [7].

IV. CLASSIFIER TRAINING WORKING PRINCIPLE

Initially, the images that are required to be recognized by the algorithm to train the classifier are called positive pictures; while the undesired pictures are called negative pictures and these pictures must be in a specific folder (Negative images should definitely not contain positive images and images that are required to be recognized).

The main frames used in classifier training are shown in the figure 2. In this classifier training, the objects in the positive pictures are scanned with the frames set in certain sizes, the pixel values in the black area inside the frame are multiplied by +1, and the pixel values in the white area are multiplied by -1.Then, specific target values are generated by controlling the sum of the pixel values in the black region and the dark luminance values in the white region sum of the pixel values.

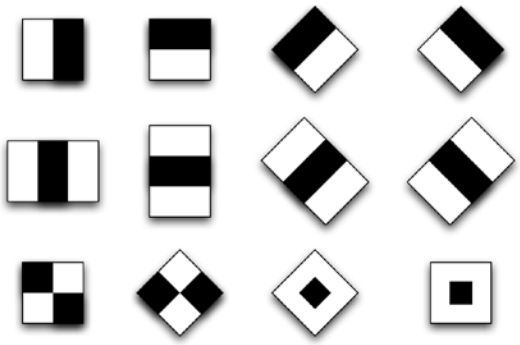


Figure 2: Commonly used frames in the Haar Cascade classifier.

It is a distinctive feature that the eye area is darker than the white area, which is shown by white figure 3. In the other picture, it is also a distinctive feature for eye recognition that the two eyes are darker than the white area with the middle. This process is repeated with each frame size of the desired frame size, every new image frame that is rotated by turning the frames at different angles. Furthermore, in the process of finding the objects, instead of being scanned repeatedly by each magnitude frame, only the parts that match the previous steps are scanned and the processing time is shortened [8]. The principle of mine detection is like the one in the face.

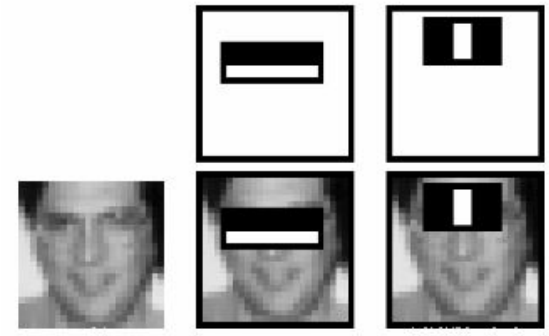


Figure 3: Scanned frame and sample positive image.

V. INTEGRAL IMAGE

In the Haar Cascade classifier, instead of calculating the individual totals of the pixel values in the image, the addition is performed by the integral method. Thus, when the real-time image is processed, the processor power is also gained. Where the integral of the view at x and y contains the sum of the pixels above and to the left of the (x, y) coordinates [9].

$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x',y') \tag{1}$$

Where, $ii(x,y)$ express integral image and $i(x',y')$ express original image.

$$s(x,y) = ii(x-1,y) + s(x,y) \tag{2}$$

$$ii(x,y = ii(x-1,y) + s(x,y) \tag{3}$$

Where, $s(x,y)$ is the cumulative sum of rows, $s(x,-1) = 0$ and $ii(-1,y) = 0$. Thus, the integral image can be calculated in a single pass over the original image. Figure 4 is the image obtained as the integral picture. The values in the boxes are pixel values. The same operation is repeated by turning the defined positive images at 45 degree angles.

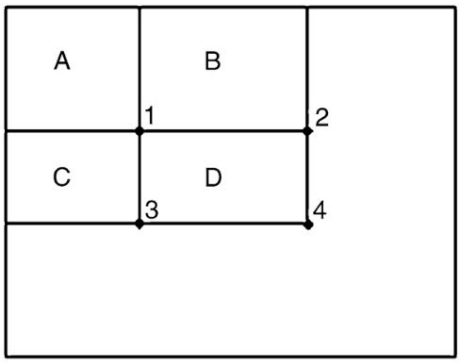


Figure 4: D is the sum of the pixels in the rectangle with four series reference calculation.

Four points are referenced to calculate the sum of the pixels in the rectangle D. First, for the 1st position, the sum of the pixels in the rectangle A is taken to obtain the integral image value. This sequence is obtained as A + B for Position 2, A + C for Position 3, and A + B + C + D for Position 4. The sum of the pixels in the rectangle D is obtained by lower right + upper left + upper right + lower left (4 + 1 - (2 + 3)).

VI. CLASSIFIER TRAINING IMPROVEMENT

The cascading training process involves two situations. In the first case, higher detection rates and lower false alarm rates are obtained. However, this requires more time to calculate the classifier. An optimization framework is defined for this. For this, an improvement is made to reduce the number of classifier stages, the number of features at each stage, the number of thresholds at each stage, and the expected number

of evaluated features [7].

When the frames received from the incoming video are scanned, positive images added to the computer are compared with the minimum values determined to detect the desired image. It is necessary to exceed this minimum value for the detection and identification of the desired object. This value is named "minHitRate" and the default value for this work is 0.995. When classifier training is taken into account,

$$F = \prod_{i=1}^{K} f_i \tag{4}$$

where, the value of the F step classifier is "minHitRate", the number of K classifiers, and f_i is the minHitRate value on the samples applied to the classifier i.

The " Maximum false alarm rate " is called the maximum false alarm value desired at each step of the classifier, indicates the detection rate, and the default value for this operation is 0.5. When a comparison is made in a trainer, the value falls below this value and the next tutorial is passed. This rate,

$$D = \prod_{i=1}^{K} d_i \tag{5}$$

Where D is the rate of detection of the digitizer classifier, K is the number of classifiers, and di is the detection rate of the classifier on the samples relative to it. The expected number of properties evaluated,

$$N = n_0 + \sum_{i=1}^{K} n_i \prod_{j<i} P_i \tag{6}$$

Where N is the expected number of properties evaluated, K is the number of classifiers, P_i is i. positive rate of classifier, n_i, i. the number of classifier's properties.

VII. MINE DETECTION

Unmanned underwater vehicle mine detection experiment is conducted in the swimming pool of Bulent Ecevit University. The camera is integrated, 128 degrees vertical, 96 degree horizontal viewing angle, and 700 TVL (TV line) resolution, the camera is located in the waterproof zone at the front of the vehicle. While the vehicle is moving in the water, video images taken by the camera are instantly detected by the mine detection algorithm developed by Phyton program.

Before the experiment, for the classifier trainer of the algorithm we used in mine diagnosis, positive images with many mines in the pool and negative images without the mines were taken with the vehicle camera and introduced to the algorithm. Each picture was taken at different angles and different minefields. The vehicle's camera was integrated into the algorithm and matched.

Videos from the camera in the car are transferred to the Operator console via cable. On the computer connected to the operator console, the mine detection algorithm instantaneously detects the images taken from the vehicle camera.

The general flow diagram of the algorithm we used for mine diagnostics is given as:

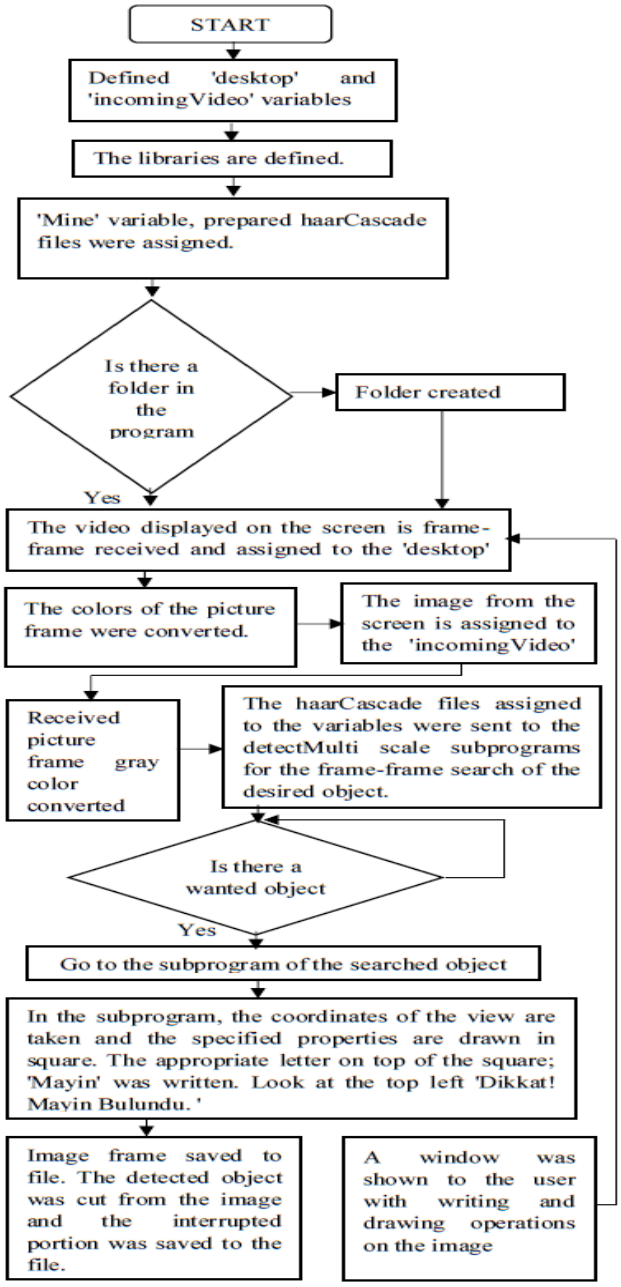


Figure 5: Flow diagram of online mine detection.

- The libraries are defined. The 'PIL' (Python Imaging Library) library supports many file formats for image processing, used for powerful image processing. The 'cv2' library was used to change color fields. The 'numpy' library is the basic package of scientific computation, necessary for the use of a strong N-dimensional matrix. The 'folder' library is used to create folders and save the resulting images to these folders.

- For the Mayin variable, prepared haarCascade files were assigned.
- Folders are created in the program directory so that when the frames of the incoming video are scanned, the desired image is recorded. This is one of the algorithm development methods to reduce the error rate when the image is received.

- On the screen the variable 'desktop' was assigned the frame-frame of images taken in the desired dimensions.
- The colors of the picture frame were converted. In the program used, the image is taken in the BGR color system. The BGR color system has been transformed into the RGB color system of which the reference values of all color codes in the nature are specified and which is the current imaging standard.
- The image frame received from the 'desktop' variable in the 'incomingVideo' variable was assigned.
- The picture frame of the reception was turned gray. In RGB (red, green, blue) color system, the result is a neutral gray tone when the color codes of the three components equal the reference values. It is easy to detect the image by converting the received image to gray color.
- The haarCascade files assigned to the variables were sent to the detectMulti scale subprograms for frame-frame search of the desired object. The desired positive and undesired negative images were introduced to the computer in advance and classifier trainers were obtained by comparing these images. Thus, the haarCascade files are created and the incoming video frames are scanned to find the desired object.
- If there is a searched object, the subprogram belonging to that object is going. Coordinates are taken and the specified properties are drawn in a square. The desired text was written on the left side of the resulting karenin and image frame. On the left side of your desktop display 'Dikkat! Mayin Bulundu'
- Image frame saved to file. The detected object was cut from the image and the interrupted portion was saved to the file. A window was shown to the user with writing and drawing operations on the image.

Before to the pool experiment, the existing mine paintings were processed into the algorithm and positive pictures were created. Algorithm's performance is tested using current mine pictures.

As shown in the figure 6 and figure 7, the mine is detected successfully under the gates. After the algorithm is tested with ready mine images, an experimental setup for mine detection is created in the pool.

Figure 6: Performance of online mine detection algorithm.

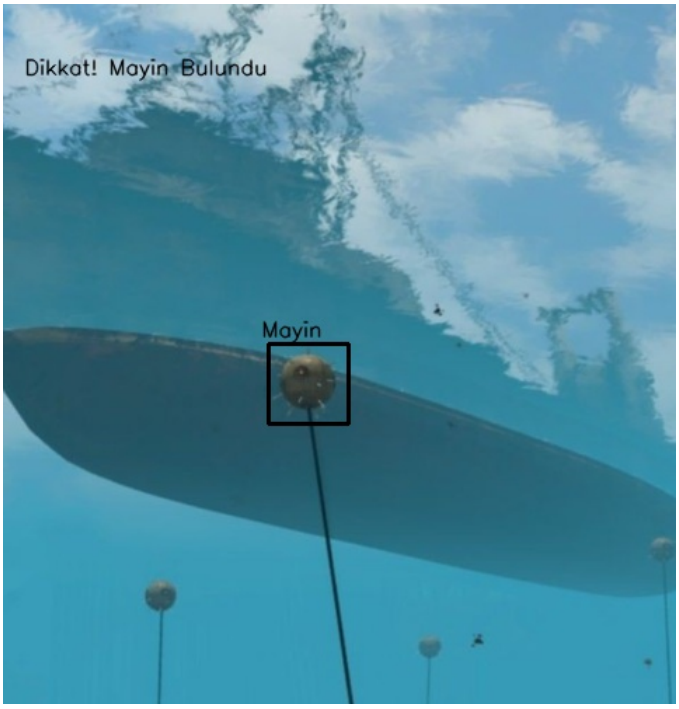


Figure 7: Performance of online mine detection algorithm.

A. Experiment Results

As a positive picture in the pool, 4 round objects which we identified as mines are placed. Around them, objects with different geometric shapes are placed inside the pool as a negative picture. Imagery placed in the pool as a mine, unmanned underwater vehicle approaching from different angles taken images are processed instantaneously. Algorithm cannot perceive objects with different geometric shapes outside mine as mine indicates the success of the algorithm. The breakage of the light was observed to affect the performance of the algorithm. As seen figure 8, 9 and 10, the online algorithm identifies mines.

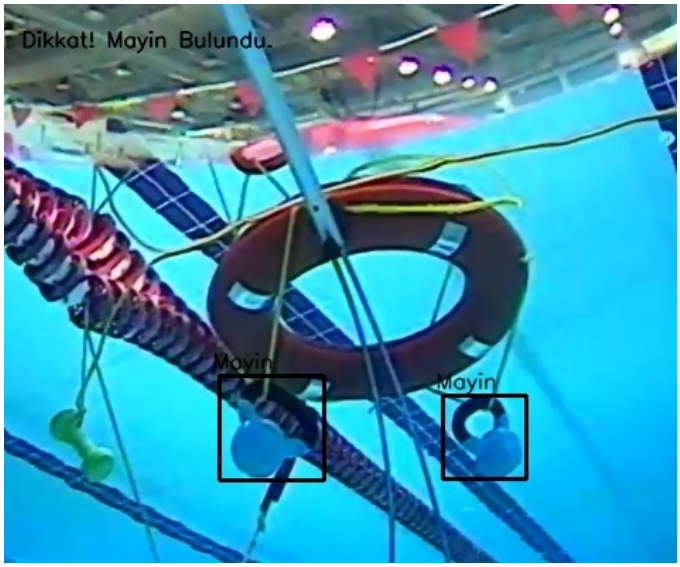


Figure 8: Experimental performance of online mine detection algorithm based on pool experiment images (combined positive and negative images) at BEU swimming pool.

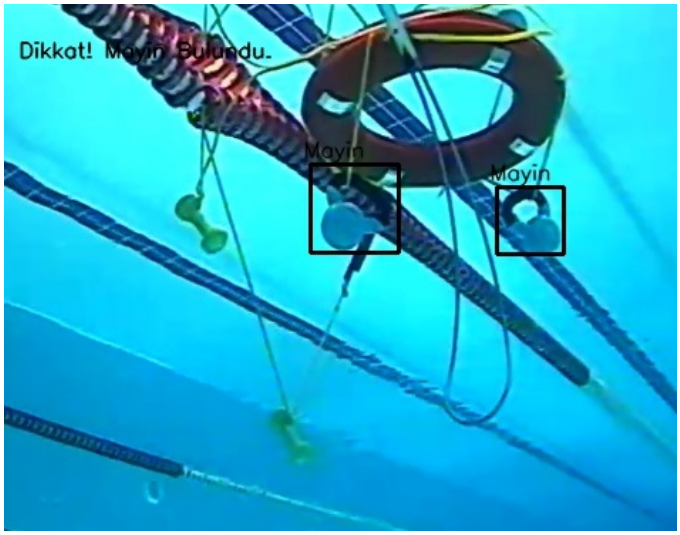


Figure 9: Experimental performance of online mine detection algorithm based on pool experiment images (combined positive and negative images) at BEU swimming pool.

VIII. CONCLUSION

Especially, particularly in the military arena, in cold and deep waters, where human life is at risk, mine detection using unmanned underwater vehicles is a reliable and cost-effective method. In this study, mine is diagnosed with an unmanned underwater vehicle. As the vehicle moves in the water, the images taken from the integrated camera are processed instantaneously and the mine is diagnosed. In the diagnosis Haar cascade classifier method is used. For this, the positive images that the object wanted to be known to the algorithm and the negative images that the object does not exist are introduced. It indicates the success of the algorithm. The incoming image is scanned with certain frames and the object is identified by comparison. Algorithms were developed using the Pyhton program. In future studies, more negative images will be loaded into the pool and the algorithm will be improved. Mine detection will also be performed by integrating sonar and camera information for mine detection at sea.

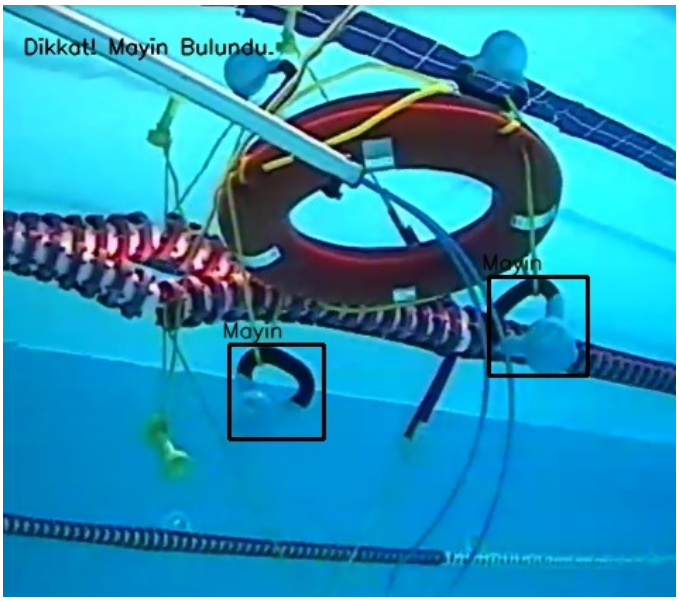


Figure 10: Experimental performance of online mine detection algorithm based on pool experiment images (combined positive and negative images) at BEU swimming pool.

ACKNOWLEDGMENT

This study is supported with 2017-47912266-01 (BEÜ-BAP). The authors are grateful for the support of BEU-BAP.

REFERENCES

[1] A. Spears, M. West, M. Meister, J. Buffo, C. Walker," Under Ice in Antarctica: The Icefin Unmanned Underwater Vehicle Development and Deployment", *IEEE Robotics Automation Magazine*, 23(4), 30-41, 2016.
[2] A. Alvarez," Redesigning the SLOCUM Glider for Torpedo Tube Launching", *IEEE Journal of Oceanic Engineering*, 35(4), 984-991, 2010.
[3] G. C. Bishop, "Gravitational field maps and navigational errors unmanned underwater vehicles", *IEEE Journal of Oceanic Engineering*, 27(3), 726-737, 2002.
[4] S. Yin, P. Ouyang, X. Dai, L. Liu; S. Wei, "An AdaBoost- Based Face Detection System Using Parallel Configurable Architecture With Optimized Computation", *IEEE Systems Journal*, 11(1), 260-271, 2017.
[5] S. Bianco; F. Gasparini; R. Schettini, "Adaptive Skin Classification Using Face and Body Detection", *IEEE Transactions on Image Processing*, 24(12), 4756-4765, 2015.
[6] R. Raghavendra, K. B. Raja, C. Busch, "Presentation Attack Detection for Face Recognition Using Light Field Camera", *IEEE Transactions on Image Processing*, 24(3), 1060-1075, 2015.
[7] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION*, 2001
[8] Jaya M. Jadhav, Deipali V. Gore, Rashmi R. Tundalwar, " Introducing Celebrities in an Images using HAAR Cascade algorithm", *International Journal of Computer Applications* (IJCA), 2013.
[9] P. Viola, M. J. Jones, "Robust Real-Time Face Detection", *International Journal of Computer Vision* 57(2), 137–154, 2004.
[10] S. Karadeniz Kartal, M. Kemal Leblebicio˘glu, E. Ege "Experimental Test of Acoustic-based Navigation and System Identification of an Unmanned Underwater Survey Vehicle (SAGA)", *Transactions of the Institute of Measurement and Control*, 2018
[11] S. Karadeniz Kartal, E. Ege, M. Kemal Leblebicioğlu , "Optimal Autopilot and Guidance of the ROV: SAGA", *ScienceDirect*, 49(3), 401-406, 2016.

The Autonomous Yaw and Depth Controller Experiment of Unmanned Underwater Vehicle

E. ERTUĞRUL[1], A. DEMİR[1], Ö. ŞENYÜREK[1] and S. KARADENİZ KARTAL[1]

Bulent Ecevit University, Zonguldak/Turkey, emreacbek@gmail.com
Bulent Ecevit University, Zonguldak/Turkey, akerimdemir35@gmail.com
Bulent Ecevit University, Zonguldak/Turkey, omer.3579@outlook.com
Bulent Ecevit University, Zonguldak/Turkey, sedakaradeniz@gmail.com

***Abstract* – In this study, the real time yaw and depth controller of unmanned underwater vehicle (UUV) are performed experimentally. Firstly, the mathematical model of UUV is obtained. The controllers are designed based on mathematical model data. Then, the controllers are performed based on experimental data. Data transmission from vehicle is provided with fiber optic cable which is connected operator console in the experiment. This operator console is connected to computer which has MATLAB inteface. Vehicle is controlled with proportional (P)-integral (I)-derivative (D) designed in MATLAB/Simulink environment. Communication between the vehicle and controller is provided using MATLAB interface during the experiment. The experimental controller responses are compared with the controller response of mathematical model.**

***Keywords* - Unmanned underwater vehicle, controller, PID, mathematical model.**

I. INTRODUCTION

Unmanned underwater vehicles are used both in civilian and military areas frequently. They are most important tools to observe underwater. The unmanned underwater vehicle used in this study is a remotely operated underwater survey vehicle specifically developed for the purpose of investigation of underwater and equipped with a camera and a two dimensional sonar (see figure1), in addition to several other custom specific sensors. It is very easy to obtain navigational data and high resolution video, as regards to underwater operation, from this vehicle.

In general, mathematical models of UUVs are nonlinear; but linear models can be preferred sometimes [1]. This study focused on a nonlinear mathematical model. There are many control techniques to design autopilots for unmanned underwater vehicles (UUV). Speed autopilots are designed by P or PI control techniques [2]. Position autopilots are also designed by the same controllers [3]. More complicated autopilots are designed with PID and sliding mode control techniques [4]. Optimal kinematic control for an autonomous underwater vehicle and a particular set of optimal motions

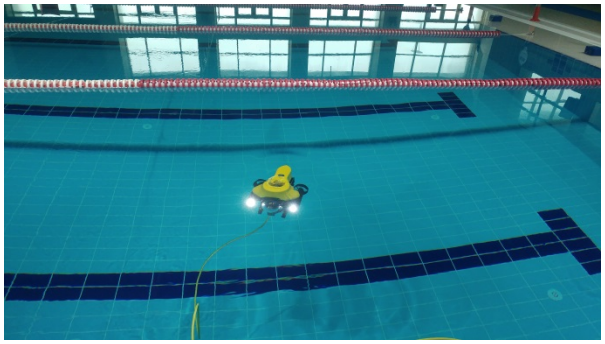


Figure 1: The unmanned underwater vehicle used in the experiment at the swimming pool of Bulent Ecevit University (BEU).

which trace helical paths are discussed in [5]. A new optimal control method based on the energy equations of the controlled system is presented in [6]. This method is applied on an under-actuated underwater vehicle. Since in many applications the controllers are integrated to guidance system, controllers play an important role to autopilot design [7]. In this study, the yaw angle and depth of the vehicle are controlled by using PID controllers.

II. MATHEMATICAL MODEL

The mathematical model of an unmanned underwater vehicle is obtained as shown in the equations below [1].

$$M(\dot{v}) + C(v)v + D(v)v + g(\eta) = u \quad (1)$$

$$\dot{\eta} = J(\eta)v \quad (2)$$

M: Total mass of the vehicle,
C: Total Centrifugal and Coriolis forces
M_{RB}: The rigid body mass of the vehicle,
M_A: Added inertia matrix,
C_{RB}: Rigid body centripetal and Coriolis matrix,
C_A: Hydrodynamic Centripetal and Coriolis matrix,
D: Damping matrix,
g: Restoring forces and moments,
τ: Input vector,
v: The linear and angular velocity vector of the vehicle,

η: The position and attitude vector of the vehicle,
J: Transformation matrix.

The motion of the vehicle in 6 DOF can be described by the following vectors:

$$\vec{\eta} = [\,\eta_1^T, \eta_2^T\,]^T \qquad \eta_1^T = [x, y, z]^T \qquad \eta_2^T = [\emptyset, \theta, \psi]^T \quad (3)$$

$$\vec{v} = [\,v_1^T, v_2^T\,]^T \qquad \overrightarrow{v_1} = [u, v, w]^T \qquad \overrightarrow{v_2} = [p, q, r]^T \quad (4)$$

$$\vec{\tau} = u = [\tau_1^T, \tau_2^T\,]^T \qquad \overrightarrow{\tau_1} = [X, Y, Z]^T \qquad \overrightarrow{\tau_2} = [K, M, N]^T \quad (5)$$

III. FORCES EXERTED BY THE THRUSTERS

The hardware elements of the motor controller-motor-propeller are collected in one place and it is called the thruster. Each thruster is driven by a DC motor. In the experiment, the input value for each motor of UUV is PWM value. Motor is stopped when PWM value reaches the value of 1500 rpm. If the applied PWM value is the higher than 1500 rpm, motor rotates positive direction. If the applied PWM value is the lower than the 1500 rpm, motor rotates the opposite direction. Since the system input for mathematical model of UUV is force value (Nm) of each thruster, applied PWM values are transformed to the force values. The relationship between the force (Kgf) value and the PWM value of the used motors is shown in figure 2 [8].

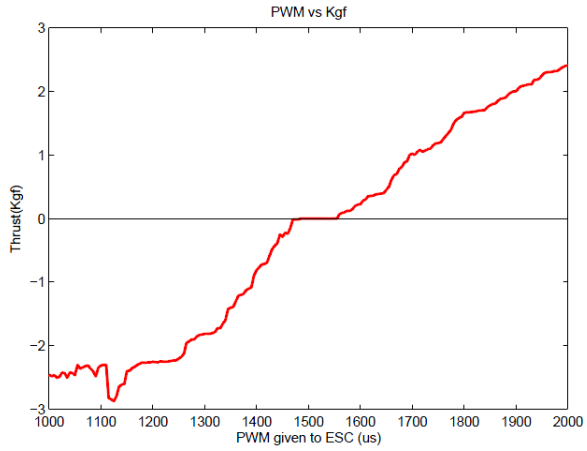


Figure 2: The relationship between PWM to force value of dc motor drives to thrusters.

In the nonlinear mathematical modeling of the vehicle, u is a column vector that consists of the moments and forces produced by the thrusters. The vehicle has three thrusters, two of which are located horizontally on the right and left sides. The last thruster is located vertically. Motion in the x-axis (surge motion) and the rotation around the z-axis (yaw angle) are accomplished by the horizontal thrusters. Motion in the z-axis (heave motion) and rotation around the y-axis (pitch angle) are managed using the vertical thruster. The resultant thrust force and moments applied on the vehicle are expressed as follow.

$$\begin{vmatrix} \sum X \\ \sum Y \\ \sum Z \\ \sum K \\ \sum M \\ \sum N \end{vmatrix} = \begin{vmatrix} T_1 + T_2 \\ 0 \\ T_3 \\ 0 \\ T_3 r_3 \\ T_1 r_1 + T_2 r_2 \end{vmatrix} \quad (5)$$

where,
T_1: Right thruster force,
T_2: Left thruster force,
T_3: Vertical thruster force,
R_1: Distance between the right thruster and the center of gravity, 119.21 mm,
R_2: Distance between the left thruster and the center of gravity, -119.21 mm
R_3: Distance between the vertical thruster and the center of gravity, 217 mm.

IV. CONTROL MODEL FOR YAW MOTION

In this study, the yaw angle and depth are controller using the by the proportional (P)-integral (I)-derivative (D) controller known as PID controller. The error between the desired state value and the output of the system is feedback to the PID controller. This error is minimized by adjusting the control input, u(t) using optimization. Control inputs are forces produced from right, left and vertical thrusters T1, T2 and T3 for UUV.

Figure 3 shows the block diagram of the yaw rate controller. As seen in figure 3, error is defined as the difference between the reference and actual yaw rates. Optimal PID gains for the rotational speed of yaw controller are obtained by minimizing the error between desired and actual yaw rate values. PID controller related to optimal gains generates produces the reference right and left thrusts for system(UUV) input.

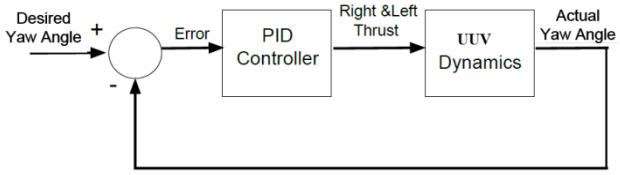


Figure 3: The block diagram of yaw controller for UUV.

Figure 4 shows the yaw rate controller response based on mathematical model. The reference yaw angle (blue line) set to -110 degree and the actual yaw angle (red line) is taken from mathematical model of UUV. As seen in figure 4, the controller reaches the desired value in 6 seconds. Negative yaw angle means that the vehicle rotates direction westward.

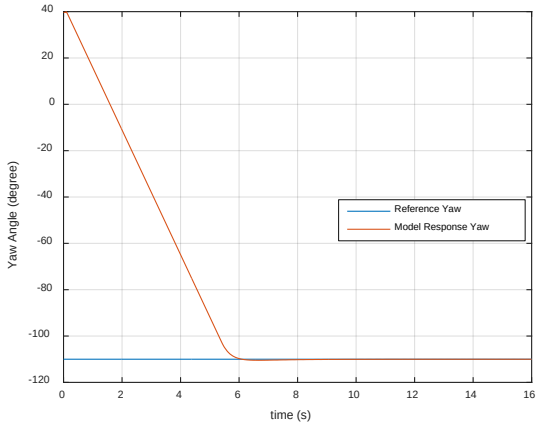


Figure 4: The yaw controller response based on mathematical model of UUV.

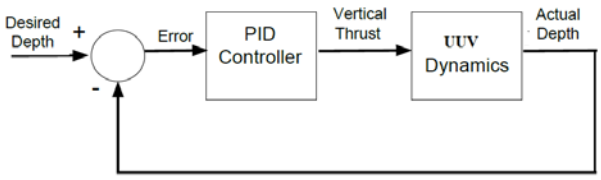


Figure 5: The block diagram of depth controller for UUV.

Depth controller response is shown in figure 6. As seen in the figure, desired depth is set to 0.6 m and the controller reaches the desired value in 10 seconds.

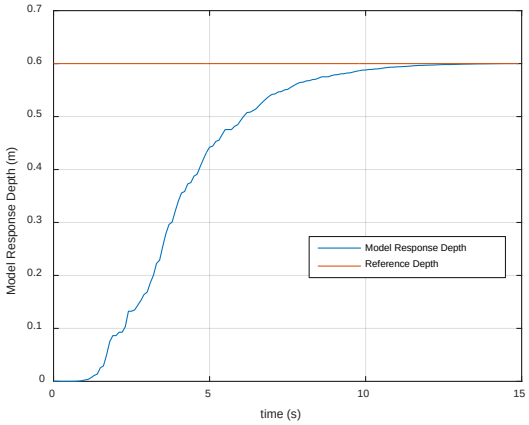


Figure 6: The depth controller response based on mathematical model of UUV.

V. EXPERIMENT

The real time controller of our unmanned underwater vehicle was performance in the swimming pool of Bulent Ecevit University. The Inertial measurement units (IMU), magnetic compass and depth sensor are located vehicle. Inertial measurement unit (IMU) used in the experiment is UM7. UM7 is a 3rd-generation attitude and heading reference system (AHRS). The UM7 combines triaxial accelerometer, rate gyro, and magnetometer data using a sophisticated Extended Kalman Filter to produce attitude (roll, pitch, yaw) and heading estimates and measures acceleration in the x,y and z-axes. The communication with the vehicle is provided by fiber optic cable is shown as figure 7.

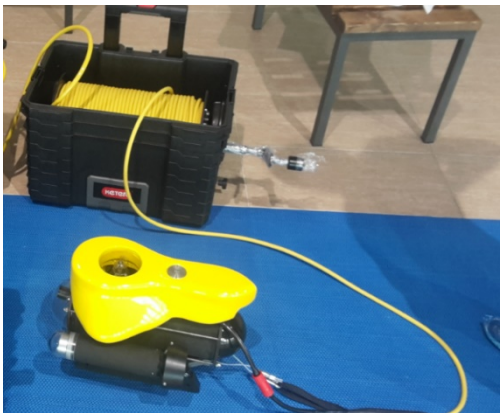


Figure 7: The connection of fiber optic cable to unmanned underwater vehicle.

In this experiment, yaw angle is measured from UM7. The depth information coming from the depth sensor is used to determine the z position of the vehicle. The data transmission controller and vehicle is provided by operator console connected computer as shown in figure 8.

The biases of all the sensors are determined by calibration before the pool experiment. The north direction of the vehicle is fixed with respect to north in the Earth frame to calibrate the UM7 (including magnetometer). The depth sensor is calibrated when the vehicle is on the surface of the water.

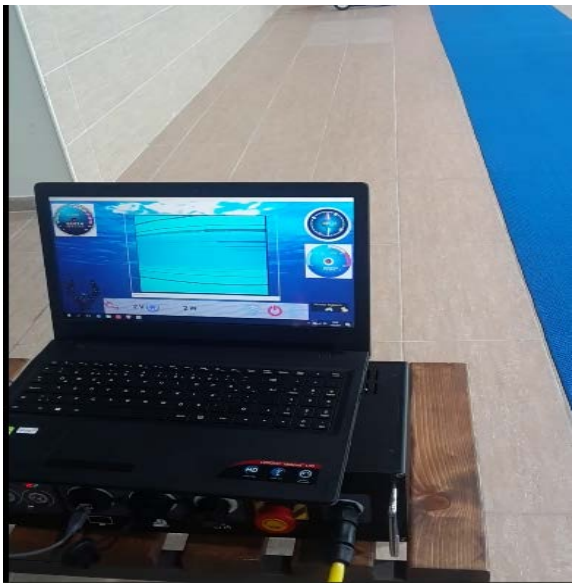


Figure 8: The connection of operator console and computer for MATLAB interface test.

A. Yaw Controller Experiment

While the vehicle moves underwater, the yaw angle is taken from the UM7 at each 0.2 seconds with MATLAB interface. Then, this measured yaw angle is input to yaw controller. The reference input signal of yaw controller is set to negative 110 degree. The error between the reference state value and the output of the system is feedback to the PID controller. This error is minimized by adjusting the gain of controller using optimization. Controller produces forces right and left thrusters T1 and T2 for UUV. The values of the optimal gains of yaw controller are obtained using optimization with PID Tune block in Simulink as $K_p = -67.1031$, $K_i = -2.75e-004$, $K_d = 2.5007$.

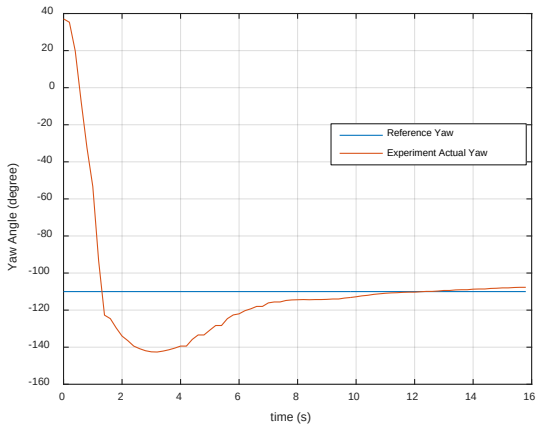


Figure 9: The yaw controller experiment response (red line) and reference yaw angle (blue line) of UUV.

The reference yaw angle (blue line) and the actual yaw angle (red line) which is obtained from controller response experimentally are shown in figure 9. As seen figure 9, controller reaches the desired (reference) value in 10 seconds. The right and left thrusters (PWM) are produced from yaw angle controller are shown in figure 10 and 11.

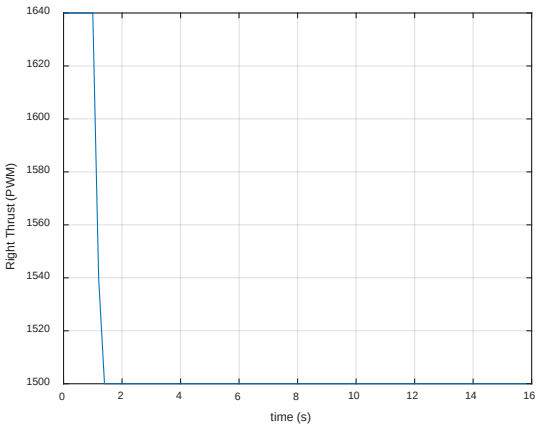


Figure 10: The right thrust produced from the yaw controller of UUV experimentally.

B. Depth Controller Experiment

While the vehicle moves underwater, the depth of the vehicle is measured from the depth sensor at each 0.2 seconds with MATLAB interface. Then, this measured depth data is input to depth controller. The reference input signal of depth controller is set to 0.6 meter. The error between the reference state value and the output of the system is feedback to the PID controller. This error is minimized by adjusting the gain of controller using optimization. Controller produces force vertical thruster T3 for UUV. The values of the optimal gains of depth controller are obtained using optimization with PID Tune block in Simulink as $K_p = 1.1614$, $K_i = -1.4275e-006$, $K_d = 0.7724$.

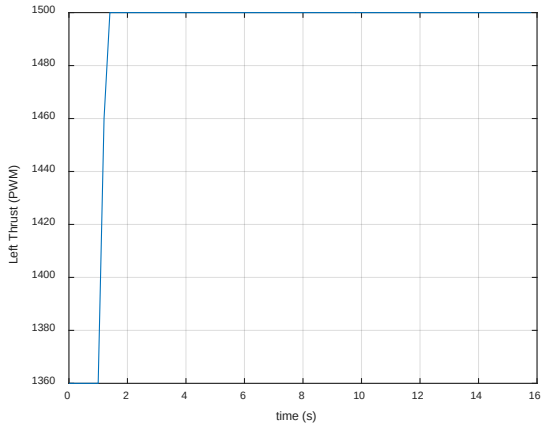


Figure 11: The left thrust produced from the yaw controller of UUV experimentally.

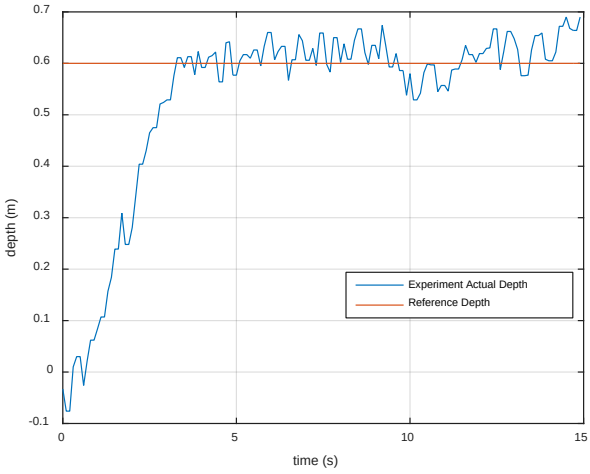


Figure 12: The depth controller experiment response (blue line) and reference depth (red line) of UUV.

The reference depth (red line) and the actual depth (blue line) obtained from controller response experimentally are shown in figure 12. As seen figure depth controller reaches the desired (reference) value in 4 seconds. The vertical thrust is produced from depth controller is shown in figure 13.

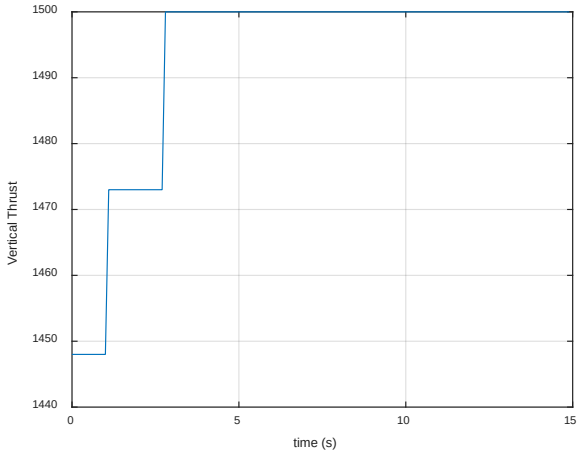


Figure 13: The vertical thrust produced from the depth controller of UUV experimentally.

VI. CONCLUSION

The real time yaw angle and depth PID controller of unmanned underwater vehicle are performance in the pool experiment. It is seen that the vehicle reaches the desired yaw angle and depth at the small time with obtained optimal gain of controller. Controllers are designed in MATLAB/Simulink environment. The actual yaw angle is reference yaw controller input is measured IMU and magnetic compass are located vehicle. The actual depth of the vehicle is input of the depth controller is taken from the depth sensor located to vehicle. This actual yaw angle and depth of the vehicle can be taken from using MATLAB interface of the vehicle. During the experiment, the weight of the fiber optic cable which allows the communication between the vehicle to operator console negatively affect the vehicle motion.

ACKNOWLEDGMENT

This study is supported with 2017-47912266-01 (BEÜ-BAP). The authors are grateful for the support of BEU-BAP.

REFERENCES

[1] T. Fossen, *Guidance and Control of Ocean Vehicles*. 1994.
[2] M. Caccia and G. Veruggio, "Guidance and control of a reconfigurable unmanned underwater vehicle", Control Engineering Practice, 8(1): 21-37, 2000.
[3] S. Zanoli and G. Conte, "Remotely operated vehicle depth control", Control Engineering Practice, 2003, 11(4): 453-459.
[4] S. K. Lee, K. H. Sohn, S.-W. Byun, and J.-Y. Kim. " Modeling and controller design of manta-type unmanned underwater test vehicle", In Journal of mechanical science and technology, 23(4): 987-990, 2009.
[5] J. Biggs and W. Holderbaum, "Optimal kinematic control of an autonomous underwater vehicle", IEEE Transactions on Automatic Control, 54(7):1623 1626, July 2009.
[6] N. Fukushima, M. S. Arslan, and I. Hagiwara, "An optimal control method based on the energy flow equation", IEEE Transactions on Control Systems Technology, 17(4):866–875, July 2009
[7] S. K. Kartal, E. Ege, and K. Leblebicioglu, "Optimal autopilot and guidance of the rov: Saga," Elsevier, ScienceDirect, 49(3):401406, 2016.
[8] D. Company. Available: http://www.desistek.com.tr/.

Information Security Risk Assessment using Bayesian Network and Fuzzy Inference System: A Case Study

S.BEKEN[1] and M.EMINAGAOGLU[2]

[1]Dokuz Eylul University, Izmir/Turkey,sevilaybeken22@gmail.com
[2]Dokuz Eylul University, Izmir/Turkey, mete.eminagaoglu@deu.edu.tr

***Abstract* - This study proposes an information security risk assessment approach based on Bayesian networks and Fuzzy Inference System in order to evaluate and calculate both qualitative and/or quantitative risks. The proposed preliminary model is developed to analyse test processes for a software services company in order to evaluate the information security risks. In order to collect data for our risk assessment model, assets, vulnerabilities, threats, and related risk values are identified and analyzed with experts and managers based on the testing process in the company. Threats, vulnerabilities, risks, and their relations are constructed with a Bayesian network and marginal probabilities for each risk are calculated. Several fuzzy membership functions are designed for assets' values, risks' probabilities, and relative risk values. Fuzzy decision rules are constructed for some of the chosen risks by using the assets' values, relevant risk probabilities, and relative risk values. In the final stage, the impacts of risk (loss) values are calculated by aggregation and defuzzification. Promising results have been obtained so far and this preliminary model will be used as a basis for an enhanced model, which can be successfully used for information security risk assessment and management with less subjectivity, more reliability, and more flexibility.**

***Keywords* - Information Security, Risk Assessment, Bayesian Network, Fuzzy Inference System**

I. INTRODUCTION

Information Security can be defined as the protection of data or information to prevent loss, unauthorized access or misuse. In the risk assessment process, the three key elements are asset, threat, and vulnerability. Assets, vulnerabilities, threats, and risk factors should be identified, analysed, and controlled within the scope of security risk assessment process [1].

The assets are the main objects for organizations that need to be protected based on information security policies. Assets can be valuable information or resources such as computers, employees, internet connection and so on. The threat is potential causes of accidents that may cause harm to systems or organizations. Vulnerability is the weak link of an asset that may be exposed by the threat [1, 2].

In this study, an information security risk assessment approach is proposed that is based on Bayesian networks and Fuzzy Inference System to evaluate and calculate both qualitative and quantitative risks. Information risk factors can be modelled in Bayesian network and risk probabilities can be calculated more accurately for business requirements with Bayesian network. Hence, Bayesian network model is selected for this study. Additionally, to obtain more reliable and less subjective approach to the risk assessment process and to combine quantitative and/or qualitative risk factors, fuzzy inference system is used.

The proposed model applied for a software services company as a case study according to their software test process. In the case study, firstly the Bayesian Model was developed to analyze their database security during the testing process. The assets, vulnerabilities, threats, risk factors and their relations were analyzed with the experts and managers in that company. Experts' opinions were collected in order to sustain needful data for assets, threats, and risk probability values in the Bayesian model. After completing the Bayesian model, fuzzy inference system was developed for risk management process. At last but not the least, risk factors were evaluated within the information security risk assessment scope based on the results for information risk factors.

II. INFORMATION SECURITY RISK ASSESSMENT

Lee, M. C. (2014) proposed that information security risk assessment process is the important prerequisite to achieve scientific and effective risk assessment. Information security risk assessment process includes following stages; preparation of risk assessment, asset identification, threat identification, vulnerability identification, and risk calculation [3]. Risk assessment process can be divided into six steps as follows [4].

1) Determining assessment objects: This step includes defining the scope of the assessment. All valuable information or resources should be considered as assets for the system.

2) Assessment performance: Evaluation plan in accordance with the requirements and assessment process should be determined. In addition, appropriate assessment methods and tools should be selected.

3) Risk identification: Especially considering the critical assets, their vulnerabilities and related risk factors should be identified.

4) Risk analysis: Combined the property of assets, the possibility, and consequences of threat used by vulnerability should be

analyzed, and the results of assessment process should be calculated.

5) Risk assessment: Results should be evaluated; risk assessment report combined with the expert's opinion should be prepared.

6) Risk control: After risk assessment process, in order to control the system risk and reduce or avoid risk, risk control plans should be prepared and applied.

In order to evaluate information security risks, quantitative or qualitative methodologies can be selected and used. Qualitative security risk assessment uses qualitative estimates while quantitative assessment uses numerical estimates. Quantitative risk analysis methods use mathematical and statistical tools to represent risk. On the other hand, risk is analyzed by the aid of subjective qualitative scalar values and methods in qualitative risk analysis [5].

Some methods of quantitative security risk analysis are such as risk value, annual loss expectancy, safeguard value, and return of investment [6]. The most common and most frequently used quantitative method of risk assessment is ALE (Annual Loss Expectancy) model, based on the idea of expected loss, which is the product of probability of occurrence of events which have negative impact on IT and values of caused by them losses [6].

Fuzzy Comprehensive Evaluation Method is one of the common qualitative methodologies, which is based on the principle of fuzzy mathematics. Fuzzy comprehensive evaluation method uses the fuzzy statistical methods through considering a combination of relative factors for evaluation to determine the weight of various factors to make the evaluation of the pros and cons of the research objects [7].

Moreover, there are some other methods in the literature that use combined qualitative and quantitative methodologies. For instance, Analysis of Hierarchy Process method is a popular approach for in security risk assessment that use qualitative methodology combined with quantitative approach. The Analysis of Hierarchy Process (AHP) can change from the qualitative index into quantitative index [3]. There are four steps to resolve problem by AHP. The first one is constructing the hierarchy structure. Secondly, the judgement matrix is constructed by use of factor comparison. Calculating the relative weight of the factors by the judge matrix is the third step and finally, the last step is calculating the whole weight of the factors at each layer [8]. Some researchers proposed models that use Bayesian networks or dynamic Bayesian networks for information security risk assessment [9, 10]. However, our model is different from these because there are no fuzzy operators or fuzzy inference methodologies in these studies.

III. BAYESIAN NETWORK

A. Bayes' Theorem

Bayes' theorem is a formula that describes how to update the probabilities of hypotheses when given evidence. It follows simply from the axioms of conditional probability, but can be used to powerfully reason about a wide range of problems [9-11]. The equation of Bayes' theorem is given in (1). In this equation, A and B denote the events; *P (A | B)* denotes the conditional probability that calculates the likelihood of A given that B is true. Similarly, *P (B | A)* represents the probability of B when A is known to have occurred.

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)} \quad (1)$$

B. Bayesian Network Model

A Bayesian network is a graphical model that encodes the joint probability distribution for a set of random variables. A Bayesian network is a way of describing the relationships between causes and effects, and contains nodes and arcs. Bayesian networks are widely used in data mining, machine learning, and artificial intelligence for representing and reasoning about problems in which probability plays a role. A Bayesian network is a directed, acyclic graph whose nodes represent random variables and arcs represent direct dependencies [12]. Each connected node's conditional probability and the joint probabilities are calculated within the network. For instance, if event A is directly connected to B, and event B is directly connected to C, then the probability of observing three events' occurrence can be calculated by the joint probability *P(A,B,C)* as follows:

$$P(A,B,C) = P(C|A,B)P(B|A)P(A) \quad (2)$$

It should be noted that *P(A)* is the probability of event A, *P(B|A) is the* conditional probability that calculates the likelihood of B given that A is observed, and *P(C|A,B)* is the conditional probability that calculates the likelihood of observing C given that both A and B is true or has occurred.

In this study, GeNIe software is used for design and implementation of our Bayesian Network model. GeNIe is a tool that can be used to design and implement several Bayesian network models that calculates the conditional probabilities and marginal probabilities. The marginal distribution of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables [13]. Marginal variables are those variables in the subset of variables being retained. These defined as "marginal" since they are calculated by summing values in a table along rows or columns and they are written the sum in the margins of the table [12, 13].

IV. FUZZY INFERENCE SYSTEM

Fuzzy inference system can use in information security risk assessment. This method is a qualitative method and its advantages are listed as below.

1) Less values, rules and decisions are required.
2) More observed variables could be evaluated.
3) Linguistic / non-numerical variables are used; making it more similar to the way humans think.

4) It relates output to input, without having to understand all the variables, enabling the design of a system that may be more accurate and stable than a conventional control system.
5) Simplicity allows the solution of previously unsolved problems.
6) Rapid prototyping is possible; a system designer does not have to know everything about the system before starting work.
7) They are cheaper to make than conventional systems because they are easier to design.
8) They simplify knowledge acquisition and representation.
9) Few rules could encompass great complexity.

V. CASE STUDY

In order to implement the proposed approach, a preliminary model was developed for a software services company to evaluate the database security during the testing process.

The selected organization for the case study is a software services company whose main business area is software testing. Company's customers are generally in the telecommunication sector. The company has seventy employees and sixty-four employees are working in software testing projects. The working place layout is an open-office type and it has two separate meeting rooms and a server room.

In the case study, firstly the Bayesian model was designed based on the testing process in the selected company. Secondly, fuzzy membership functions were constructed for both assets and risk factors. After these steps, risk factors were evaluated within the information security risk assessment scope based on the results for information risk factors. Details about implemented approach were explained in the Design and Implementation section.

A. Materials

In order to collect expert opinions about probabilities in the Bayesian model, questions about threats, risks, and assets were asked to the experts in the company. Fifty-one experts who knows the system very well and are responsible for company's software test process contributed for analysing the system. The questions were arranged within four main categories, which are given below, and the answers were collected from the experts.

- Assets – what do we have?
- Vulnerabilities – is the asset at risk?
- Threats – what / who will attack / damage / destroy it?
- Risk factors – what are the main risks?

According to the scope of our risk assessment, several different assets, vulnerabilities, threats, and risks were identified and used as below.

1) Seven assets; "Data stored in Computers/Laptops", "VPN Connection", "Customer data stored in database", "Test data", "IT Users", "Test Users", "Test Support Users".

2) Twelve vulnerabilities; "Computers Vulnerable to Malfunctions", "VPN Connection Vulnerable to Malfunctions", "Vulnerable to Physical Problems/Damages", "Vulnerable to Malicious Codes", "Might Be Easily Lost/Stolen", "Lack of Experience or Training", "Lack of Awareness", "Computers Vulnerable to Unauthorized Access", "VPN Connection Vulnerable to Unauthorized Access", "Disgruntled IT Users", "Disgruntled Test Users", "Disgruntled Test Support Users".

3) Eleven threats; "Technical Problems", "Malicious Code", "Blackout/Brownout", "Fire", "Earthquake", "Thieves", "Human Error/ Failure by IT Users", "Human Error/Failure by Test Users", "Human Error/ Failure by Test Support Users", "Hackers/Cyber Criminals", "Resignation and Expel with Client Data".

4) Nine risks; "Unavailability of Computers", "Connection Loss", "Loss of Test Data", "Database Crash", "Disclosure of Confidential Information", "Unauthorized Change or Damage of Data", "Discontinuation of Testing Processes", "Inaccurate Test Results", "Prestige Loss".

Twenty-one questions were prepared to define the possibilities of threats in the system. Three different levels (Low, Medium, and High) of scale were used in this phase. In order to analyse risk values, thirty-four questions were asked to the experts and the answers were evaluated by using a scale with five levels (Very Low, Low, Medium, High, and Very High). The final phase contains the questions for asset values in the system. Experts were asked to evaluate seven different information assets based on either their approximate monetary values or a scale of ten points (one, two … ten) where "one" represents the lowest value, "ten" represents the highest value.

Collected data was analysed for calculating the probability of related threat, risk, and asset values. For each question and its corresponding criticality, the mean value was calculated based on the answers of experts. All these mean values were used in the Bayesian model to calculate the conditional probabilities of related threat or risk.

B. Design and Implementation

After collection of data, threats and risk factors are evaluated one by one. Conditional and marginal probabilities are calculated in the GeNIe software according to marginal probability distribution. The marginal probabilities for the threat "Resignation and Expel with Client Data" are shown in Figure 1 and the results for the risk "Unauthorized Change or Damage of Data" are shown in Figure 2.

After defining all assets, vulnerabilities, threats, risk factors and their relations, Bayesian network model is constructed to calculate and retrieve the probabilities of each of information security risks in our model. A small part of the Bayesian network architecture is denoted in Figure 3. The entire preliminary model is shown in Appendix section.

Because of the uncertainty of the risk factor, the fuzzy logic method, and a fuzzy inference system is used in this study. First, membership functions are determined for both assets and risk values. Hence, it could be deduced that membership function is a curve showing a point mapping points of inputting data into membership values, whose interval is between zero and one [14].

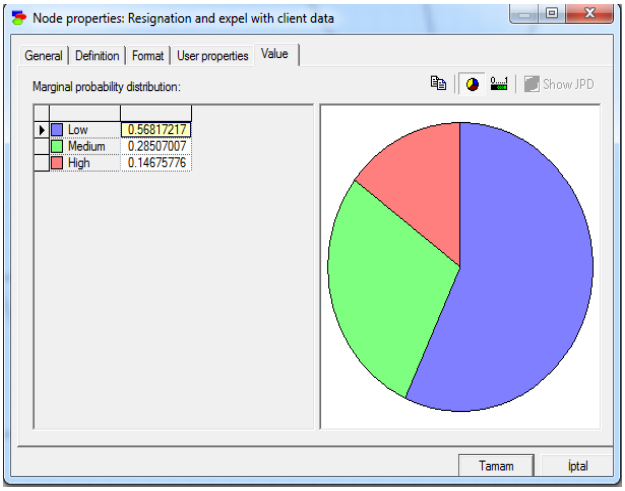


Figure 1. Marginal probabilities for the threat titled as "Resignation and expel with client data".

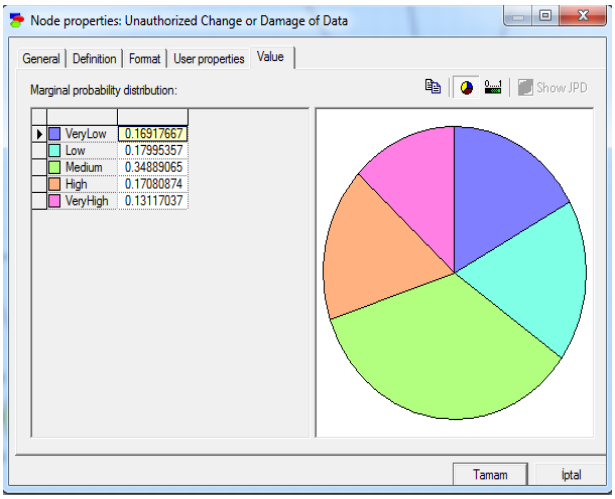


Figure 2. Marginal probabilities for the risk titled as "Change or Damage of data".

We used Mamdani Fuzzy Inference System (FIS) for fuzzification, defuzzification, and aggregation operations in our risk assessment model [15]. For each fuzzy rule, the minimum or maximum of the fuzzy membership values in the antecedent part is taken according to ''and'' or ''or'' operators in the antecedent part of that fuzzy rule. The propagated membership value from operations on the antecedents then truncates the membership function for the consequent for that rule. This truncation or scaling is established for each rule, and then the truncated membership functions from each rule are aggregated [15].

In order to create fuzzy membership function for assets and risk values, Trapezoidal membership function was selected and Gaussian membership function was used for risk probabilities. It should be noted that, "the final risk / impact value" in our model represents the overall risk or impact that is produced as the outcome of the relevant asset value and that risk's probability value. Hence, the outcome of a fuzzy rule in our model gives the fuzzy risk value using Mamdani FIS. Using these membership functions and adjusting the parameters, the fuzzy membership values for crisp asset values, risk values, and risk probabilities are calculated and they are given in tables Table1, Table 2, and Table 3 respectively.

Table 1. Fuzzy membership values for assets.

Asset value	Fuzzy membership values				
	Very Low	Low	Medium	High	Very High
1	1.000	0.800	0.000	0.000	0.000
2	1.000	1.000	0.000	0.000	0.000
3	0.893	1.000	0.143	0.000	0.000
4	0.536	0.862	0.857	0.404	0.000
5	0.179	0.517	1.000	0.617	0.167
6	0.000	0.172	1.000	0.830	0.500
7	0.000	0.000	0.938	1.000	0.833
8	0.000	0.000	0.313	1.000	1.000
9	0.000	0.000	0.000	0.938	1.000
10	0.000	0.000	0.000	0.625	1.000

Table 2. Fuzzy membership values for risks.

Risk Value	Fuzzy membership values				
	Very Low	Low	Medium	High	Very High
1	1.000	0.878	0.000	0.000	0.000
2	0.778	1.000	0.288	0.000	0.000
3	0.556	1.000	0.663	0.057	0.000
4	0.333	0.852	1.000	0.343	0.106
5	0.111	0.481	1.000	0.629	0.319
6	0.000	0.111	1.000	0.914	0.532
7	0.000	0.000	0.946	1.000	0.745
8	0.000	0.000	0.495	1.000	0.957
9	0.000	0.000	0.045	0.938	1.000
10	0.000	0.000	0.000	0.625	1.000

The formula for Trapezoidal membership function that is used for calculating asset values and risk values is shown as in (3).

$$\mu_F(x,a,b,c,d) = \begin{cases} 0, & if\ x < a \\ \dfrac{x-a}{x-b}, & if\ a \le x \le b \\ 1, & if\ b < x < c \\ \dfrac{d-x}{d-c}, & if\ c \le x \le d \\ 0, & if\ d < x \end{cases} \quad (3)$$

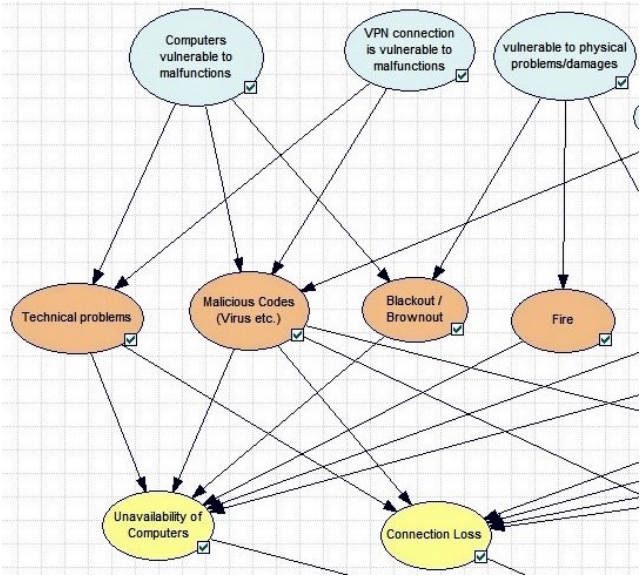


Figure 3. Partial view of our Bayesian network model.

Table 3: Fuzzy membership values for risk probabilities.

	Fuzzy membership values				
Risk probabilities *(Retrieved from Bayesian Network)*	Very Low	Low	Medium	High	Very High
0.000	1.000	0.946	0.135	0.018	0.044
0.011	0.999	0.957	0.147	0.020	0.047
0.093	0.953	1.000	0.266	0.042	0.077
0.129	0.912	0.995	0.332	0.056	0.093
0.154	0.877	0.984	0.383	0.068	0.107
0.190	0.818	0.956	0.464	0.089	0.129
0.203	0.795	0.942	0.495	0.098	0.138
0.250	0.707	0.882	0.607	0.135	0.172
0.290	0.627	0.818	0.703	0.175	0.207
0.333	0.539	0.739	0.801	0.227	0.249
0.351	0.504	0.704	0.837	0.251	0.268
0.366	0.474	0.674	0.867	0.273	0.285
0.389	0.430	0.627	0.908	0.308	0.312
0.411	0.392	0.585	0.938	0.342	0.338
0.427	0.362	0.551	0.959	0.371	0.359
0.450	0.324	0.506	0.980	0.411	0.389
0.500	0.249	0.411	1.000	0.506	0.458
0.591	0.144	0.263	0.936	0.688	0.592
0.667	0.085	0.168	0.801	0.830	0.707
0.675	0.079	0.159	0.782	0.844	0.720
0.730	0.052	0.110	0.654	0.923	0.797
0.750	0.044	0.096	0.607	0.946	0.823
0.820	0.024	0.056	0.441	0.995	0.904
0.834	0.021	0.050	0.409	0.999	0.918
0.887	0.013	0.032	0.301	0.992	0.961
0.902	0.011	0.028	0.274	0.985	0.971
0.950	0.007	0.018	0.198	0.946	0.992
0.980	0.005	0.014	0.158	0.910	0.999
1.000	0.004	0.011	0.135	0.882	1.000

Gaussian Fuzzy membership function is shown in equation (4).

$$\mu(x,a,b) = e^{\frac{-(x-b)^2}{2a^2}} \quad (4)$$

After defining fuzzy membership functions, twenty-seven fuzzy rules were constructed for the inference process. Some of the rules in our model are given below:

1) If Asset is *Very High* and Probability is *High*, then Risk is *Very High*.
2) If Asset is *High* and Probability is *High*, then Risk is *High*.
3) If Asset is *Low* and Probability is *High*, then Risk is *Medium*.
4) If Asset is *Low* and Probability is *Low*, then Risk is *Low*.
5) If Asset is *Very Low* and Probability is *Very Low*, then Risk is *Very Low*.

For instance, suppose that a specific asset's crisp value is given as 4, and the risk probability is calculated as 0.5 for a specific risk related with that asset. If we use the fourth fuzzy rule in our model (*"If Asset is Low and Probability is Low, then Risk is Low"*), then according to Mamdani FIS, the antecedent part of this rule will give a fuzzy value as 0.411. This is achieved by finding the minimum of the fuzzy membership values (0.411 and 0.862) of asset and risk probability.

After defining the fuzzy membership functions and rules, aggregation was processed based on the fuzzy rules for each related risk. For calculating the crisp risk values during the defuzzification process, Center of Gravity method has been used. If the output fuzzy set has at least two convex sub-regions, then the center of gravity (i.e., z^* is calculated using the centroid method) of the convex fuzzy sub-region with the largest area is used to obtain the defuzzified value $z*$ of the output [15]. This is given algebraically in Equation (5).

$$Z^* = \frac{\int \mu C_m(z) z dz}{\int \mu C_m(z) dz} \quad (5)$$

However, it should be noted that Eq. 5 is used when the membership values are continuous. If the values are discrete, the center of gravity method can be simply calculated as follows:

$$z^* = \frac{\sum_{i=1}^{n} x_i \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)} \qquad (6)$$

For instance, we can assume that a specific asset's crisp value is given as 5, and the risk probability is calculated as 0.73. Using the fuzzy membership tables in this study, and are aggregating all of the fuzzy rules mentioned in the previous page, then this risk's defuzzified value (by using the center of gravity method) can be calculated as follows:

$$R_d = \frac{(1x0.052) + (2x0.11) + (3x0.11) + (4x0.517) + \cdots + (8x0.617}{(0.052 + 0.11 + 0.11 + 0.517 + 0.517 + 0.517 + 0.617 +} = 6.064$$

C. Results

We implemented our model by using the methodologies, and making the calculations that were described in the previous section. The asset values were obtained by averaging the expert opinions' evaluation scores, which are given in Table 4. These asset values were given as crisp inputs to the fuzzy inference system. The marginal probabilities of risks were calculated from the Bayesian network and they are given in Table 5. These probabilities were also fed into the fuzzy inference system as the crisp input values. The fuzzy membership values of assets and risk probabilities were obtained by the membership functions. For each asset, relevant risk or risks were grouped, and then for each of these groups, all the fuzzy rules in the rule base were applied. Finally, aggregation and defuzzification operations were executed and the crisp risk / impact values were obtained, which is given in Table 6. It could be seen from Table 6 that the risk with highest score was "Discontinuation of Test Processes" and the risk with the lowest value was "Unauthorized Change or Damage in Data". It was also observed that there were relatively significant differences between the higher and lower risk values, which shall be useful for the risk evaluation and assessment processes.

Table 4. Crisp asset values evaluated by expert opinions.

Asset name	Asset value
Data stored in computers/laptops	6.90
IT users (experience / knowledge)	7.74
Customer data stored in Database	7.55
Test data	7.21
VPN connection	8.40
Test users (experience / knowledge)	8.20
Test support users (experience / knowledge)	8.01

Table 5. Risks' marginal probability values.

Risk name	Risk level	Marginal probability	Marginal probability (normalized)
Unavailability of computers	VL	0.138	0
	L	0.195	0.452
	M	0.238	0.786
	H	0.265	1
	VH	0.164	0.207
Connection Loss	VL	0.134	0.227
	L	0.228	0.701
	M	0.261	0.866
	H	0.288	1
	VH	0.088	0
Loss of Test Data	VL	0.162	0.141
	L	0.213	0.629
	M	0.226	0.763
	H	0.251	1
	VH	0.148	0
Database Crash	VL	0.144	0.351
	L	0.267	0.97
	M	0.272	1
	H	0.244	0.856
	VH	0.074	0
Disclosure of Confidential Information	VL	0.192	0.505
	L	0.235	0.859
	M	0.189	0.479
	H	0.252	1
	VH	0.132	0
Unauthorized Change or Damage in Data	VL	0.169	0.175
	L	0.180	0.224
	M	0.349	1
	H	0.171	0.182
	VH	0.131	0
Discontinuation of Test Processes	VL	0.074	0
	L	0.098	0.086
	M	0.201	0.448
	H	0.358	1
	VH	0.269	0.689
Inaccurate Test Results	VL	0.053	0
	L	0.145	0.344
	M	0.226	0.644
	H	0.321	1
	VH	0.254	0.751
Prestige Loss	VL	0.026	0
	L	0.085	0.167
	M	0.137	0.315
	H	0.378	1
	VH	0.375	0.99

Table 6. The final risk / impact values after defuzzification.

Risk name	Final risk / impact values
Discontinuation of Test Processes	8.79
Connection Loss	8.06
Database Crash	7.45
Prestige Loss	7.13
Unavailability of computers	5.84
Inaccurate Test Results	5.32

Loss of Test Data	5.28
Disclosure of Confidential Information	4.91
Unauthorized Change or Damage in Data	4.53

VI. CONCLUSION

In this study, we proposed a new information security risk assessment model based on Bayesian networks and Fuzzy inference system in order to evaluate and calculate both qualitative and/or quantitative risks in a more reliable, flexible, and objective manner. This information security risk assessment approach is different from other methodologies in the literature by combining the Bayesian network and fuzzy inference system.

The proposed preliminary model is developed to analyse test processes for a software company in order to evaluate the information security risks. First, the system's assets, threats, and vulnerabilities have been thoroughly analysed with the experts in the selected company. Then, information risk factors, threats, vulnerabilities, and their relations have been modelled in Bayesian network. Data is collected for our risk assessment model, with experts and managers based on the testing process in the company. Assets, vulnerabilities, threats, and related risk values are identified and analysed. Vulnerabilities, risks, and their relations are constructed with a Bayesian network and marginal probabilities for each risk are calculated. After Bayesian network is constructed, fuzzy membership functions are designed for assets' values, risks' probabilities, and risk values. In order to obtain more reliable and less subjective approach to the risk assessment process, fuzzy inference system has been used in this preliminary model. Twenty-seven fuzzy decision rules are constructed for some of the chosen risks by using the assets' values, relevant risk probabilities, and relative risk values. Finally, the risk / impact values are calculated in the aggregation and defuzzification processes. Based on the final risk values, the risks are evaluated and ranked for information security risk assessment process. This preliminary model will be used as a basis for an enhanced model, which can be successfully used for information security risk assessment.

As a future study, we will design and develop an alternative Bayesian network model will be developed to compare risk assessment results in different Bayesian models. This alternative model will be constructed based on the opinions of third party information security experts and consultants. Their feedback about the preliminary model will be evaluated for creating an enhanced model. Furthermore, in order to execute Bayesian network model and fuzzy inference system together and establish the integration between them, an interface might be developed with an appropriate programming language. We will also enhance this preliminary model with more fuzzy rules, weighted fuzzy rules, and alternative fuzzy membership functions with different parameters.

REFERENCES

[1] International Organization for Standardization. (2005). ISO/IEC 27001:2005.

[2] GB/T 29084-2007. Information security technology Risk assessment specification for information security.

[3] Lee. M. C. (2014). Information security risk analysis methods and research trends: AHP and fuzzy comprehensive method. *International Journal of Computer Science & Information Technology*. *6*(1). 29.

[4] Fu S. Xiao Y. (2012). "Strengthening the research for Information security risk assessment". International Conference on Biological and Biomedical Science Advanced in Biomedical Engineering. Vol. 9; pp. 386-392.

[5] Denys. P. (2006. February). Efficiency of Risk Assessment Methods. In Modern Problems of Radio Engineering, Telecommunications and Computer Science. 2006. TCSET 2006. International Conference (pp. 353-354). IEEE.

[6] Landoll, D. J. (2006). *The Security Risk Assessment Handbook*. Auerbach Publications.

[7] Yuhan. H. Xiaoyan. C. Linqiao. D. Songsong. Z. Min. W. & Yanxiong. H. (2013). The Reclamation Soil Suitability Study of the Highway Dumping Site Based on Fuzzy Comprehensive Evaluation Method. *Nature. Environment and Pollution Technology*. *12*(1).

[8] Zhao. D. M. Wang. J. H. Wu. J. & Ma. J. F. (2005. August). Using fuzzy logic and entropy theory to risk assessment of the information security. In *Machine Learning and Cybernetics. 2005. Proceedings of 2005 International Conference on* (Vol. 4. pp. 2448-2453). IEEE.

[9] Foroughi. F. (2008. July). Information security risk assessment by using Bayesian learning technique. In *Proceedings of the World Congress on Engineering* (Vol. 1. p. 133).

[10] Wang. J. Fan. K. Mo. W. & Xu. D. (2016. July). A Method for Information Security Risk Assessment Based on the Dynamic Bayesian Network. In *Networking and Network Applications (NaNA). 2016 International Conference on* (pp. 279-283). IEEE.

[11] Bayraktarli. Y. Y. Ulfkjaer. J. P. Yazgan. U. & Faber. M. H. (2005. June). On the application of Bayesian probabilistic networks for earthquake risk management. In *9th International Conference on Structural Safety and Reliability (ICOSSAR 05)* (pp. 20-23).

[12] Barber, D. (2011). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

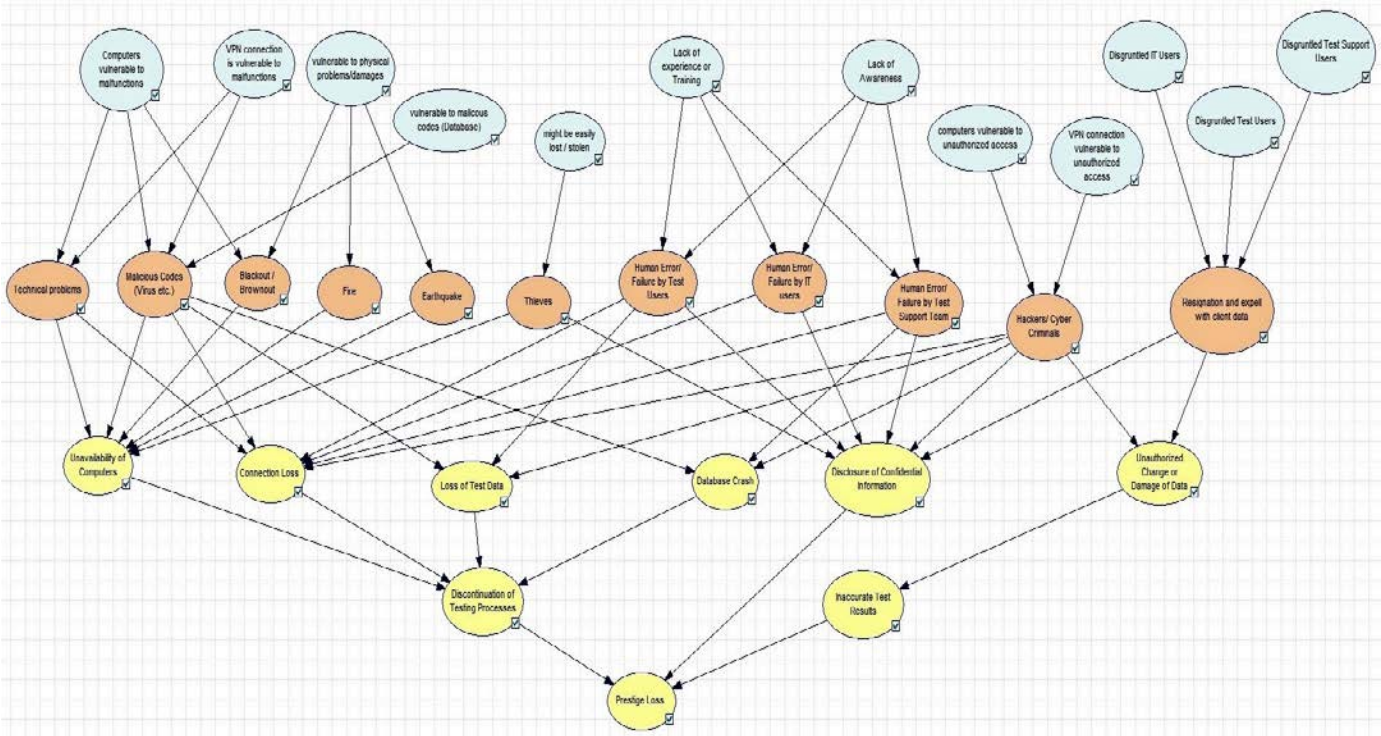
[13] Pitman Jim. (2016). *Stochastic Processes. An Introduction: Statistics. Statistics. 1st edition*. Content Technologies.

[14] Ariyanti. R. D. Kusumadewi. S. & Paputungan. I. V. (2010. January). Beck depression inventory test assessment using fuzzy inference system. In *Intelligent Systems. Modelling and Simulation (ISMS). 2010 International Conference on* (pp. 6-9). IEEE.

[15] Ross, T. J. (2004). *Fuzzy Logic with Engineering Applications, 2nd edition*. John Wiley & Sons Ltd.

APPENDIX

Figure 4. Preliminary Bayesian network model used in this case study.



Increasing the security of Mobile Communication with Steganography

C. MISMAN [1] and M. HACIBEYOGLU[2]

[1]Necmettin Erbakan University, Konya/Turkey, mismancelaleddin@gmail.com
[1]Necmettin Erbakan University, Konya/Turkey, hacibeyoglu@konya.edu.tr

Abstract **- Hiding and securing information is a basic demand throughout humanity. People have used their own bodies, languages, writings, etc. to provide this need. Steganography is acknowledged by science and art that researches hiding information methods. Steganography consists of two element basically that cover and secret information. In the past, people used their bodies and poems, diaries for cover and used tattoo and acrostic methods for secret information. In recently thanks to the developments of technology, Steganography has widened its methods and study areas. There are four Steganography methods which are text, image, audio and video in computer science. All types of Steganography methods have distinctive different ways to hide information. But if we want to mention the most used ones, we can say that changing characteristic of text (like color, font size) in text Steganography and changing Least Significant Bit(LSB) way for other types of Steganography methods. The LSB is a way that we overwrite the LSB of each byte of the cover (image, video, audio) with secret information binary representation. In our project, we are developing an android mobile application that allows user to hide a secret information inside any image. The image can be captured instantly or selected from user gallery. We are using LSB image Steganography method in order to hide secret information in image. Beside this, we encrypt the secret information with an encryption algorithm before inserting it in image. At the end, user can save the result image for the future or share with somebody who able to see the secret information only with this application**

Keywords **- Steganography, mobile communication, information security.**

I. INTRODUCTION

Steganography is a hiding information technique that anybody can't notice the message send by pairs has another secret message (information), except the sender and receiver. Therefore, we call the seen message cover media.

In the World War 2, the Germans used Microdot technique. We can define this technique as dividing large information into smaller pieces and hide them in the cover message. They send the cover message over unsafe channel, although it is really hard to detect hidden information [1]. For another steganography method, we can regard invisible inks as the method that have been used since ancient times.

In present, there are lots of studies in the digital steganography area. The steganography algorithm can work independently from cover media. This boundless situation gives people broad perspective and open mind in order to work different rare subjects with steganography like military, secure communication, and water marking. Moreover, to increase the security of the secret information, two or more different steganography methods can be combined. For example, when we consider image steganography, after embedding the secret information with Least Significant Bit (LSB) method to the cover image, we can change the spatial domain of image to the frequency domain with Discrete Cosine Transform (DCT) method. Subsequently, we can process a compress algorithms to the stego-image [2].

In 2003, Wu and Tsai designed an image steganography method that proposes pixel-value difference (PVD). One of the reason of this development is that LSB steganography in image can be detected when statistical steganalysis methods are applied. The PVD method changes the gray values of an image. Changed bits mostly gathered in edges than LSB method, in order to decrease detection of steganography by human eyes. The embedding secret message process starts with splitting of the cover image into non-overlapping blocks of sequential pixels. After difference value is calculated from these values of the two pixels in each block, all potential distinct values categorized into a number of range. In choice of the range intervals, the qualification of human visions sensitivity to gray value variations from smoothness to contrast is based. Then the difference value is altered by a new value to hide the value of a sub-stream of the secret message. The embeddable bit count in a pixel pair is calculated by the width of the range that the difference value belongs to [3].

When we research for the mobile applications that have been developed using steganography, we encounter "AUCH". "AUCH" is a realtime chat application that developed client - server architecture. The name stands for combination of the first two letter of "AUdio" and "CHat" words. The application allows user to chat with text and audio. While recording audio, users can hide information behind the audio and send to another user. Thus, secret messages can be transferred to other users although the application seems like a simple chat application [4].

For another usage of the steganography, we can count the digital watermarks. Digital files, images, music, videos can be stamped with digital watermarks by using different kinds of steganography methods.

Beside embed and hide secret information to cover objects, there are researches, that called steganalysis, to detect a digital media whether have secret message or not.

General steganalysis methods in detection of LSB steganography conclude with equations, histograms and statistical data that are obtained by comparing patterns, set of bits with similar images, patterns, close color pairs, and etc. [5].

At the end of the comparison and statistics, steganalysis decides whether a secret message is embedded in image or not. Discovering the secret information is not the subject of steganalysis.

II. STEGANOGRAPHY METHODS

When we look at the digital world, steganography has three elements in general and has been used in four main areas depending on the cover media. The steganography methods text, audio, video, image are shown in Figure 1.

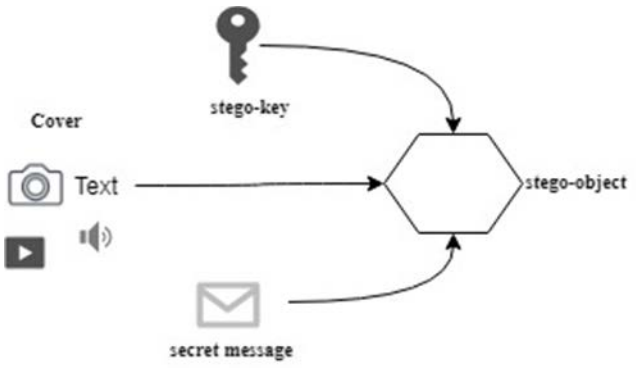


Figure 1: General work strategy of steganography

A. *Text Steganography*

Thanks to the existing of inscription from based on the human history, text steganography has inventive methods and ways. Changing characteristic of text (color, size, and etc.), line shifting, word shifting, acrostic methods are some of them.

In the below example, you will see a paragraph that has different word spaces. The spaces can be decoded 0 and 1 in respect of the spaces one and two.

```
happy␣families␣␣are␣all␣␣alike␣every␣unhappy␣␣family␣is␣unhappy
in␣its␣␣own␣way␣every␣␣thing␣was␣in␣␣confusion␣in␣the␣oblonskys
house␣the␣␣wife␣␣had␣discovered␣that␣the␣␣husband␣␣was␣carrying
on␣an␣␣intrigue␣␣with␣a␣french␣girl
```

Figure 2: Word shifting example

In the below example, when you combine the first words of lines you will reach the secret message **"Come to our place at midnight".**

Come see how the swallows fly
to the south island helped by
our fine weather. They locate the
place in the immense ocean by navigating
at night, following the clouds which form around
midnight above the shore.

B. *Audio Steganography*

Embedding and hiding secret message into digital sound is known as audio steganography. The message can be embedded into WAV, AU and MP3 sound files.

- LSB coding
- Phase coding
- Parity coding
- Spread spectrum are the methods that are commonly used in audio steganography.

C. *Video Steganography*

A video is a composite of image and audio elements. Therefore, most of the steganography methods that applied to image and audio covers, can be also applied to a video file. Thanks to faster frame rate of the videos, it is really hard to detect secret information. The secret message can be covered by H.264, Mp4, MPEG, AVI, and etc. video formats.

- LSB method
- Non - Uniform rectangular partition
- Masking and filtering
- Compressed video steganography

D. Image Steganography

Image steganography is a method to hide and embed secret information inside a digital image. The information can be embedded into JPEG, BMP, GIF image files.

- Least Significant Bit
- Masking and Filtering
- Algorithms and Transformations are the methods mostly used in image steganography

Figure 3: Cover image

Figure 4: Stego-object

In the Figure 3, there is a cover image and in the Figure 4 there is a stego-object that contains a Napoleon Hill quote: "Before success comes in any man's life, he's sure to meet with much temporary defeat and, perhaps some failures. When defeat overtakes a man, the easiest and the most logical thing to do is to quit. That's exactly what the majority of men do."

III. LEAST SIGNIFICANT BIT METHOD

LSB stands for Least Significant Bit that is the lowest bit in a series bits string of a byte. For example, in the binary number: 10111001, the least significant bit is the far right 1.

In digital world images consists of pixels that are defined as 8-bit or 24-bit color scheme. In 8-bit color scheme, every pixel defined by 1 byte and 24-bit color scheme, every pixel defined by 3 byte that are red, green, blue. (RGB) As a result of this,

- 8-bit = 2^8 = 256 different colors
- 24-bit = 2^{24} = 16777216 different colors can be produced

In our mobile application, we have used images as cover media. User selects the cover image from his/her image gallery or takes an instant picture.

Table 1: A color representation of a pixel.

	Red	Green	Blue
Binary	11111000	11001001	00000011
Numeric	248	201	3
Hex	F8	C9	03

In the example above; binary, numeric and hex representations of the color **RGB(248,201,3)** = **#F8C903** = **11111000 11001001 00000011** are shown. Because of that hiding information in image is a bitwise operation, we will be interested in binary representations of the pixel. If we give a simple example to LSB method, we will show how to embed a 'C' character to 3 pixels of a picture.

C' character binary representation is **'01000011'**

Table 2: First 3 pixels of an image.

00100111	11101001	11001000
00100111	11001000	11101001
11001000	00100111	11101001

Table 3: After embedding 'C'.

0010011**0**	11101001	11001000
0010011**0**	11001000	1110100**0**
1100100**1**	00100111	11101001

After embedding operation we have changed 4 bits of 9 bits. 5 bits have remained same value.

IV. STEGMIS

In this study, we propose an android application which called STEGMIS and designed with client-server architecture. The client side of the STEGMIS is built in Java with the development environment Android Studio Integrated Development Environment. Clients communicate with each other over Openfire Server which is configured with MySQL database management system. Before embedding secret information in images, secret information is encrypted by The Advanced Encryption Standard (AES) algorithm with a 128-bit key. LSB method have been used for last 3-bits of pixels.

A. Android Studio

Android Studio is a sophisticated developed tool for developers to build apps on every type of Android device. Android Studio provides elegant code editing, debugging, and performance tooling. It also has an instant build/deploy system that allow developers to focus on generating advanced quality apps. Android Studio is the official Integrated Development Environment (IDE) for Android app development, based on IntelliJ IDEA [6].

Android Studio has special features which make developers comfortable in code completion, refactoring. Beside these; developers can trace network activity, code analysis, real-time statistics for app's CPU and memory usage of the application on the IDE.

B. Openfire Server

Openfire is a real time collaboration (RTC) server licensed under the Open Source Apache License. It uses XMPP (also called Jabber) widely adopted open protocol for instant messaging. XMPP (Extensible Messaging and Presence Protocol) is a protocol based on Extensible Markup Language (XML) and developed for instant messaging [7].

Openfire server is available for windows, linux and mac operating system. Openfire has powerful security and performance beside the ease of installation. Openfire can be configured with most of database management systems (MySQL, Oracle, PostgreSQL, IBM DB2, and etc.) thanks to its own database scripts. Openfire has some plugins to provide various demands. Beside this, there is a client application called Spark and a client API (Application Program Interface) called Smack in order to supply the communication between client and Openfire server.

Smack is an open source XMPP client library build in Java. It is available for Java SE compatible, JVMs and Android. Smack is a naive Java library, it can be imported into your applications to make XMPP instant messaging client. Smack enables to establish simple XMPP integrations such as sending notification messages, presence-enabling devices. Smack and XMPP allows you to easily exchange data, in various ways e.g. fire-and-forget, publish-subscribe, between human and non-human endpoints [8].

C. MySQL

MySQL is a popular Open Source SQL database management system build up, distributed, and improved by Oracle Corporation [9]. Beside MySQL databases are relational, the MySQL Database Server provides fast, reliable, scalable data transaction operations.

D. Advanced Encryption Standard (AES)

The Advanced Encryption Standard is canonical algorithm to encrypt digital data. It is accepted by American government. We can define the features of AES

- Symmetric key
- Symmetric block cipher
- 128/192/256-bit keys can be used
- More efficient than Triple-DES
- Can be implementable with most of the programming languages

In our android application, we use Openfire android Smack API for communication with server. To use STEGMIS, users just need to signs up the application with a unique username and password. After the user enter the application, there are 3 screen for the user

1) Generate an image with embedded secret message.

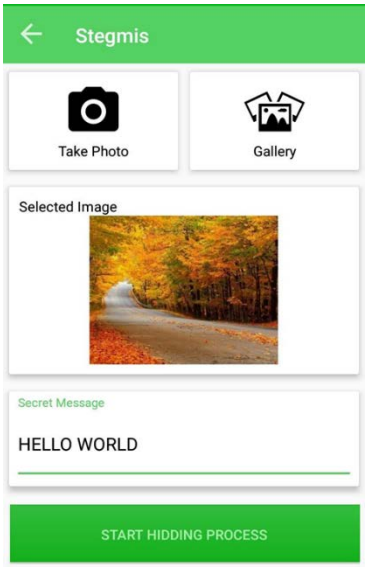


Figure 5: Embed secret message screen

As you see from Figure 5, the user firstly selects the cover image by capturing or adding from gallery. After entering the secret message, user can start the encoding process.

2) Share stego-object

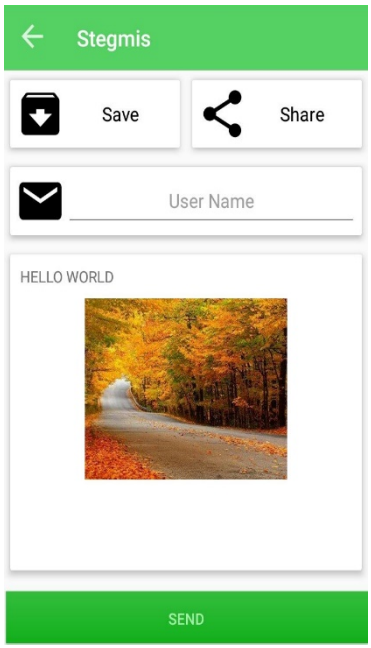


Figure 6:Save-share encoded image screen

After encoded process is finished, user can save or share the image. If the user wants to share the image with a STEGMIS user and knows his/her username, the user can send the image by typing his/her username and clicking the send button.

3) Decode stego object and reveal embedded secret message

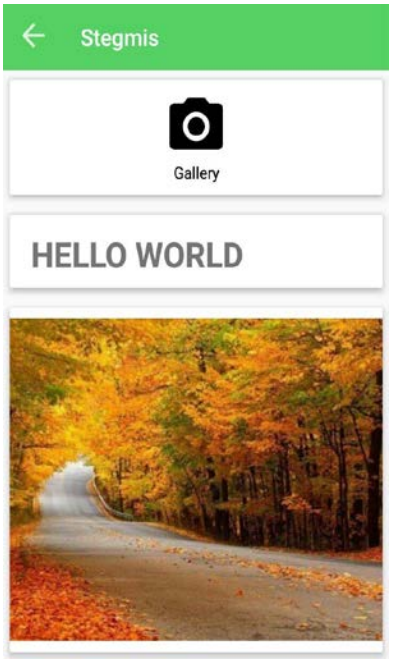


Figure 7: Decode image and reveal secret message screen

In the Figure 7 screen, after user select image from his/her gallery, decode operation starts automatically.

As an example of the proposed STEGMIS, we embedded the Gelir poem by Necip Fazıl Kısakürek to a nature image. The original image and the stego-image are shown in Figure 8 and 9, respectively.

Figure 8: Nature image

Figure 9: Stego-image

V. CONCLUSION

All in all, steganography can be really sophisticated method to hide the secret information and secure privacy. Therefore, it can arise in open-mind and clever people hands. People can send or share secret, private messages inside their own selfie, nature, animal images with our mobile application. In addition to this, people can store secret and private information in images with steganography and cryptography. In our mobile application, we combine AES and LSB. Thus we don't store the secret message as plain text bytes, but AES encrypted cipher text generated with 128-bit key. For the new versions of our mobile application, we plan to use asymmetric encryption algorithms instead of AES and allow user to select how many least significant bit used to embed the secret message.

REFERENCES

[1] Jamil T., "Steganography: The art of hiding information is plain sight",IEEE Potentials 18:01, 1999.

[2] Raja, K. B., Chowdary, C. R., Venugopal, K. R., & Patnaik, L. M., "A Secure Image Steganography using LSB, DCT and Compression Techniques on Raw Images" In Intelligent Sensing and Information Processing, 2005. ICISIP 2005. Third International Conference on. IEEE, 2005.

[3] Wu, D. C., & Tsai, W. H. "A steganographic method for images by pixel-value differencing", Pattern Recognition Letters 24 (9–10), pp.1613-1626, 2003

[4] Ali Tarık Gürkan. "Gerçek Zamanlı Mobil Steganografi Uygulaması Geliştirilmesi", Şubat 2016

[5] Luo, X., Liu, B., & Liu, F, "Detecting LSB steganography based on dynamic masks In Intelligent Systems Design and Applications, 2005. ISDA'05. Proceedings. 5th International Conference on (pp. 251-255). IEEE, 2005

[6] Android Studio, https://developer.android.com/studio/index.html

[7] Openfire, https://www.igniterealtime.org/projects/openfire/

[8] Smack, https://github.com/igniterealtime/Smack

[9] MySQL Database Management System, https://dev.mysql.com/doc/refman/5.7/en/what-is-mysql.html

[10] AES (Advanced Encryption Standard) ,https://www.tutorialspoint-.com/cryptography/advanced_encryption_standard.htm

Rule-Based Performance Measurement in Open Source IDS Systems

M.COŞAR[1] and H. E. KIRAN[2]

[1]Hitit University, Çorum/Turkey, mustafacosar@hitit.edu.tr
[2]Hitit University, Çorum/Turkey, harunemrekiran@hitit.edu.tr

Abstract **- In recent years, intrusion detection (IDS) and prevention systems (IPS) are vital in small and medium-sized computer networks where data security is important. Examples of these systems are Raspberry Pi (RasPi) computer with open source IDS software. It is necessary to develop a rule-based architecture in order to detect the attack by analyzing the traffic with this system on a network. In this study, Snort IDS module has installed on RasPi v3 computer and it has tried to measure alert performance value according to the number of rules of developed system. The scenario has that a total of an attack of 1 million data packets, including one packet attack in 50 microseconds, to a server on the network has been found to result in a nearly 15% decrease in alert performance after about 7500 rules.**

Keywords **– Network security, IDS, Raspberry Pi, Snort, Performance measurement.**

I. INTRODUCTION

IN today's digital world, security has become extremely important in the course of production, transmission and storage of data. The basis of this security concept is the security of personal computers and communication network. Data traffic need to be monitored and analyzed to ensure network security. In this case, a firewall can be used to detect and prevent unusual traffic.

The provision of security systems, especially in small-scale home-office computer networks, it has brought considerable rate time, cost and information load. Therefore, the development and use of low cost and practical solutions are highly demanded [1].

On the network, embedded microprocessor systems with low-energy, sensor network capabilities that can carry the characteristics of a computer at minimum scale can be located. Example of these systems are Raspberry Pi and Arduino. These systems can be placed as a security system by with some software components with detection and processing algorithms [2].

RasPi, a firewall implementation via a virtual switch [3] prevented an IP from generating IP-based malicious traffic. Raspberry Pi can be useful tool in IT systems and be an inexpensive way to perform for network monitoring [4].

With the help of Raspberry Pi, a simple, efficient and low cost home network monitoring application, has allowed the listing and control of IoT devices [5]. In addition, malicious packages were detected with a programmable embedded IDS prepared in [6,7].

Although previous studies have measured the performance of CPU, RAM, and capturing attack packets, performance measurements have not found depending on the number of rules. In this study, a firewall was created with RasPi and Snort IDS on a local network. According to the scenario, a user has attacked to a server on the network with synflood packets for 50 microseconds. It has been tried to measure the performance of this firewall, which is operated according to different rule numbers, to catch attack packets.

II. CONCEPTUAL FRAMEWORK

A. Raspberry Pi

Low cost and easy to use, Raspberry Pi [8] is called modular, portable and manageable computers running on Linux operating system. These systems, which are shown in Figure 1, are preferred in the analysis on network by listening to traffic, collecting data and calculating operations with a simple user screen and powerful processors [9].

Figure 1: Raspberry Pi-3 overview

B. Snort

First developed by Martin Roesch [10] Snort is Linux-based, open source and free software. Snort is a signature-based software that is used as an intrusion detection and prevention system and can perform protocol and anomaly analysis on the network, and users can write their own rules [11-13]. It is also widely preferred by researchers because of its advantages [14]. Snort can be a very good IDS if you understand network attacks

and know what to look for [15]. An analysis of network traffic with Raspberry Pi and Snort can be made to make the network more secure [16].

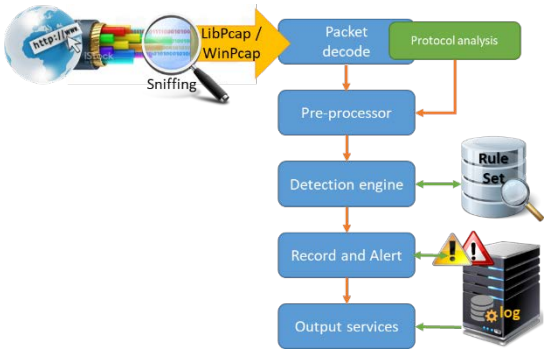


Figure 2: Snort IDS structure

In Figure 2, a general architecture of Snort is tried to be given. As you can see, the first step is to listen to the network traffic and catch the packages depending on the operating system's structure. These packets are processed by protocol and port based analysis. Then the attack starts to be detected depending on the rule sets. The log server records the warning if it is clear that is an attack.

III. System Architecture

In the study, as hardware components, Cisco 2960 x switch, Raspberry Pi 3, laptop, desktop and server computers have used. An architecture of the system prepared in Figure 4 is given.

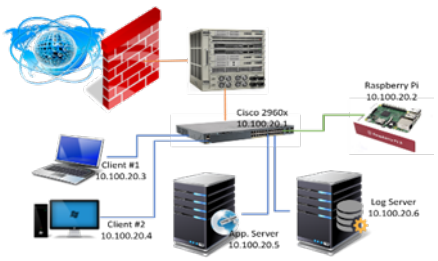


Figure 3: Architecture of the prepared system

As software components; Debian, Ubuntu and Windows operating systems, Snort IDS, hping3, SynFlood and log server were actively used.

A. Preparation of Components

After installing Debian operating system on Raspberry Pi, Snort IDS was installed. For the installation of Snort, prerequisite packages *(build-essential libpcap-dev libpcre3-dev libdumbnet-dev zlib1g-dev)* have been installed. Then the source code was downloaded from snort.org site and installation was continued. In Figure 4, the next screen view is shared since the installation was completed.

IV.

```
pi@raspberrypi:/etc/snort/rules $ snort -V

        ,,_     -*> Snort! <*-
       o"  )~   Version 2.9.11.1 GRE (Build 268)
        ''''    By Martin Roesch & The Snort Team: http://www.snort.org/contact#team
                Copyright (C) 2014-2017 Cisco and/or its affiliates. All rights reserved.
                Copyright (C) 1998-2013 Sourcefire, Inc., et al.
                Using libpcap version 1.8.1
                Using PCRE version: 8.39 2016-06-14
                Using ZLIB version: 1.2.8

pi@raspberrypi:/etc/snort/rules $
```

Figure 4: Snort installation process completion screen.

To register the warning list the folder *(sudo mkdir / var / log / snort)* must be activated. In addition, the SQL database must be created for all the data to be saved *(mysql> create database snort;)*.

```
pi@raspberrypi:/etc/snort/rules $ tree /etc/snort
/etc/snort
├── attribute_table.dtd
├── barnyard2.conf
├── classification.config
├── file_magic.conf
├── gen-msg.map
├── preproc_rules
├── reference.config
├── rules
│   ├── iplists
│   │   ├── default.blacklist
│   │   └── default.whitelist
│   ├── local1.rules
│   ├── local3500.rules
│   ├── local500.rules
│   └── local.rules
├── sid-msg.map
├── snort.conf
├── so_rules
├── threshold.conf
└── unicode.map

4 directories, 16 files
pi@raspberrypi:/etc/snort/rules $
```

Figure 5: Installation summary

After all process are complete, the installation summary should be as shown in Figure 5.

B. Adding IDS Rules

As can be seen in Figure 6, some of the rules examples have inserted into the Snort. The next it was tried to have monitored the attack packages from the users in the group of 10.100.20.0 network according to the scenario.

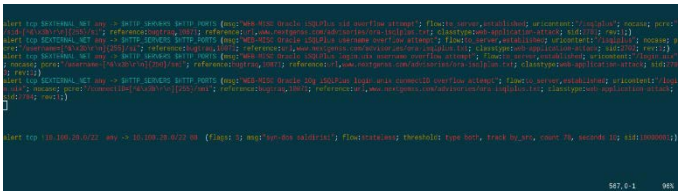


Figure 6: Entering the rules to Snort

Snort-IDS rules performance evaluations perform by means of detection rate, which can calculate by detection number per total attack of each attack type [12].

C. Scenario

After the system components were assembled, two different scenarios have been developed depending on the number of rules.

Scenario: SYN Flood with 16 byte + header pack

With this attack method, Snort's performance was tried to be measured depending on the number of rules ranging from 500 rules to 12500 rules with each one packet attack per 50 microseconds. During this period, 1 million packet attacks were attempted.

V. RESULTS

The packet capture values of the developed system depending on the number of rules at the time of the attack have tried to be given in Fig.7.

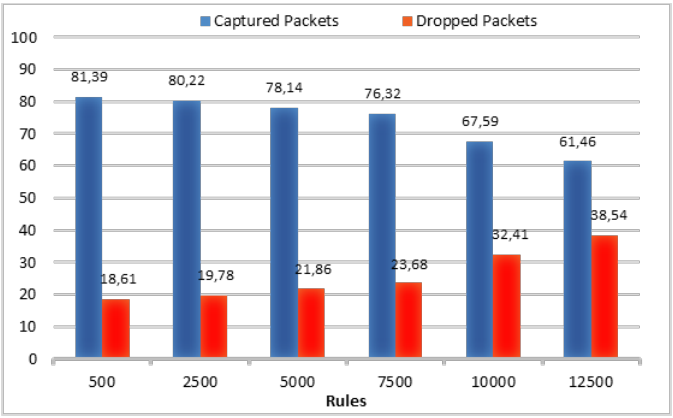


Figure 7: Packets performance values

At the time of the attack, performance of the IDS system started with 500 rules and performance decreased by 1-2% in 2500 and 5000 rules analysis, and performance decreased by 9% when 10000 rules were executed after 7500 rules.

Finally, with the implementation of 12500 rules, packet capture performance has been reduced by up to 61%.

VI. CONCLUSION

According to research findings, when an attack on the network, it was determined that the IDS developed by the number of rules the capture performance of the attack packets after 7500 rules is decreased by 15%. Therefore, the more rules that are processed during the use of such a system, the more loss of performance. Among the reasons for this are the CPU and Memory capacity of the RaspPi.

Different performance values can be obtained according to different rule sets. For this reason, after network traffic analysis, IDS systems should be positioned according to network characteristics.

REFERENCES

[1] Nabi, Z., A $35 Firewall for the Developing World, arXiv, https://arxiv.org/abs/1405.2517, May 2014.

[2] Ferdoush, S., Li, X., Wireless Sensor Network System Design using Raspberry Pi and Arduino for Environmental Monitoring Applications. The 9th International Conference on Future Networks and Communications (FNC-2014), Procedia Computer Science 34 (2014) 103–110.

[3] Rolbin, M., Early detection of network threats using Software Defined Network (SDN) and virtualization, Master Thesis, Carleton University, Ottawa, Canada, 2013.

[4] Hogg, S., Raspberry Pi as a Network Monitoring Node, Network World, Oct 30, 2013.

[5] Kutukian, G., Raspberry Pi-3 Home Network Monitoring Tool, Master Thesis, Computer Science in California State Polytechnic University, Pomona, 2016.

[6] Tuncer, T., Tatar, Y., Fpga Tabanlı Programlanabilir Gömülü Saldırı Tespit Sisteminin Gerçekleştirilmesi, Journal of the Faculty of Engineering and Architecture of Gazi University, Vol 27, No 1, 59-69, 2012.

[7] Coşar, M., Karasartova, S., Raspberry Pi ve Snort ile SOHO Networklerde Bir Güvenlik Duvarı Uygulaması, 2nd International Conference on Computer Science and Engineering-UBMK2017, Antalya-Turkey, October 5-8, 2017.

[8] Kim, H., Kim, J., Ko, Y., Developing a Cost-Effective OpenFlow Testbed for Small-Scale Software Defined Networking, ICACT2014, February 16-19, 2014.

[9] Hentschel, K., Jacob, D., Singer, J. and Chalmers, M., Supersensors: Raspberry Pi Devices for Smart Campus Infrastructure (short paper), IEEE 4th International Conference on Future Internet of Things and Cloud, 2016

[10] Roesch, M., Snort – Lightweight Intrusion Detection for Networks, Proceedings of LISA '99: 13th Systems Administration Conference, Seattle, Washington, USA, November 7–12, 1999.

[11] Bilgem, Snort'a Genel Bakış, https://www.bilgiguvenligi.gov.tr/saldiri-tespit-sistemleri/snort-2.9.2-kurulumu-1.html, Erişim Tarihi: 25 Haziran 2017.

[12] Khamphakdee, N., Benjamas, N. and Saiyod, S., Improving Intrusion Detection System Based on Snort Rules for Network Probe Attack Detection, Internatiınal Conference on Information and Communicaton Technology, 2014.

[13] Sharma, M., Kaushik, A. and Sangwan, A., Performance Analysis of Real Time Intrusion Detection and Prevention System using Snort, International Journal of Engineering Research & Technology (IJERT), Vol. 1 Issue 5, July – 2012.

[14] Sforzin, A., Conti, M., Marmol, F.G. and Bohli, J-M., RPiDS: Raspberry Pi IDS A Fruitful Intrusion Detection System for IoT, DOI 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.114.

[15] Theodor, F.J. and Hjalmar, W., Network Intrusion and Detection-An evaluation of SNORT, Linköping University Department of Computer and Information Science, Bachelor thesis, 2017.

[16] Kyaw K.A., Chen, Y., Joseph, J., Pi-IDS: Evaluation of Open-Source Intrusion Detection Systems on Raspberry Pi 2, 2015 Second International Conference on Information Security and Cyber Forensics (InfoSec), ISBN: 978-1-4673-6988-6 ©2015 IEEE.

Fusion of Full-Reference and No-Reference Anti-Spoofing Techniques for Ear Biometrics under Print Attacks

İ.TOPRAK[1] and Ö. TOYGAR[1]

[1] Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10, Turkey, imren.toprak@emu.edu.tr

[1] Computer Engineering Department, Faculty of Engineering, Eastern Mediterranean University, Famagusta, North Cyprus, via Mersin 10, Turkey, onsen.toygar@emu.edu.tr

Abstract **- In this paper, we propose an anti-spoofing method that employs the fusion of various full-reference and no-reference image quality assessment techniques to detect fake and real ear images presented to biometrics systems under print attacks. In this context, full-reference image quality assessment measures such as Error Sensitivity Measures, Pixel Difference Measures, Correlation-Based Measures, Edge-Based Measures, Spectral Distance Measures, Gradient-Based Measures, Structural Similarity Measures and Information Theoretic Measures are used. Additionally, no-reference image quality assessment measures such as Distortion Specific Measures, Training Based Measures and Natural Scene Statistics Measures are implemented to distinguish fake and real ear images. A comparative analysis of the performance of these quality metrics and the proposed method using decision-level fusion of all aforementioned measures are performed. The experimental results are presented using AMI and UBEAR ear databases by creating print attack counterparts of the ear images used in these databases.**

Keywords **- Ear biometrics, Spoofing, Image quality measures, Print attacks.**

I. INTRODUCTION

NOWADAYS, protection of the biometric systems against spoof attacks becomes popular research area in the biometric community. Many studies have been performed to evaluate different types of biometric systems security [1-9].

Ear is one of the biometric traits that satisfies the requirements which are uniqueness, universality, permanence and collectability of recognition systems. Since ear biometric trait has some advantages, it is also used in the implementation of recognition systems. When it is compared with the other biometric traits such as iris, fingerprint or palmprint, collaboration with the user is less i.e. ear sample is acquired without contacting to the sensor. Specifically, it is useful in the systems that are used in surveillance applications. Additionally, ear recognition systems are implemented to distinguish the identical twins [10]. In contrast to face recognition systems, occlusion of the face does not affect the performance of the ear recognition systems [11-13].

In this paper, we propose a novel method which employs the fusion of full-reference and no-reference image quality measures to detect real and fake images in ear recognition systems. In order to show the performance of the proposed method, we used two databases that are AMI and UBEAR. The type of attack that is implemented is print attack.

The rest of the paper is organized as follows: Related works are explained in Section II. Section III and IV describe the spoofing and image quality assessment, respectively. Proposed method is introduced in Section V. Experimental results are demonstrated in Section VI. In the last section, the study is concluded.

II. RELATED WORKS

Recently, image quality assessment (IQA) methods are applied in order to protect the biometric systems against fraudulent attempts. In these studies, various types of biometric traits are used. In [1], multiple biometric traits that are iris, fingerprint and face are used to show the efficiency of IQA techniques against multiple attacks. The authors used 21 full-reference and 4 no-reference image quality measures to detect fake images. Another novel approach is proposed in [2] in order to detect spoofs by using multiple biometric traits that are face and palmprint. In the proposed method, Local Binary Patterns (LBP), Difference of Gaussians (DOG) and Histograms of Oriented Gradients (HOG) are used for feature extraction of an image and Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are used to reduce the high dimensionality of an extracted feature vector. In addition to this, 7 full-reference image quality measures are applied to evaluate the image quality. In [3], an anti-spoofing method is proposed for iris, face and palmprint by using IQA techniques. Iris, face and palmprint images are fused in that study to get a single image. After this step, image quality measures are obtained for fused image. Finally, Support Vector Machine (SVM) classifier is used for decision making (i.e. real or fake).

On the other hand, some studies are based on one biometric trait for distinguishing real and fake images. Unimodal biometrics approaches are proposed in [4] and [5] for

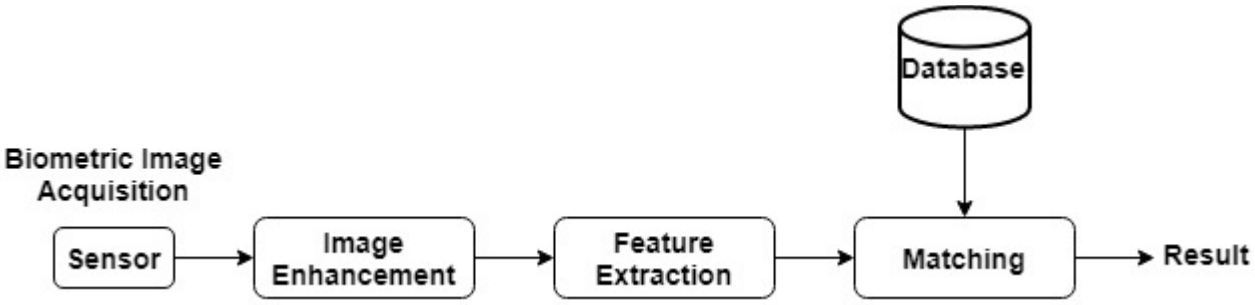


Figure 1: Block diagram of biometric system.

fingerprint and iris liveness detection, respectively. In these studies, IQA technique is applied with new parameterization in order to detect the liveness of a fingerprint and an iris. Two fingerprint databases (LivDet and ATVS) and one iris database (BioSec) are used in [4] and [5], respectively. Furthermore, liveness detection method under varying illumination condition for iris biometrics is studied in [6]. In the proposed method, in order to overcome illumination problem, Self Quotient Image (SQI) is applied. In fact, the remaining steps of the method are the same as in [3]. In the experiments, two different databases that are ATVS-Flr and IRISBASE are used. On the other hand, a new method is proposed in [7] for palmprint liveness detection. In that method, Binarized Statistical Image Features (BSIF) technique is implemented to extract the texture feature and get histogram of the image. Eight full-reference image quality measures are implemented and the resulting features are combined. SVM classifier is employed to distinguish genuine palmprint and imposter palmprint. The experiment results show the high accuracy of the proposed method. Aishwarya et. al. in [8] is another recent study that is based on palmprint biometric trait. In that study, 25 image quality measures, where 21 of them are full-reference and 4 of them are no-reference, are implemented to extract quality measures. Before extracting quality measures, image enhancement is performed by using Weber's Local Descriptor (WLD). In the final step, Euclidean distance is used as the distance measure. According to the results, the proposed method is error free, less complex and cost effective.

III. SPOOFING

Biometric systems are widely used technologies in order to verify or identify the living individual based on physical or behavioral characteristics. Figure 1 depicts the general structure of a biometric system. In the first step, the biometric image is acquired by the sensor or camera. The quality of the acquired biometric image is enhanced for further processing in image enhancement step. Next, the features that are important for recognizing the identity are extracted from the raw biometric image by feature extraction module. The extracted features are stored in the database as a template. In the matching step, the enrolled templates are compared with the new acquired biometric image to make a decision. While all these steps are in progress, the security vulnerability of the biometric system attracts the fraudulent attacks. Especially, the intruders can easily use the sensor module as a way to fool the biometric system. For this purpose, fake biometric image can be presented to the sensor. Fooling the biometric system by using fake biometric trait is called spoofing. Distinguishing a real biometric trait from the fake biometric trait is called spoof detection [14].

Many types of intruders may attempt to fool the biometric system for many reasons. For example, an individual may conceal his/her identity to enter another country by using artificial fingerprint, mask or contact lens. The second example may be an individual who wants to have access to another individual's account by imitating his/her biometric trait. Another example is an individual who is enrolled to the biometric system by using an artificial biometric trait, and then shares that identity with another individual so that multiple individuals have access to the system by sharing the same fake biometric trait [14].

Recently, spoof detection or anti-spoofing became a popular field of research. There are different types of anti-spoofing methods that are explained in Section 2. Anti-spoofing methods have to satisfy some requirements. Firstly, it has to be non-invasive that means it should not harm the users or require too much contact with the user. Secondly, it must be a user-friendly method. Thirdly, the method is required to be fast. The proposed method has to be as fast as possible in order not to let the user wait. Another requirement is that the method needs to be low cost. Finally, as spoof detection methods' performance increases, the recognition performance of the system should not decrease [1].

Anti-spoofing techniques are divided into three categories. The first category includes sensor-level techniques that are hardware-based. In this category, specific devices are added to the sensor to measure face thermogram, blood pressure, fingerprint sweat, or reflection of eye of the biometric trait to decide whether the biometric trait is live or not. The second category includes feature-level techniques that are software-based. Features that are used to detect fake biometric trait are extracted from raw biometric image. These techniques are integrated in feature extraction module in Figure 2. The third category includes score-level techniques. These are focused on the study of biometric systems at score-level in order to propose fusion strategies that increase their resistance against spoofing attempts [1].

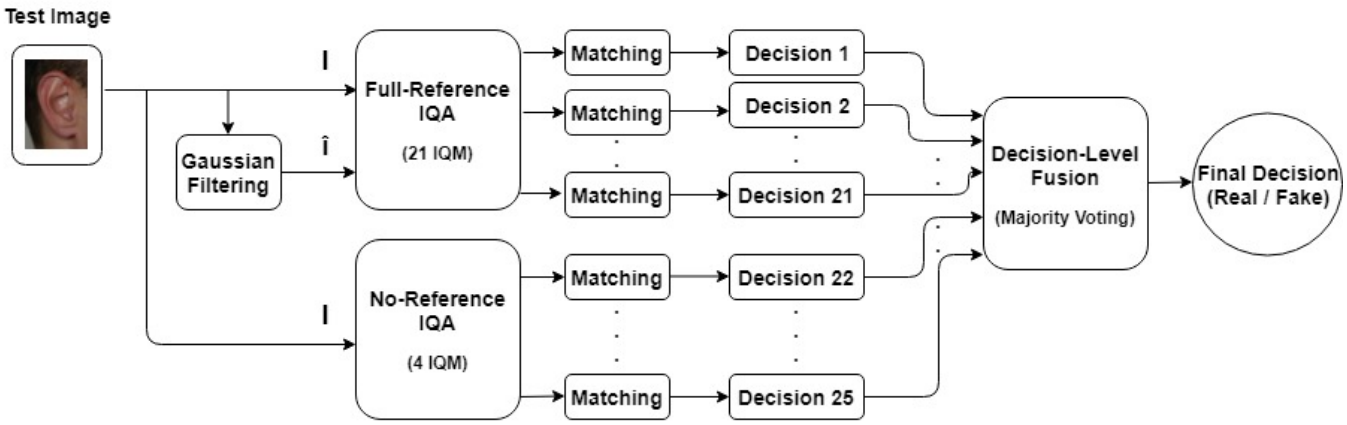


Figure 2: Block diagram of proposed method.

IV. IMAGE QUALITY ASSESSMENT

Image quality assessment (IQA) plays an important key role to detect the distortions on the image [15]. Since the image quality measures (IQM) are sensitive to some kinds of image distortions, they are used to measure the quality of an image for steganalysis. In addition to this, IQMs are applied efficiently on iris, fingerprint and face biometric traits for the detection of fake images in [1]. As a summary, IQMs are able to measure the quality difference between real and impostor images in terms of blur, noise, changes of color and sharpness, etc.

IQMs are classified into two groups as full-reference (FR) and no-reference (NR). FR measures are applied when there are two images to be compared. In this group, there is a reference image which is assumed to be perfect quality. Reference image is compared with the distorted image to analyze the differences between them by using FR measures. On the other hand, NR measures are used to assess the quality of an image without comparing it with any reference image [1].

Full-reference measures (FR) are collected in three different titles. Firstly, error sensitivity measures are used to measure the error between the reference image and distorted image. These measures can be introduced in five categories. The first category is pixel difference measures which include Mean Squared Error (MSE), Peak Signal to Noise Ratio (PSNR), Signal to Noise Ratio (SNR), Structural Content (SC), Maximum Difference (MD), Average Difference (AD), Normalized Absolute Error (NAE), R-Averaged Maximum Difference (RAMD) and Laplacian Mean Squared Error (LMSE). The second category is named as correlation-based measures and includes Normalized Cross-Correlation (NCC), Mean Angle Similarity (MAS) and Mean Angle Magnitude Similarity (MAMS). The third category is edge-based measures which include Total Edge Difference (TED) and Total Corner Difference (TCD). Next category is spectral distance measures and includes Spectral Magnitude Error (SME) and Spectral Phase Error (SPE). Finally, the last category is gradient-based measures which include Gradient Magnitude Error (GME) and Gradient Phase Error (GPE) [16-18]. The second group of FR measures is structural similarity measures which consider image degradations as perceived changes in structural information variation. It is named as Structural Similarity Index Measure (SSIM) [19]. The last group for FR measures is information theoretic measures. Since the image information between real and impostor image differs, information theoretic measures are developed. These measures are Visual Information Fidelity (VIF) [20] and Reduced Reference Entropic Difference (RRED) [21].

On the other hand, no-reference (NR) measures are divided into three categories. In the first category, quality of an image is measured according to its specific distortions. JPEQ Quality Index (JQI) [22] and High-Low Frequency Index (HLFI) [23] measures are used for this purpose. The second category is based on training approaches. In this category, different kinds of distortions are analyzed to have a general quality score. Blind Image Quality Index (BIQI) [24] is applied within this category. The last category is based on natural scene statistics. Some statistical properties which alter in the distorted image are available on natural scenes. While measuring the quality of an image with this approach, Natural Image Quality Evaluator (NIQE) [25] is used for the measurement.

V. PROPOSED METHOD

In this section, our proposed method is explained in detail. Figure 2 shows the system diagram of the proposed spoof detection method which is based on decision-level fusion. In the system diagram, I and Î are both gray-scale representations of the original test image and filtered test image, respectively. The input test image of ear biometrics is filtered by using Gaussian kernel filtering with σ value of 0.5 and size 3x3 for the application of FR IQA. Additionally, 21 FR IQMs are employed to measure the quality difference between reference image (**I**) and smoothed image (**Î**). Moreover, 4 NR IQMs are employed to measure the quality of reference image (**I**). In the matching part of the system, Nearest Neighbor classifier is applied to make a decision (real or fake) for each IQM. After obtaining the results for all IQMs, the final decision is found by applying decision-level fusion using Majority Voting technique.

VI. EXPERIMENTAL RESULTS

In order to show the performance of the proposed spoof detection method, several experiments are performed on two ear databases. These databases and the experimental work are explained in the following subsections.

A. Databases

The first database used to evaluate the performance of the proposed method is AMI [26] ear database which is a publicly available database. In our study, 50 subjects and their 2 ears (left and right), which makes totally 100 real ear images, are used from AMI database. The original size of the images is 492x702 pixels.

On the other hand, UBEAR [27] ear database is used as the second database in the experiments. We used 50 subjects and their 2 ears (left and right), which makes totally 100 real ear images, are selected randomly from UBEAR database. The original size of the images is 300x370 pixels.

Since spoof database for ear biometric trait is not available, ear spoof databases are constructed for both databases by taking the photograph of the printed original images from the aforementioned ear databases. Firstly, 100 real ear images are printed by using Olivetti d-color mf223 printer for each database. Next, HD photographs of the printed images are taken by IPhone 6s camera which is 12 megapixels with size of 1150x1640 and stored in our spoof database. In order to have the same size for real and fake images, photographs of the printed images are resized to 492x702 and 300x370 pixels for AMI and UBEAR databases, respectively. As a summary, there are 100 real images and corresponding 100 fake images for each database.

In this study, train and test sets are constructed for the experiments. Train set of one database contains 50 real images and corresponding 50 fake images. The test set of one database contains 50 remaining real images and corresponding 50 fake images.

B. Results

Experimental results are shown in Table 1 and Table 2 in terms of False Fake Rate (FFR) and False Genuine Rate (FGR). FFR is the number of real images that are classified as fake. FGR is the number of fake images that are classified as real. Half Total Error Rate (HTER) is computed as (FFR+FGR)/2.

In the first experiment, 25 IQMs are analyzed one by one. According to the results, SSIM has the minimum HTER with 4.5 for AMI database as shown in Table 1. However, Table 2 shows that HTER of SSIM for UBEAR database is not the minimum. Therefore, we cannot make a decision according to the result of one specific IQM. Additionally, the result of one specific IQM may not give the best result for every database.

On the other hand, in order to show the effectiveness of the proposed method, fusion of the features that are extracted by using 25 IQMs is analyzed. The experimental results using feature-level fusion (FLF) method and our proposed decision-level fusion (DLF) method are demonstrated in the last two rows of Tables 1 and 2. The proposed method achieves better performance compared to the feature-level fusion of 25 IQMs.

Table 1: Results (in percentage) for AMI database

#	Type	IQM	AMI database		
			FFR	FGR	HTER
1	FR	MSE	24.00	13.00	18.50
2	FR	AD	23.00	24.00	23.50
3	FR	GME	26.00	18.00	22.00
4	FR	GPE	2.00	8.00	5.00
5	FR	LMSE	9.00	4.00	6.50
6	FR	MD	20.00	26.00	23.00
7	FR	NAE	18.00	26.00	22.00
8	FR	NCC	27.00	15.00	21.00
9	FR	PSNR	24.00	13.00	18.50
10	FR	RMD	28.00	21.00	24.50
11	FR	SNR	24.00	15.00	19.50
12	FR	SME	24.00	8.00	16.00
13	FR	SPE	23.00	15.00	19.00
14	FR	SC	24.00	15.00	19.50
15	FR	SSIM	7.00	2.00	4.50
16	FR	TCD	20.00	31.00	25.50
17	FR	TED	25.00	28.00	26.50
18	FR	VIF	27.00	33.00	30.00
19	FR	RRED	16.00	15.00	15.50
20	FR	MAMS	24.00	19.00	21.50
21	FR	MAS	29.00	24.00	26.50
22	NR	JQI	10.00	3.00	6.50
23	NR	HLFI	24.00	27.00	25.50
24	NR	BIQI	9.00	2.00	5.50
25	NR	NIQE	21.00	13.00	17.00
Our study		FLF of 25 IQMs	24.00	8.00	16.00
Proposed Method		DLF of 25 IQMs	11.00	6.00	8.50

Table 2: Results (in percentage) for UBEAR database.

#	Type	IQM	UBEAR database		
			FFR	FGR	HTER
1	FR	MSE	18.00	23.00	20.50
2	FR	AD	24.00	26.00	25.00
3	FR	GME	21.00	20.00	20.50
4	FR	GPE	24.00	26.00	25.00
5	FR	LMSE	23.00	24.00	23.50
6	FR	MD	20.00	25.00	22.50
7	FR	NAE	22.00	21.00	21.50
8	FR	NCC	23.00	24.00	23.50
9	FR	PSNR	18.00	23.00	20.50
10	FR	RMD	18.00	29.00	23.50
11	FR	SNR	25.00	27.00	26.00
12	FR	SME	20.00	15.00	17.50
13	FR	SPE	14.00	29.00	21.50
14	FR	SC	25.00	27.00	26.00
15	FR	SSIM	23.00	14.00	18.50
16	FR	TCD	3.00	39.00	21.00
17	FR	TED	21.00	28.00	24.50
18	FR	VIF	24.00	2.00	13.00
19	FR	RRED	25.00	28.00	26.50
20	FR	MAMS	26.00	26.00	26.00
21	FR	MAS	12.00	23.00	17.50
22	NR	JQI	14.00	25.00	19.50
23	NR	HLFI	26.00	22.00	24.00
24	NR	BIQI	19.00	18.00	18.50
25	NR	NIQE	17.00	31.00	24.00
Our study		FLF of 25 IQMs	20.00	15.00	17.50
Proposed Method		DLF of 25 IQMs	9.00	22.00	15.50

The feature-level fusion method shows a HTER of 16.00% and 17.50% on AMI and UBEAR ear databases, respectively. However, the proposed decision-level fusion method demonstrates a HTER of 8.50% and 15.50% on AMI and UBEAR ear databases, respectively which shows the superiority of the proposed method over the feature-level fusion method on both databases.

VII. CONCLUSION

In this paper, spoof detection of ear biometrics is focused under print attacks. Our method is based on decision-level

fusion of different types of full-reference and no-reference image quality measures. According to the comparisons of our proposed method with individual image quality measures and feature-level fusion of these metrics, the proposed method achieves better results on both AMI and UBEAR databases. Half Total Error Rates of the proposed decision-level fusion method on AMI and UBEAR datasets used in the experiments are 8.5% and 15.5%, respectively whereas the Half Total Error Rates of the feature-level fusion approach are 16% and 17.5% on the aforementioned datasets under print attacks. Consequently, the proposed decision-level fusion method outperforms the feature-level fusion approach under print attacks on both ear datasets.

REFERENCES

[1] J. Galbally, S. Marcel and J. Fierrez, "Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process*, vol. 23, no. 2, pp. 710–724, 2014.

[2] M. Farmanbar and Ö. Toygar, "Spoof detection on face and palmprint biometrics," *Signal, Image and Video Processing*, vol. 11, no. 7, pp. 1253-1260, October 2017.

[3] P. Pravallika and K. S. Prasad, "SVM Classification for Fake Biometric Detection Using Image Quality Assessment: Application to iris, face and palmprint," *International Conference on Inventive Computation Technologies*, pp. 1-6, 2016.

[4] J. Galbally, F. Alonso-Fernandez, J. Fierrez and J. Ortega-Garcia, "A high performance fingerprint liveness detection method based on quality related features," *Future Generation Computer Systems*, vol. 28, pp. 311-321, 2012.

[5] J. Galbally, J. Ortiz, J. Fierrez and J. Ortega-Garcia, "Iris Liveness Detection Based on Quality Related Features," *5th IAPR International Conference on Biometrics*, pp. 271-276, 2012.

[6] A. Malhotra and R. Gupta, "Iris anti-spoofing under varying illumination conditions," *1st India International Conference on Information Processing*, pp. 1-6, 2016.

[7] X. Li, W. Bu, X. Wu, "Palmprint Liveness Detection by Combining Binarized Statistical Image Features and Image Quality Assessment," *Lecture Notes in Computer Science*, vol. 9428. Springer, 2015.

[8] D. Aishwarya, M. Gowri and R. K. Saranya, "Palm print recognition using liveness detection technique," *Second International Conference on Science Technology Engineering and Management*, pp. 109-114, 2016.

[9] C. Sousedik and C. Busch, "Presentation attack detection methods for fingerprint recognition systems: a survey," *IET Biometrics*, vol. 3, no. 4, pp. 219-233, December 2014.

[10] Ö. Toygar, E. Alqaralleh and A. Afaneh, "Symmetric Ear and Profile Face Fusion for Identical Twins and Non-Twins Recognition," *Signal, Image and Video Processing*, DOI: 10.1007/s11760-018-1263-3, 2018.

[11] E. Alqaralleh and Ö. Toygar, "Ear Recognition Based on Fusion of Ear and Tragus under Different Challenges," *International Journal of Pattern Recognition and Artificial Intelligence*, DOI: 10.1142/S0218001418560098, 2018.

[12] Ž. Emersic, V. Štruc, P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26-39, 2017.

[13] I. Omara, F. Li, H. Zhang, W. Zuo, "A novel geometric feature extraction method for ear recognition," *Expert Systems with Applications*, vol. 65, pp. 127-135, 2016.

[14] K. A. Nixon, V. Aimale, R. K. Rowe, "Spoof Detection Schemes," *Handbook of Biometrics*, Springer, 2007.

[15] İ. Avcibas, "Image Quality Statics and Their Use in Steganalysis and Compression," *Ph.D. dissertation*, Dept. Elect. Eng., Bogazici Univ., Istanbul, Turkey, 2001.

[16] I. Avcibas, B. Sankur, K. Savood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 11, no. 2, pp. 206-223, 2002.

[17] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959-2965, Dec 1995.

[18] A. Liu, W. Lin and M. Narwaria, "Image Quality Assessment Based on Gradient Similarity," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1500-1512, April 2012.

[19] Z. Wang, A. C. Bovik, H. Rahim Sheikh and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

[20] H. R. Sheikh and A. C. Bovik, "Image Information and visual quality," *IEEE Transactions on Image Processing*, vol. 15, no.2, pp. 430-444, February 2006.

[21] R. Soundararajan and A. C. Bovik, "RRED Indices: Reduced Reference Entropic Differencing for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 21, no.2, pp. 517-526, February 2012.

[22] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-Reference perceptual quality assessment of JPEG compresses images," *International Conference on Image Processing*, vol.1, pp. 477-480, 2002.

[23] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," *International Workshop on Quality of Multimedia Experience*, San Diego, CA, pp. 64-69, 2009.

[24] A. K. Moorthy and A. C. Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350-3364, December 2011.

[25] L. Zhang, L. Zhang and A. C. Bovik, "A Feature-Enriched Completely Blind Image Quality Evaluator," *IEEE Transactions on Image Processing*, vol. 24, no. 8, pp. 2579-2591, August 2015.

[26] AMI Ear Database. Available: http://www.ctim.es/research_works/ami_ear_database.

[27] R. Raposo, E. Hoyle, A. Peixinho, and H. Proença. "UBEAR: A dataset of ear images captured on-the-move in uncontrolled conditions", *IEEE Workshop on Computational Intelligence in Biometrics and Identity Management*, Paris, France, pp. 84-90, 2011.

Data Security on Virtual Private Networks

S.BÖGE[1] and A.ÖZTÜRK[2]

[1] KTO Karatay University, Konya/Turkey, Seyitboge@gmail.com
[2] KTO Karatay University, Konya/Turkey, ali.ozturk@karatay.edu.tr
Havelsan A.Ş., Ankara/Turkey, aliozturk@havelsan.com.tr

Abstract **- Network security has become one of the most important issues in today's world. Along with the widespread use of the Internet, corporations and companies share important confidential information over networks. The protection of this information, which may lead to serious harm if third-parties have access to it, is vital. Moreover, with increased cyber attacks, corporate or private networks are under serious threat. Increasing the security of networks is of vital importance, because corporate networks are especially critical to national security. In this study, necessary protocols, equipment, technologies and necessary precautions have been investigated in order to make data communication in virtual private networks, which is one of the most used network technologies today, to be done safely.**

Keywords **- Virtual private network, Data security, Encryption, Tunneling, Verification.**

I. INTRODUCTION

Network interception emerged in 1962 with the request for communication between computers in different locations. It was first implemented in 1965 by means of communication between two computers. As a result of the widespread use of the Internet and increased security breaches, the Point-to-Point Tunneling Protocol (PPTP) was developed by Microsoft in 1996, laying the foundations of today's virtual private networks. Today, large corporations and some individual users who care about privacy use virtual private networks. The main purpose of a virtual private network is to provide confidentiality and security in the communication between two points. In this study, it is aimed to examine data security measures in virtual private networks and to offer new suggestions.

II. LITERATUR SEARCH

There have been many studies on network security in recent years. The studies that describe the measures to be taken against emerging threats are helping to keep defense techniques up to date.

In the study conducted by Zeynep YÜKSEL, all components related to network security were examined and the firewall, VPN and NAT applications were examined. General marking concepts in study, modulation techniques and transmission environments, TCP / IP and layer security in local area networks are examined. Security policies, security wall architectures and components related to the management of every component that concerns the security of the computer network have been investigated and the measures taken on network devices have been investigated and the necessity of reviewing security as a whole has been emphasized. In the conclusion, the importance of network security is emphasized. [1]

In the study conducted by İlker SÖGÜT, the virtual private network setup is examined using the MPLS method, a virtual private network is established between the works and the fields, and the communication with the labeling method is started. With the developed method, MPLS is used to solve the problems of known routing protocols in virtual private network construction between domains. At the end of the study, a label was generated for each network definition in the routing table in each router for the inter-domain virtual private network in the network topology prepared as the application. Thus, the routing was performed using the multi-protocol label switching between the created private virtual network and the private virtual network structure was realized in accordance with the study. [2]

In the study conducted by Yunus Emre SEYYAR, "Top of site virtual private network" topology was prepared using GNS3 simulation application. Prepared topologies are made to work with virtual computers using VMware virtual application. In both applications, there was no difference in terms of device costs. However, when the performance is examined in detail, it is observed that the dynamic multi-point virtual private network configuration is more easily and less than the virtual private network configuration and relieves the general network traffic. As a result, for multi-point users, the virtual private network application from the site to the site would benefit from the multi-point virtual private network application. [3]

The study by Taha ALJADIR focuses on identifying system security problems for various types of attacks against computer networks. The study examines the remote access with VPN on site and the attitude of the VPN against attack. Two types of VPN network, SSL and PPTP, have been installed and different types of attacks have been implemented to observe security vulnerabilities. Recommendations for security vulnerabilities of these protocols have been developed. [4]

Analysis and Application of Remote Access Computer Communication in the work done by Mouath SALIM. General information about virtual private networks has been examined in the study. Information about network installation methods has been provided, and three different types of remotely accessed Virtual Private network protocols (PPTP, SSL and 2TP / IPSec) have been virtually provided for use with VMware and GNS3 tools. The performance of these protocols was tested by measurement tests performed under similar conditions. [5]

In a study by Shaneel Narayan et al., The performance of

virtual private network protocols was evaluated on Windows 2003 operating system. Bandwidth and CPU utilization times of IPSec, PPTP and SSL tunneling protocols were measured in the study, and it was observed that the VPN server changed its CPU performance depending on the tunneling protocol to be selected. [6]

In the study conducted by Zhao Aqun and colleagues, virtual private network protocol features were investigated, and the pros and cons of the protocols were determined. Tunnel protocols configurations and installation are explained. It provides recommendations on the management and maintenance of tunnel protocols. [7]

In the above studies there are different methods for securing virtual private networks. The purpose of this study is to provide data flow over SSL-TLS using SSTP protocol for data security in virtual private networks and to investigate the contributions of these protocols to data security and to determine the methods to be used with test results on protocols.

III. VIRTUAL PRIVATE NETWORK PROTOCOL

A. Point-to-Point Protocol (PPP)

The Data Link Layer protocol, PPP, provides bidirectional communication between two points. The development of the PPP protocol, the compression and editing features of the SLIP protocol, has emerged as a result. Both protocols can establish a serial connection to a modem or other network device. While PPP is a standard Internet protocol for these protocols, the SLIP protocol is not a standard protocol. PPP protocol continues to be used today with data compression, authentication, addressing and error correction features. PPP, in contrast to SLIP, also works with synchronous lines in addition to asynchronous lines [8].

B. Point-to-Point Tunneling Protocol (PPTP)

It is a networking protocol that enables a remote client to transfer data securely to a remote server by creating a virtual private network on TCP/IP-based data networks, using a client-server design running on Layer 2 of the OSI model. A PPTP tunnel provides communication using the TCP port 1723. Today, there are many reasons why PPTP is one of the most popular and widely used VPN protocols. PPTP is the first VPN protocol supported by Windows, and it is one of the biggest factors that every Microsoft Windows operating system supports from 1995 that it supports this protocol. Other than Windows, operating systems such as Linux and OS X have begun to support this protocol. Today, almost every desktop and mobile platform supports PPTP [9].

C. Layer 2 Tunneling Protocol (L2TP)

The second layer tunneling protocol is a tunneling protocol that supports encryption of the protocol traffic developed by Cisco and subsequent point-to-point data transfer. It does not contain any privacy or encryption within itself; It utilizes the ipsec protocol, which provides information transfer via tunnel structure. Because it uses the UDP 500 port, it can sometimes have problems with the firewall. L2TP, PPTP, and L2F, and collects the best properties of these two protocols on their own. L2TP uses the Internet Protocol (IPSec) security Encapsulating Security Payload (ESP) for encryption. Encapsulation of L2TP packets consists of two layers. The first layer is encapsulated with the PPP frame L2TP and UDP header. The second layer is encapsulated with the IPsec authentication footer and the IP header. The IP header contains the VPN server IP address. [6]

D. Secure Socket Tunnel Protocol (SSTP)

The Secure Socket Tunneling Protocol (SSTP) is a tunneling protocol that allows traffic to pass through firewalls that prevent PPTP and L2TP / IPSec traffic. SSTP provides a mechanism for encapsulating the PPP traffic over the SSL channel of the HTTPS protocol. The use of PPP provides support for strong authentication methods such as EAP-TLS. The use of HTTPS flows over the TCP 443 port, and this connection is typically a port used for Web access. Secure Sockets Layer (SSL) provides transport-level security with advanced key negotiation, encryption and integrity checking. [10]

IV. PERFORMANCE RESULTS

The RC4 algorithm is still one of the oldest ciphers used in TLS today. RC4, a streaming code, is very useful for performance on small systems. It is fast but weak for security reasons.

The most commonly used algorithm is AES. There are 3 types of AES, 128, 192 and 256 bits. At present, the most common use of 128 bits is to spread the view that 128 bits do not need the necessary cost for the 256 bits that provide adequate security.

TripleDES or 3DES is an older utility developed for Windows XP. Since it is not as powerful as AES, it is recommended that 3DES be placed behind AES in the form of a symmetric password preference. Figure 1 shows the results for the data lengths of 1 MB coded by different encryption algorithms.

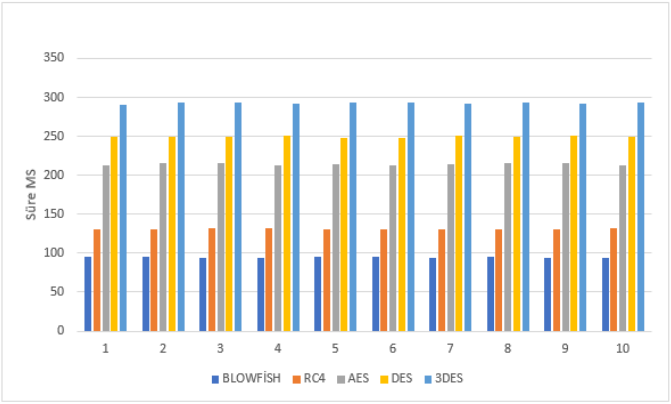


Figure 1: 1 MB Data File Encryption Times.

In Figure 1, although there are fluctuations in the measurement made on the 1 MB data, the order of the algorithms does not change. Blowfish algorithm gives the best performance according to the measurement results. This algorithm follows RC4, AES, DES and 3DES algorithms respectively.

Hash algorithms are known as One-Way Algorithms. The data encoded with the one-way algorithm can not be reverted back to the original. Encryption algorithms can convert the generated result back to its original form. Hash algorithms can not convert the generated result to its original form. The contribution of these algorithms to data security is that during the data transfer between two points, the sender receives the file's hash code, receives the hash code again when the recipient receives the file. Then checks if the first hash code and comparison code are same. If even a single letter is changed from say a 1 GB data file, the hash code also changes.

MD5 has lost credibility in recent years. SHA takes the place of MD5 algorithms. As the number of bits increases, the algorithm is more secure (as long as it is in the same algorithm family). SHA1 outputs 160 bits long. SHA2 can output 256,384 and 512 bits long. It can be said that SHA2 provides better protection than SHA1.

In Figure 2, the hash code generation times of different algorithms for a 1 MB data file are given.

Figure 2: 1 MB Data File Hash Code Generation Times.

The Point to Point Tunneling Protocol has become the standard protocol for virtual private networks. Since it is the first VPN protocol supported by Windows, it is supported by many platforms and installation is very easy. By underestimating the protocol, its operation makes it one of the fastest protocols available. Although data communication is done with 128-bit encryption, there are serious security vulnerabilities.

The Layer 2 Tunneling Protocol, unlike other VPN protocols, does not provide any privacy or encryption of traffic flow. For this reason, it is usually implemented with a protocol suite known as IPSec for encrypting the data before transmission and securing privacy and security for the user. Most of today's platforms support L2TP / IPSec protocol. Although the installation of the protocol is as quick and easy as PPTP, some firewalls can cause problems due to UDP 500 port usage. There is no significant security vulnerability in IPSec cryptography, and it can provide security when properly applied. Since the protocol is encapsulated twice, the rate of operation is lower than the others.

The Secure Socket Tunneling Protocol is supported by Windows and Linux. Because it uses SSL and TLS, it does not cause security wall problems. Mainly because it is a Windows specific protocol, it is stable and easy to use. SSTP is very secure in terms of security as it supports multiple authentication methods with passwords, smart cards, and certificates.

The download speeds of Standard Connection, L2TP, PPTP and SSTP for 25 trial sessions are given in Figure 3, Figure 4, Figure 5 and Figure 6, respectively.

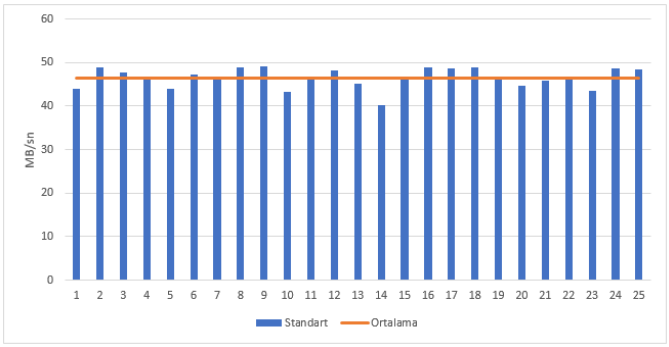


Figure 3: Standard Connection Download Speed MB \ sec.

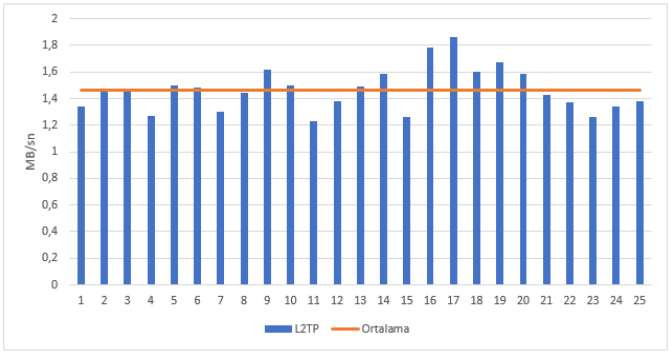


Figure 4: L2TP Connection Download Speed MB \ sec.

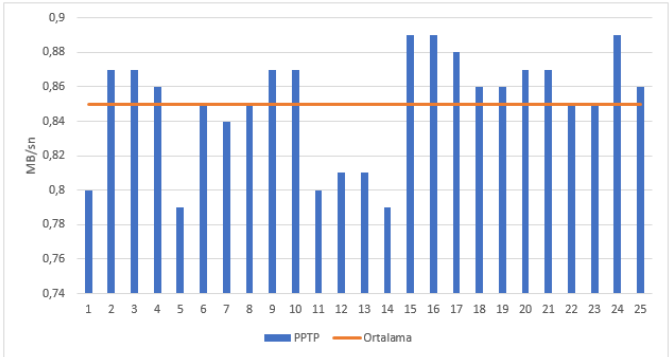


Figure 5: PPTP Connection Download Speed MB \ sec.

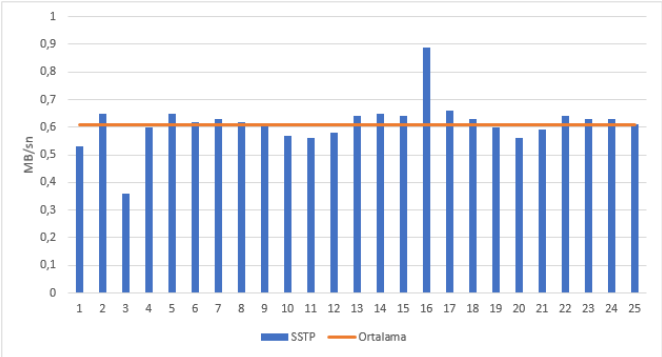


Figure 6: SSTP Connection Download Speed MB \ sec.

Table 1: Mean and Standard Deviation Values of Tunnel Protocols in MB / sec.

	Mean Value	Standard Deviation
Standard	46,54	2,274
L2TP	1,46	0,163
PPTP	0,85	0,032
SSTP	0,61	0,083

As can be seen in Table1, the connection speed decreases about 98% when the tunnel protocols are used. L2TP gave the best performance in terms of the connection speed. This protocol is followed by PPTP and SSTP respectively.

V. CONCLUSION

Today, almost everywhere in our lives, the importance of internet is increasing day by day. Many public internet services are now being used from banking transactions to government agencies. The internet, which has entered into our lives, is increasing the appetite of malicious people. Security measures are being developed to protect our personal and corporate information against third parties. The most popular of these security measures are virtual private networks. In our country, government departments and private companies use VPN service for data transfer within their branches. We need to securely handle our business on the Internet and protect our data privacy.

In this study, a virtual private network is set up with Softether application. Tunneling protocols have been tested on the installed virtual private network. As a test result, L2TP gave the best result at the download speed and this protocol was followed by PPTP and SSTP, respectively. The order of the test results for the loading speed was observed as PPTP, L2TP and SSTP from best to worst. The L2TP protocol, which was the first in the download test, dropped to second place in the load test. The reason for this is that L2TP reduces the speed of operation because the data is encapsulated twice. According to the test results and the investigations, the SSTP protocol has been used more steadily than the others and the use of this protocol is due to the use of up-to-date technologies for data security.

SSL-TLS uses encryption sets for security. Encryption sets come in a variety of algorithms and protocol configurations. In encryption sets, key exchange algorithm, certificate key, data encryption algorithm and summarization algorithm are standard. In my study, I found that the DH and ECDHE algorithms provide sufficient security in today's conditions, that the Diffie-Hellman algorithm gives about 10 times better results than the Diffie-Hellman algorithm and Elliptic curve variant, and provides sufficient security without overloading the hardware components. There is a similar situation in the certificate key. It has been observed that ECDSA performs better than RSA. In encryption algorithms, symmetric encryption algorithms do not provide sufficient security against the best time Blowfish algorithm according to asymmetric cryptography algorithms, according to tests on symmetric encryption algorithms with the reason that it provides sufficient security without using system resources too much. The AES algorithm is good both in terms of time and security, and because of its widespread use today, it is the result of using this algorithm in data encryption.

As a result, deciding which hardware, software and methods are required for data security in virtual private networks and how to use these tools will provide benefits in terms of data security.

REFERENCES

[1] Zeynep YÜKSEL, 2007, Network Security and Firewall VPN and NAT Applications, Graduate thesis, Yıldız Teknik University Graduate School of Natural and Applied Sciences, İstanbul
[2] İlker SÖGÜT, 2009, Multi-Domain Virtual Private Network (VPN) Setup Using Multi-Protocol Label Switching (Mpls), Graduate thesis, Muğla University Graduate School of Natural and Applied Sciences, Muğla
[3] Yunus Emre SEYYAR, 2013, Virtual Private Network and Dynamic Multipoint Virtual Private Network Comparative Survey on Site, Graduate thesis, Kırıkkale University Graduate School of Natural and Applied Sciences, Kırıkkale
[4] Taha ALJADIR, 2015, Measurement Of System Securıty Issues Of Prıvate Computer Networks For Dıfferent Types Of Attacks, graduate thesis, Çankaya University Graduate School of Natural and Applied Sciences, Ankara
[5] Mouath SALIM, 2015, Analysıs And Implementatıon Of Remote Access Computer Communıcatıon, graduate thesis, Çankaya University Graduate School of Natural and Applied Sciences, Ankara
[6] Shaneel Narayan, Samad S. Kolahi, Kris Brooking, Simon de Vere, Performance Evaluation of Virtual Private Network Protocols in Windows 2003 Environment, UNITEC New Zealand
[7] Zhao Aqun, Yuan Yuan, Ji Yi, Gu Guanqun, Research on tunneling techniques in virtual private networks, 21-25 Aug. 2000
[8] "Point to Point Protocol (Noktadan Noktaya Protokolü)" access address: http://bidb.itu.edu.tr/seyirdefteri/blog/2013/09/07/point-to-point-protocol-(noktadan-noktaya-protokolü) access date: July 2017
[9] "Understanding PPTP", access address: https://technet.microsoft.com/en-us/library/cc768084.aspx, access date June 2017
[10] "SSTP Remote Access Step-by-Step Guide: Deployment" access address: https://technet.microsoft.com/en-us/library/cc731352.aspx access date: July 2017

Creating of Probability Maps of Earthquake Occurrences Using Kriging Method With the Geographic Information Systems (GIS): Estimates for 3 Section of the NAFZ (Western, Central, Eastern)-Part 2

Tuğba TÜRKER[1] and Yusuf BAYRAK[2]
[1] *Geophysical Engineering, Karadeniz Technical University, Trabzon, tugbaturkerktu@gmail.com*
[2] *Geophysical Engineering, Karadeniz Technical University, Trabzon ,bayrak@ktu.edu.tr*

Abstract - **In this study, we created probability maps of earthquake occurrences using Kriging method from Geostatistical techniques with the Geographic Information Systems (GIS) in the NAFZ (western, central, eastern). Geostatistical techniques had both the capability of producing a prediction surface and provide some measure of the certainty or accuracy of the predictions. Kriging method depended on mathematical and statistical models. Kriging was an interpolation that can be exact or smoothed depending on the measurement error model. Kriging used for statistical models that allow a variety of output surfaces including probability. We used an instrumental catalog for $M_S \geq 4.0$ magnitude between 1900-2017 period. Additionally, Kriging method fitted a mathematical function to a specified number of earthquakes or all earthquakes within a specified radius, to determine the output value for each region. We have used $Z(s) = \mu(s) + \varepsilon(s)$ basis formula for all the different types of Kriging method. We applied polygonal declustering method for the database. We used probability Kriging type for estimations of probability output surface. We selected to approximation method both Multiplicative Skewing type and Gamma distribution. After, we estimated spatial covariance variables and selected K-Bessel model. As a result, probability maps of earthquake occurrences created with Kriging method in the NAFZ especially to western NAFZ, central NAFZ and eastern NAFZ. We determined the high and low probability estimates of earthquakes with Kriging method for 3 sections of the NAFZ. This study, the scientists will be lead to probabilities estimate of earthquake occurrences in Turkey.**

Keywords **- Kriging method, The Geographic Information Systems (GIS), The NAFZ (Western-Central-Eastern)**

1. INTRODUCTION

The North Anatolian Fault Zone (NAFZ) was 1200-km long strike-slip fault which estabilished right-lateral motion of the Anatolian plate notional to the Eurasian plate between eastern Anatolia and the northern Aegean Sea (Barka and Kadinsky-Cade, 1988). The NAFZ was a tight structure along its eastern and central sections, but, it two branchs into several edges in the eastern Marmara region (McKenzie, 1972). These faults were important and easily definable based on the different topography along the fault strands, seismic activity, and GPS-reproduce slip rates (McClusky et al., 2000).

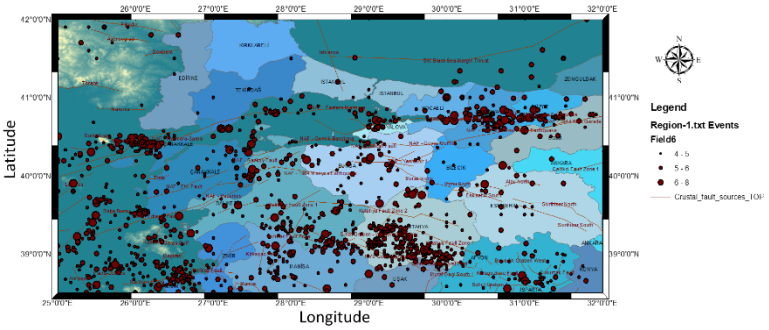


Figure 1. The epicentr distribution and tecnonic map plotted in the Western NAFZ.

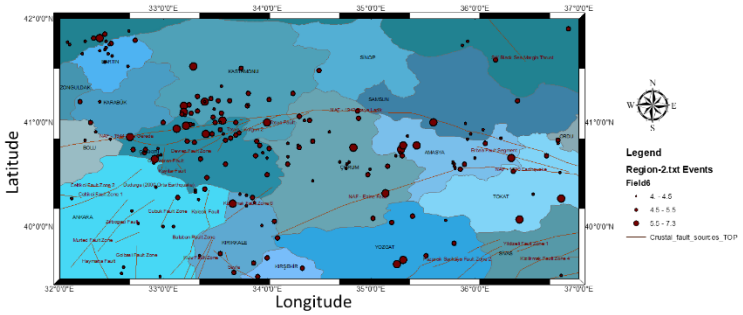


Figure 2. The epicentr distribution and tecnonic map plotted in the Central NAFZ.

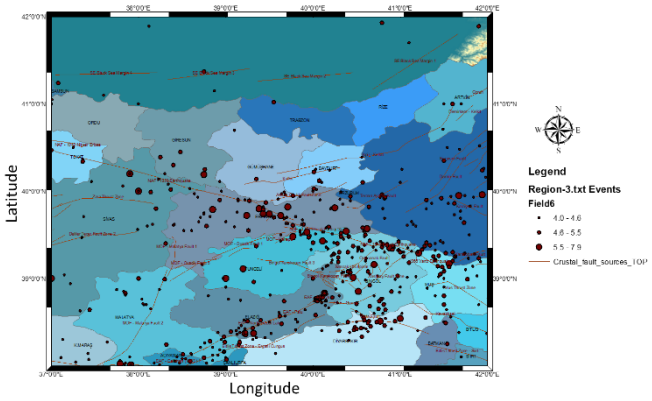


Figure 3. The epicentr distribution and tecnonic map plotted in the Eastern NAFZ.

We created probability maps of earthquake occurrences using Kriging method from Geostatistical techniques with the Geographic Information Systems (GIS) in the NAFZ (Fig. 4, 5 and 6). So, we divided into 3 sections of the NAFZ (Western, Central, Eastern) (Fig. 1, 2 and 3). Earthquake catalogue contained longitude, latitude, depth, historic, hour, magnitude, etc. pieces of information. All analysis estimated for $M_s \geq 4.0$ earthquake occurrences taking into account variables of cutoff magnitudes (M_c) of the catalogue in 1900-2017 period.

2. Kriging Method

The kriging method of interpolation was one of from geostatistical technique. İt considered both the distance and the degree of variation among known data points when predicting values in undetermine areas. In kriging, the first step determined the data with identify the spatial structure. Also, it often represented by the empirical semivariogram (Isaaks and Srivastava, 1989). The BLUE of unknown realization *Un'* denoted as U~, is constructed linearly in terms of *(n* - 1) known realizations Joumel and Huijbregts (1989). This process of evaluating BLUE is the original meaning of kriging. Formulation may also be presented in the estimator form involving corresponding stochastic variates:

$$U_n^* = \sum_{i=1}^{n-1} \lambda_{in} U_i$$

The kriging weights are determined based on the unbiased condition:

$$E(U_n^* - U_n) = E\left(\sum_{i=1}^{n-1} \lambda_{in} U_i - U_n\right) = \sum_{i=1}^{n-1} \lambda_{in}\mu - \mu = 0$$

and on the minimum estimation variance:

$$E[(U_n^* - U_n)^2] = E\left[\left(\sum_{i=1}^{n-1} \lambda_{in} U_i - U_n\right)^2\right]$$

$$= \sum_{i=1}^{n-1}\sum_{j=1}^{n-1} \lambda_{in}\lambda_{jn} R_{ij} - 2\sum_{i=1}^{n-1} \lambda_{in} R_{in} + R_{nn}$$

where = denotes "by definition;" and *[K]* =kriging matrix. When the stochastic process has zero mean:

$$\begin{Bmatrix} \lambda_{1n} \\ \lambda_{2n} \\ \vdots \\ \lambda_{(n-1)n} \\ \gamma_n/2 \end{Bmatrix} = [\mathbf{K}]^{-1} \begin{Bmatrix} R_{1n} \\ R_{2n} \\ \vdots \\ R_{(n-1)n} \\ 1 \end{Bmatrix}$$

$$= \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1(n-1)} & 1 \\ R_{12} & R_{22} & \cdots & R_{2(n-1)} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ R_{1(n-1)} & R_{2(n-1)} & \cdots & R_{(n-1)(n-1)} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \begin{Bmatrix} R_{1n} \\ R_{2n} \\ \vdots \\ R_{(n-1)n} \\ 1 \end{Bmatrix}$$

Minimum estimation variance known as the kriging variance, for the nonzero mean stochastic process:

$$\sigma_k^2 = \min\{E[(U_n^* - U_n)^2]\} = R_{nn} - \sum_{i=1}^{n-1} \lambda_{in} R_{in} - \frac{1}{2}\gamma_n$$

and for the zero mean stochastic process:

$$\sigma_k^2 = \min\{E[(U_n^* - U_n)^2]\} = R_{nn} - \sum_{i=1}^{n-1} \lambda_{in} R_{in}$$

3. Results

Kriging method fitted a mathematical function to a specified number of earthquakes or all earthquakes within a specified radius, to determine the output value for each region. We have used $Z(s) = \mu(s) + \varepsilon(s)$ basis formula for all the different types of Kriging method. We applied polygonal declustering method for the database. We used probability Kriging type for estimations of probability output surface. We selected to approximation method both Multiplicative Skewing type and Gamma distribution. After, we estimated spatial covariance variables and selected K-Bessel model. As a result, probability maps of earthquake occurrences created with Kriging method in the NAFZ especially to western NAFZ, central NAFZ and eastern NAFZ. We determined the high and low probability estimates of earthquakes with Kriging method for 3 sections of the NAFZ. This study, the scientists will be lead to probabilities estimate of earthquake occurrences in Turkey.

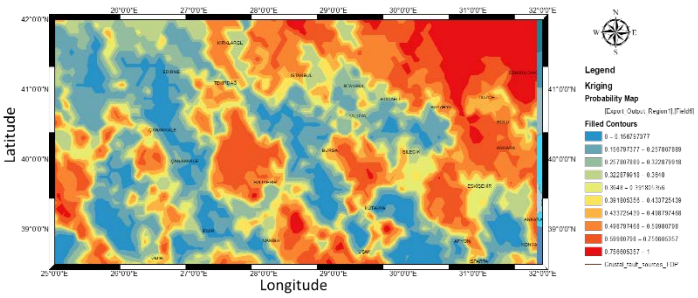


Figure 4. The Kriging method probability maps plotted in the Western NAFZ.

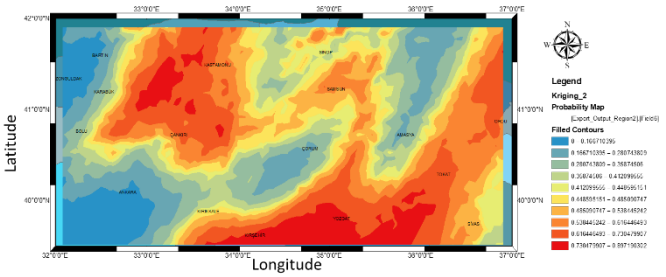


Figure 5. The Kriging method probability maps plotted in the Central NAFZ.

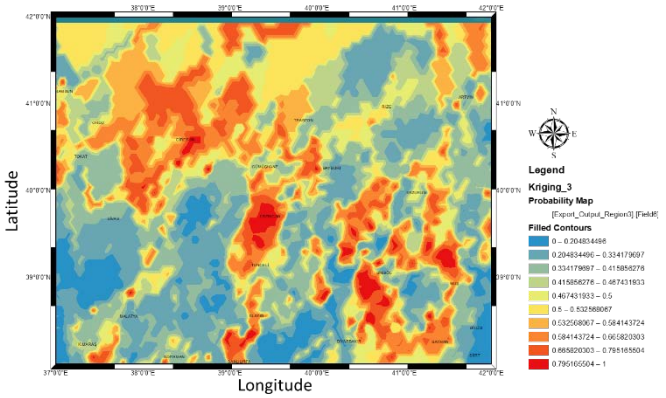


Figure 6. The Kriging method probability maps plotted in the Eastern NAFZ.

REFERENCES

[1] Barka A.A. Kandisky-Cade K., 1988. Strike-slip fault geometry in Turkey and its influence on earthquake activity, Tectonics , 7, 663–684.

[2] Isaaks, E. and Srivastava, R. (1989): An Introduction to Applied Geostatistics, Oxford University Press Inc., New York, 561 p.

[3] Journel, A. G. and Huijbregts, C. J. (1978): Mining Geostatistics, Academic Press, London, UK, 600 p.

[4]McClusky, S., Balassanian, S., Barka, A., Demir, C., Ergintav, S., et al., 2000. Global Positioning System constraints on plate kinematics and dynamics in the eastern Mediterranean and Caucasus. J. Geophys. Res. 105, 5695–5719

[5]McKenzie, D.P., 1972. Active tectonics of the Mediterranean region. Geophys. J. R. Astron. Soc. 30, 109–185.

A study on obtaining rectified photographs for architectural applications

A.SOYCAN[1] and M.SOYCAN[1]

[1] Yildiz Technical University, Istanbul/Turkey, topbas@yildiz.edu.tr
[1] Yildiz Technical University, Istanbul/Turkey, soycan@yildiz.edu.tr

Abstract **- In this study, geometric methods used to obtain rectified photographs that will form a base for drawings of building facades are examined. Mathematical models are presented by informing briefly about the methods evaluated.**

The subject is exemplified by a test study. In the test study, firstly, the behavior of the related methods on the created artificial image was examined and the applications on the real photographic sets of the models obtained with different types and different cameras were performed.

The obtained results were evaluated statistically and various conclusions were drawn and the suggestions were made for those who want to apply the method and want to do the study in this subject.

Keywords **– Photography, Architectural Surveying, Photogrammetry, Image Transformation, Rectification**

I. INTRODUCTION

In architectural studies, building facade drawings is an important component of the project. For this purposes, complicated techniques such as laser scanning, orthophotography, classical or modern photogrammetric approaches and stereo drawing can be used today if a model is needed for objects with more complex details.

On the other hand, the geometrically corrected (rectified) facade photographs are a cost-effective method for drawing details on objects such as relatively flat structures, facades of building, floors, glass windows and wall paintings. According to the all approaches described above,, rectified photographs have the potential to provide much more information than the drawings obtained with other conventional approaches.

However, it can be said that the approach is only an effective method for relatively flat structures and objects. Such photographs can be used to create a multipurpose drawing and are considered a data source that can be easily read and interpreted for different disciplines in architectural works.. It also provides detailed photo recording of the object or structure and serves as an important inventory record.

Rectification can also be considered as the work of transforming images into geometrically distorted images with numerical methods. As a result of the geometric transformation, the pixel values (gray level, brightness, radiometric or color values) must be resampled to obtain the brightness values of the pixels in the new image. Various interpolation methods are used for this process. In this study, the geometric image transformations used to perform the rectification process are described with several examples.

In practical applications, standard image correction procedures are recommended for correcting oblique photographs. On the other hand, conventionally, in architectural photogrammetric applications, geodetically measured control points are needed to create an image record in the object coordinate system.The measurement of architectural works is also important when it is examined from the perspective of the historical background of architectural structures.

Generally, in the plans made within the scope of architectural works, front elevations, sections, profiles and, in some cases, the coordinates of certain regions and points of the structure are desired. Today, due to its easy and wide use, cameras are also included in other devices such as mobile phones and tablets. This digital technology has become widespread and, in a sense, an integral part of our daily lives. Digital technology, which finds itself as a concept that affects more individuals and society in everyday life, also comes with its problems as well as the solutions it creates. Today, research and development studies intensely ongoing to further develop digital photography technology. In practice, there are many factors to consider in the photo; Photo Techniques, Camera features, Lens types, Drag modes, Settings related to shooting environment, Photo Processing and Editing etc. The simplest way to edit a photo is probably to play the photos we take on or after the machine, and make adjustments as desired. This arrangement sometimes requires fine tuning, simple toning, and sometimes even more work. While basic adjustments can be made on the camera and / or via Mobile Photo Editing Applications, Photo Editing Software developed for more complex edits can be used.

II. ARCHITECTURAL PHOTOGRAPHY

Architectural photography includes all photographic techniques used to demonstrate the geometry and aesthetics of architectural structures.

In professional architectural photography, the perspective and therefore the control of horizontal architectural features, especially the aesthetic settlement, is very important for the purpose of making architectural photographs suitable and correct. The main tasks of architectural photographers are to present three-dimensional architectural themes in a realistic and impressive way in two-dimensional photography. In this respect, specializing in this area also requires having technical knowledge and experience, as well as general photographic

information. There are two different approaches to architectural photography. The first approach is that architects and photographers can gain a different view and perspective on the architectural structure by using their original perspective and photographic techniques. This approach expresses artistic side in a sense of the matter. The second approach is to document the technically important features and functions of the architectural work being drafted. On the other hand, in commercial sense, architectural photography works seem to have been applied within a broad perspective. These shootings are performed for different purposes.

One of the most important issues that should be paid attention to architectural photography is the reflection of the perspective of architectural structure and the effect of the third dimension correctly on the photo. The ability to deliver these effects in a technically correct way depends primarily on the point of view the photographer will prefer during shooting, the height of the point of view, and the distance between the subject and the camera. However, it is important in technical tools such as cameras and lenses that the photographer will use. When photographs of architectural structures are taken, the most common problem is the occurrence of deformations due to the height of the structures, the lack of suitable angled places and the use of wide angle lenses. In such a photograph, the vertical lines of the architectural structure are not parallel to the vertical edge lines of the photograph but are inclined inwards. To avoid from this mistake, you need to use a technical camera with a bellows, use a perspective-controlled (PC) lens, and shoot at an appropriate angle and height. However, with the development of digital photography technology and software today, these mistakes can also be eliminated through software.

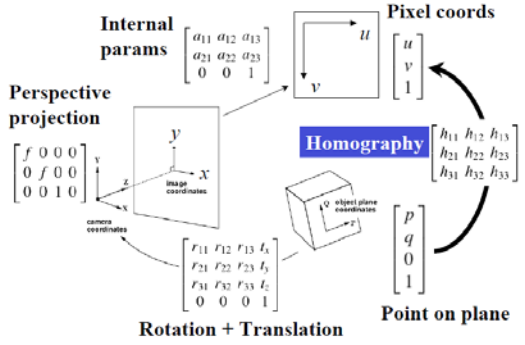


Figure 1: Relationship with coordinate systems [9].

III. IMAGE RECTIFICATION WITH GEOMETRIC TRANSFORMATION METHODS

Architectural objects are made up of a series of large planes. A single perspective view is enough to obtain a corrected and scaled view, even if the plane of the object being examined is oblique to the image or the photograph.

However, sometimes an object may appear different in appearance than it actually looks. The reason of this discrepancy is perspective deformation. Images of the same object captured from different camera distances and viewing angles may exhibit different perspective distortions. Using a wrong lens or the camera's trembling can cause the perspective in the photographs to deteriorate. As it may be a perspective deformation, vertical continuous lines or geometric shapes may become more prominent in photographs.

Distortions in the vertical lines due to the lens used in some photographs (especially wide angle) or gaze height appear to be a technical flaw. Especially in architectural photographs, this problem is solved by using a technical camera or tilt-shift (perspective control lens) lens, but these defects can also be corrected by software. In architectural photography, the fact that the vertical columns are not curved or the horizontal symmetry can be preserved is a more beautiful element in a photograph. It is also possible to arrange these perspective distortions later with Photoshop or similar photo editing programs, since it is not possible to adjust the framing in situations where the photograph needs to be drawn quickly.

Perspective rectification can be referred to as image correction to measure from digital or traditional photographs. With geometric correction it is possible to define horizontal lines and correct and scale a flat surface photograph with two scale-defined scale information.

In the geometric correction process with the reference points, it is necessary to make measurements with the topographical instruments directly for the correction and scaling of the photographs. These reference points must be defined points on the building surface or must be mounted to the surface.

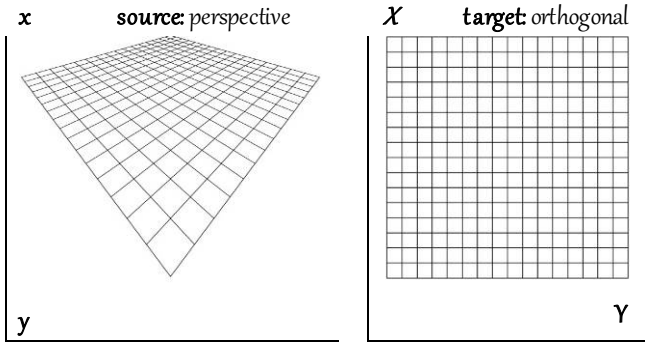


Figure 2: Perspective and orthogonal planes.

Thus, images expressed in corrected, scaled, and even geographic coordinate systems can be edited for further processing, drawing, framing or archiving, and can even be used as a background for future work on CAD software.

Although measurement with topographical instruments is not always necessary, it is one of the most ideal options for perspective correction, distortion modeling, scaling and geo-referencing. Because they can be used directly with the data they provide without the need for other procedures and applications.

IV. 2D HOMOGRAPHIES TO REMOVE PERSPECTIVE DISTORTION

The digital image correction process consists of the following steps;

1. The camera is positioned so that the architectural object surface to be documented looks as close to the center as possible

2. Correction of the building planes with the help of software on the obtained digital images.

In the literature, it is seen that various methods of rectification are used. Polynomial rectification, projective rectification and differential rectification are three commonly used methods (Novak, 1992). The first two of these techniques are performed with analytical transformations that are thought to exist between image and orthophoto, without adding the array of sensor arrays, the projection geometry, and the external orientation of the sensor systems. These are considered approximate solutions of the problem. However, these approximate solutions are sufficient for processing images of some applications.

Projective transformation is used to eliminate the perspective effect resulting from the central projection in the images. It is a popular geo-referencing technique used worldwide. It is based on quite complex geometric and mathematic concepts, known as "homogeneous coordinates" and "projective planes". An orthogonal image is obtained by removing the perspective effect in the images. That is, the object is rendered parallel to the plane as if it were obtained by parallel projection. This method is usually used to retrieve defects found in aerial photographs of flat areas and images of building surfaces, applying homographies to remove perspective distortion, homographies for bird's-eye views and homographies for mosaicking.

2D homography can be defined as a projective transformation h is an invertible mapping from P^2 to P^2 that preserves collinearity between points (x_1, x_2, x_3 on same line0 $h(x_1), h(x_2), h(x_3)$ on same line). Given a set of points x_i in IP^2 and a corresponding set of points x_i' likewise in IP^2, compute the projective transformation that takes each x_i to x_i'. In a practical situation, the points x_i and x_i' are points in two images (or the same image), each image being considered as a projective plane IP^2. In this scenario every 2D point can be projected in any other plane in the space. Based on these concepts, it is defined the "homography between 2 planes" which, simply speaking, means that given 4 points in a plane, there always exists a relationship that transforms them into the corresponding 4 points in another plane.

The homography transformation method has been used to rectify a perspective image, for example to generate a "plan" view of a building from a "perspective" photo. The transformation equation can be defined Eq. 1 for x,y source system from X,Y target for in this type of process the homography.

$$X = \frac{a.x + b.y + c}{g.x + h.y + 1}, Y = \frac{d.x + e.y + f}{g.x + h.y + 1} \quad (1)$$

If the equation (1) multiplying through by denominator, and then rearrange as equation (3).

$$\begin{aligned} a.x + b.y + c &= X.(g.x + h.y + 1) \\ d.x + e.y + f &= Y.(g.x + h.y + 1) \end{aligned} \quad (2)$$

$$\begin{aligned} a.x + b.y + c - X.g.x - X.h.y - X &= 0 \\ d.x + e.y + f - Y.g.x - Y.h.y - Y &= 0 \end{aligned} \quad (3)$$

Where X-Y are the coordinates to be calculated in the second reference system (target), given coordinates x-y in the first reference system (source) in function of 8 transformation parameters a, b, c, d, e, f, g, h. So, having these 8 unknowns, at least 4 known points in both systems are required. If the number of points is more than 4, it can be realized by solution (LSA). The matrix system of transformation equations can be defined as;

$$\begin{vmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1.X_1 & -y_1.X_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -x_1.Y_1 & -y_1.Y_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2.X_2 & -y_2.X_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -x_2.Y_2 & -y_2.Y_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3.X_3 & -y_3.X_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -x_3.Y_3 & -y_3.Y_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4.X_4 & -y_4.X_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -x_4.Y_4 & -y_4.Y_4 \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_n.X_n & -y_n.X_n \\ 0 & 0 & 0 & x_n & y_n & 1 & -x_n.Y_n & -y_n.Y_n \end{vmatrix} \cdot \begin{vmatrix} a \\ e \\ b \\ f \\ c \\ g \\ d \\ h \end{vmatrix} = \begin{vmatrix} X_1 \\ Y_1 \\ X_2 \\ Y_2 \\ X_3 \\ Y_3 \\ X_4 \\ Y_4 \\ X_n \\ Y_n \end{vmatrix} + \begin{vmatrix} v_{X_1} \\ v_{Y_1} \\ v_{X_2} \\ v_{Y_2} \\ v_{X_3} \\ v_{Y_3} \\ v_{X_4} \\ v_{Y_4} \\ v_{X_n} \\ v_{Y_n} \end{vmatrix}$$

$$\underline{A}.\underline{X} = \underline{L} + \underline{v} \quad (4)$$

- a = fixed scale factor in X direction with scale Y unchanged.
- b = scale factor in X direction proportional to Y distance from origin.
- c = origin translation in X direction.
- d = scale factor in Y direction proportional to X distance from origin.
- e = fixed scale factor in Y direction with scale X unchanged.
- f = origin translation in Y direction.
- g = proportional scale factors X and Y in function of X.
- h = proportional scale factors X and Y in function of Y.

The unknown coefficients of the models with their covariance information are simply determined according to Least Squares Adjustment (LSA) principles which is minimizing the sum of squares of the residuals v_i, as follow with the equal weights.

$$\underline{X} = \left(\underline{A}^T.\underline{A}\right)^{-1}.\left(\underline{A}^T.\underline{L}\right), \underline{v} = \underline{A}.\underline{X} - \underline{L}, \sigma_0^2 = \left(\underline{v}^T.\underline{v}\right)/(2n - r) \quad (5)$$

L is the observation vector, which consists of coordinates in target reference system, $\underline{X}$ is the estimated value of the unknown transformation coefficients vector (a, b, c, d, e, f, g, h), $\underline{A}$ is the design matrix, $\underline{v}$ is the residual vector. Once calculated, these 8 parameters can easily be used to transform any point from the first reference system to the second. σ_0^2 is the unit weight of variance, q is the number of common points used for transformation, r is the unknown parameter number (number of transformation coefficient).

V. TEST STUDY

In the test study, firstly, the behavior of the related methods on the created artificial image was examined and the applications on the real photographic sets of the models obtained with different types and different cameras were performed. The obtained results were evaluated statistically and various conclusions were drawn and the suggestions were made for those who want to apply the method and want to do the study in this subject. Software such as Photoshop ArcGIS, Surfer, Didger, Matlab have been used in the study for the purpose of editing photographs and so on. The technical details of the photographs in the datasets used in the study are given in Table 1.

Table 1: Test Data.

	Resolution	DPI	Bit depth	Camera	Focus Length
Artificial Image	3840*3840	600	24	-	-
Building facade	3072*4608	300	24	Nicon D3100	26.3mm
Consecutive photographs	2592*1728	72	24	HTC C501e	4mm

At the first stage of the test study, an artificial image formed from the square grids. The artificial image has been transformed into perspective views and the success of geometric transformation methods in correcting these perspective views has been examined. The created image is of resolution and size in standard photograph features and is defined separately in orthogonal and perspective geometries as given in Figure 2. The aim is to convert perspective view into orthogonal image with 5 defined control points. The differences between the squared positions in the transformed and original orthogonal images (Figure 3) provides a significant comparison in terms of the suitability of the transformation model.

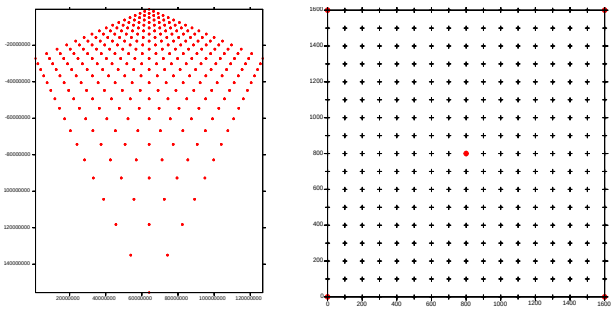


Figure 3: Grid positions in source and target coordinate systems.

For this purpose, the differences between the grid positions and the original positions were statistically analyzed. In addition, evaluations were made visually. The artificial image was created with 16 * 16 = 256 squares and 289 grid points. These points are expressed in a 2D plane coordinate system and 5 control points are selected. The remaining 284 points are accepted as comparison points. The success of the geometric transformation methods discussed in the study was explored using various variations of the points in the control point group.

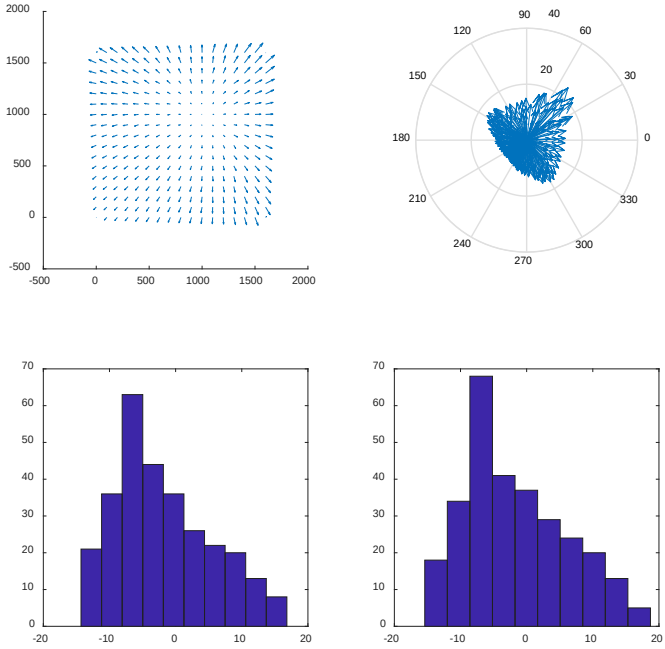


Figure 4: Differences between the grid positions in the transformed and original orthogonal images.

The vector plots for the differences between the grid positions in the transformed and original orthogonal images, the compass plots and the histograms for the distributions of the differences are given in Figure 4. The results demonstrate that the differences are maximum 16.32mm in the X-axis, minimum -14.31mm, average 6.03mm and RMS value is 7.5mm. On the Y axis, maximum 18.71mm, minimum -15.40mm, average 6.39mm, RMS value is 7.84mm (Table 2).

Table 2: The statistics for grid position differences for first data.

	X Differences (mm)	Y Differences (mm)
Number of values	284	284
Minimum	-14.31	-15.40
Maximum	16.82	18.71
Mean	-1.81	-1.57
Standard deviation	7.29	7.69
Average deviation	6.03	6.39
RMS	7.50	7.84

The RMS value obtained from statistical data can be considered as the most important value that gives an idea about the accuracy of the conversion models. The RMS is calculated by summing the mean squares of the errors for X and Y coordinate differences. RMS indicates how closely model estimates the measured values. The smaller this error, the better estimation could be performed.

The second data is the surface of a smooth building ceiling in the shape of a tile. Photos taken from different locations, different viewpoints and directions were used for this facade based on single photo resection. Since the tile sizes (about

595mm) are also visible on this surface, the scaling is done using their dimensions. The locations of the junctions of the tiles are compared by calculating the dimensions and areas of the tiles (Figure 5).

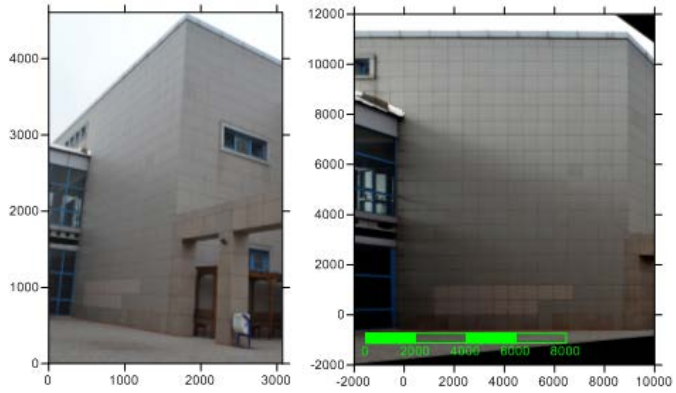


Figure 5: Original and transformed building facade images.

It has been seen that the view has been transformed quite successfully in the examinations made on the facade due to the factors related to the vertical and horizontal control points, the tiles' overlaps, the tile sizes, facades and tile surface area (Table 3). RMS value is calculated as 2.5 mm for transformed image.

Table 3: Analysis of second test data results (Units: mm).

Tile Size Comparison				
Number of Tile: 252	Min.	Max.	Avr.	Meas.*
Tile length	585.8	615.3	599.9	595
Perimeter	2365.5	2451.4	2407.6	2380
Tile areas(mm^2)	349620	375169	361468	354025
Facade Size Comparison				
Height	Left	Right	Avr.	Meas.**
	10704.2	10690.5	10697.4	10710
Length	Top	Bottom	Avr.	Meas.**
	8321.5	8336.9	8329.2	8330
Diagonal	Bottom-Left to Top-Right	Bottom-Right to Top-Left	Avr.	Meas.**
	13541.3	13531.1	13536.2	13568
Area(mm^2)	-	-	89100784.08	89214300
*Tape measurement, **Laser distance meter measurement				

As the third data of test study, a panoramic image formed from the four photographs taken from the historic barracks building located in the Yildiz Technical University Davutpasa campus was used (Figure 6). A geo-referencing approach based on the control points on the facade was used to rectify this image. The locations of the control points are determined by geodetic measurements.

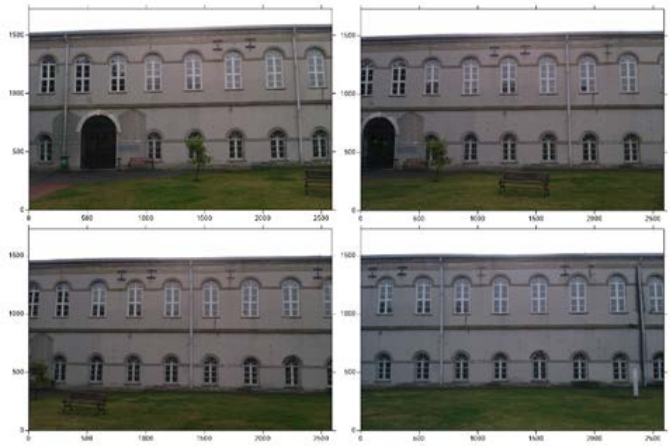


Figure 6: Consecutive photographs for historic barrack building.

Figure 7: Panoramic image formed from the consecutive photographs for historic barrack building.

The four superimposed photos taken from the inner garden of the historic barrack building were combined with the "image mosaic" approach (Figure 7). In this approach, consecutive photographs are paired with each other and combined in the coordinate system of the first photograph. The image obtained from the combined 4 photographs was transformed based on the 6 control points measured geodetically on the surface (Figure 8).

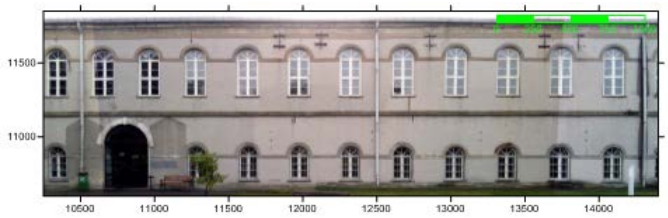


Figure 8: Geo-referenced image for historic barrack building.

The result of this transformation is that the perspective errors found in the merged image are eliminated and an orthogonal view is obtained. The resulting RMS value is around 5mm. Similar to other data sets, the edges defined in different directions and sizes on the facade were compared in terms of dimensions. In addition, the lengths between the detail points on the facade, the vertical and horizontal controls, and the results of the examinations made in the window details have resulted in a sufficient level of image in terms of architectural works (Table 4).

Table 4: Analysis of third test data results (Units: mm).

Window Size Comparison				
1.Floor	Min.	Max.	Avr.	Meas.*
Perimeter	5923.5	6029.9	5973.8	5960
Areas	22105.8	22851.4	22469.1	22550
2.Floor	Min.	Max.	Avr.	Meas.*
Perimeter	8229.5	8105.6	8171.6	8185
Areas	38089.6	37505.93	37836.6	38215
Facade Size Comparison				
Height	Left	Right	Avr.	Meas.**
	11473.3	11419.9	11413.5	11445
Length	Top	Bottom	Avr.	Meas.**
	37974.4	37983.3	37953.8	37960
Diagonal	Bottom-Left to Top-Right	Bottom-Right to Top-Left	Avr.	Meas.**
	39715.2	39689.4	39702.3	39689
Area	-	-	433185696.3	
*Tape measurement, **Laser distance meter measurement				

VI. CONCLUSION

Along with the developing computer and information technologies, conventional method has been replaced with digital photography and terrestrial photogrammetry method. This method has many advantages such as automatic directing and measuring processes, digital three-dimensional vector data, digital orthophoto, digital surface and the production of terrain models. The resultants are numerical, allowing these products to be used in different areas of application, such as three-dimensional modeling outside the photogrammetric survey, visualization of the three-dimensional data, management and presentation in the GIS environment. As an important component of terrestrial photogrammetry, the images obtained with this approach can provide significant benefits for the projects; Survey, Restitution and Restoration, Preparation of Environmental Plan, Landscape Projects and Urban Design, Street Health Projects and Preparation of Facade Rehabilitation, Preparation of photogrammetric silhouettes of straits and coastal areas, Preparation of Electrical, Mechanical, Static and Strengthening Projects of Historical Buildings, 3D City Modeling, Animation and Urban Modeling Preparation, Construction Application Services in terms of applicability, prevalence, sustainability and practicality.

REFERENCES

[1] B. Ladd, Gabriel & Nagchaudhuri, Abhijit & J. Earl, Tracy & Mitra, Madhumi & Anne, Princess & Bland, Geoff. (2006). Rectification, georeferencing, and mosaicking of images acquired with remotely operated aerial platforms.

[2] Visintini, Domenico. (2003). High-efficiency building surveys by dynamic and static digital images oriented with" mixed models".The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXIV, Part 5/W12

[3] Hamamcıoğlu Mine Turan, Mimari Fotogrametri Alanındaki Çağdaş Gelişimlerin Değerlendirilmesi, J. Fac. Eng. Arch. Gazi Univ., Cilt 19, No 1, 43-50, 2004.

[4] William B. Hockey, Scaled-Rectified Photography on Site, Bulletin of the Association for Preservation Technology, Vol. 7, No. 3 (1975), pp.37-78, Association for Preservation Technology International (APT)

[5] Agarwal, Anubhav & V Jawahar, C & J Narayanan, P. (2008). A Survey of Planar Homography Estimation Techniques.

[6] Prakash Duraisamy et all, Choosing Appropriate Homography Transformation for Building Panoramic Images, International Journal of Computer Vision and Signal Processing, 2(1), 29-37(2013)

[7] Baráth, Dániel & Hajder, Levente. (2016). Novel Ways to Estimate Homography from Local Affine Transformations.

[8] V. Tsironis, A. Tranou, A. Vythoulkas, A. Psalta, E. Petsa*, G. Karras, Automatic Rectification of Building Facades, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W3, 2017 3D Virtual Reconstruction and Visualization of Complex Architectures, 1–3 March 2017, Nafplio, Greece

[9] Collins Robert, Lecture 16: Planar Homographies, CSE486, Penn State

[10] M. S. Temiz, S. Doğan, Dijital Görüntülerin Rektifikasyonu: Sensör Modelleri, Geometrik Görüntü Dönüşümleri ve Yeniden Örnekleme, TMMOB Harita ve Kadastro Mühendisleri Odası, 10. Türkiye Harita Bilimsel ve Teknik Kurultayı, 28 Mart - 1 Nisan 2005, Ankara

[11] Alizadehashrafi, Behnam & Rahman, Alias. (2018). Enhancing Textures of 3D Building Facade.

[12] http://www.buildingconservation.com/articles/rectified/rectified.htm

[13] https://www.ldv.ei.tum.de/fileadmin/w00bfa/www/content_uploads/Vorlesung_3.2_SpatialTransformations.pdf

[14] Soycan A., "Digital cameras derived raster image transformation of old map sheets", SCIENTIFIC RESEARCH AND ESSAYS, vol.5, pp.4011-4017, 2010

[15] Soycan M., Soycan A., Tunalioğlu N., "Transformation of Distorted Geodetic Networks to New Coordinate Reference Systems : A Case Study for ED50-ITRFxx Transformation in Turkey", Geodetski Vestnik, vol.61, pp.58-75, 2017

Prediction of Aboveground Carbon Storage in Forest Areas Using Remote Sensing Data

C. CAKIR[1], S. YILDIRIM[1], M. MISIR[2], N. MISIR[2]

[1]Ondokuz Mayıs University, Samsun/Turkey, cemile.cakir@omu.edu.tr
[1] Ondokuz Mayıs University, Samsun/Turkey, secil.erkut@omu.edu.tr
[2]Karadeniz Technical University, Trabzon/Turkey, mmisir@ktu.edu.tr
[2]Karadeniz Technical University, Trabzon /Turkey, nuray@ktu.edu.tr

Abstract **- Forest ecosystems play an increasingly important role in the global carbon cycle that storing CO_2 in the form of different chemical compounds (lignin, cellulose, etc) from atmosphere by photosynthesis. Determination of the amount of carbon is extremely important in reducing greenhouse effect and preventing global warming. Two different methods are used for determining the amount of carbon stored in forests. In both cases, control points necessary to determine the parameters of the different stand. However, these studies are time consuming and costly. But using geographical information systems and remote sensing techniques can provide more accurate information in a shorter time. In this study, we tried to estimate the amount of aboveground carbon stored in pure scoth pine stands using various vegetation indices. For this purpose Kunduz Forest Management Chiefdom selected for study area. Forest management plans, topografic maps and satellite image were used. The amount of carbon of pure scoth pine stands were determined using single tree carbon allometric equations. Relation between determined carbon data and various vegetation indices obtained from Sentinel 2A satellite images in 2017 were examined. In order to determine relations various of vegetation indices were tested. Relationship between results and the band combinations investigated by using correlation analysis, relationships between the variables modelled with help of regression analysis.**

Keywords - Carbon, Scotch pine, Remote sensing, Geographic information systems

I. INTRODUCTION

In the last quarter of the twentieth century, there are many negative effects on the world such as an increase in average temperature in the atmosphere, decrease in rainfall in arid and semi-arid areas, and decrease in water resources due to increase in drought. Since the 1950s, when the population grew rapidly with the industrial revolution, various greenhouse gases left to the atmosphere by human activity and industrial systems, and CO2 accumulation, which has the largest share of 53.2%, are causing these problems. It has been determined that 2/3 of the CO2 in the atmosphere is caused by fossil fuel consumption, 1/3 by land use change and deforestation**[1].**

In addition to carbon-dense contributions to forests, biomass fuels from renewable energy sources to meet growing energy needs suggest that forest ecosystems are the only source of concern. Forest ecosystems have an important contribution to mitigating the negative effects of global warming and to preserving regional and even global climate stability.

Biomass, which is a source of renewable energy, which releases less CO2 to the atmosphere, alternative to fossil fuels and less polluting the environment, is gaining importance all over the world. Forest ecosystems, which have the biggest share in biomass, are the major carbon pools that hold half of the 100Gt CO2 they receive every year from the atmosphere. The ocean, which is the second carbon pool on Earth, gives back 100Gt of 104Gt CO2. **[2].** 74% of carbon in forest ecosystems is above ground and 26% below ground. Thirty-five percent of the above-ground area is consistently kept in the ecosystem, 32.5% is returning to the atmosphere with normal decay and decomposition, and the remaining 32.5% is in forest products produced from seeds. **[1].**

The effects of forest ecosystems on the CO2 exchange in the atmosphere are demonstrated by biomass surveys and carbon sequestration studies. For this purpose, many methods are used to determine carbon accumulation. In addition to determining the amount of carbon storage through the coefficients in the guidelines prepared by the Intergovernmental Panel on Climate Change, many studies have been carried out on the basis of tree species in order to obtain clearer results. By determining the allometric biomass models and carbon models of each component, biomass and carbon contents are estimated by easily measurable parameters such as chest diameter and height. With these methods, Uğurlu et al., 1976; Atmaca, 2008; Yavuz et al., 2010; Aydın, 2010; Ülker, 2010; Çömez, 2011; Tolunay, 2011 for Pinus sylvestries, Mısır et al, 2012; Özkaya, 2004; Bülbül, 2012 for Picea orientalis, Mısır vd, 2012 for Abies and Saraçoğlu, 2000; Makineci et al., 2011; Mısır et al., 2013; Erkut, 2013 for Fagus orientalis carried out studies on carbon storage. Asan, 2011 examining the annual carbon stock change trends in Turkey's forests, Mısır et al., 2011 which aims to determine the amount of carbon storage, It is available in studies which the total carbon stock forest ecosystems in Turkey and live trees are also available studies aimed at calculating the carbon accumulation in vegetative mass made by Tolunay, 2011; Tolunay and Çömez, 2007; Tolunay, 2011; Tolunay, 2013; Mısır et al., 2017.

When all these studies are carried out, local measurements are obtained. However, data collection and evaluation with terrestrial measurements requires time consuming, costly and labor-intensive work. Therefore, the use of remote sensing methods provides a great deal of benefits in order to make more practical and cost effective studies. In order to determine carbon storage capacity by remote sensing methods using satellite images in our country, Mısır et al., 2011; Gülsunar, 2011; İnce,

2011; Bulut, 2012 was done by the work. In this study, it is aimed to determine the amount of carbon stored on soil using remote sensing techniques and geographic information systems.

II. MATERIALS AND METHODS

Research Area

Amasya Forest District Directorate Vezirköprü Forest Management Directorate Kunduz Forest Management Directorate was selected as the research area in the study. The topographic maps of 1/25000 scale (Samsun F-33c2, F-33c3, F-34d1, F-34d2, F-34d3, F-34d4) covering the research area are obtained by the field work carried out in 2009 in the forest management plan draft stand map and Sentinel 2A satellite displays are the basic bases used for working.

Land Studies

In this study, the data obtained from the inventory study carried out in 2009 within the scope of forest management plan renewal study was used. In the said inventory study, sampling areas were determined in the digital environment with a distance of 300x300 meters. A total of 500 sample areas were measured after the points indicating openness and agricultural areas were removed (Figure 1).

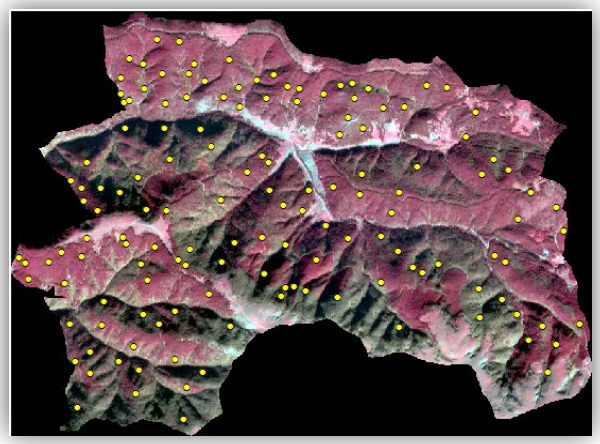


Figure 1: Distribution of sample areas to research area.

In the sample areas of different sizes (400-800 m2) according to the occlusion, the species of all the trees within the sample area boundaries were determined and age and height measurements were made in the number of trees determined according to the diameter of the chest height and the sample area size.

Calculation of the Amount of Carbon stored in the Stands

Carbon storage quantities which constitute the purpose of the study are calculated by different methods. In this study, carbon models developed according to tree species were used to determine the amount of carbon stored in the sample areas.

In the study;

For the beech $C = 0{,}025672xd^{2{,}775}$ [3].

For the oak $C = 2277{,}6xd^{1{,}1333}$ [4].

For the scoth pine $C = 0{,}413xd^{2} - 28{,}36$ [5].

The amount of carbon stored in each tree in each sample was calculated and collected and the necessary carbonation amounts were obtained in the hectares of the stems.

Creating Slope and Aspect Maps

Located Earth objects electromagnetic energy from the sun, storage, including migration from on-site mirroring and uses three different ways. The subject of remote sensing is the energy reflected by the objects. This reflected amount of energy, the structure of the object, the size of the surface, the place on the ground, etc. depending on the factors. These reflection quantities in trees, vary according to the type of trees, age, growth environment, yield power, elevation and aspect. In this study, average slope and aspect values of sample areas were calculated to reflect this change. For this purpose the help of contour lines obtained from digital topographic map of the study area, digital terrain model was produced. The slope and aspect maps of the study area were created with the help of the slope (%) and the aspect (°) values calculated from the digital terrain model. The slopes and aspect of all sample areas in the study area with the help of slope and aspect maps were calculated as weighted average (Figure 2).

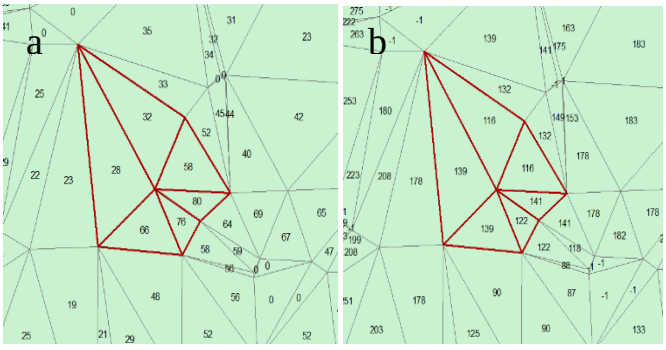


Figure 2: Sample area average slope calculation (%) (a) and average aspect calculation (◦) (b)

Obtaining Reflection Values from Satellite Images

Sentinel 2A satellite images were corrected using ERDAS Imagine 10.0 software. After this process, with the "Raster to Polygon" transformation, area feature layers including reflection values for each band are obtained, including blue, green, red and infrared bands. These layers were used to obtain the reflection values of the sample areas (Figure 3).

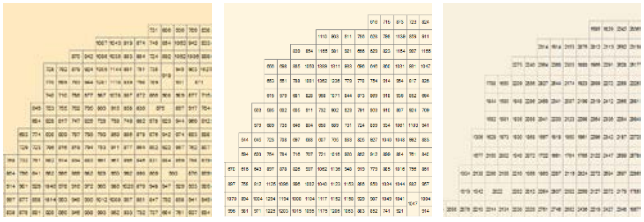


Figure 3: Satellite image reflectance values for the 250 sample area

The spatial resolutions of the satellite images used in the study are 10x10 m. Accordingly, there are approximately 4 pixels within the smallest sample area size of 400 m² in the study area. In the neighborhood analysis, "Mean (mean = mean reflection value calculated as weighted average for sampling area)" option was used.

Relationship Amount of Carbon Storage with Remote Sensing Data

Remote sensing data can provide detailed information about the composition of the stand structures and the species forming the stand. Different stand structures have different spectral reflectance values in different wave lengths. In addition, the topographic structure on the stands affects their reflection values. Therefore, the relationship between stand parameters and remote sensing data also varies. In this study, dependent variable of carbon storage amount; reflectance values for satellite image bands, vegetation indices obtained from slope and aspect data and different band combinations were using as independent variables, the relationship between the amount of carbon storage in the stands and the remote sensing data was investigated.

III. RESULTS

The statistical values of certain characteristics of the sampling areas taken in the research area are given in Table 1.

Table 1: Statistical values table

	Minimum	Maximum	Mean
Carbon Amount (ton/ha)	44.6	832.2	271.079
Slope(%)	8	87	45
Aspect(◦)	6	358	201
Band2	764	912	847
Band3	455	730	620
Band4	269	599	453
Band8	366	2025	1372

The amounts of carbon storage stands, the relationship between the values of the reflection band satellite images and vegetation index values are tried to be obtained. As explained above, the reflection values of the bands, the vegetation index values and the sample area average slope and aspect values are used as independent variables in the models in which the amount of carbon storage is a dependent variable. The obtained models are given in Table 2.

Table 2: Carbon models

Band	Models	R^2
2	$C = e^{-3,623 + 0,011x\,(Band2)}$	0.473
3	$C = 2302,122 - 8,576x(Band3) + 0,008x\,(Band3)^2$	0.547
	$C = -702,665 + 1,168x(Band3) + 0,973x(Aspect) + 412,257xNDVI$	0.742
	$C = -1018,349 + 1,030x(Band3) + 1,054x\,(Aspect) + 693,752xNLI$	0.750
4	$C = 0,001 + (Band4)^{2,163}$	0.630
8	$C = 250,262 - 0,309x\,(Band8) + 0,000209x(Band8)^2$	0.611

IV. SUGGESTION

Turkey climate in complex structure, particularly due to global warming, is one of the countries most affected by climate change will be seen. For these reasons, it is of great importance to determine the amount of carbon storage in our forests, how changes in carbon storage amount over time, how carbon storage amount changes according to planning units, and how interventions affect the amount of carbon storage in the fastest and most practical way. In this study, it was aimed to determine the amount of carbon storage of the stands easily and quickly for the entire study area by taking advantage of the inventory samples. For this purpose, the band values of color digital satellite images with 10 m local resolution and the terrestrial measurement data and vegetation indices and the carbon storage amounts of the stands were used. Many analyzes were performed using this data with the IBM SPSS program. Result of the analyzes and examinations made, the most suitable model which gives the amount of carbon storage of the stands is obtained. Determination of the amount of carbon stocks in forests is becoming more important day by day because it is one of the most important factors affecting global climate change. Since we have a quick knowledge of the amount of carbon stocks in our forests, it is very important that these and similar studies are carried out using different satellite images.

REFERENCES

[1] N. Saraçoğlu, Küresel İklim değişiminin Yavaşlatılmasında Ormanların Rolü, Bilim ve Aklın Aydınlığında Eğitim, S.135, Mayıs 2011, ss. 60-71

[2] K. İnce, Uzaktan Algılama Yöntemiyle Karbon Depolama Miktarının Belirlenmesi(Artvin Örneği), KTÜ Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Trabzon,2011

[3] Mısır, M., Mısır, N., Ülker, C., Erkut, S., Saf kayın meşçerelerinin karbon depolama miktarının belirlenmesi (Trabzon Orman Bölge Müdürlüğü Örneği), Bilimsel Araştırma Projesi, 2013.

[4] Makineci, E., Özdemir, E., Çalışkan, S., Yılmaz, E., Kumbaşlı, M., Keten, A., Beşkardeş, V., Zengin, H., Yılmaz, H., Ecosystem carbon pools of coppice-originated oak forests at different development stages. European Journal of Forest Research 134 (2): 319-333, 2015.

[5] Yavuz, H., Mısır, M., Mısır, N., Tüfekçioğlu, A., Karahalil, U., Küçük, M., Karadeniz Bölgesi saf ve karışık Sarıçam (Pinus sylvetris L.) meşcereleri için mekanistik büyüme modellerinin geliştirilmesi, biyokütle

ve karbon depolama miktarlarının belirlenmesi. Tübitak Projesi, Proje no: 106O274, 2010.
[6] Uğurlu, S., Araslı, B., Sun, O., Stebe geçiş yörelerindeki sarıçam meşcerelerinde biyolojik kütlenin saptanması, Ormancılık Araştırma Enstitüsü Yayınları, 80:48, 1976.
[7] Aydın, Ç., Artvin Orman Bölge Müdürlüğü Borçka Orman İşletme Müdürlüğü Sarıçam Biyokütle Tabloları. Karadeniz Teknik Üniversitesi Fen Bilimleri Enstitüsü, Basılmamış Yüksek Lisans Tezi, Trabzon, 2010.
[8] Çömez, A., Sündiken dağlarında sarıçam (Pinus sylvestris l.) meşcerelerinde karbon birikiminin belirlenmesi, Orman Mühendisliği Anabilim Dalı, 2010.
[9] Tolunay, D., Total carbon stocks and carbon accumulation in living tree biomass in forest ecosystems of Turkey. Turkısh Journal of Agrıculture And Forestry, 35:265-279, 2011.
[10] Mısır, N., Mısır M., Erkut, S., 2012. Estimations of total ecosystem biomass and carbon storage for fir (Abies Nordmanniana s. Subsp. Bornmülleriana (mattf.)) forests (Western Black Sea Region), Tam Metin Bildiri, 14th International Fir Symposium, 12-14 Eylül 2012.
[11] Özkaya, S., Artvin-Genya dağı yöresi doğu ladini (Picea orientalis (L.) Link.) ormanlarında toprak üstü biyokütlenin belirlenmesi, Yüksek Lisans Tezi, Artvin, 2004.
[12] Bülbül, E., K.T.Ü Orman Fakültesi Eğitim ve Araştırma Ormanı saf ladin meşcerelerinin karbon depolama miktarının belirlenmesi. Yüksek Lisans Tezi, Trabzon, 2012,
[13] Mısır, N., Mısır M., Erkut, S., Estimations of total ecosystem biomass and carbon storage for fir (Abies Nordmanniana s. Subsp. Bornmülleriana (mattf.)) forests (Western Black Sea Region), Tam Metin Bildiri, 14th International Fir Symposium, 12-14 Eylül 2012.
[14] Saraçoğlu, N., Sakallı Kızılağaç (Alnus glutinosa (L.) Gaertn subsp. barbata (C.A. Mey.) Yalt.) biyokütle tabloları, Turk. Jour. of Agriculture and Forestry, 24: 147-156, 2000.
[15] Mısır, M., Mısır, N., Ülker, C., Erkut, S., Saf kayın meşçerelerinin karbon depolama miktarının belirlenmesi (Trabzon Orman Bölge Müdürlüğü Örneği), Bilimsel Araştırma Projesi, 2013.
[16] Erkut, S., Giresun Orman Bölge Müdürlüğü Akkuş Orman İşletme müdürlüğü saf kayın meşcerelerinin ekosistem bazında karbon depolama kapasitesi, Trabzon, 2013,
[17] Asan, Ü., Türkiye ormanlarındaki yıllık karbon stok değişimi trendinin irdelenmesi ve 2023 yılındaki durumun kestirilmesi. I. Ulusal Akdeniz Orman ve Çevre Sempozyumu, Kahramanmaraş, s. 930-944, 26-28 Ekim 2011.
[18] Mısır, N., Mısır M., Ülker C., Karbon depolama kapasitesinin belirlenmesi. I. Ulusal Akdeniz Orman ve Çevre Sempozyumu, 26-28 Ekim, Kahramanmaraş, s. 524-531, 2011.
[19] Mısır, M., Mısır, N., Bulut, A., Karbon depolama kapasitesinin landsat etm+ uydu görüntüsüyle belirlenmesi. I. Ulusal Akdeniz Orman ve Çevre Sempozyumu, Kahramanmaraş, s. 532-538, 26-28 Ekim 2011,
[20] Tolunay, D., Türkiye'de artım ve ağaç servetinden bitkisel kütle ve karbon miktarlarının hesaplamasında kullanılabilecek katsayılar. Ormancılıkta Sektörel Planlamanın 50. yılı Uluslararası Sempozyumu, pp:240-251, Antalya, 2011.
[21] Tolunay, D. ve Çömez, A., "Türkiye ormanlarında toprak ve ölü örtüde depolanmış organik karbon miktarları." Hava Kirliliği ve Kontrolü Ulusal Sempozyumu Bildiri Kitabı, Hatay, s. 750-765, 2008.
[22] Gülsunar, M., Ormanların karbon depolama kapasitesinin uzaktan algılama yöntemi ile belirlenmesi(Düzdağ Orman İşletme Şefliği örneği), Yüksek Lisans Tezi, Trabzon, 2011.
[23] Bulut, A., Ormanların karbon depolama kapasitesinin üç farklı uydu görüntüsü kullanılarak uzaktan algılama yöntemi ile belirlenmesi (Alacadağ Orman İşletme Şefliği Örneği), Yüksek Lisans Tezi, Trabzon, 2002.
[24] Ülker, C., Amasya Orman Bölge Müdürlüğü Sarıçam (Pinus sylvestris L.)Meşcerelerinin Biyokütle Tablolarının Düzenlenmesi (Kunduz Örneği), Yüksek Lisans Tezi, K.T.Ü., Fen Bilimleri Enstitüsü, Trabzon, 2010.
[25] Atmaca, S., Erzurum Orman Bölge Müdürlüğü Sarıçam Biyokütle tablolarının düzenlenmesi, Yüksek Lisans Tezi, Z.K.Ü., Fen Bilimleri Enstitüsü, Zonguldak, 2008.
[26] Mısır, N., Yıldırım, S., Mısır M., Türkiye Ormanlarının Karbon Depolama Kapasitesi, IV Ulusal Ormancılık Kongresi, Cilt: 1, ss: 381-392, Antalya, 2017.

Relationship with GIS of Neighborhood Features belong to Konya

S. YALPIR[1] and F. B. UNEL[2]

[1] Selcuk University, Konya/Turkey, sarici@selcuk.edu.tr
[2]Konya/Turkey, fatmabunel@gmail.com

Abstract **- Appraisal of real estates which have a large important for the country economics is pretty a hard occupation field. The value of real estate indicates important level changes in accordance with location and neighborhood features of their places apart from legal and physical factors. Using value prediction by explaining as mathematical these changes is a complex problem. It is appropriate to use Geographical Information Systems (GIS) application because of the size of the study area in real estate valuation and the need to integrate the attribute information into the map.**

The purpose of this study is to provide using neighborhood features affecting the value of real estate for mass appraisal. In Centre Neighborhood of Konya, the neighborhood features were taken into account, standardized with the help of GIS software by producing the prediction and thematic maps in form of simple and easy and made the ready form for valuation analysis. These maps will be able to use as the base in all practices of valuation from taxation to expropriation.

Keywords **– Geographical information systems, mass appraisal, real estate valuation, neighborhood features.**

I. INTRODUCTION

REAL estate is very important for people to continue their lives. For this reason, it should be managed by including valuation in a good way. Considering all aspects of a city; each of real estate which exists agricultural, treasury and municipal land, plot, residential, commercial, industrial and public building etc. should be assessed separately step by step.

Land management is the process by which the resources of land are put to good effect. It covers all activities concerned with the management of land as a resource both from an environmental and from an economic perspective [1] *towards sustainable development* [2]. There is a need for a database in which all kinds of information about real estates in large and small scale are included in order to make city planning suitable for geographical, geological and topographic conditions in land management and to make optimum decisions about land. For the economic dimension of land management, real estate value maps should be produced and added to the database. For this reason, it is necessary to model by using computer-aided mass appraisal in order to calculate the values of the real estates unknown to the known values of the real estates. The real estate value is the market value of real estate that is suitable for normal purchasing and selling conditions in the stable economy. *Market value–the estimated*

amount for which an asset or liability should exchange on the valuation date between a willing buyer and a willing seller in an arm's length transaction, after proper marketing and where the parties had each acted knowledgeably, prudently and without compulsion [3, 4]. Individuals that may be purchasers and sellers need to determine the market values of the real estate sectors, such as private companies, government agencies, banks and investment companies at local and worldwide. There is also a need for a number of other processes to the value, such as managing land well, in commercial purchasing-selling for investment purposes, and making risk analysis for project development.

In literature, it was investigated that general level of neighborhood quality, demographic characteristics, the income level of population and prestige, planning and urban development characteristics, educational level of population and crime level are applied for mass valuation of urban land [5, 6] and environmental aesthetics effect on house prices [7]. In the two localities, differences of economic, demographic, and environmental factors which influence the market value of urban land were demonstrated [8]. 67 Different neighborhoods defined by the property tax assessment office of the Town of Amherst were used in order to estimate prices of the houses [9]. The effects on the real estate value have been detected in studies where natural risks such as flood, avalanche [10], and in studies where special criteria such as traffic noise pollution [11, 12], air pollution [13, 14] and shale gas [15, 16]. It is seen that the thematic maps produced by Konya Metropolitan Municipality under the titles of environment, tourism, residence, social, demographic, infrastructure, geology, safety, transportation, and education are not used for statistical purposes and they are not produced in a standard structure to be used in the real estate valuation [17].

Urban areas; healthy, prosperous, regular development and suitable construction for people's life, it is necessary to find development plans suitable for these purposes and then to construct according to the plans. All these structures and zoning plans of their basis require the plots and plot production. It is essential that these plans are executed meticulously according to temporal predictions in a broad perspective because of the structure and location of the site cannot be changed. Production of incorrect plots and defective zoning plans are taken a long time and large financial burdens are brought about after completion of lower and upper structure studies with zoning plans. Determination of the value of plot is an important step in the transformation from land to the plot when new land parts become zoned areas for use as urban areas. In addition to applications such as taxation and expropriation, which are related to more than one real estate, besides privatization and evaluation of the nationalization, first of all, the properties of

the plot where the real estate is located are examined and then the features of the building on it are examined.

In the previous studies [18-20] of the plot features, were gathered in four main headings as Legal, Physical, Locational and Neighborhood features. This work is organized taking into account only neighborhood features. The purpose of the work is to standardize the neighborhood features of Konya province to be associated with Geographical Information Systems (GIS) and to be a base for mass real estate appraisal. Konya city center neighborhoods are examined from the sub-titles which are consisting of criteria such as population, level of education, prestige, structure situation, development potential, purchasing-selling mobility, slope, geological situation, climate, air and noise pollution and Konya city is examined in all aspects of Konya city.

II. MATERIAL AND METHOD

A. Study Area

Konya province is Turkey's largest city with a survey area of 40,814 km^2 (excluding the lakes 38,873 km^2) and it is 7th in the crowded cities of Turkey with a total population of 2,161,303 in 2016 [21] . Turkey Statistical Institute (TUIK) in 2016, according to the Statistical Region Classification groups that making up the TR52 Konya and Karaman province, according to the household labor force survey results; 25.8% of the employed population is working in the agriculture sector, 28.3% in the industrial sector and 45.9% in the services sector. The current area of the province has 4,081,353 hectares of land, 46.6% (1,910,639 hectares) of agricultural land, 18.7% (761,461 hectares) meadow-pasture, 13.2% (540,189 hectares) forested-nursery and 21.4% (869,064 hectares) of other land [22]. Besides, with the production method, the gross domestic product estimate has increased by 24.2% at current prices to 827 billion 230 million ₺ in the third quarter of 2017 [21].

The province of Konya consists of three central districts Karatay, Meram and Selçuklu (Figure 1). Central districts separated by administrative borders from the center of Konya city; has positioned as Karatay on the east, Meram on the southwest and Selçuklu on the north. The intersection of the three cities is the exact midpoint of the city center. These districts were examined by using SWOT Analysis (Table 2).

The population of Karatay in 2016 is 308,983 and it is the least crowded district of Konya by forming 24% of the total population of the central districts. The net migration rate in the district where the regional migration is concentrated is three per thousand. A large part of the district consists of in the old settlement areas. Mevlana Museum, historical mosques, madrasahs, baths, and fountains in the existence of urban site areas. With the old and irregular structuring, the diversity of social texture and the low price of houses and rents cause the majority of outside migration to settle in this district.

Figure 1: Central districts Karatay, Meram and Selçuklu

Industrial areas are mostly located in the Selçuklu and Karatay districts; it has been detected that the industrial and residential areas in the Karatay district have entered one within the other. The ratio of primary school graduates to the population has the highest value with 70.80%, while the ratio of the university and postgraduate graduates to the population has the lowest value of 5.79% and 0.39% respectively. In addition to these, it is usually this area because of the preference of fattening areas and the preference of those who want to deal with agriculture and animal husbandry (Table 1 and Table 2).

Table 1: Data related to Karatay, Meram and Selçuklu districts

THE DISTRICTS	Karatay	Meram	Selçuklu
The population and percentage	308,983	346,366	622,846
	%24	%27	%49
In-migration	5,151	7,142	20,900
Out- migration	4,256	6,359	19,034
Rate of net migration	%0.30	%0.20	%0.30
Primary school graduate and rate	209,098	220,252	336,928
	%70.80	%64.62	%57.63
High school graduate and rate	29,237	45,356	104,584
	%9.90	%13.31	%17.89
University graduate and rate	17,100	31,522	69,196
	%5.79	%9.25	%11.84
M.Sc/PhD graduate and rate	1,136	4,014	8,054
	%0.39	%1.18	%1.37
The other	38,761	39,673	65,882
	%13.12	%11.64	%11.27

The city of Meram is famous for its vineyards and gardens, and it is a city with an elite social texture. Some part of the district including to the natural site area due to Meram stream and the presence of many centuries-old trees. Some of the residential areas of the district center can be prevented from urban development in that direction due to the fact that they are site areas and mountainous areas. However, due to the small number of industrial areas in the district and zoning plan implementations, the settlement is more homogeneous and the population ratio is lower (Table 1 and Table 2).

The Selçuklu district is a district where a large number of the new buildings are located and thus the majority of it is zoned and regular. It is observed that the education level is also high in line with the fact that Konya is the most crowded district with a population of 622,846. Because it has a flat and wide area, it

is a district where open for improvements and high-rise houses are predominantly. One of the biggest reasons why the net migration rate is three per thousand is the increase of the number of civil servants and students. Although the net migration rate in the Selçuklu and Karatay districts is the same, the situation and conditions of migrants show the difference (Table 1 and Table 2).

As a market sampling area, central districts near to the city center of Konya were preferred. As you move away from the city center, you are encountered neighborhoods called formerly villages. Therefore, the density of the plot will decrease in remote neighborhoods. Due to the fact that this study was used data of plots in the verification step, Karatay, Meram and Selçuklu central districts of Konya were taken and the neighborhoods in the village were excluded (Figure 3).

Table 2: Positive and negative aspects of Karatay, Meram and Selçuklu districts (SWOT Analysis)

CENTRAL DISTRICTS	*POSITIVE ASPECTS*	*NEGATIVE ASPECTS*
	STRENGTHS	**WEAKNESSES**
KARATAY	• Mevlana Tomb and due to ample its historical sites are very important advantages in terms of tourism. • Having flat and wide areas suitable for development • The Palace of Justice is located in this district. • Konya Chamber of Commerce Karatay University • Urban Site Area	• Foreign immigration (Out of city, abroad) • Skewed and disordered structuring • Industrial areas predominantly in this district • Lots of agriculture and animal husbandry • Social tissue variety • Difficulties in zoning implementations due to the density of the site areas in the district center
MERAM	• It is the district where it has the most natural greenery. • It has Meram stream. • Having the clean air • Necmettin Erbakan University • Having a homogeneous and elite social texture • Natural Site Area • The low population density	• Due to the fact that it is a site area, structuring with a zoning plan for protection intentional. • The Military Region is predominantly located in this district • Infertility of district center development due to geographical situation and sites
SELÇUKLU	• The area to be developed is large and flat • To be done major investments-projects • Selçuk University • Regular structuring due to new settlement • Facilitation of investments due to a lot of public lands	• Being a cosmopolitan social structure
	OPPORTUNITIES	**THREATS**
KARATAY	• Opportunity to zoning up flat and wide areas • Due to the multiplicity of the old urban structure, a district where urban transformation is mainly applied	• The presence of garbage collection areas • Strong air circulation and dense air pollution since it is in pit as topographical • The entrance of one industrial area within residential areas
MERAM	• The direction of Natural Tourism can be improved • Producing fruit and vegetables since the soil is fertile • Having a curative for the disease • Low density of population and structure, more homogeneous social environment	• It has the water basin • Risks of reduction of greenery
SELÇUKLU	• Opportunity to zoning up flat and wide areas • Open to developments in terms of urban and industrial areas	• Rapid increase in population density • For future traffic and air pollution problems

B. Method

Since the definitional data were obtained in different formats from different places, they were combined with geographic data by editing in a matrix format suitable for GIS software. By using neighborhood based data, it was generated thematic and prediction maps in ArcGIS.

III. NEIGHBORHOOD FEATURES

Considering the central districts on Karatay, Meram and Selçuklu in Konya, map bases where neighborhood borders located, were arranged on the small scale (Figure 3).

A. Population Density

Population data were used not by raw but by converting the number of people belonging to the neighborhood to the area of the area in terms of hectares and by calculating the population density. Thematic maps were obtained from these transformation results.

B. Education Level

Educational status; It was taken into consideration as primary school, high school, university, and postgraduate. Educational status changes according to the population density of the neighborhood. Because the rate of literacy is also high in places with high population, the education situation will show better than when the population is low. For this reason, it has been

ensured that the values of the education level of the neighborhood are compared to the numbers of the population (A=Educational Status/Population). The thematic map was obtained from the values detected as percent (A*100) of the population and how many of them and from where graduates are graduated.

C. Prestige

According to Article 72 of the Tax Procedure Law No: 213 established by "Valuation Commission" and According to Article 29 of the Law on Real Estate Tax, No. 1319 the minimum unit values of the streets in terms of the street, road or value are detected every four years for each neighborhood. The minimum unit values of the plots in Konya in 2016 cannot be used due to the change in tax system of Selçuklu Municipality in 2014. The minimum unit values of the year 2013 of the plots, which are approximately the same in the Karatay, Meram and Selçuklu districts, have been taken into consideration [23]. In particular, the unit values plot on the street are shown and associated with a separate point for each neighborhood in the map, and an estimation map has been produced.

D. Residential Density

The number of residential does not mean the density of structure. In order to convert the number of residential into density, the total number of residential needs to be proportioned to the total number of parcels. Thus, the neighborhood-based distribution of residential numbers can be observed in balancedly and used in the data set for valuation work. However, since the number of parcels could not be reached, the density of structure was calculated as the number of residential in hectares and the thematic maps were obtained by comparing the number of houses to the area in hectares.

E. Development Potential

This situation, which involves the process from the production of the plots to the completion of the residential, is directly related to the development potential. First of all zoning plans by applying to the land, the land is being produced and prepared for housing construction. Roads are identified and infrastructure works to supply basic needs such as water, electricity, sewage, natural gas and telephone are started. On the other hand, residential construction starts and after the infrastructure works are completed, the construction of the residential is continued. Roads, pavements and landscaping are started while residential constructions are in progress. Social facilities such as schools, healthcare facilities, social facilities and parks are completed within the period from the implementation of the zoning plan to the construction of the last residential (Figure 2).

The criteria consisting of four sections, which the percentages are evaluated by gathered, the developmental status of the neighborhoods were detected over 100% and the thematic map was obtained.

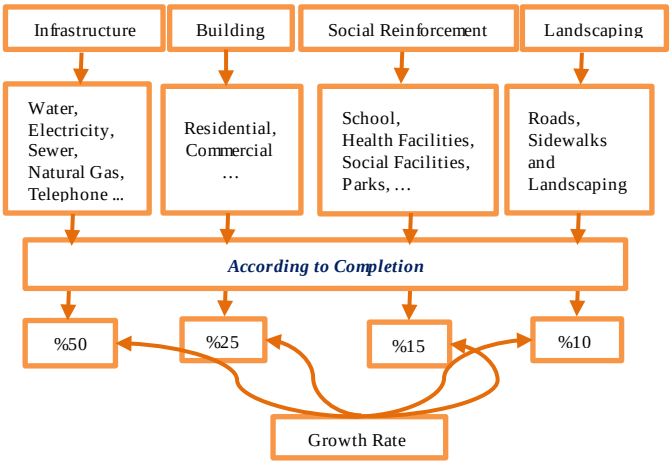


Figure 2: Development rate

F. Purchasing and Selling Mobility of Real Estate

The general real estate purchasing and selling mobility in the Directorate of Land Registry was attained from previous studies and the thematic map was obtained deficiently because the distribution in the map was not related to title deeds and municipal boundaries.

G. Slope

A global data bank, accessible via the internet as "Global Data Explorer" supported by NASA (EarthData), by preparing and it has presented to the user. In the databank there are files with ".tif" extensions belonging to the Digital Elevation Model and the data of Konya is downloaded from the internet and arranged in GIS and registered with the layer of the neighborhoods and the slope map is arranged.

H. Geological Condition

Geological Map; The General Directorate of Mineral Research and Exploration (MTA) is being produced in the Presidency of Geological Studies Department and has the map in 1/25,000; 1/50,000; 1/100,000 and 1/500,000 scales. The map, in general terms; there are seven zones in the form of Appropriate Area, Inappropriate Area, Precautionary Area 1, Precautionary Area 2, Precautionary Area 3, Precautionary Area 4 and Precautionary Area 5. In the Netcad program, each of the regions was re-created in ArcGIS by being translated as the field in a separate layer.

I. Climate Condition

There are three stations of Konya Meteorology 8th Regional Directorate, Bahri Dağdaş numbered 9029, Konya numbered 17244 and Regional Station numbered 17245 [24]. Data about temperature, humidity and wind speed are gathered every two seconds from the points where the automatic meteorological stations are installed.

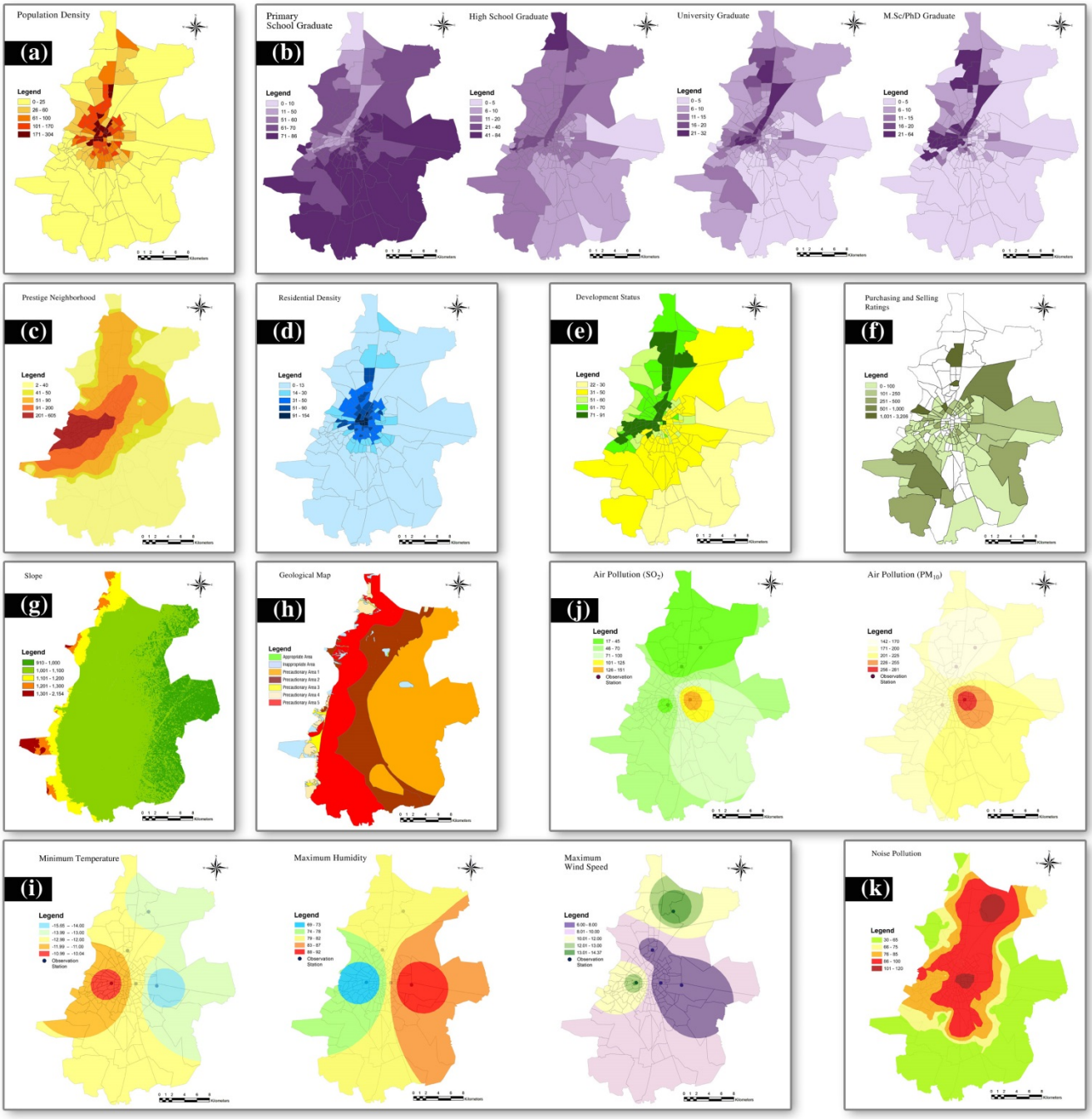


Figure 3: Develop Thematic Map in Konya

Since it is not appropriate to use all the data separately in terms of both valuation and statistics, it is decided to take into consideration the monthly average, maximum and minimum values of the data of recent years. The temperature of the air has affected the warmth felt by the presence of humidity and wind. In January the air felt by the presence of the wind was colder, and in July the air has become warmer due to the presence of the humidity [24]. Therefore, the maximum speed of the wind in January, the maximum humidity averages in July were taken into account and prediction maps were created.

J. *Air Pollution*

There are four fixed air quality monitoring stations in Konya, two of which belong to the National Air Quality Monitoring Network belonging to the Ministry of Environment and Urban Planning, and two of Konya Metropolitan Municipality's Meram, Mevlana, Aydınlık and Horozluhan stations and hourly average values are obtained by measuring with automatic devices [25].

In view of the most uncomfortable conditions related to air pollution, measures were taken on the winter calendar between October 15, 2015, and March 15, 2016, and the maximum values of SO_2 and PM_{10} for each day were by being arranged

in a matrix format and arithmetic average was calculated. The average values of SO_2 and PM_{10} in relation to the table with 4 station points in the map were associated and an air pollution map was produced as an estimation map.

K. Noise Pollution

The previous studies had traffic and railway noise data belonging to the province of Konya. The average of the maximum values of these data was taken and the prediction map was produced.

IV. CONCLUSION

It is important to reveal the historical, natural and social texture, topographic, the environmental and geological structure of a city, settlement plan in general terms, at real estate valuation. The neighborhood features have been analyzed with small-scale maps of the city and an objective viewpoint has been presented for real estate valuation. In addition, the thematic and prediction maps of the local features were standardized and to the mass real estate appraisal base was established.

In real estate valuation; a system should be designed which can model by taking into consideration usage purposes such as taxation, expropriation, insurance, privatization, implementation of zoning plans. Locality and other properties determined for use in the modeling should be flexible in the form of the attachable-removable system according to the extraordinary situations. Below the local roof, properties to be taken into consideration as the scale grows, should be more detailed and accurate to the value of the real estate according to the purpose of use.

The next study will be concerned with which time interval at which these data should be updated of necessity for a sustainable mass real estate valuation system. In addition, the system is directly involved in retrieving updated data and incorporating it into the valuation process as well other issues to be investigating.

ACKNOWLEDGMENT

The authors would like to thank pollsters who made efforts and participants who patiently completed the questionnaire in the data collection phase of the survey. This study is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) with 115Y769 Project Number and the Coordinator of Selçuk University's Scientific Research Projects with 15101008 Project Number.

REFERENCES

[1] UN, *Land Administration Guidelines-With Special Reference to Countries in Transition*. United Nations Publication: New York and Geneva, 1996.

[2] FIG, *The Bathurst Declaration on Land Administration for Sustainable Development*. International Federation of Surveyors (FIG), Australia: Publication No 21, 1999.

[3] UDES, *Sermaye Piyasasında Uluslararası Değerleme Standartları Hakkında Tebliğ*. Seri: VIII, No: 45, Tarih: 1/5/2006.

[4] IVS, *International Valuation Standards (IVS)*. 2011 (March 5, 2015); Available: http://www.ivsc.org/content/about-international-valuation-standards-council-ivsc.

[5] M. Kryvobokov, "Mass valuation of urban land in Ukraine: From normative to a market-based approach", PhD, Real Estate Planning and Land Law Department, Real Estate and Construction Management School of Architecture and The Built Environment Royal Institute of Technology, Stockholm, 2006.

[6] M. Kryvobokov, "What location attributes are the most important for market value?", *Property Management*, vol.25, pp. 257–286, 2007.

[7] G. E. Çetintahra, and E. Çubukçu, "Çevre estetiğinin konut fiyatlarına etkisi", *itüdergisi/a, mimarlık, planlama, tasarım*, vol.10, pp. 3–12, 2011.

[8] N. Kheira, and B. A. Portnov, "Economic, demographic and environmental factors affecting urbanland prices in the Arab sector in Israel", *Land Use Policy*, vol.50, pp. 518–527, 2016.

[9] C. C. Lin, "Critical analysis and effectiveness of key parameters in residential property valuations", PhD, Department of Civil, Structural, and Environmental Engineering, State University of New York, New York, 2010.

[10] C. Casas, "Essays in applied industrial organization", PhD, Economics, University of Wisconsin, Madison, 2014.

[11] El-Gohary, M., "Property valuation model effect of traffic noise on property value", *ECE 557 PROJECT Member, IEEE*, 2004.

[12] A. Szczepanska, A. Senetra, and M. Wasilewicz-Pszczołkowska, "The effect of road traffic noise on the prices of residential property–A case study of the polish city of Olsztyn", *Transportation Research Part D.*, vol.36, pp. 167–177, 2015.

[13] R. G. Ridker, and J.A. Henning, "The determinants of residential property values with special reference to air pollution", *The Review of Economics and Statistics*, vol.49, pp. 246–257, 1967.

[14] S. Zheng, J. Cao, M. E. Kahn, and C. Sun, "Real estate valuation and cross-boundary air pollution externalities: Evidence from Chinese Cities", *J Real Estate Finan Econ.*, vol.48, pp. 398–414, 2014.

[15] H. A. Klaiber, and S. Gopalakrishnan, "The impact of shale exploration on housing values in Pennsylvania", in *The Agricultural & Applied Economics Association's*, 2012, Washington.

[16] L. Muehlenbachs, E. Spiller, and C. Timmins, "The housing market impacts of shale gas development", Seminar, Duke University, Washington, New York, 2014.

[17] KonyaBel. *Konya CBS İstatistikleri*. 2018 (March 12, 2018); Available: http://www.konya.bel.tr/sayfadetay.php?sayfaID=121.

[18] F. B. Unel, and S. Yalpir, "Positional determination of real estates with Analytic Hierarchy Process", in *Proceedings of the Fourth International Conference on Mathematical and Computational Applications*, 2013, Manisa, pp. 326–336.

[19] Ş. Yalpır, and F. B. Ünel, "Türkiye'de ve Uluslararası çalışmalarda arsa değerlemede kullanılan kriterlerin irdelenmesi ve Faktör Analizi ile azaltımı", *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, vol.16, pp. 303–322, 2016.

[20] F. B. Ünel, "Taşınmaz Değerleme Kriterlerine Yönelik Coğrafi Veri Modelinin Geliştirilmesi", Doktora Tezi, Harita Mühendisliği A.B.D. Selçuk Üniversitesi, Konya, 2017.

[21] TUIK. *Ana Sayfa*. 2018 (March 12, 2018); Available: http://www.tuik.gov.tr.

[22] Valilik, *Ekonomik Durum*. 2015, Türkiye İstatistik Kurumu (TUIK) ve Konya Valiliği: Konya.

[23] KMSbel. *Arsa Rayiç Bilgileri*. 2013 (April 20, 2016); Available: https://tahsilat.karatay.bel.tr/EmlakRayicDeger/syRayicDegerParam; https://ebelediye.meram.bel.tr/web/guest/5; https://tahsilat.selcuklu.bel.tr/PublicServices/RayicBedel.

[24] Meteoroloji. *İklim özellikleri ve Konya'daki gözleme istasyonları*. 2016 (April 10, 2016); Available: https://www.mgm.gov.tr/genel/sss.aspx; http://www.konya.mgm.gov.tr/gozlem-sebekesi.aspx.

[25] Havaizleme. *Ulusal Hava Kalitesi İzleme Ağı*. 2016 (March 25, 2016); Available: http://www.havaizleme.gov.tr/Default.ltr.aspx.

Real Estate Valuation Using Artificial Neural Networks Method

Ş.YALPIR, O. ORHAN, H. ULKU, G. SARKIM and G. ERVURAL

Department of Geomatics Engineering, Engineering Faculty, Selcuk University, 42250 Konya/TURKEY, (sarici@selcuk.edu.tr, osmanorhan44/gamzesarkim@gmail.com, harika_ulku/guneriervural@hotmail.com)

Abstract - The real estate has an important position economically in the world. Proper valuation of the real estate is important for the country's economy. For the real estate valuation, it is necessary to know well the concept of value related to real estates and the factors, which affect the real estate value around the area. With the development of computer technology, it is possible to reach quick and accurate results by making detailed analyzes. In recent years, developments in artificial intelligence technologies have made artificial intelligence methods more attractive in real estate valuation. Also, advanced geographical information system(GIS) technology has started to use extensively in the real estate valuation. Thus, the creation of the databases, which has predominantly spatial information, for real estates increased the role of GIS in real estate valuation methods.

In this study, positional analysis of agricultural data in the GIS environment was conducted and the factors affecting depreciation were examined. Artificial neural networks model is developed by data that are prepared in GIS environment. Subsequently, the success of the predicted outcome was assessed. As a result of this work is aimed to obtain accurate information about the value of agricultural land using mathematical modeling. The results show that real estate prediction study using ANN was in good agreement with absolute success value of 93% and correlation coefficient (R^2) value of %76.

Keywords - **Artificial Neural Networks, Real Estate Valuation, Geographic Information Systems.**

I. INTRODUCTION

In real estate valuation, Each property has its own characteristics due to its location [1]. In our country, while many goods are priced at a single price, this is not the case for real estate that is considered as an investment instrument. It is possible to see the same real estate in different values in the market [2]. As a result, accurate real estate valuation is important to mortgage lenders, investors, insurers, and buyers and sellers of real property. In this way, trade or business real estate markets will ensure clarifying standards and securing. For this reason, the real estate should be evaluated objectively and scientifically.

Real property is defined as all the benefits, rights and liabilities inherent in the ownership of the physical real estate, where real estate is the land together with all developments that are permanently attached to it and all appurtenances associated to that place [3]. Real estate valuation, property appraisal or land valuation is the process of developing an opinion of value for real estate (generally market value). However, for accurate real estate appraisal, it is very important to use accurate techniques besides adequate information about the real property and the environment [4]. In the literature, the majority of the valuation methods will rely on some form of comparison to estimate market value. Comparable method, contractor's method/cost method, multiple regression method, investment/income method, profit method are some of these traditional methods. On the other hand, other models try to analyze the market by directly mimicking the thinking processes of the actors in the market. Artificial neural networks (ANNs) is the powerful tool among these advanced models. ANNs models have been presented as a possible solution to many problems in real estate valuation [3].

The aim of the study is to evaluate the real properties by using the ANNs method, unlike the valuation methods in the real estate market. The values related to the agricultural area in the Karapınar district of Konya are determined by ANNs method and the results are analyzed.

One of the important contributions of the study is to design a new and more advanced methodology for valuation of real estate and incorporate uncertainty into the valuation process. Since one of the criteria affecting the valuation is the risk factor, the model is addressed under the risk and uncertainty factor.

The occurrence of the sinkhole is the risk factor for all real estate in the region. This risk factor is taken into account when evaluating real estate properties in the related area and it is one of the criteria that affect value. To our best knowledge, there is no study in the literature dealing with risk factors in the real estate valuation for Turkey.

II. METHODS

A. Real Estate Valuation Methods

Each country has a different culture and experience, which will determine the methods adopted for any certain valuation. There is no exact method for real estate evaluation. The majority of the methods relies on comparison to evaluate market value. Other models or methods try to analyze the market by directly simulating the thinking processes of the actors in the market [3].

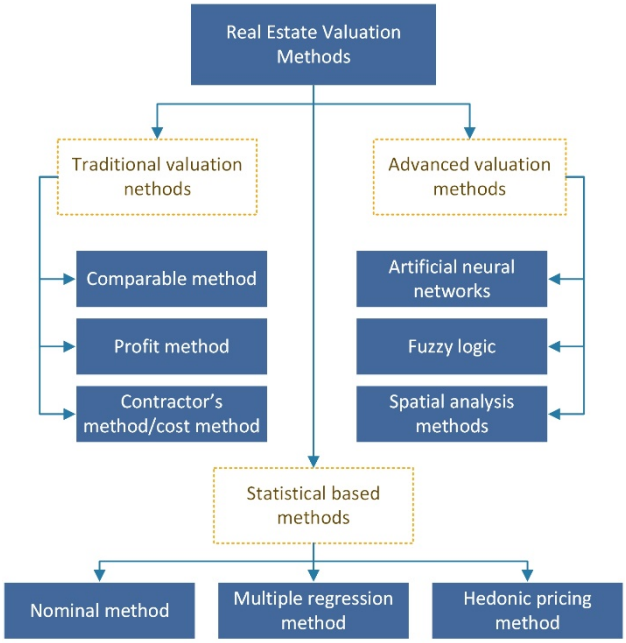


Figure 1: Real Estate Valuation Methods [5]

For real estate valuation, comparable method, investment/income method, and cost method are mostly used. However, these methods can deal with a limited number of criteria. Therefore, it is insufficient in today's conditions and does not show a realistic approach [5].

Factors affecting the value of real estate in agricultural areas and real estate values constitute the data set. According to the literature, the factors affecting the value of real estate in agricultural areas are as follows:

- Location of the real estate,
- Geological structure,
- Geographical structure,
- Distance to city center,
- Property security

B. Geographic Information Systems

With the development of geographical information system (GIS) technologies, the concept of GIS has been used in the real estate appraisal in many countries in recent years. A GIS is a system designed to capture, store, manipulate, analyze, manage, and present spatial or geographic data. Also, the use of GIS technologies in real estate valuation has emerged. The importance of GIS has been increased by the fact that the number of data dependent on the location that helps in determining the real estate values. Thus, the changes in the factors affecting the value of the real estate can be easily correlated and more accurate values can be obtained for the buy-sell price [6]. In this study, the factors affecting the value of real estate in agricultural areas were examined in GIS environment and the factors affecting the real estate value were correlated and analyzed.

C. Artificial Neural Networks (ANN)

ANNs network model is a method that can make it possible to solve many problems encountered during the real-estate valuation. Neural networks or with its known name artificial neural networks have been developed by being inspired by biological neural networks. As in a biological neural system, there are neurons and neural networks connecting these neurons to each other, i.e. briefly it imitates the human brain. In a neural network, the data set is trained first and then the method (model) is used to estimate the newly updated values from the same market, and each node corresponding to the neuron in the biological neural network works connected to the other node with a certain weighted coefficient. After the appropriate weights were determined according to the given inputs and outputs, the neural network became trained [7]. In Figure 2 shows the ANNs structure having 7 criteria used in the study.

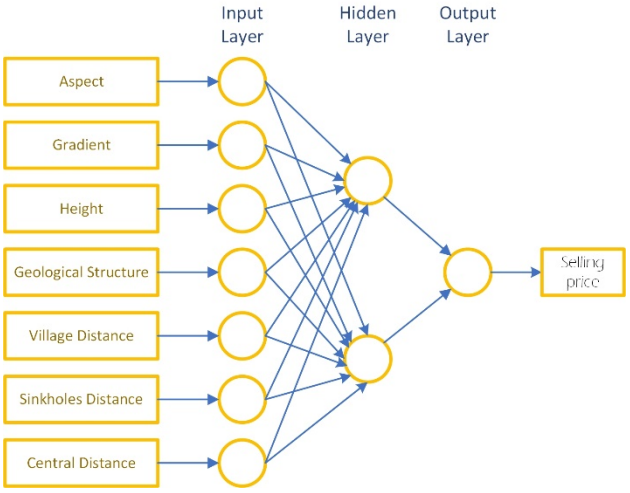


Figure 2. Neural network model

1. Input data layer
2. Hidden layer(s) is usually referred as black box, and
3. The output layer(s)

is the layer at which the value or the values of the land real-estate is calculated. Hidden layer(s) is formed from two sections namely weights' cumulative function and transformation function. Both of these functions are the connections of the values between the input data (factors affecting properties of the land real estate) and the output data (selling prices). The cumulative function of weights used as the neural network model in forwarding and backward feeding propagation is;

$$Y_j = \sum X_i W_{ij} \qquad (1)$$

In the above equation, Xi is the input value and Wi is the weight allocated to the input values in each of the hidden layer nodes. A transformation function depending on the values of output variable(s) or the cumulative value(s) of hidden layer(s) at Yi can be in terms of various types like linear functions, linear threshold functions, graded linear functions, sigmoid functions or Gaussian functions. Most of the software programs use regular sigmoid transformation function. For instance:

$$Y_T = \frac{1}{1+e^{-y}} \qquad (2)$$

This function is preferred due to its non-linearity, continuity, stability properties and characteristics presenting continuous variability [8].

III. APPLICATION

In this study, the artificial neural networks method was used in the evaluation of agricultural areas. Estimates of the value of real estate are made using artificial neural networks method. The reason for using ANNs in this study is the ability to solve complex problems. However, the method has some disadvantages. It is difficult to obtain successful results if the data set and the used parameters are not selected correctly With this method, the importance of the data set is great to achieve successful results. Data collected from the study area were generated in a homogeneous distribution and the data set was obtained.

Karapınar District of Konya was chosen as the study area. The selected region consists of approximately 8 neighborhoods. Approximately 150 plots on sale were found in this area and 120 data used in the model were selected. Land that does not fit the value of the region is not used in modeling. The information about these land has been obtained from real estate agents and landowners of the region. The data used are the purchase and sale values for the year 2017.

As a result of the regional surveys, the data set has been formed with the factors affecting the value of the agricultural lands and the unit price of the land in the purchase and sale. In the dataset used in the model, the criterion affecting 7 values is used as the input value and 1 unit of the unit price is used as the output value. The aspect, gradient, height, geological structure, distance (m) to villages, distance to Karapınar (m) and distance to sinkholes (m) were used as input values. Estimated selling price (m2) is also used as the output value. Table 1 also shows the maximum and minimum values of the inputs and outputs.

Table 1: Maximum and Minimum Values of Inputs and Outputs

Input/Output	MIN.	MAX.
Aspect	1	9
Gradient	3	12
Height	1000	1120
Geological Structure	1	6
Central Distance	0	33000
Village Distance	0	6000
Sinkholes Distance	0	15000
Real Estate Unit Value	6	18

The feature information of these criterion affecting the value is related to the geographical information system and the map of each criterion is created and analyzed through the maps. Factors affecting the value of agricultural land in the study area are shown in Figure 3.

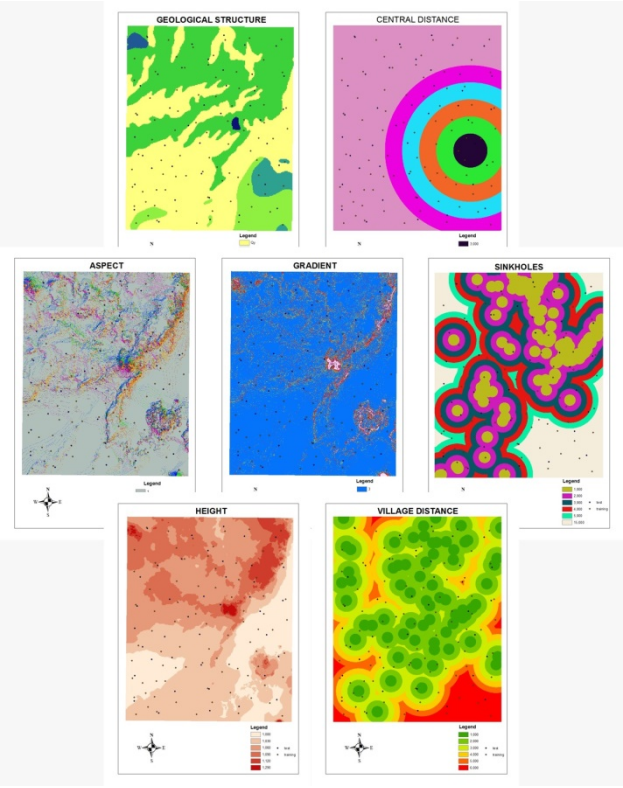


Figure 3. Map of the Factors Affecting the Immovable Value

The attribute data for each criterion affecting the value are normalized between 1 and 2 for use in the ANN method. The normalization between 1 and 2 required for does not seem to be worthy of the relevant property in the value estimation. As shown in (3) normalization equation is,

$$Xn = \frac{Xr - Xmin}{Xmax - Xmin} + 1 \qquad (3)$$

Here, Xr is the actual value of the feature. $Xmax$ and $Xmin$ are also the smallest and largest values of this feature. All normalized values are divided into two subgroups: 80 training and 40 test data for use in modeling. This dataset was prepared in the matrix format and applied in ANN method.

The accuracy of the method has been examined according to AA% and R^2

$$AA\% = \frac{\sum_{i=1}^{n}\left(1-\frac{|Xp-Xi|}{Xp}\right)}{n} \times 100 \qquad (4)$$

$$R^2 = 1 - \frac{\sum_i^n (x_p - x_i)^2}{\sum_i^n (x_i - \bar{x})^2} \qquad (5)$$

IV. Conclusion And Recommendations

The best fit curve (y = x line) between the sales values of the data used in the study area and the estimated sales value obtained from the model is shown. In a successful model, R^2 must be close to 1.The best fitting curve and the R^2 value for the model generated in this study are given in Figure 4.

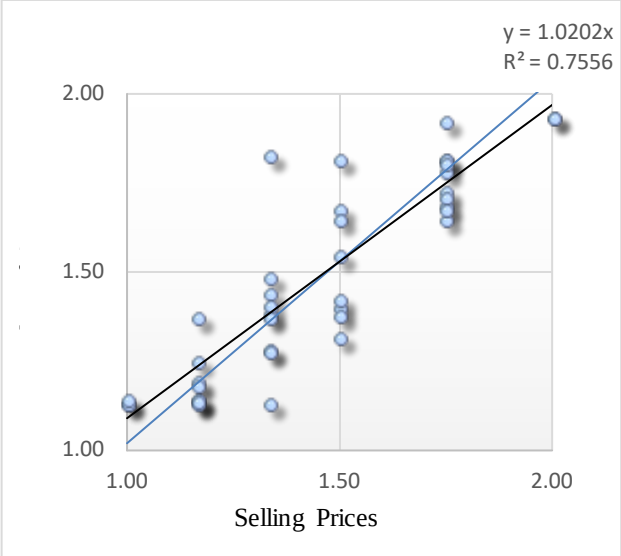


Figure 4. y = ax equation and R^2 value of ANN model

The mean affinity obtained in this model, which is applied by ANN method, was obtained with an absolute success of 93%. The training and test rates obtained from the model in the ANN result are also shown in Figure 5.

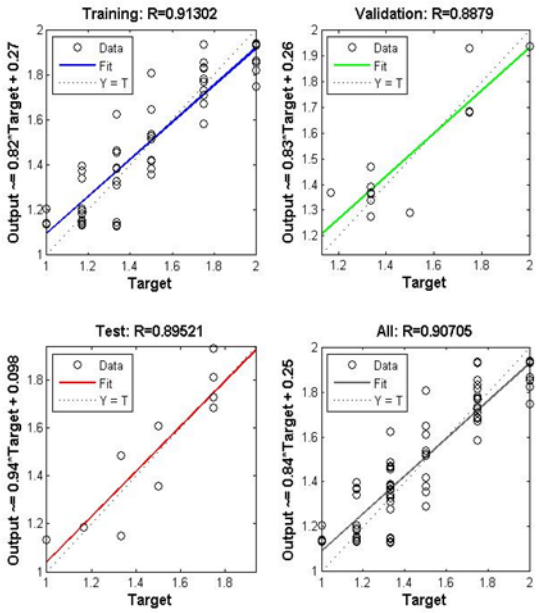


Figure 5. Return values of ANN model

As a result of this work, the people living in the region will be made conscious and the awareness that the formation of the sinkhole is also a worthy influence. At the same time, the

risk factor with this kind of valuation of other risk factors has also been shown to be subject in Turkey.

The method that will be used for the valuation of the real-estates should be selected by considering the habits of the region, types of the data and market conditions.

In this study, rapid and accurate results for the valuation of the real-estates were obtained by analyzing many data together with ANN method. The used method is accepted to be the preferable method for the valuation of the real-estates. However, the values obtained with ANN occurred closer to the market values in this study.

REFERENCES

[1] B. Vanderporten, "Timing of bids at pooled real estate auctions," *The Journal of Real Estate Finance and Economics*, vol. 5, pp 255-267, 1992.

[2] B. Nas Bulut, "YSA Ve DVM Yöntemleri İle Taşınmaz Değerlemesi İçin Bir Yaklaşım Geliştirme," M.Sc dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 2011.

[3] E. Pagourtzi, V. Assimakopoulos, T. Hatzichristos and N. French, "Practice Briefing Real Estate Appraisal: A Review of Valuation Methods," *Journal of Property Investment & Finance*, vol. 21-4, 2003.

[4] E. Güngör, "Gayrimenkul Değerlemesi ve Türkiye'de Sermaye Piyasalarında Gayrimenkul Ekspertiz Şirketlerine Yönelik Düzenlemeler Yapılmasına İlişkin Öneriler," T.C. Başbakanlık Sermaye Piyasası Kurulu Kurumsal Yatırımcılar Dairesi, Ankara, 1999.

[5] Ş. Yalpır, "Bulanık Mantık Metodolojisi ile Taşınmaz Değerlenme Modelinin Geliştirilmesi ve Uygulaması: Konya Örneği," PhD dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 2007.

[6] E. Deveci and I. Yılmaz, "Coğrafi Bilgi Sistemleri Yardımıyla Taşınmaz Mal Değerlemesi: Afyonkarahisar İl Merkesi Örneği," *Harita Teknolojileri Elektronik Dergisi*, vol. 1-1, pp. 33-47, 2009.

[7] M. K. Ölgen, "Yapay Zeka ve Coğrafya", International XII. Turkish Symposium on Artificial Intelligence and Neural Networks – TAINN, İzmir Türkiye, 2003.

[8] R.R. Trippi, E. Turban, "Neural Networks in Finance and Investment: Using Artificial Intelligence to Improve Real- world Performance," Neural Networks in Finance and Investment, Pobus, Chicago., 1993.

The Effects of Urban Transformation on Real Estates and Land Valuation at City Plans

S. YALPIR[1], S. SISMAN[1] and A. AKAR[1]

[1]Selcuk University, Konya/Turkey, sarici@gmail.com
[1]Selcuk University, Konya/Turkey, suleyman.sisman1@gmail.com
[1]Selcuk University, Konya/Turkey, utkuakar42@gmail.com

Abstract - The importance of immovable value in urban areas increases day by day. The immovable valuation is a planned, disciplined and a wide-ranging subject. In this study, the immovable value changes created by the changes in the properties of the buildings and the immovable properties of the urban structures from the past to the present day are examined. Plants and buildings, which form the basis of urban construction of the application, have been used together.

Keywords - Geographical Information Systems, Remote Sensing, Immovable Valuation

I. INTRODUCTION

There have been many developments in the field of technology in recent years. Computer technology is one of these areas. Innovations in computer technology have created new working environments and different application areas for many scholars. Geographic Information Systems (GIS) and Remote Sensing Systems (UA) are these areas. With the widespread use of GIS, current spatial data problems have arisen. The data collection methods currently in use are both costly and time consuming. Because of this reason, remote sensing data is an important resource for GIS. It is important that the two systems are used together in the decision support phase for the problems that need to be solved in the spatial sense. In order to be able to see the effect of planning, the position information and the values of the immovables must be determined objectively. Immovables have been subject to many discussions over the years and are considered to be one of the main areas of individual and public property in all countries. The fact that immovable assets are a long-term investment instrument, the ability to generate income, to use it as a security or as a counterpart, makes the immovables attractive (Erdoğdu, 2012). As a matter of fact, 56% of capital resources in the world is based on immovables (Bender, 1997). The immovable has passed through various phases from past to present. A land that was previously land has completed the development of urban land, urban land, and structured land.

Figure 1: Change of soil over time

When we look at these four stages, every universe eventually comes to value increments.

The land, is the piece of land that does not pass the zoning application with the shortest expression. The boundaries of the land are determined by hills, mountains, roads, rivers, walls, woods, roads or signs.

When this concept is analyzed from different sources, land according to TDK; part of the earth, land, land. A plot can be described as land in the general scope, but the land should not be considered as plot. (Ünel, 2017).

The land is the part of the land that enters the municipality's zoning practice area. According to TDK 's land, it is defined as a place reserved for construction on it. According to the Law on Reconstruction No.3194, the zoning plan (land) is the cadastral parcels in zoning islands, arranged according to zoning law and zoning regulations. municipal land is regarded as a land plots of land.

The properties of the land do not change and the usage rights recognized by the zoning remain stable as long as it is not unusual. However, this situation is different in the buildings. (Ünel, 2017). In this study, immovable exchanges between 1986-2005-2010-2014 years were examined using Nominal Valuation Method.

Selçuk, Şeker, Kılınçarslan and Hacıkaymak neighborhoods were selected as the study area in the Selçuklu province of Konya province.

II. IMMOVABLE VALUATION AND METHODS USED

Immovable valuation; residence, land, workplace, field, vineyard, garden, etc. real estate is the process of finding market value for a certain time under economic conditions.

Methods used in immovable valuation;

- Revenue Method
- Cost Method
- Peer Comparison Method
- Nominal Valuation Method

ıt shape.

Revenue methods, value estimation is done by analyzing the relations between the values of the immovables and the expected income of the holdings. The basic data in the method are net income and the return of the investment and the expectation of the buyer are calculated accordingly. In the case of the commercial real estate and rentable houses, the income method is preferred in order to determine the price.

The Cost Method is a method used in the evaluation of structures. The costs of construction and acquisition are calculated taking into consideration the buildings and other economic values on the immovable.

In the Peer Comparison Method, this comparative approach takes into account the sale of similar or substitute properties and related market data and makes comparative value appraisal. A generally valued immovable property compares to the sales of similar properties in the open market.

(UDES, 9.2.1.1). It is widely preferred to use the precedent comparison method in the evaluation of unconstructed land.
In areas with a large number of immovables, it is necessary to establish the value distributions between the value to be valued and the corresponding immovable assets. The unit to be taken as a basis for this value distribution may be the market value or the parametric values to be obtained with a scoring. By formulating the value criteria to be considered by this method, base and selling points are determined and a value coefficient expressing each immovable is calculated. These value coefficients show the status of the immovables relative to each other and are easily converted to fair value when necessary. Such an assessment has a much different approach than the "traditional" methods given above and is called "nominal" evaluation (Yomralıoğlu, 1995).

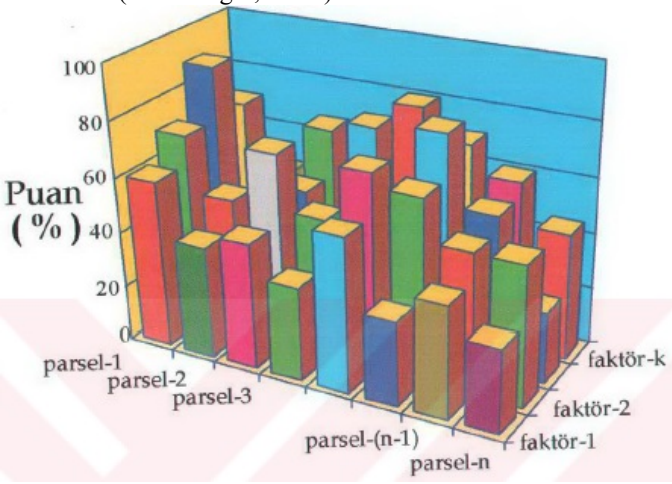


Figure 2: Nominal Valuation Method

In this study, nominal unit values for 6193 cadastral parcels and 1025 zoning parcels were calculated using ARCGIS software.
The goal of the project; taking into account the economic conditions in which our country is located, determine the physical development of the cities and the values of the immovables for the years 1986-2005-2010-2014.
In urban areas with remote sensing data; commercial area, industrial area, transportation network, recreational facilities, planned and unplanned areas and areas of equipment will be determined.

III. MATERIAL AND METHODS

I.MEETING OF THEIR WORK IN THE WORK

In this project, remote sensing data and vector data were used together.
Orthophoto images from 1986-2005 to 2010-2014 were used to review the past. The data were provided from Konya Metropolitan Municipality.
Using the ARCGIS software, buildings, roads, parks, education areas and chapels of the above-mentioned years have been drawn on individual layers with the help of orthophoto images.
Application Development Plan is also a geographical data provided by the Metropolitan Municipality of Konya. Implementation Plan has land parcels in the land-use areas. In the areas where application plan is not applied, it is only in the form of a construction island. In this study, it will be processed with the parcel parsel. Thus, only the layer of the parcel plot has been taken into consideration.
The development and cadastral parcels, which are important for the planning effect, are also prepared in different layers.

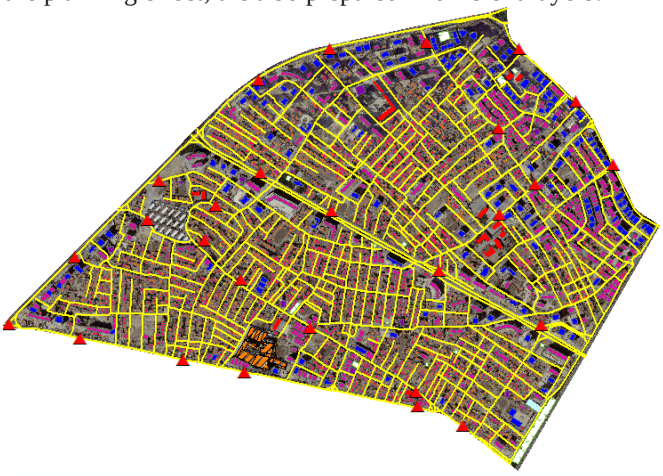


Figure 3: Drawing buildings and roads from orthophoto

II. DETERMINATION OF CRITERIA AND NEGLIGENCE ANALYSIS TO BE USED IN TEMPORARY VALUES

Geographic Information Systems allow you to work with data in the raster data structure. The smallest element of the data is the pixels. The values assigned to the pixels are the basis of immovable value maps.
Many criteria can be used in immovable valuation studies. The number of criteria can be increased in proportion to the available data. In this study, 12 criteria have been used for the nominal valuation method. Proximity analysis has been done for each of the criteria.
A literature review was performed for the clarity of proximity analysis. Commonly accepted distances have been used in the service domains of urban equipment (Accessible Distances).
Proximity analyzers are divided into five separate classes. Pixels score between 20 and 100 points.
When making a rating, a value of 0 to 100 for each pixel is assigned, depending on the distance to each pixel that affects the value. This value indicates the degree or quality of the pixel value factor.
For example, those who are 50 m away from the road are assigned a score for each pixel, such as 90 points for those with 100 points between 50-100.
In a sense, which measures 0 to 50 m in pixels, it means the value or quality of that pixel compared to other pixels. However, for negative factors this value is reversed
(Zeng ve Zhou, 2001).
For example, a scoring center will be scored from 0 to 100. So, places close to the criteria used take low values while receiving distances.
The points assigned to the Pixels as above have been loaded on cadastral and parcel plots. For each of these four years, a reclassified value map of each criterion has been produced (Figure 4)

Figure 4: Reclassified value maps of criteria for proximity analysis

IV. PRODUCTION AND RESULTS OF MORTGAGE VALUES WITH NOMINAL METHOD

The immovable value maps of the 4 selected ones were produced in ARCGIS software by nominal valuation method. Later on, these generated maps were used to analyze the percentage increase in value between years. (Figure 5).
When the differences between the years are examined, it is seen that the increase between 1986 and 2005 is 86%. This increase is due to the improvement of the economic situation as well as the emergence of new equipment areas over the years.

The value increase between 2005 and 2010 is 98%. When the increase is examined, it is seen that there is a value increase by one hundred percent. The most important reason for this is the emergence of development plans and the addition of new equipments.
The increase between 2010 and 2014 was set at 22%. The increase is less than the other years because of the fact that the years are close to each other as a timing. The hem is connected to the fact that the region has reached a saturated state as a reinforcement and has not shown a great improvement.
Taking into account the generally generated value maps, there is an ever-increasing value increase from 1986 to 2014.
However, this value increase is not evenly distributed for each parcel, and it is more in some areas.
In this study, the results obtained from the academic studies for the weights of analyzes and criteria are used. With this study, immovable value changes created by changes in urban constructions, reinforcements and immovable properties from past to present day are determined.
The valuation studies that are based on the mentioned points will be the important consequences for both the citizen and the state. An objective assessment can be made so that the criteria used to make healthier decisions and the standards for the advantages and disadvantages of these criteria will be provided. Urban area plans to be made up to and including this time will provide great benefits both for the city and for the country in terms of development. In order to increase these benefits and to be able to use them effectively, the values of the properties should be determined with scientific methods up to date and high.

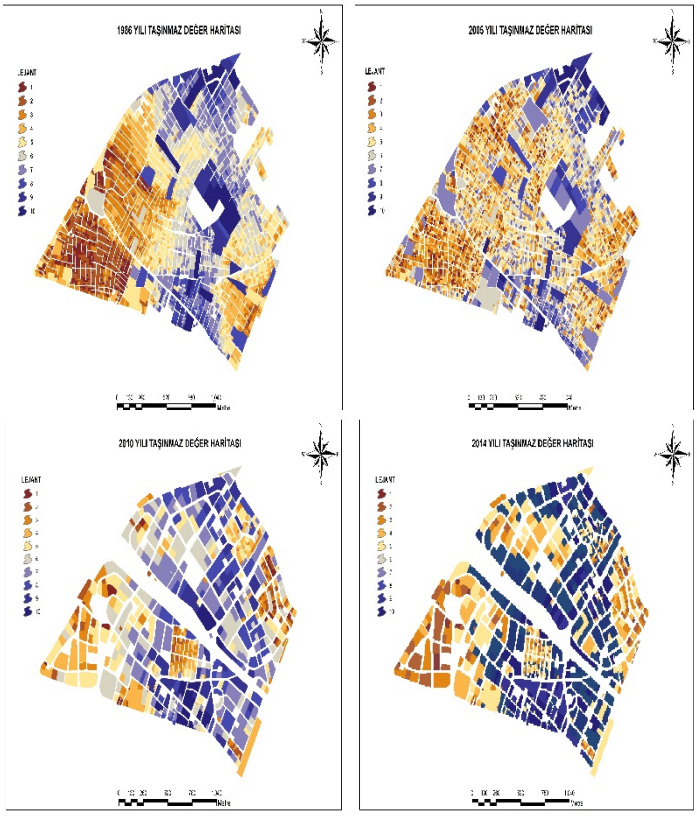


Figure 5: Immovable Value Maps

IV. CONCLUSION

There are studies for the purpose of immovable valuation in our country. The valuation studies that are performed are the basis for many studies and investments such as taxation, expropriation, privatization, land consolidation. However, the criteria used in these studies and the weights of these criteria differ from each other. This situation can cause an objective evaluation to be impossible. In other words, there must be standards about the criteria used to make healthier decisions and the advantages and disadvantages of these criteria.

In this study, the results of the academic studies done for the weights of the analyzes and the criterions are used. In this study, immovable value changes caused by changes in the urban structures, equipment and immovable properties from the past to the present day are determined. The valuation studies made by taking into consideration the above mentioned points will be important consequences for both the citizen and the state.

VI. ACKNOWLEDGMENT

Made in this study, Konya Metropolitan Municipality, which helps to work on the provision of data, we would like to thank the City Information System Branch.

REFERENCES

[1] 2499 (Mülga) Sayılı Sermaye Piyasası Kanunu.

[2] 3194 Sayılı İmar Kanunu. Kabul Tarihi: 3/5/1985, Yayımlandığı R. Gazete; Tarih: 9/5/1985, Sayı: 18749, Yayımlandığı Düstur; Tertip: 5, Cilt: 24, Sayfa: 378.

[3] Bender, A., Din, A., Favarger, P., Hoesli, M., & Laakso, J. (1997). An analysis of perceptions concerning the environmental quality of housing in Geneva. Urban Studies, 34(3), 503-513..

[4] Nişancı, R., 2005. CBS ile nominal değerleme yöntemine dayalı piksel tabanlı kentsel taşınmaz değer haritalarının üretilmesi. Doktora Tezi, Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Trabzon, 217.

[5] Ünel ,F. (2017) Taşınmaz Değerleme Kriterlerine Yönelik Coğrafi Veri Modelinin Geliştirilmesi(Doctoral dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü)

[6] Yalpır, Ş. (2007). Bulanık mantık metodolojisi ile taşınmaz değerleme modelinin geliştirilmesi ve uygulaması: Konya örneği(Doctoral dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü)

[7] Yomralıoğlu, T. (1995). Taşınmazların Değerlendirilmesi, Ders notaları. KTÜ, Trabzon.

[8] Zeng, T. Q., & Zhou, Q. (2001). Optimal spatial decision making using GIS: a prototype of a real estate geographical information system (REGIS). International Journal of Geographical Information Science, 15(4), 307-321. Motorola

Application Artificial Neural Network in Mass Real Estate Appraisal for Centre Neighborhood of Konya

F. B. UNEL[1] and S. YALPIR[2]

[1]Konya/Turkey, fatmabunel@gmail.com
[2]Selcuk University, Konya/Turkey, sarici@selcuk.edu.tr

Abstract **- Mass appraisal of real estate is a complex problem because of multiple criteria. Developed technologic methods provide a solution by transforming into the form of simple and easy. Multiple Regression Analysis (MRA) is frequently used in academic and practice studies in the world and our country. New method quest continues according to recent advances in computer technology. Artificial Neural Network (ANN) which is one of the artificial intelligence methods should investigate in order to use in the valuation because it can imitate human brain.**

The aim of this study is to estimate with ANN method the value of real estate. The study data consisted of the market samples concerning the plots in Centre Neighborhood of Konya. The data of the 558 samples were collected as main headings in the form of Legal, Physical, Locational and Neighborhood Features. The data set was separated for 70% training and 30% test. The market values of the plots were forecasted by using 70% training data with ANN and MRA methods and compared by testing 30% data in both methods. According to the results, it has seemed that success of ANN method is higher than the success of MRA method.

Keywords **- Mass appraisal, real estate valuation, artificial neural network, multiple regression analysis.**

I. INTRODUCTION

FULFILMENT of all needs from the world scale to the household scale demand on the basis of needs depends on the supply cycle in a measurable amount. Land, which is a scarce resource shouldn't use up quickly by providing supply-demand balance. That is, it is necessary to plan the land usage and to evaluate every aspect of these plans by considering the cost/benefit and without disrupting the balance of natural life. Thus, the valuation of land is of great importance as the land is the basis for the need for sheltering apart from agriculture, industry, and underground resources. While traditional valuation methods used today producing momentary solutions, the mass appraisal is needed by using modern valuation methods [1].

There is no mass real estate appraisal system and infrastructure in our country. Establishment of this structure depends on the determination of the criteria affecting the real estate value at the optimum level and standard. Because criteria affecting the value of real estates are matters of the existence of objective and subjective elements and there are many criteria that vary in micro and macro scale [1]. In the previous study, it was determined criteria as standard in optimum level and the base of mass appraisal was created.

In the literature, it has been observed that ANN is used for estimation of residential values and MRA methods are generally used for verification, as well [2-8]. It was used ANN and MRA as mass valuation models for land consolidation in South Cyprus [9, 10]. These studies were compared and combined with a geographical information system (GIS) [11]. In addition, it was seen that data set for ANN and MRA was divided into four datasets due to circles of supply and demand and studied in four different regions [12]. It is determined that some development countries such as Northern Ireland, Spain used ANN methods [13, 14].

The purpose of this study is to carry out ANN one of the advanced valuation methods to the data of the market samples coming from the plots in the central districts of Konya. The output is the value of market samples and input consists of the criteria for Legal, Physical, Locational and Neighborhood features with relation to these samples. While the Legal and physical futures of the criteria take part in the dataset as standardized separately, the criteria of Locational and Neighborhood Features are standardized and simplified as Neighborhood Index and Location Index. These index values were added to the data set considering the weighted average of each market sample. ANN and MRA methods have compared under equal conditions. It has seemed that success of ANN method is higher than the success of MRA method.

II. MATERIAL AND METHOD

A. Study Area and Market Samplings

The central neighborhoods of Konya province are the study area and, the plot is addressed as real estate type. In the given neighborhoods, the number of samples from plots which are mentioned the purchase and sale was calculated as 382 in the 95% confidence interval and 5% the margin of error, and 558 market sampling has reached (Figure 1).

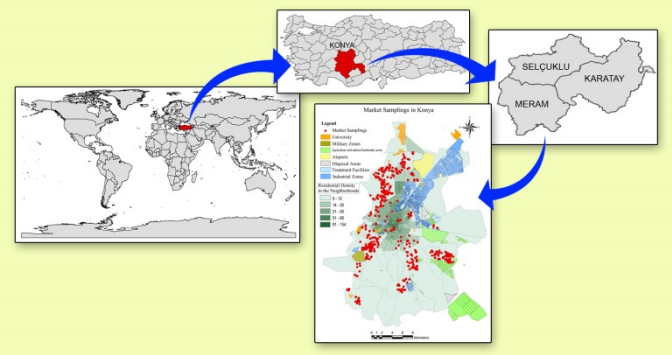


Figure 1: Market samplings in study area.

B. Data

Land, plot, residential, etc. real estates have many different values from each other. Market value, tax value, expropriation value are diverse like the value determined by the court. To be used in the verification phase of this study, there is a need for the reasonable (fair) value that the dependent variable, rent, which is stable, that with the free will of purchaser and seller and that about the plot is mutually known everything [15]. Institutions and organizations that produce and keep record plot values in our country;

- Directorate General of Land Registry and Cadastre (Title deed value),
- Municipalities (Tax value, expropriation value),
- Appraisers Association of Turkey (Market and hypothec value),
- TR Central Bank (Market and hypothec value),
- Courts (Market value),
- Local Real Estate Agents (Market value),
- General Directorate of National Real Estate,
- Revenue Administration,
- Directorate of Execution

and in addition to all the other expropriation agencies, valuation and insurance companies can be sorted. Despite the fact that there are many institutions that produce and record value, it is very difficult to reach healthy value knowledge. The purchaser and the seller must pay taxes every year with the fees and insurance in transfer transactions over the percentages of the reasons stated by the law, which is the biggest reason why the seller does not declare the true value that agreed between them. Moreover, the fact that the official value record based on the sales history does not exist, the ones that do not reflect the reality, the ones that reflect the truth cannot be shared due to the confidentiality procedures, and the decisions about the deficient of the administration and management regarding the real estate have caused to could not be given as objectively. In addition to these, mass appraisal studies have brought to a halt.

Features and values of real estates are not available in their database in the current, accurate and reliable database. In the development of advanced valuation methods for mass real estate appraisal, there is a need for a database which is composed of data in digital and standard format. In this study, the values of 558 market samples which occurred of the purchase and sale prices in the market and they were obtained from the local real estate purchase and sale offices and the announcements on the internet in June-December 2016 period.

Criteria affecting the plot value were grouped [1, 16, 17] in previous studies as legal, physical, locational and neighborhood features, and the data of the plots where the values by collected were determined and arranged in a matrix format.

C. Method

The methods of real estate valuation are classified as traditional (comparison, income and cost), statistically (multiple regression, nominal and hedonic), spatial analysis and advanced valuation methods (artificial intelligence and hybrid techniques. Advanced Valuation Methods [18];

- Artificial Intelligence Techniques
  - Artificial Neural Networks (ANN)
  - Genetic Algorithms (GA)
  - Rough Set Theory (RST)
  - Expert systems
  - Fuzzy Logic (FL)
- Hybrid Techniques
  - Neuro-Fuzzy System
  - Genetic-Fuzzy System

Artificial Neural Networks (ANN)

ANNs are computer systems whose microprocessors, rather than laid out in series as in traditional computers, are connected in parallel, forming layers and making multiple connections, imitating the way the neuronal network is organized in the brain [19].

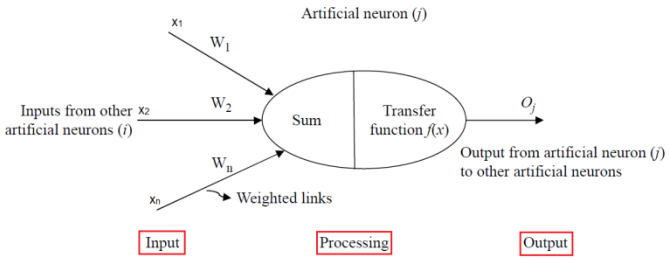
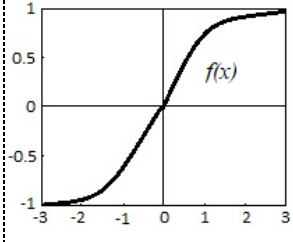


Figure 2: The process structure of ANN [20].

ANN was used to estimate market values of the plots and each neuron was represented by a circle. ANN has an input layer, a hidden layer, and an output layer (Figure 2). Signals in the ANN feed forward from left to right [20].

In this study, activation function [2, 21, 22] was used Equation (1) and the number of inputs and hidden layers were 15.

Tangent Hyperbolic	***Tansig(x)***
	$f(x) = \frac{1 - e^{-2x}}{1 + e^{2x}}$ (1)

Multiple Regression Analysis (MRA)

As the model verification method, Multiple Regression Analysis (MRA), which is the most prevalent, most used and statistical methods in the literature, is used. MRA is a statistical analysis technique that derives and predicts a linear mathematical equation between dependent and multiple independent variables [18, 23]. The mathematical model of MRA is represented by Equation (2) in the form of [24];

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i \quad (2)$$

y_i : The dependent variable *(Market* values*)*,
$x_{i1}, x_{i2}, \ldots x_{ik}$: The independent variables [ownership stake, area, Basement Area Coefficient (BAC), Floor Area Coefficient (FAC), the number of floors,…],
u_i : The disturbance or error term,
β_0 : Constant coefficient and
$\beta_1, \beta_2, \ldots \beta_k$: Variable coefficients.

III. APPLICATION

Mapping was done in ArcGIS software by multiplying the values of the neighborhood and locational features where in itself of the criteria in the map with the weights they obtained from the previous survey study. Separate values were obtained for the standard neighborhood and location that will be the base for the real estate valuation and named Location Index and Neighborhood Index.

The data set consisted of the samples that plots were known the purchasing-selling values in the market. The data set is arranged in form of Legal (ownership stake, BAC, FAC, the number of floors, building order and area), Physical [parcel status (corner/break), length of the frontage, number of the frontage, geometric shape, infrastructure, road type and road width), Locational (Location Index) and Neighborhood (Neighborhood Index) features belong to market samples. The number of independent variables/inputs were 15 and the number of dependent variable/output was single.

The 558*16 matrix size dataset by normalized and the value ranges and units of all criteria have standardized. 70% of 558 market samples have allocated for training, 30% for testing. While the value is considered as output, other criteria are included in the analysis as input. Both of ANN and MRA methods were produced with training data. Using this model in test data, the market value was estimated and performance analysis has carried out.

A. ANN of Training Data

ANN structure was created in Matlab R2013a software according to criteria and values for this study area. The best results in ANN method were obtained with 1000 epoch. It was observed that these data output variable is linear. R values of training, validation, test and all data were calculated and presented in Figure 3.

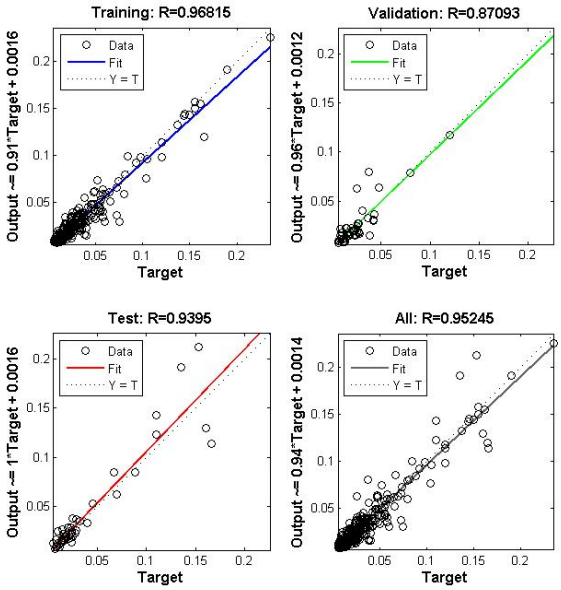


Figure 3: R values and output distribution in ANN.

B. MRA of Training Data

The data, which were separated 70% of the data set for training in ANN, were used in order to generate the mathematical model in MRA. The other data, which were separated 30% of the data set for testing in ANN, were used in order to test mathematical model of MRA. The model and coefficients can be seen in the following Equation (3).

MRA output=(-.103)+.027* ownership stake-.049* BAC+.110 * FAC+.032* the number of floors+.012* building order+ 9.996* area-.012* parcel status-.028* length of the frontage +.033* the number of the frontage+.026* geometric shape+.016* infrastructure-.010* road type+.057* road width +.027* Neighborhood Index+.012* Location Index (3)

In accordance with the coefficients, it was seemed that the area of plot was very important. It is also the case in literature and practices. BAC and FAC were found reverse sign in Equation (3), because they were related to each other. It was calculated output of MRA by testing the model by Equation (3) and applied performance analysis.

C. Performance Analysis of test data in ANN and MRA

The mean average percentage error [MAPE in Equation (4)], the root mean square error [RMSE in Equation (5)] and the mean absolute error [MAE in Equation (6)] were calculated in order to determine the models' performances comparatively [2, 20]. MAPE, RMSE, MAE, R^2 and linear equations were presented as performance analyses of test data for ANN and MRA (Figure 4).

$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\lvert y_i - \hat{y}_i\rvert}{y_i}$ (4)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$ (6)
$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$ (5)	y_i : Market values, $\hat{y}_i$: Model values, $i = 1, 2, 3, \ldots . n$ n : The number of the samples

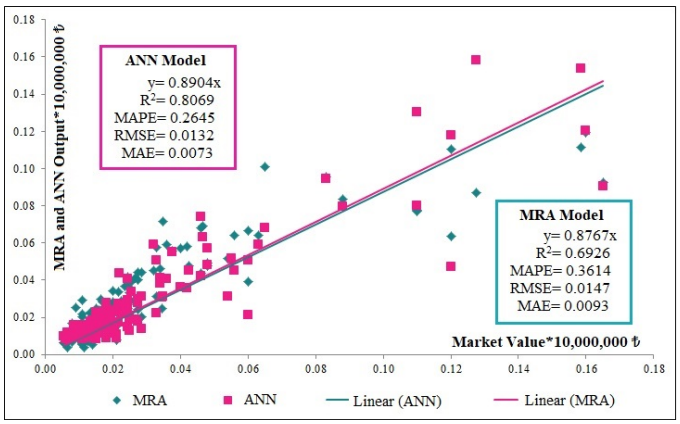


Figure 4: Performance analysis.

Point distributions and the linear line between the market values of plots in test data and outputs of ANN and MRA models were examined. In accordance with R^2 values of test data, ANN had 81% accuracy and MRA had 69% accuracy. Error rates of ANN were lower than error rates of MRA. For this reason, it is determined that ANN was successful according to MRA (Figure 4).

IV. CONCLUSION

Real estate has an important place for countries, institutions and citizens. It is strong on the direction of the economic because of processes of taxation, expropriation, planning, mortgaging, purchasing and selling. It needs to know value of real estate in order to provide especially social justice in these processes. Mass real estate appraisal should made by using the criteria and the market values which are far from subjectivity.

In this study, the plots, which are one of real estate types, were handled and the mass plot valuation was applied with ANN and MRA methods. The most important difference from other applications is to perform ANN with the criteria included Location Index and Neighborhood Index. When found under the same condition, performances of ANN and MRA were examined; ANN results were more successful than MRA results. In ANN method, the criteria reflected 80% of the market value, residue 20% come from changed economic, politic status and other reasons.

In next study, value maps that enable visual analyses will generated with ANN and MRA outputs for Centre neighborhood in Konya. It is suggested applying hybrid techniques by reducing the criteria.

ACKNOWLEDGMENT

The authors would like to thank pollsters who made efforts and participants who patiently completed the questionnaire in the data collection phase of the survey. This study is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) with 115Y769 Project Number and the Coordinator of Selçuk University's Scientific Research Projects with 15101008 Project Number.

REFERENCES

[1] F. B. Ünel, "Taşınmaz Değerleme Kriterlerine Yönelik Coğrafi Veri Modelinin Geliştirilmesi", Doktora Tezi, Harita Mühendisliği A.B.D., Selçuk Üniversitesi, Konya, 2017.

[2] S. Yalpir, S. S. Durduran, F. B. Unel and M. Yolcu, "Creating a valuation map in GIS through artificial neural network methodology: A case study", *Acta Montanistica Slovaca*, vol.19, pp. 89–99, 2014.

[3] E. Saraç, "Yapay sinir ağları metodu ile gayrimenkul değerleme", Yüksek Lisans Tezi, İnşaat Mühendisliği A.B.D., İstanbul Kültür Üniversitesi, İstanbul, 2012.

[4] A. Khalafallah, "Neural network based model for predicting housing market performance", *Tsinghua Science And Technology*, vol.13, pp. 325–328, 2008.

[5] P. Rossini, "Application of Artificial Neural Networks to the Valuation of Residential Property", in *Third Annual Pacific-Rim Real Estate Society Conference*, 1997, New Zealand.

[6] H. Selim, "Determinants of house prices in Turkey: Hedonic regression versus artificial neural network", *Expert Systems with Applications*, vol.36, pp. 2843–2852, 2009.

[7] N. Nguyen, and A. Cripps, "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks", *The Journal of Real Estate Research*, vol.22, pp. 313–336, 2001.

[8] A. Mimis, A. Rovolis, and M. Stamou, "Property valuation with artificial neural network: the case of Athens", *Journal of Property Research*, vol.30, pp. 128–143, 2013.

[9] D. Demetriou, "A spatially based artificial neural network mass valuation model for land consolidation", *Environment and Planning B: Urban Analytics and City Science*, vol.44, pp. 864–883, 2017.

[10] D. Demetriou, "The assessment of land valuation in land consolidation schemes: The need for a new land valuation framework", *Land Use Policy*, vol.54, pp. 487–498, 2016.

[11] D. Demetriou, "Automating the land valuation process carried out in land consolidation schemes", *Land Use Policy*, vol.75, pp. 21–32, 2018.

[12] I. C. Yeh, and T. K. Hsu, "Building real estate valuation models with comparative approach through case-based reasoning", *Applied Soft Computing*, vol.65, pp. 260–271, 2018.

[13] A. Barańska, "Real estate mass appraisal in selected countries – functioning systems and proposed solutions", *Real Estate Management and Valuation*, vol.21, pp. 35–42, 2013.

[14] Ü. Yıldız, "Gayrimenkul bilimlerinde kitlesel değerleme uygulamaları ve Türkiye için model önerisi", Yüksek Lisans Tezi, Taşınmaz Geliştirme A.B.D., Ankara Üniversitesi, Ankara, 2014.

[15] UDES, *Sermaye Piyasasında Uluslararası Değerleme Standartları Hakkında Tebliğ*. Seri: VIII, No: 45, Tarih: 1/5/2006.

[16] F. B. Unel, and S. Yalpir, "Positional determination of real estates with Analytic Hierarchy Process", in *Proceedings of the Fourth International Conference on Mathematical and Computational Applications*, 2013, Manisa, pp. 326–336.

[17] Ş. Yalpır, and F.B. Ünel, "Türkiye 'de ve Uluslararası çalışmalarda arsa değerlemede kullanılan kriterlerin irdelenmesi ve Faktör Analizi ile azaltımı", *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, vol.16, pp. 303–322, 2016.

[18] F. B. Unel, and S. Yalpir, "Grouping and analyzing of real estate valuation approaches" *International Journal of Multidisciplinary Thought*, vol.3, pp. 171–182, 2013.

[19] J. G. Mora-Esperanza, "Artificial intelligence applied to real estate valuation: An example for the appraisal of Madrid", *CT-Catastro*, pp. 255–265, 2004.

[20] C. C. Lin, and S. B. Mohan, "Effectiveness comparison of the residential property mass appraisal methodologies in the USA", *International Journal of Housing Markets and Analysis*, vol.4, pp. 224–243, 2011.

[21] B. N. Bulut, "YSA ve DVM yöntemleri ile taşınmazdeğerlemesi için bir yaklaşım geliştirme", Yüksek Lisans Tezi, Elektronik ve Bilgisayar Sistemleri Eğitim A.B.D., Selçuk Üniversitesi, Konya, 2011.

[22] S. Canan, "Yapay sinir ağları ile GPS destekli navigasyon sistemi", Doktora Tezi, Elektrik-Elektronik Mühendisliği A.B.D., Selçuk Üniversitesi, Konya, 2006.

[23] E. Pagourtzi, V. Assimakopoulos, T. Hatzichristos, and N. French, Real estate appraisal: A review of valuation methods", *Journal of Property Investment & Finance*, vol.21, pp. 383–401, 2003.

[24] R. Altunışık, R. Coşkun, S. Bayraktaroğlu, and E. Yıldırım, Sosyal Bilimlerde Araştırma Yöntemleri SPSS Uygulamalı. Sakarya: Sakarya Yayıncılık, 2010.

Comparison of Pixel Based And Object Based Classification Methods on Wetland Areas: Example of Aslantaş Dam Lake

M. H. KESIKOĞLU[1], S. Y. ÇIÇEKLI[2], T. KAYNAK[1]*

[1]Erciyes University, Kayseri/Turkey, hayrikesikoglu@erciyes.edu.tr
[2]CukurovaUniversity, Adana/Turkey, yoturanc@cu.edu.tr
*Corresponding Author, e-mail address: tolgakaynak@erciyes.edu.tr

Abstract **- By the development of technology, image classification algorithms frequently use to identify land use and land cover of any area in remote sensing studies. Due to the diversity and complexity of land cover on the wetland areas, it is quite difficult to obtain accurate results related to the earth's surface. The main purpose of this research is to compare the overall accuracies of object based and pixel based image classification methods. Arslantaş Dam Lake is structured on Ceyhan River for irrigation, flood control and electricity generation in Osmaniye province. In this study, Landsat-8 LDCM satellite image of Aslantaş Dam Lake with spatial resolution of 30m, acquired on December 29, 2017 was used. Firstly, image was classified by pixel based classification with support vector machines (SVM) method. After that, image was reclassified by object based classification with K-nearest neighbour (KNN) method. Five classes namely lake, agricultural area, soil, vegetation and building area were determined by using these algorithms. Ground truth data were gathered from aerial photographs, available maps and personal informations. Finally, overall accuracies of these methods were compared. It is observed from the classification results that object based KNN method provide higher accuracy than the other classification method.**

Keywords **- Support vector machines, k-nearest neigbour, remote sensing, Landsat 8 LDCM**

I. INTRODUCTION

IMAGE classification algorithms are often used in many disciplines. Remote sensing is one of the most preferable discipline on determination of land use and land cover, wetland management, ecological studies, and monitoring of deformations.

In this study, object-based and pixel based classification methods are compared on wetland area. There are many studies about comparing image classification methods in literature. Cengiz et al. [1] compared object based with pixel based image classification. Object based image classification gave better results. Gholoobi et al. [2] performed pixel-based and object-based classification to determine land use / land cover in mountainous regions, and object-based classifications gave more accurate results. Zhang et al. [3] performed pixel based and object based classification method in the Heine River basin by using Spot-5 images. They observed that object-based classification gave more accurate results. Al Fugara et al. [4] made pixel based and object based classification in the Klong valley of Malaysia by using Landsat-7 images. As a result of the work, object-based classification provided better results than pixel-based classification.

The aim of this study is performing object based image classification with K-nearest neighbour (KNN) and pixel based image classification with support vector machines (SVM) methods. We also aim to compare the accuracy of these methods at Aslantaş Dam Lake in the Turkey based on analysis of Landsat 8 LDCM imagery.

II. STUDY AREA

Aslantaş Dam is located in Osmaniye and it is onto the Ceyhan River. The dam was built between 1975 and 1984 for irrigation, flood control and electricity generation. The dam supplies irrigation to a very large area, and the hydroelectric power plant with a capacity of 138 MW supplies 569 GWh of electricity per year [5].

Figure 1: Aslantaş Dam Lake Landsat-8 LDCM image

III. MATERIAL AND METHOD

A. Materials

In the study, Landsat 8 LDCM satellite image of Aslantaş Dam Lake in 2017 was used. Satellite image was taken on December 29h. In Table 1, Landsat 8 bands and their wavelength and resolution values are given in detail.

Table 1: Landsat-8 LDCM OLI / TIRS Satellite Bands and Features [6]

Band	Wavelength (micrometers)	Resolution (meters)
Band-1 Ultra Blue (Coastal Aerosol)	0.43-0.45	30
Band-2 Blue	0.45-0.51	30
Band-3 Green	0.53-0.59	30
Band-4 Red	0.64-0.67	30
Band-5 Near Infrared (NIR)	0.85-0.88	30
Band-6 Shortwave Infrared (SWIR-1)	1.57-1.65	30
Band-7 Shortwave Infrared(SWIR-2)	2.11-2.29	30
Band-8 Panchromatic	0.50-0.68	15
9 Cirrus	1.36-1.38	30
Band-10 Thermal Infrared (TIRS-1)	10.60-11.19	100*30
Band-11 Thermal Infrared(TIRS-2)	11.50-12.51	100*30

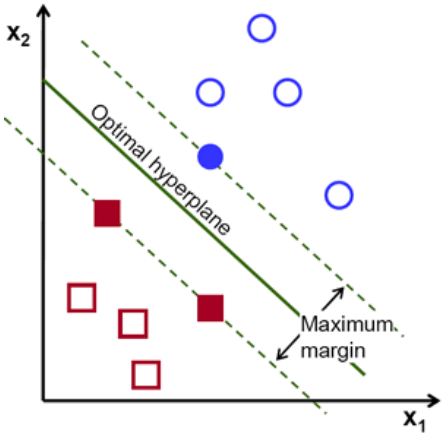
B. Methods

In this study support vector machines and k-nearest neighbour classification methods were used.

1) Support Vector Machine Method (SVM)

In this study pixel based classification was made by using support vector machines method. Support vector method is supervised classification method and learning technique.

Support vector machine, which separates the two class lines by a boundary is an extended linear class. The SVM passes a linear boundary from a plane and provides all optimizations [7].



. Figure 2: Support vector machines method[8]

Datasets are $\{x_i y_i\}$ $(i = 1, \ldots . k)$ for $x \in R^N$ is an N-dimensional space and $y \in \{-1 + 1\}$. For x is point on hyperplane, this hyperplane is defined as

$$w \times x_i + b = 0 \quad (1)$$

A separating hyper plane can be defined for two classes as [9]:

$$w \times x_i + b \geq 1 \text{ for all } y = +1 \quad (2)$$

$$w \times x_i + b \leq 1 \text{ for all } y = -1 \quad (3)$$

In literature, many studies were made about support vector machines. Bahrambeygi et al. compared support vector machines and artificial neural network classification method in ophiolitic regions. In the result of the classifications, 52% accuracy rate was obtained in support vector machine and 65% accuracy rate was obtained in neural network method [7]. Pham et al. determined changes of biomass mangrove forests in the Cangio region in Vietnam by using support vector machine classification method. The overall classification accuracy for the images were 77.1% and 82.9% [10].

2) K-Nearest Neighbour Method (KNN)

In this study object based classification was made by using K-Nearest Neighbour method.

K-nearest neighbors (KNN) method is a non-parametric

approach to estimate on the basis of similarities between the point and the observed other variables [11].

A database measured (D) has been established. K is set to the nearest neighbor number. The value is usually determined experimentally. Feature vector of M (m1, m2 ... mn) are calculated for each point and created. The distances between each point of M' and D collection are (M): dist (M', M). Euclidian Distance widely used is shown in equation [12].

$$d(M',M) = \sqrt{\sum_i^n (m_i - m_j)^2} \quad (4)$$

According to the calculated distance, the nearest K point, D, is selected and a new data is collected according to the above equation (D_k). The output value of M is used the following equation according to D_k is calculated.

$$M' = \frac{\sum_{(i=1)}^{K} M_j}{K} \quad (5)$$

In literature, there are many studies about K-nearest neigbour classification method.

Finley and McRoberts used K-nearest neigbour method for multi-source forest attribute mapping.[13]. Thessler et al. used k-nearest neigbour for classification Landsat images. They analysed to classify rain forest types in Costa Rica with overall accuracy 91% [14].

IV. Application and Results

The flowchart of the study was shown in Figure 3.

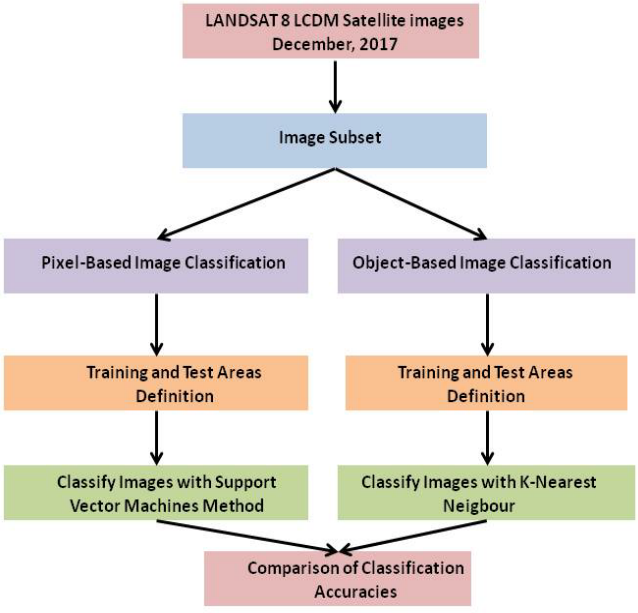


Figure 3: Flowchart of this study

In this study, subset of Landsat-8 LDCM image was made in first step

Figure 4: Subset images of Aslantaş Dam Lake image

After that Pixel based image classification was done. Support vector machines method was used for this. By this classification, five classes were created which were called lake, agricultural area, soil, vegetation and building area. Classified images of Aslantas Dam Lake are shown in Figure 5.

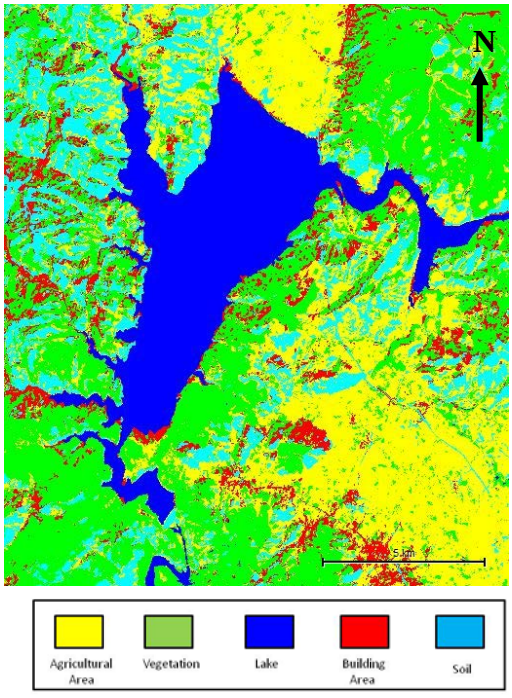


Figure 5: Classification images of Aslantaş Dam Lake image by Support Vector Machines

Segmentation is the first step for object based classification so segmentation was made before k-nearst neigbour classification method. Multiresolution segmentation method

was used.

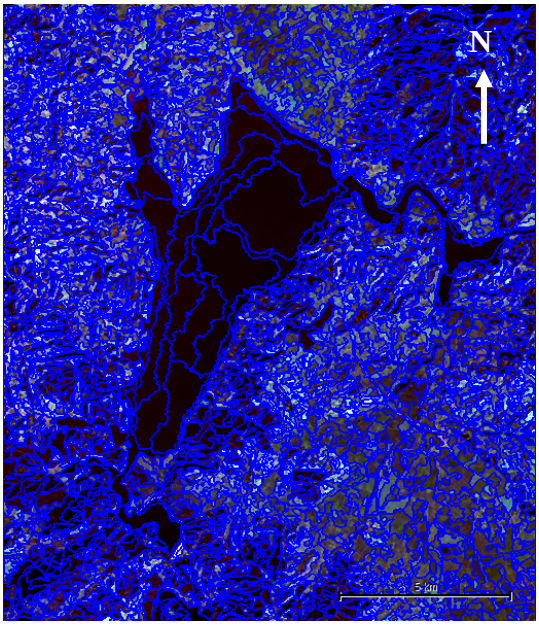


Figure 6: Segmentation images of Aslantaş Dam Lake Multiresolution Segmentation method

After segmentation object based image classification was done. K-nearest Neighbour classification method was used. By K- nearest neighbour classification, five classes were created which were called lake, agricultural area, soil, vegetation and building area. Classified images of Aslantas Dam Lake are shown in Figure 7.

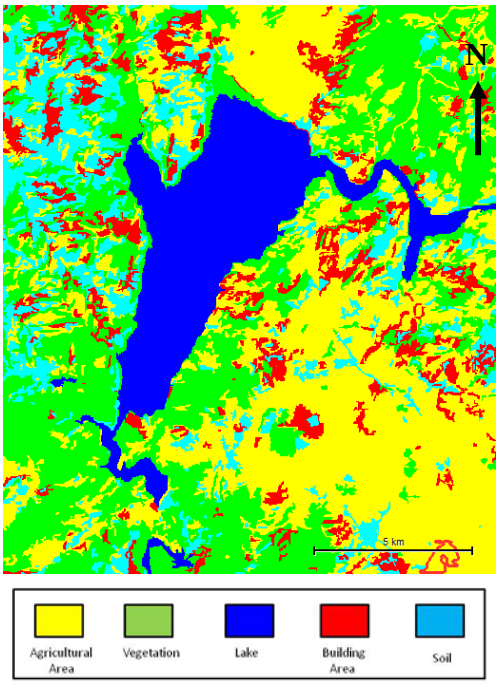


Figure 7: Classification images of Aslantaş Dam Lake image by K-Nearest Neighbour Classification Method

In the third step, the image classification accuracies were determined for support vector machines method and K-nearest Neighbour methods. It was observed 87% overall accuracy for support vector machines method. It was observed 99% overall accuracy for K-nearest neigbour method.

V. CONCLUSION

In this study, both pixel-based with SVM and object-based with KNN image classification methods were used to determine the water areas in the dam and the results were evaluated.

The classification accuracies obtained with the KNN and SVM classification methods. KNN method provided higher accuracy than SVM method.

REFERENCES

[1] S. Cengiz, S. Görmüş and A. Ateşoğlu, "Uzaktan algılama aracılığıyla tarımsal peyzaj karakterizasyonu," *5. UZAL-CBS Sempozyumu*, İstanbul, 2014.

[2] M. Gholoobi , A. Tayyebi, M. Taleyi, and A. H. Tayyebi "Comparing pixel based and object based approaches ın land use classification ın mountainous areas," *International Archives of the Photogrammetry. Remote Sensing and Spatial Information Science*, Volume XXXVIII., Part 8, Kyoto, Japan, 2010.

[3] H. Zhang, Q. Li, J. Liu, X. Du, T. Dong, and H. McNairn, C. Champagne, M. Liu, J. Shang, "Object-based crop classification using multi-temporal SPOT-5 ımagery and textural features with a random forest classifier", *Geocarto Interratical*, 2017.

[4] A. M. Al Fugara, B. Pradhan, and T. A. Mohamed, "Improvement of land-use classification using object-oriented and fuzzy logic approach", *Applied Geomatics*. 1(4), 111, 2009.

[5] Aslantaş Barajı [online] avaiable: http://www2.dsi.gov.tr/baraj/detay.cfm?BarajID=86

[6] What are the band designations, [online] avaiable: http://landsat.usgs.gov/band_designations_landsat_satellites.php

[7] B. Bahrambeygi, and H. Moeinzadeh, "Comparison of support vector machine and neutral network classification method in hyperspectral mapping of ophiolite mélanges–A case study of east of Iran", The *Egyptian Journal of Remote Sensing and Space Science*, 20(1), 1-10,2017.

[8] (2018) Introduction To Support Vector Machines . [online]. Available: https://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

[9] T. Kavzoğlu and I. Çölkesen, "A kernel functions analysis for support vector machines for land cover classification", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 11, pp.352–359,2009.

[10] L. T. Pham, and L. Brabyn, L. , "Monitoring mangrove biomass change in Vietnam using SPOT images and an object-based approach combined with machine learning algorithms", *ISPRS Journal of Photogrammetry and Remote Sensing*, *128*, 86-97,2017.

[11] R. E. McRoberts, M. D. Nelson, and D. G. Wendt, "Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique", *Remote Sensing of Environment*, *82*(2-3), 457-468,2002.

[12] L. Sang, Y. X, R. Cao, Y. Chen, Y. Guo, and R. Xu, "Modeling of gan hemt by using an improved k-nearest neighbors algorithm", *Journal of Electromagnetic Waves and Applications*, *25*(7), 949-959, 2011.

[13] A. O. Finley, and R. E. McRoberts, "Efficient k-nearest neighbor searches for multi-source forest attribute mapping.", *Remote Sensing of Environment*, *112*(5), 2203-2211,2008.

[14] S.Thessler, S. Sesnie, Z. S. R. Bendaña, K. Ruokolainen, E.Tomppo, and B. Finegan, "Using k-nn and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica", *Remote Sensing of Environment*, *112*(5), 2485-2494,2008.

Artificial Neural Network Model Design for Daily Demand Prediction in Mass Meal Production

D. YERGÖK[1], C. T. GÜVEN[1] and M. ACI[1]

[1] Mersin University, Mersin/Turkey, deryayergok@gmail.com
[1]Mersin University, Mersin/Turkey, ctamerguven@gmail.com
[1] Mersin University, Mersin/Turkey, maci@mersin.edu.tr

***Abstract* – In this study, four artificial neural network models are designed for the daily meal demand in places where mass meals are produced. It is aimed to avoid the cost of overproduction in the places where mass meals are produced and at the same time to prevent the ending of the meals with less food production. In the study, the number of people eating food was estimated by using the data obtained from the university dining hall. Feed-forward neural network, function fitting neural network, cascade-forward neural network, and multilayer perceptron neural network models and linear regression methods are used in the developed prediction models. The best results were obtained with multilayer perceptron with 93% accuracy and cascade-forward neural network models with 85% accuracy.**

***Keywords* – artificial neural network; regression; meal demand prediction**

I. INTRODUCTION

Today, with the development of the industry, the globalization of the world, the increase of large and small enterprises, the opening of new educational institutions and shopping centers, the food sector is growing day by day. Based on this growth, the influence of food companies also increases. In the dining area, many companies are showing respect to their domains. Catering companies are at the forefront in regular works and organizations, also restaurants are in shopping malls and dining halls in schools and workplaces. As the food market continues to develop, the most important problem is the cost for the companies. Every catering company wants to keep costs down to the lowest level. In this case, the exact number of people who eat food plays the most important role in costs. If the number of people is not known correctly, costs and wastes may increase and problems such as not having food can be encountered.

The motivation of this study is to predict the number of people who will eat according to the meal to be prepared, using the data obtained from university dining hall. It is aimed to prevent the problems that will arise when too much or little meal is produced in this view. This problem has been tried to be solved by using the proposed artificial neural network models.

Artificial neural networks (ANN) are data processing systems that are inspired by biological neural networks. A neural network is a parallel distributed processor consisting of simple units that have a natural tendency to store information. These networks are used to store information about the inter-neuron connections known as synaptic weights. Similarly, ANNs are composed of many operators and are interconnected by weights carrying numerical representations of different forms [1].

The dataset used in the study was divided into two sets as test and training dataset, and the predictions were obtained by test data after being trained by the proposed model with training data. These operations are repeated separately using different network structures and parameters in the MATLAB software and the Python language to get the best results.

Four different ANN models were designed in the study. These are feed-forward neural network (FFNN), function fitting neural network (FITNET), cascade-forward neural network (CFNN) and multilayer perceptron neural network (MLP). Estimates were also made using linear regression (LR) which is a statistical method, and these statistical results were used to compare the accuracy of the results obtained from artificial intelligent based models.

Three different neural networks (FFNN, FITNET, and CFNN) are modeled in the MATLAB environment. Different network structures are used in each of the models. For CFNN, FITNET and FFNN models, 3, 2 and 1 hidden layers are used respectively. Bayesian Regularization (TRAINBR) was used as training function, Hyperbolic Tangent Sigmoid (TANSIG) and Log-Sigmoid (LOGSIG) were used as transfer functions in each of these methods. The accuracy of the proposed models was determined by Mean Squared Normalized Error (MSE), Mean Absolute Error (MAE) and Multiple Correlation Coefficient (R).

In the models proposed using the Python language, the Scikit Learn library, specially designed for machine learning, has been used. Logistic Sigmoid Function (LOGISTIC) was used as the activation function in the MLP model. Finally, the results were also estimated by the LR method. The accuracy of the proposed models was determined by R.

In order to train the models, k-fold cross-validation, one of the methods of segmentation of the data set, has been used. In this method, all the data are divided into k equal group. One of these groups is used for training and the other remaining groups for testing. This operation is repeated every time until k is reached. Thus, all groups are used for both testing and training. With k-fold cross-validation, mistakes arising from incorrect grouping of data are prevented. The value of k in this study was set as 5 [2].

II. Related Works

In recent years, many studies have been done on artificial neural networks and demand prediction.

Kılıç has trained the data of Pamukkale University refectory to predict daily demand of the refectory by using LR, techniques of ANN such as multi-layer neural network and radial basis function neural networks. The study has showed that by using different statistical and artificial intelligence methods daily food demand can be predicted well than experts' decisions [3].

The future demand quantities of the dried apricot product of Malatya province were estimated by Karahan using ANN. The developed model has five input and one output variables and the network has one hidden layer with 5 neurons. The estimates produced by the developed neural network model showed that the model is reliable and consistent [4].

Blecher has determined which forecasting model would most accurately predict meal demand in Title IIIc congregate lunch programs designed for serving older adults. Forecasting techniques including naive, moving average (three versions) and simple exponential smoothing were applied to data collected over a 4-month period from seven meal sites located in a large urban area. An analysis of the forecasting models using mean absolute deviations and MSE indicated that simple mathematical forecasting techniques provided better predictions of meal demand than did the naive method for all sites [5].

III. Dataset and Models

A. Dataset

Before the dataset is used, the text inside is converted to numeric values. In addition, a grouping operation has been performed for the similar data. For example, fried potatoes, aubergine fries, and mixed fries are collected in the frying group. In this way, the learning ability of the network is increased. There are 6 attributes in the dataset. 5 (days of the week, soups, main courses, side meal, and exam week) of them are input and 1 (number of people) is output data. Table 1 shows a sample of the attributes content. A dataset sample converted to numerical values is given in Table 2.

Table 1: A sample of the dataset content.

Attribute	Description
Days of the Week (1-5) (A)	Monday, Tuesday, Wednesday, Thursday, Friday
Soups (0-9) (B)	Noodle soup, chicken soup, lentil soup, etc.
Main Courses (0-14) (C)	Beef meal, meal with beef meat, chicken meal, meal with chicken meat, fish, etc.
Side Meal (0-5) (D)	Rice, pasta with sauce, spaghetti, etc.
Exam Week (0-1) (E)	Exam week or not (each day at the exam week marked as 1)
Number of People (0-9) (F)	Number of people eating at the dining hall

Table 2: A sample of the dataset.

A	B	C	D	E	F
3	3	3	1	0	1
4	3	12	0	0	2
5	5	2	4	0	2
1	0	5	4	1	0
2	3	10	1	1	0
3	3	5	1	1	0
4	5	14	5	1	0
5	3	3	2	1	0
1	3	8	1	0	0
3	3	2	1	0	0

B. Cascade-Forward Neural Network Model

The network structure used in the CFNN model is shown in Figure 1.

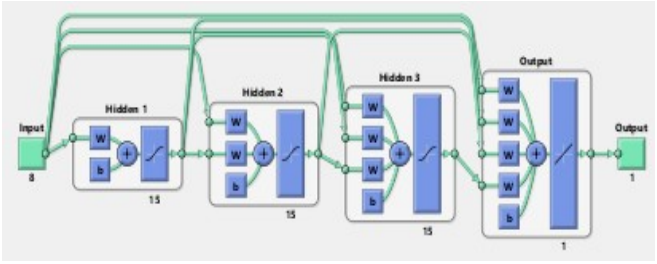


Figure 1: Structure of the CFNN.

The key features of the model are listed below:

- Number of hidden layers: 3
- Training function: Bayesian Regularization
- Transfer functions: tansig and logsig
- Performance criteria: MSE, MAE, and R

C. Function Fitting Neural Network Model

The network structure used in the FITNET model is shown in Figure 2.

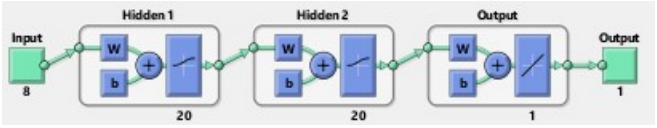


Figure 2: Structure of the FITNET.

The key features of the model are listed below:

- Number of hidden layers: 2
- Training function: Bayesian Regularization
- Transfer functions: tansig and logsig
- Performance criteria: MSE, MAE, and R

D. Feed Forward Neural Network Model

The network structure used in the FFNN model is shown in Figure 3.

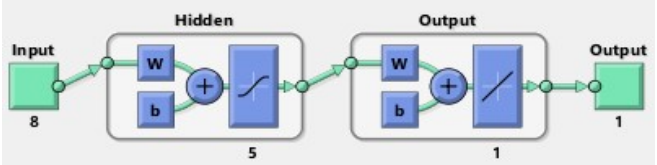


Figure 3: Structure of the FFNN.

The key features of the model are listed below:

- Number of hidden layers: 1
- Training function: Bayesian Regularization
- Transfer functions: tansig and logsig
- Performance criteria: MSE, MAE, and R

E. Multilayer Perceptron Neural Network Model

The key features of the model are listed below:

- Number of hidden layers: 50
- Activation function: logistic
- Number of iterations: 50.000

F. Linear Regression Model

In this method, data were divided into two groups, 80% as training data and 20% as test data. The accuracy is calculated using the functions of the Scikit learning library.

IV. Results And Discussions

All prediction models were evaluated in terms of three performances measured, (1) R value is used for measuring the correlation between target and predicted values, (2) MSE measures the average of the squares of the errors, (3) MAE measures the closeness of the predictions to the target values. Equations of these performance measures are given in equations (1), (2) and (3), respectively [6]

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}(O_i - O_m)^2}} \quad (1)$$

$$MSE = \frac{1}{n}[\sum_{i=1}^{n}(O_i - P_i)^2] \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \quad (3)$$

where n is the number of data points used for testing, P_i is the predicted value, O_i is the observed value and O_m is the average of the observed values.

According to the results given in Table 3, 4, 5 and 6, the following results are obtained:

- Best results are achieved with the CFNN model for MSE as 1.500 at second fold, MAE as 1.0875 at fifth fold and R as 0.8883 at fifth fold.
- For the CFNN model, average values of 2.0478, 1.1272 and 0.8682 were obtained for MSE, MAE and R, respectively.
- Best results are achieved with the FITNET model for MSE as 1.8697 at first fold, MAE as 1.0334 at second fold and R as 0.8640 at second fold.
- For the FITNET model, average values of 2.2958, 1.1956 and 0.8215 were obtained for MSE, MAE and R, respectively.
- Best results are achieved with the FFNN model for MSE as 2.0564, MAE as 1.1363 and R as 0.8616 at second folds.
- For the FFNN model, average values of 2.8657, 1.3287 and 0.7887 were obtained for MSE, MAE and R, respectively.
- The best results are achieved with the MLP model for R as 0.9639 at second fold.
- For the MLP model, an average value of 0.9171 was obtained for R.

Table 3: The results obtained with the CFNN model.

Fold #	MSE	MAE	R
1	1.8304	1.0955	0.8562
2	**1.5000**	1.1721	0.8485
3	2.5294	1.1551	0.8771
4	2.1873	1.1262	0.8712
5	2.1921	**1.0875**	**0.8883**
Average	2.0478	1.1272	0.8682

Table 4: The results obtained with the FITNET model.

Fold #	MSE	MAE	R
1	**1.8697**	1.0602	0.8309
2	1.8852	**1.0334**	**0.8640**
3	2.4574	1.2867	0.8516
4	2.5608	1.3293	0.7632
5	2.7061	1.2687	0.7979
Average	2.2958	1.1956	0.8215

Table 5: The results obtained with the FFNN model.

Fold #	MSE	MAE	R
1	3.3025	1.4333	0.7367
2	**2.0564**	**1.1363**	**0.8616**
3	3.2741	1.3813	0.6997
4	2.7124	1.3318	0.8532
5	2.9833	1.3609	0.7925
Average	2.8657	1.3287	0.7887

Table 6: The results obtained with the MLP model.

Fold #	R
1	0.8136
2	**0.9639**
3	0.9629
4	0.9518
5	0.8935
Average	0.9171

Figures 4, 5 and 6 show performance graphs for MSE of CFNN, FITNET, and FFNN models. The best performance results were reached in the 34th, 21st and 10th epochs for the CFNN, FITNET, and FFNN models, respectively.

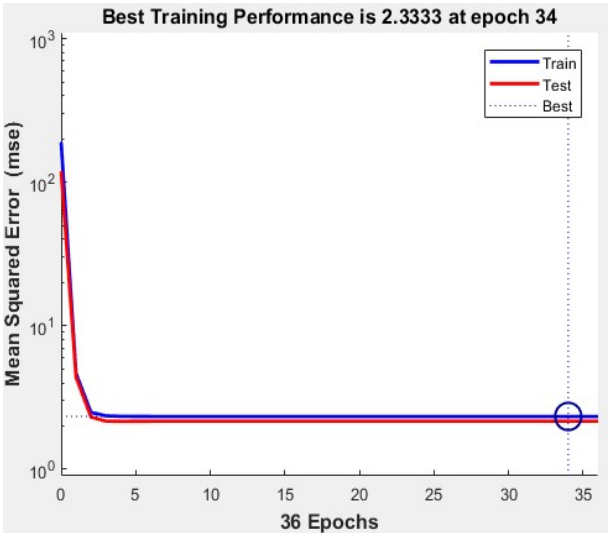


Figure 4: Performance graph of CFNN model.

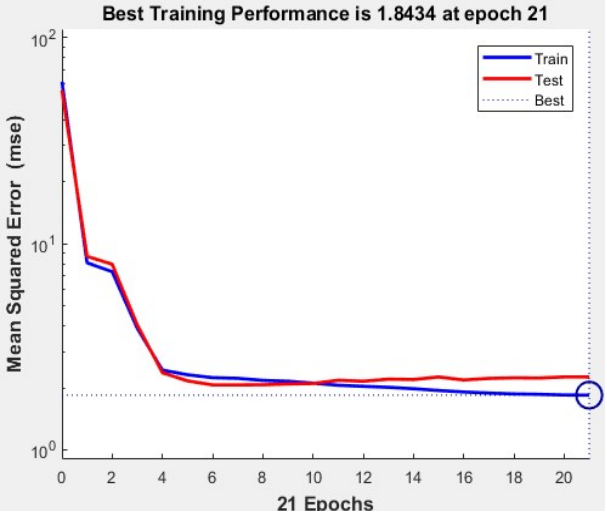


Figure 5: Performance graph of FITNET model.

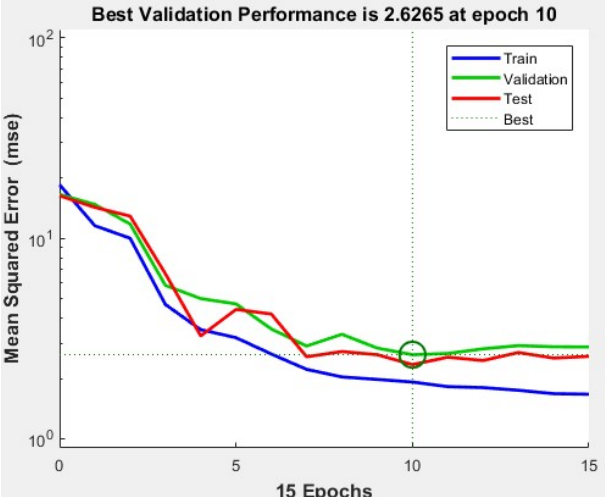


Figure 6: Performance graph of FFNN model.

V. CONCLUSION

In this study, four different artificial neural network models were designed to predict the number of people who will eat according to the meal to be prepared, using the data obtained from university dining hall, and satisfactory results have been achieved. The results show that the MLP and CFNN models are more successful than the other designed models.

The successful results show that existing methods can be improved and better results can be achieved using different methods/models.

VI. REFERENCES

[1] Ö. Aydın, "Yapay sinir ağlarını kullanarak bir ses tanıma sistemi geliştirilmesi," MSc. Thesis, Bilgisayar Mühendisliği ABD, Trakya Üniversitesi, 2005.

[2] B. İşçimen, Y. Kutlu, A. N. Reyhaniye, C. Turan, and D. Bilimleri,"Balık tanınmasında görüntü analiz yöntemleri - Image analysis methods on fish recognition," *IEEE 22nd Signal Processing and Communications Applications Conference (SIU 2014)*, pp. 1411-1414, 2014.

[3] G. Kılıç, "Yapay sinir ağları ile yemekhane günlük talep tahmini," MSc. Thesis, Fen Bilimleri Enstitüsü, Pamukkale Üniversitesi, 2015.

[4] M. Karahan, "İstatistiksel tahmin yöntemleri: Yapay sinir ağları ile ürün talep tahmini uygulaması." PhD. Thesis, Sosyal Bilimler Enstitüsü, Selçuk Üniversitesi, 2011.

[5] L. Blecher, "Using forecasting techniques to predict meal demand in Title IIIc congregate lunch programs," *Journal of the American Dietetic Association*, vol. 104 (8), pp. 1281-1283, 2004.

[6] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*. Burlington: Morgan Kaufmann, 2005.

Determination Of Coastline Changes at Kozan Dam Lake By Using Artificial Neural Networks Method

T. KAYNAK[1], S.Y. ÇİÇEKLİ[2] and M. H. KESİKOĞLU[1*]

[1]Erciyes University, Kayseri/Turkey, tolgakaynak@erciyes.edu.tr, hayrikesikoglu@erciyes.edu.tr
[2]CukurovaUniversity, Adana/Turkey, yoturanc@cu.edu.tr
*Corresponding Author, e-mail address: hayrikesikoglu@erciyes.edu.tr

Abstract **- With the development of technology, remote sensing is commonly used for ecological studies and monitoring wetland and management. Artificial Neural Networks are extremely simplified model of the brain occurring by neurons and layers connecting to neurons so artificial neural networks method is frequently used to classify satellite images. In this study, Landsat-5 satellite image with spatial resolution of 30m, acquired on October 29, 2007 and Landsat-8 acquired on November 27, 2017 were used to identified the coastline changes at Kozan Dam Lake. The lake is used as drinking and irrigation water. Therefore, it is very important to examine the coastline changes of the lake. In first step, image to image registration was made to conform image coordinate systems of images to each other. Second step, images were classified by artificial neural networks method. Four classes namely lake, agricultural area, soil, and vegetation area were determined. Third step, image classification accuracies were determined. Finally, the changes in coastline of Dam Lake were calculated by post classification comparison method. Coastline change of Dam Lake was calculated as 0.6 km^2 increase and the change image map was created. At the end of the study Kozan Dam Lake coastline changes were monitored by using remote sensing methods.**

Keywords **- Artificial neural networks, coastline change, remote sensing, Landsat**

I. INTRODUCTION

NEW image classification methods were developed in recent year. By development of the technology, remote sensing became one of the most preferable technique wetland studies, environmental studies and coastline monitoring.

There are many studies about artificial neural networks in literature. One of them is classification of remote sensing imagery. In this study, artificial neural networks were used for image classification. Özçelik and Arısoy [1]detected water depths of west coast of Foça by using artificial neural networks classification method on Aster and Quickbird satellite images. Artificial neural networks method provided higher accuracy for Quickbird satellite image than Aster image. Mas and Flores [2] reviewed artificial neural networks for remote sensing studies and the study shown that artificial neural network have wide range of task in remote sensing studies. Ingram et al.[3] used remote sensing to mapping tropical forest and used artificial neural network to predict basal area. This study shown that artificial neural networks were useful to estimate and produce basal area. Hong et al. [4] used artificial neural network for classification of satellite cloud images and used classification images for estimating rainfall.

The aim of this study is to determine of coastline changes on Kozan Dam Lake by using pixel based image classification with artificial neural networks (ANN) method. We also aim to detect the accuracy of this method at Kozan Dam Lake in the Turkey based on analysis of Landsat 8 LDCM imagery.

II. STUDY AREA

The Kozan Dam Lake is structured on the Kilgen River in Adana Province, Turkey. It is nearly 8 km of Kozan.

Figure 1: Images of Kozan Dam Lake[5]

The Dam is constructed between 1967 and 1972 for flood control and irrigation. Today it is also used for generating electricity. [6]

Figure 2: Landat-8 images of Kozan Dam Lake

III. METHODS

A. Material

In the study, Landsat 8 LDCM satellite image of Kozan Dam Lake in 2017 was used. Satellite image was taken on November 27th. In Table 1, Landsat 8 bands, their wavelength and resolution values are given in detail.

Table 1. Landsat-8 LDCM OLI / TIRS Satellite Bands and Features [7]

Band	Wavelength (micrometers)	Resolution (meters)
Band-1 Ultra Blue (Coastal Aerosol)	0.43-0.45	30
Band-2 Blue	0.45-0.51	30
Band-3 Green	0.53-0.59	30
Band-4 Red	0.64-0.67	30
Band-5 Near Infrared (NIR)	0.85-0.88	30
Band-6 Shortwave Infrared (SWIR-1)	1.57-1.65	30
Band-7 Shortwave Infrared(SWIR-2)	2.11-2.29	30
Band-8 Panchromatic	0.50-0.68	15
9 Cirrus	1.36-1.38	30
Band-10 Thermal Infrared (TIRS-1)	10.60-11.19	100*30
Band-11 Thermal Infrared(TIRS-2)	11.50-12.51	100*30

In the study, Landsat 5 TM satellite image of Kozan Dam Lake in 2007 was used. Satellite image was taken on October 29th. In Table 2, Landsat 5 bands, their wavelength and resolution values are given in detail.

Table 2. Landsat-5 TM Satellite Image Bands and Features [7]

Bands	Wavelength (micrometers)	Resolution (meters)
Band-1 Blue	0.45-0.52	30
Band-2 Green	0.52-0.60	30
Band-3 Red	0.63-0.69	30
Band-4 Near Infrared (NIR)	0.76-0.90	30
Band-5 Shortwave Infrared (SWIR) 1	1.55-1.75	30
Band-6 Thermal	10.40-12.50	120*30
Band-7 Shortwave Infrared(SWIR-2)	2.08-2.35	30

B. Methods

In this study artificial neural networks and post classification comparison methods were used. Information on the methods are given in the subheadings.

1) Artificial neural networks

Artificial neural networks (ANN) are mathematical modeling tools that are useful for forecasting and prediction in complex environments [8]. Basic component of artificial neural network is called the neuron [9].

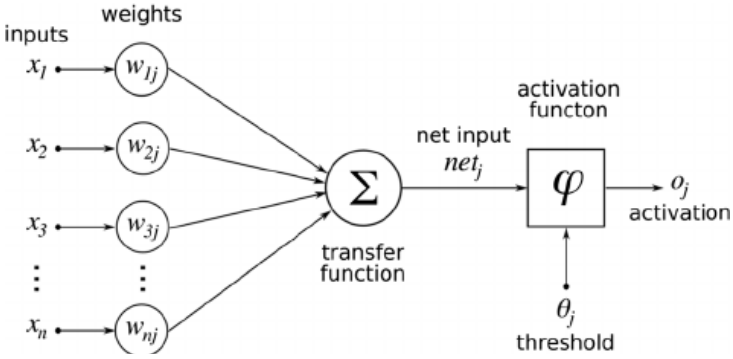


Figure 3: Neuron Structure [10]

The structure of neuron is shown in Figure 3. $x_1, x_2, x_3 \ldots . x_n$ are inputs. $w_{1j}, w_{2j}, w_{3j} \ldots . w_{nj}$ are weights of inputs.

$$Net\ input = \sum x_i \times w_{nj} \qquad (1)$$

The activation function is identified as given in equation 2.

$$\varphi = F(x) \qquad (2)$$

The structure of artificial neural network is shown in Figure 4.

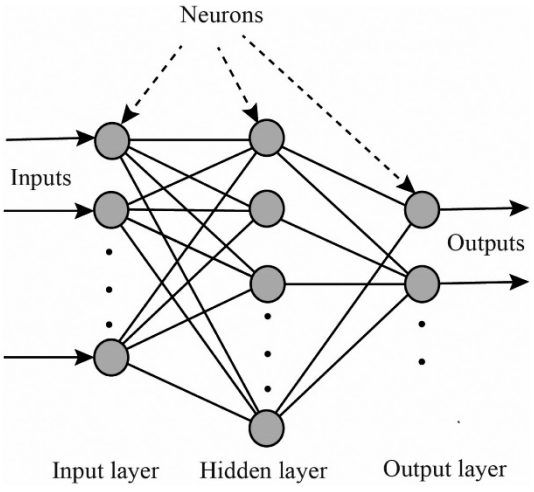


Figure 4: Structure of artificial neural networks [11]

2) Post Classification Comparison

In this study, post classification comparison was done for detecting changes of coastline at Kozan dam lake. In this method, classified satellite images which was taken on two different times were compared with each other. Finally, direction and amount of the change were calculated.

IV. APPLICATION AND RESULTS

The flowchart of the study was shown in Figure 5.

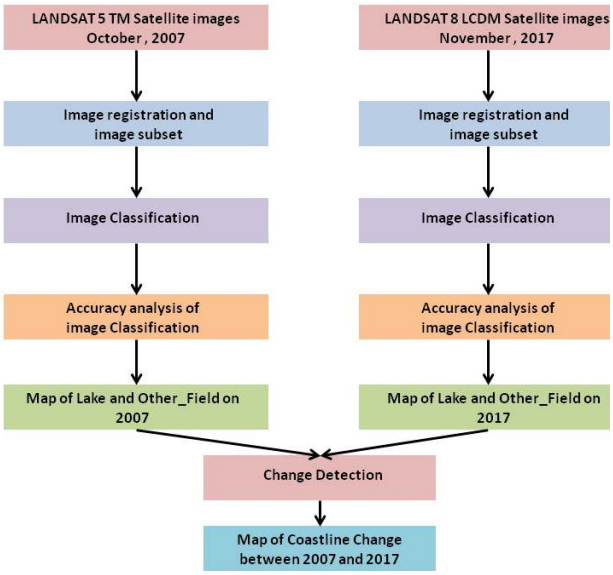


Figure 5: Flowchart of this study

In first step, "Image to Image Registration" was done for all satellite images. Satellite image which was taken on October 29th, 2007 was selected as reference image. The other satellite image which was taken on November 27th, 2017 was registered to reference image to correct to meaningful nonsystematic errors which were occurred because of satellites position. As a result of this process, the images were overlapped. The registered and subset images are shown in Figure 6.

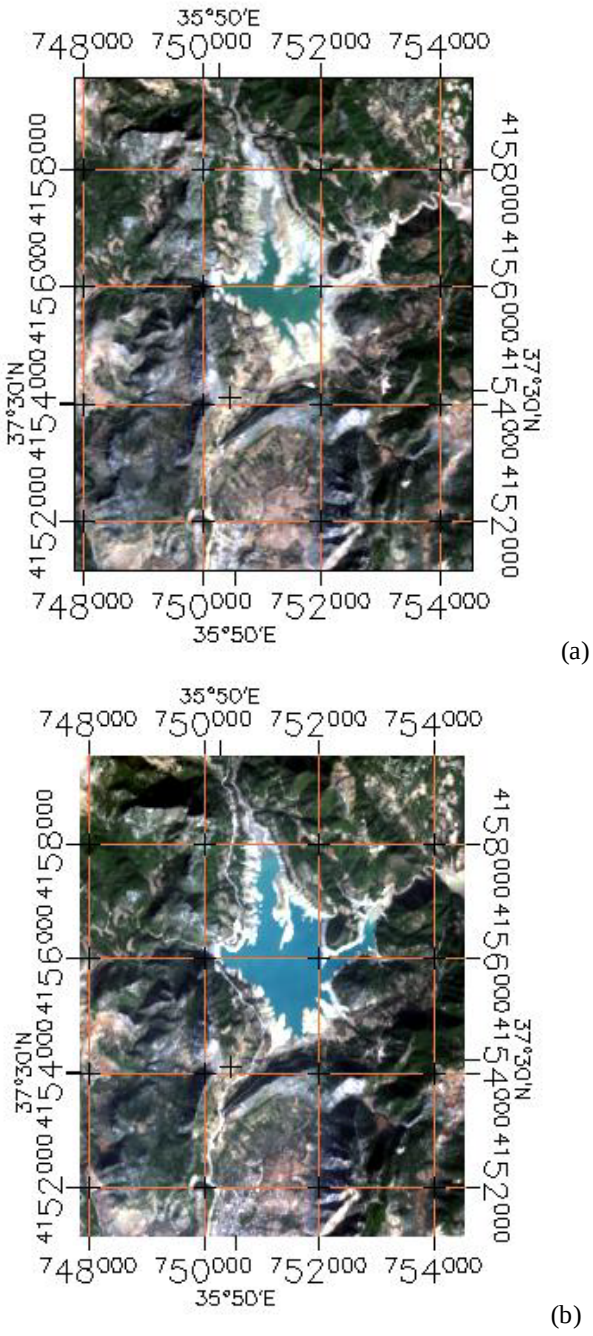


Figure 6: (a) 2007, (b) 2017 registration images of Kozan Dam Lake

In this study, artificial neural network method was used. Four classes namely lake, agricultural area, soil, and vegetation area were determined. Classification images are shown in Figure 7.

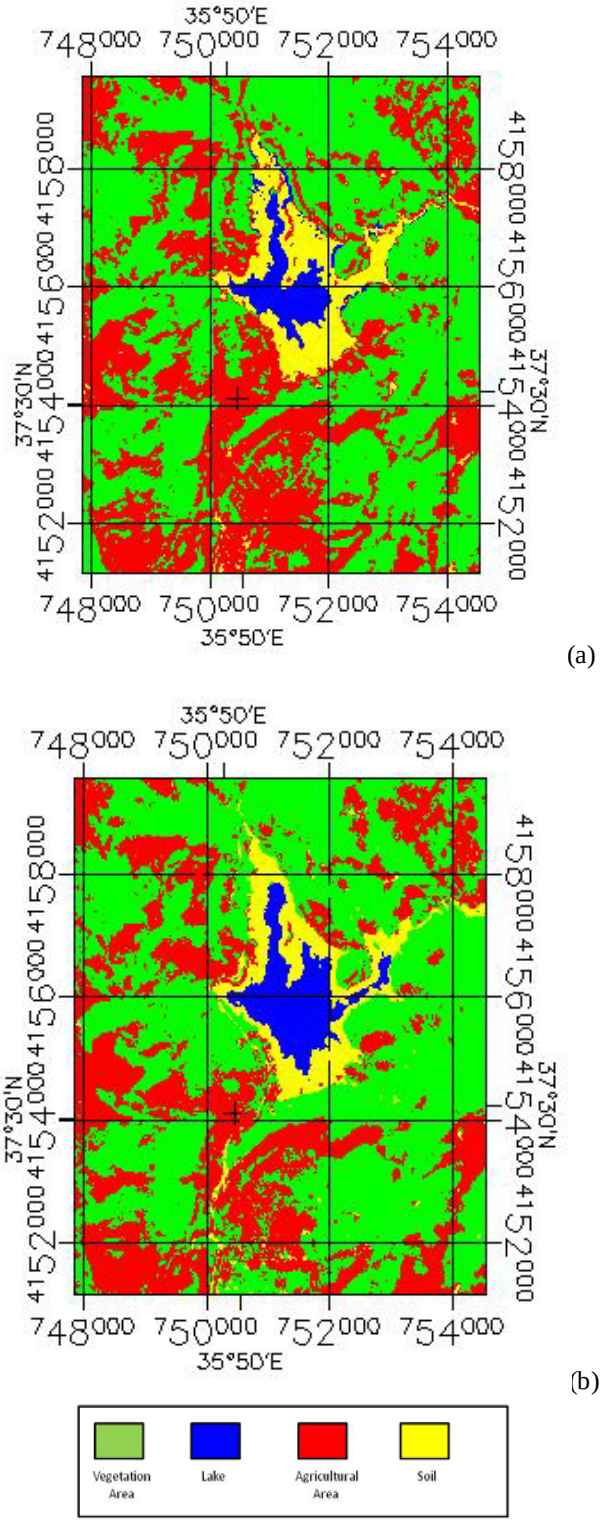


Figure 7: (a) 2007, (b) 2017 Classification images of Kozan Dam Lake

The image classification accuracies were determined for artificial neural network method. Overall accuracy 86.12% and kappa coefficient 0.7610 were determined for 2017 image. Overall accuracy 91.5% and kappa coefficient 0.8529 determined for 2007 image.

Change was determined by using post classification comparison method. Change detection map is shown in Figure 8.

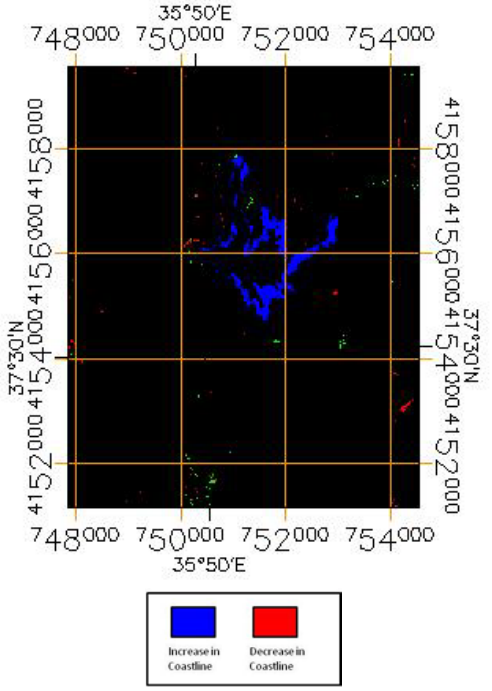


Figure 8: Change Detection Map

In 2007 image, 1807 pixel lake area was detected corresponding to 1.626 km^2 lake area. In 2017 image, 2569 pixel lake area was detected corresponding to 2.232 km^2 lake area. 0.626 km^2 increase in the lake area was occurred.

V. CONCLUSION

Coastline change has a big importance for environmental studies and wetland area management. Coastlines can change because of climatic conditions and use of water such as agricultural irrigation. Thus, coastline areas should be observed regularly.

In this study, image registrations, classification and change detection analysis were done. After that, post classification comparison method was used to determine the changes between 2007 and 2017 years by using classified Landsat 8 LDCM and Landsat 5 TM satellite images with artificial neural networks. Finally, the amounts of the change in land cover were detected. Finally, the amount of change in land cover was determined. This study showed that artificial neural networks (ANN) method produce meaningful results in satellite images of different years.

REFERENCES

[1] C. Özçelik, and Y. Arisoy, "Sığ sularda su derinliklerinin yapay sinir ağları kullanılarak uzaktan algılanması", *I. Uzaktan Algılama ve CBS Çalıştay ve Paneli*, İstanbul, 2006.

[2] J. F. Mas, and J. J. Flores, "The application of artificial neural networks to the analysis of remotely sensed data", *International Journal of Remote Sensing*, Vol. *29*(3), pp. 617-663, 2008.

[3] J.C. Ingram, T. P. Dawson, and R. J. Whittaker. "Mapping tropical forest structure in southeastern Madagascar using remote sensing and artificial neural networks." *Remote Sensing of Environment*, Vol. 94.4, pp. 491-507, 2005.

[4] Y. Hong, K. L. Hsu, S. Sorooshian, and X. Gao, "Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system", *Journal of Applied Meteorology*, Vol. 43(12), pp. 1834-1853, 2004.

[5] Kozan Belediyesi [Online], Available: http://www.kozan.bel.tr/?/kozan-genel-fotograflar/3

[6] Kozan Barajı [Online], Available: www2.dsi.gov.tr/baraj/detay.cfm?BarajID=44

[7] What are the band designations, [online] avaiable: http://landsat.usgs.gov/band_designations_landsat_satellites.php

[8] M. M. Hamed, M. G. Khalafallah, and E. A. Hassanien, "Prediction of wastewater treatment plant performance using artificial neural networks", *Environmental Modelling & Software*, Vol. *19*(10), pp. 919-928, 2004.

[9] M. S. Nasr, M. A. Moustafa, H. A.-Seif, and G. El Kobrosy, "Application of Artificial Neural Network (ANN) for the prediction of EL-AGAMY wastewater treatment plant performance-EGYPT", *Alexandria engineering journal*, *51*(1), pp. 37-43,2012.

[10] How to improve performance of Neural Networks [online] avaiable: https://d4datascience.wordpress.com/2016/09/29/fbf/

[11] Y. Lin, H.Wen, and S. Liu "Surface runoff response to climate change based on artificial neural network (ANN) models: A case study with Zagunao catchment in Upper Minjiang River, Southwest China", *Journal of Water and Climate Change*, Vol 9(1), 2018.

Particle Swarm Optimization Based Determination of Learning Parameters in Artificial Neural Networks with Backpropagation Learning Algorithm

E. ÇELİK[1], N. ÖZTÜRK[1], A. DALCALI[2]

[1] Gazi University, Ankara/Turkey, emrecelik@gazi.edu.tr
[1] Gazi University, Ankara/Turkey, ozturk@gazi.edu.tr
[2] Karabük/Turkey, dalcali.adem@gmail.com

Abstract **- In this paper, a particle swarm optimization (PSO) algorithm is introduced to determine learning parameters required for the backpropagation (BP) learning algorithm, which is used for training of a feed-forward neural network (FFNN). PSO algorithm utilized within the paper works slightly different compared to conventional PSO (CPSO) algorithm in such a way that each particle adjusts its position based on the best mid-position of all particles and its group's previous best. The major reason of such a change is to enhance the performance of CPSO algorithm, which is explained in detail in the study suggested by Tamer, S and et.al. To test the proposed method, a FFNN with three layers is designed for function interpolation. Learning parameters of the designed NN are determined by both conventional error and trial method and the proposed method. Afterwards, using these two groups of learning parameters, the NN is trained and tested under the same conditions. According to the test results, learning parameters determined by the PSO provide a better performance and interpolating capability for the NN than those determined by the conventional method.**

Keywords **- feed-forward neural networks, learning parameters, function interpolation, particle swarm optimization.**

I. INTRODUCTION

Particle swarm optimization (PSO) method is a member of a wider class of swarm intelligence methods used for solving global optimization problems. The method was originally proposed by Kennedy and Eberhart as a simulation of social behavior of bird flock and was initially introduced as an optimization method in 1995 [1].

In PSO, candidate solutions named particles in the search space form a population. The population is initially generated at random. The position of each particle representing the potential solutions to the given problem is updated through the generations, and each of them moves through a multi-dimensional search space with a designated velocity to seek the optima or sub-optima. PSO has the capability of handling large multidimensional problems due to its fast convergence property and it has less computational complexity due to fewer no. of parameter adjustments [2, 3]

During the last decade, artificial neural networks (ANNs) have been the focus of a great deal of attention, due to their capabilities in solving nonlinear problem by learning [4]. ANN is an information processing system constituted by an assembly of a large number of simple processing elements that are interconnected to perform a parallel distributed processing in order to solve specific task, such as pattern classification, function approximation, clustering (or categorisation), prediction (forecasting or estimation), optimisation and control [5]. The backpropagation (BP) supervised learning algorithm has been developed for training feed-forward neural networks (FFNNs), which iteratively adjusts the network weights considering the derivatives of the cost function with respect to those weights and it is by far the most preferred learning method [6]. In the BP, selection of the learning parameters, which are learning rate and momentum rate, is an important task for training of the network more efficiently. If the learning rate is too low, learning speed of the network will be very slow and if it is chosen too high, then the objective function and the weights are likely to diverge. Keeping the momentum rate small dampens the oscillations during the learning, but can also serve to slow down the learning process. Taking these cases into account, learning parameters should be chosen to ensure global convergence which tends to be difficult task to achieve [7].

Recently, some successful studies have been proposed in order to determine the optimum learning parameters. Üstün and Yıldız used genetic algorithms (GAs) to adjust the learning rate and momentum rate efficiently [8]. Sheell and et al. set the learning parameters by a heuristic approach depending upon the sign of the network output error [9]. Ribeiro and Schlansker focused on the wholly optimized ANN constructed for reactive power control using the conventional PSO [10]. Yu and et al. obtained the learning parameters dynamically using the first and second order partial derivatives of the learning and momentum rate according to the error criteria [11]. In these studies, the dynamic calculation of the learning and momentum rate has a relatively high computational load, though only one learning and momentum rate were used. Hence, these approaches are difficult to implement in practice. Also, it has

already been proved that PSO can be considered as a powerful alternative to GA, and besides, it has some superiorities over GA such as having fewer parameters to control and less complex algorithm, which makes it easier to implement.

In this paper, PSO algorithm has been introduced to determine the optimum learning parameters required for the BP learning algorithm, so that the trained FFNN can learn a non-linear function more efficiently. To explore the applicability of PSO, a FFNN with three layers has been designed, and the two learning rates and two momentum rates are coded into a string called *particle*. During the optimization, by simulating the behavior of a flock of birds, PSO algorithm seeks the optimum learning parameters. According to the simulations performed under the same conditions, the proposed method is found to achieve better approximation ability than the conventional one, where the learning parameters are obtained based on the error and trial method after many tedious trials, which also does not guarantee to provide a global convergence unlike our proposal.

II. THE DESIGNED NEURAL NETWORK

A three-layer NN arranged in the paper is given in Fig. 1. The network has single input and single output, and the hidden layer is composed of 5 neurons, which form the roof of the network and perform the actual calculations of the network [9]. Activation functions of the hidden neurons are selected as *tanh* and *linear* for the output neuron. w_{ih}, w_{bh}, w_{ho} and w_{bo} are the connection weights. The training data set which consists of inputs and expected outputs is represented by $\{x_i, y_i\}$, and y and e_i are the network output and network output error between y_i and y, respectively.

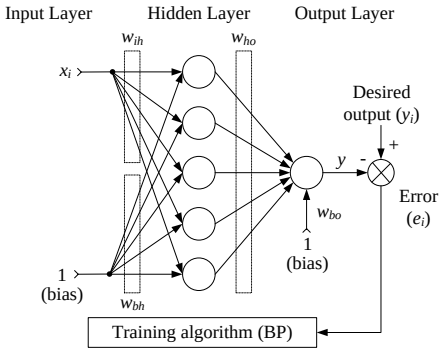


Figure 1: The designed NN architecture.

III. PARTICLE SWARM OPTIMIZATON

PSO is a population-based random search optimization technique proposed by Kennedy and Eberhart. This algorithm is inspired by simulation of social psychological expression of birds and fishes [12]. The population of particles starts to move in a search space by following current optimum particles and changing their positions based on the internal velocity in order to find the optimum [13-16]. PSO algorithm used in this paper works slightly different compared with the conventional PSO (CPSO) in terms of velocity updating mechanism of a particle, in that, the particles update their velocities and positions based on the best mid-position of all particles unlike CPSO and its group's previous best, as stated in (1) and (2) [17]. Thus, particles do not search between the previous best positions of its neighbors and its own. Instead, there is a search between the previous best positions of particles' neighbors and the best mid-position of all the particles in the swarm, allowing to reach the optimal regions faster.

$$v_{i,j}(k+1) = w(k) \times v_{i,j}(k) + \lambda_1 \times r_1 \times (p_j^* - p_{i,j}(k)) + \lambda_2 \times r_2 \times (p_{i,j}^{global\ best} - p_{i,j}(k)) \quad (1)$$

$$p_{i,j}(k+1) = p_{i,j}(k) + v_{i,j}(k+1) \quad (2)$$

where $v_{i,j}$: particle velocity, $p_{i,j}$: each variable in particle, r_1, r_2 : uniform random numbers ranging between 0 and 1, λ_1, λ_2 : learning parameters, $p_{i,j}^{global\ best}$: best solution achieved so far, w : inertia weight, k : iteration index. If the best position for the i-th particle is given by $p_i^{local\ best} = \left[p_{i1}, p_{i2}, \cdots, p_{i,j} \right]$, then the particles' average best position is calculated as,

$$p_j^* = \left(p_{1j} + p_{2j} + \cdots + p_{Nj} \right) / N \quad (3)$$

where j is the column index and N is the number of particles in the swarm. In (1), instead of using the best position found by each particle called as $p_{i,j}^{local\ best}$, the average of $p_{i,j}^{local\ best}$ positions for all of the particles, p_j^*, is used. In this scheme, each particle will fly towards a new position with respect to the particles' average best position and the swarm's best position. Thus, each individual benefits from not only the swarm's best particle experience but from the experiences of all the particles in the swarm, which yields higher efficiency and performance than CPSO. At this point, the used PSO algorithm differs from the regular one.

In the study, particles' velocities are limited by the user-specified value, maximum velocity $V_{\max}$ to prevent the particles from flying too far from a target [18]. In (1), the inertia weight, which varies with iteration, is calculated by the pre-determined boundary values $w_{\max}$ and $w_{\min}$ as,

$$w(k) = w_{\max} - \frac{w_{\max} - w_{\min}}{iter_{\max}} \times k \quad (4)$$

The concept of the PSO is also briefly illustrated in vector form in Fig. 2.

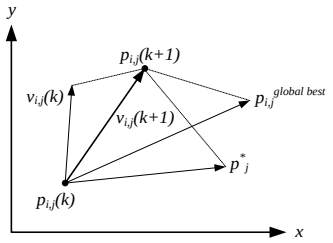


Figure 2: Flying concept of a particle in the PSO.

The flowchart of PSO procedure is presented in Fig. 3.

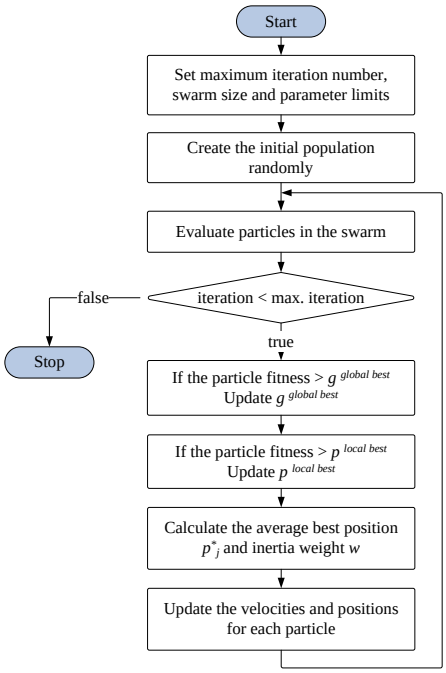


Figure 3: The PSO procedure.

IV. SIMULATION RESULTS

These studies were followed in three steps. In the first step, the test setup were developed and prepared for the computer simulations. In the second step, the test functions were selected to compare the proposed approach with the conventional one. Simulation results of comparisons were given and the effects of learning parameters on the training success in the NN model were also analyzed in the third step. These three steps were explained under two subsections in the following sentences.

A. Test setup

In the testing phase, to find out the superiority of the paper's proposed method, four nonlinear functions were selected as subjects of problems and the performance of our contribution was compared with the conventional method through the selected test functions given below.

1. $y = f^{(1)}(x) = (x^2 + x)\sin(x),\ \ x \in [-1, 9]$
2. $y = f^{(2)}(x) = \dfrac{\sqrt{x}}{(\sin(x) + x)},\ \ x \in [1, 20]$
3. $y = f^{(3)}(x) = \sin(x) + 0.5\sin(2x) + 0.25\sin(4x),\ \ x \in [-5, 5]$
4. $y = f^{(4)}(x) = \begin{cases} 1, & x \in [1,3) \\ 4.5, & x \in [3,5) \\ 7, & x \in [5,7) \\ 4, & x \in [7,10] \end{cases}$

To train the PSO based NN, 100 data points (patterns) in the specified ranges are used and before the patters are presented to the NN, they are standardized into the dataset with a mean of 0 and standard deviation of 1.

After the network is trained, in order to test the trained NN, 300 independent patterns, which are also called the test data set, are selected from the specified ranges. During the training and testing, the NN performance is evaluated by the equation defined in (5), where E_{ave} is the mean square error.

$$E_{ave} = \frac{1}{2}\frac{1}{I}\sum_{i=1}^{I}(y_i - y)^2 \qquad (5)$$

As there are 2 learning rates and 2 momentum rates in the NN, a particle is composed of 4 parameters, which are in the range [0, 1]. It is observed that randomly initialized population may cause the NN not to learn and in such a case, connection weights and objective function value are likely to diverge. To avoid this phenomenon, initial population is checked for the particles that cause NN not to converge. After recognizing these particles, they are replaced by randomly generated new particles. Also to render a fair comparison between the proposed and the conventional method, the same random initial connection weights for the NN are used in both methods. The collection of the PSO and NN parameters which are set by the error and trial method according to the NN and PSO performance is included in Table 1.

Table 1: The collection of the PSO and NN model parameters.

PSO parameters	Swarm size	15
	Maximum iteration	750
	Particle length	4
	Maximum search space range	[0, 1]
	Maximum velocity, $V_{\max}$	10% of the SS
	Learning parameters, λ_1, λ_2	2.0
	$[w_{\min}, w_{\max}]$	[0.05, 1]
NN parameters	Maximum iteration	3000
	Training type	Incremental
	Number of layers	3
	Number of hidden-layer neurons	5

SS: search space

B. Test results

In this section, computer simulations have been realized to verify the efficiency and applicability of the PSONN algorithm using the test setup explained in Section A. The PSO and NN model trained by the BP are run for 750 and 3000 iterations, respectively.

Fig. 4 shows the progress of E_{ave} against the number of iterations with the proposed approach (solid line) and with conventional method (dashed line) during the training. It should be noted that for Fig. 5a, although the proposed approach seems to be unstable at the beginning of the algorithm due to the high values of learning parameters, after around 500 iterations, it converges to a more global solution compared to the conventional method. As for the other three cases, better training curves are obtained in terms of the convergence speed and solution quality using the proposed approach

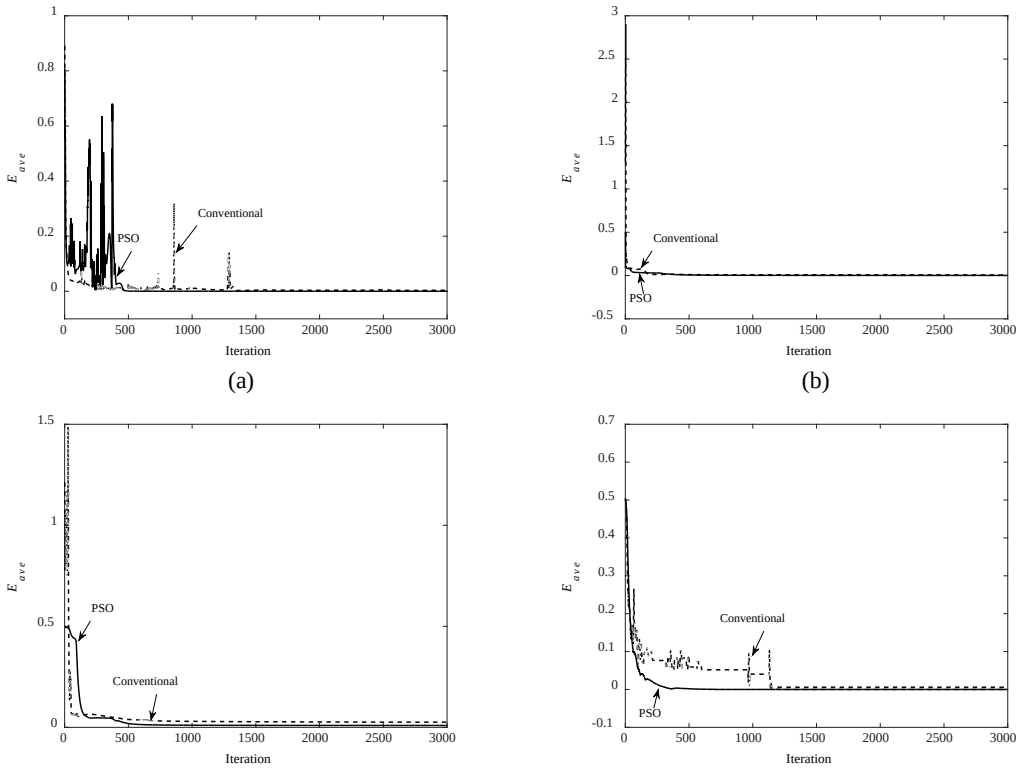


Figure 4: E_{ave} curves of the NN during the training with optimized learning parameters by PSO and conventional method with fixed parameters for (a) $f^{(1)}$ (b) $f^{(2)}$ (c) $f^{(3)}$ (d) $f^{(4)}$.

Table 2 shows the calculated E_{ave} values for the test results and the optimized learning parameters produced by running the PSO where the results of interests for E_{ave} are **bold-faced**. It also provides the learning parameters obtained by the error and trial method at the end of long-time trials.

Table 2: Test results for the proposed and conventional method.

Test functions	Learning parameters determined by PSO				E_{ave}	Learning parameters determined by conventional method				E_{ave}
	α_{ho}	γ_{ho}	α_{ih}	γ_{ih}		α_{ho}	γ_{ho}	α_{ih}	γ_{ih}	
$f^{(1)}$	0.174	0.479	0.399	0.337	**0.2495**	0.09	0.3	0.2	0.09	3.0825
$f^{(2)}$	0.037	0.0	0.347	0.001	**1.5×10^{-5}**	0.355	0.159	0.296	0.589	8.1×10^{-5}
$f^{(3)}$	0.001	0.315	0.140	0.488	**0.0055**	0.242	0.324	0.750	0.023	0.0156
$f^{(4)}$	0.001	0.446	0.9	0.934	**0.0423**	0.002	0.210	0.951	0.999	0.0756

In Table 2, it is apparent that for $f^{(1)}$, the calculated E_{ave} value for the conventional method is 3.08 while it is about 0.25 for the proposed method. For other test functions, similar improvements are achieved as well. As a result, compared with the conventional method, the proposed method comes with better training performance. The resulting approximations to $f^{(1)}$, $f^{(2)}$, $f^{(3)}$ and $f^{(4)}$ are shown in Figs. 5-8, respectively, where, it is demonstrated that the NN model trained with the optimized learning parameters approximates the nonlinear functions in a better way compared with the case where the learning parameters are chosen based on trial and error.

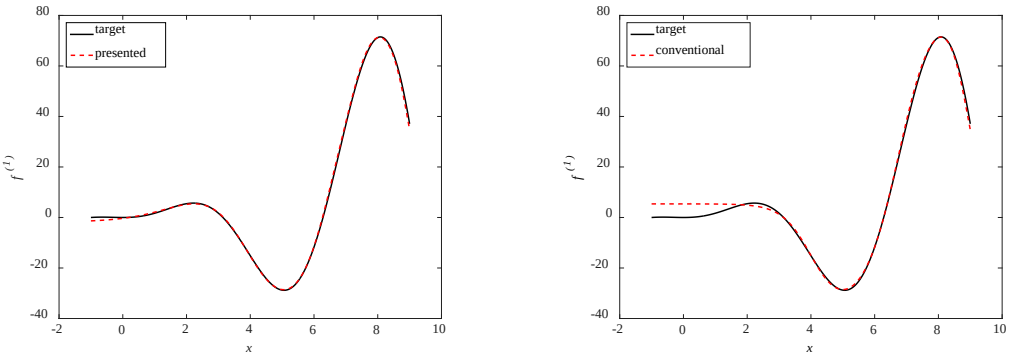


Figure 5: The results for $f^{(1)}$.

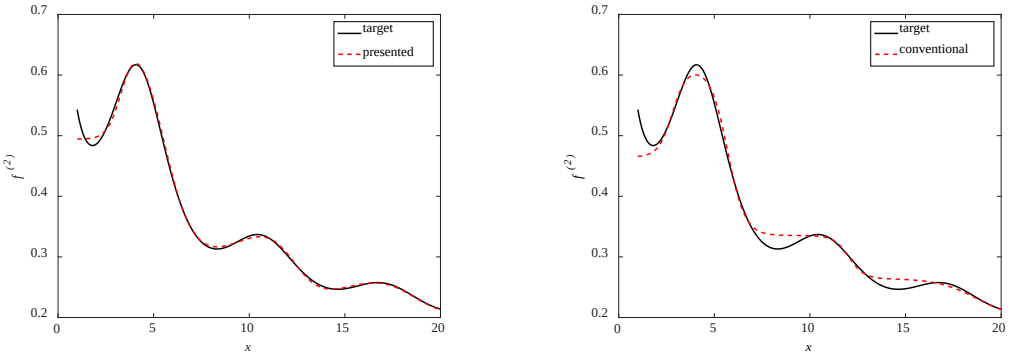


Figure 6: The results for $f^{(2)}$.

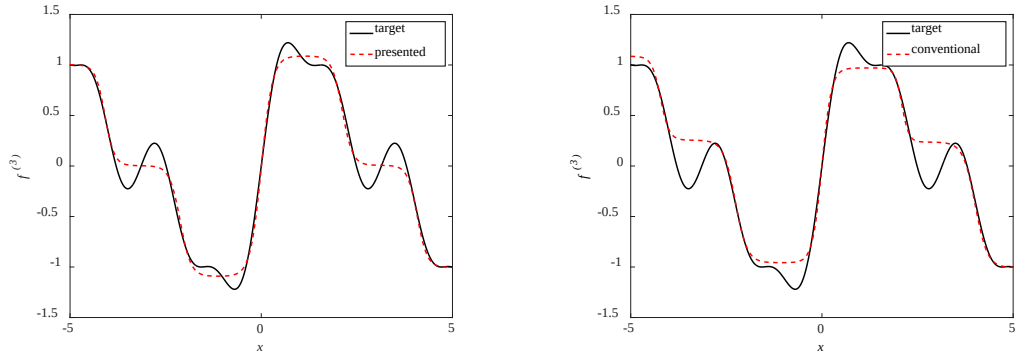


Figure 7: The results for $f^{(3)}$.

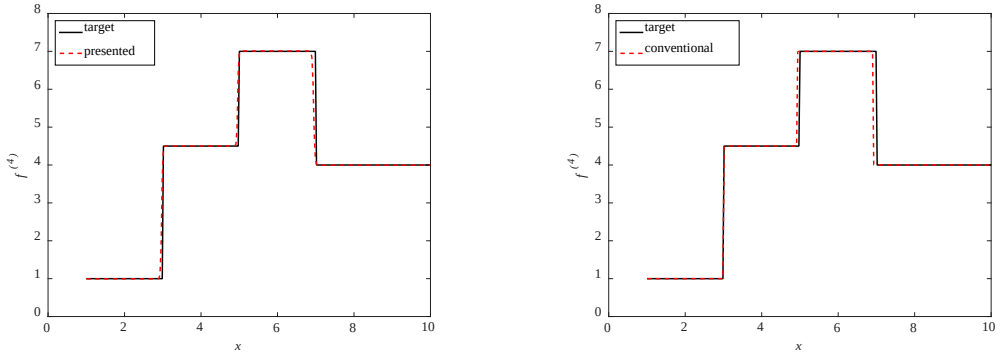


Figure 8: The results for $f^{(4)}$.

The optimization process of the learning parameters for learning $f^{(1)}$ is also visualized in Fig. 9. Note that around 350 generations, PSO slowly becomes less random as it converges to its final values reported in Table 2

.

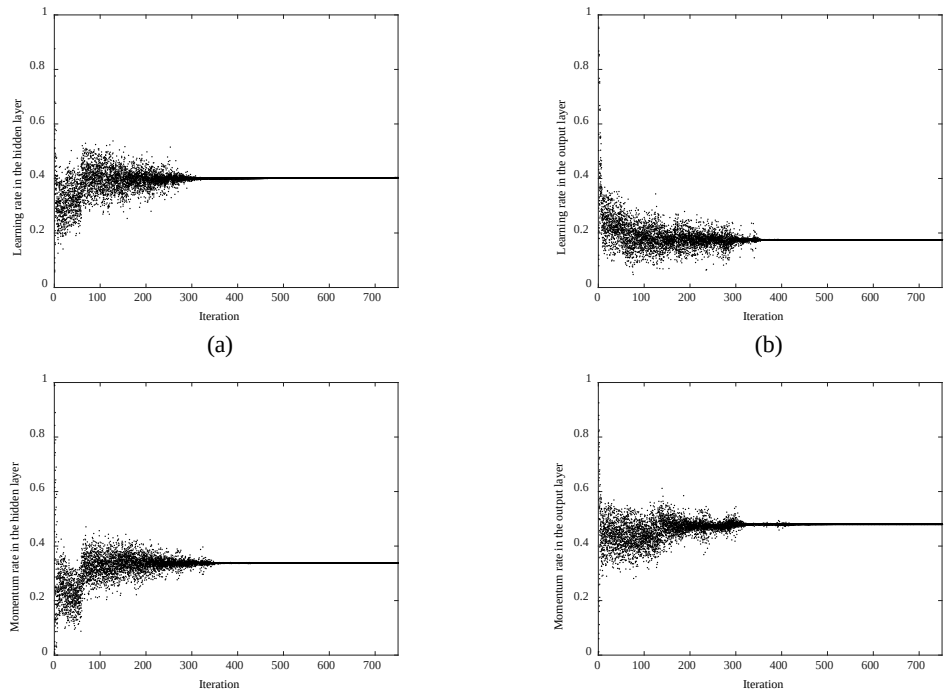


Figure 9: Learning parameters for learning $f^{(1)}$ in successive iterations of the PSO algorithm (a) α_{ih} (b) α_{ho} (c) γ_{ih} (d) γ_{ho}.

V. CONCLUSION

In this paper, PSO algorithm has been successfully introduced to determine the learning parameters required for the BP learning algorithm. To test and to compare the proposed method with the conventional one, a three-layer NN model is designed and it is separately trained to learn nonlinear four test functions. Fair comparisons made between the two methods have shown that the proposed method comes with a relatively high efficiency to learn the test functions over the conventional method for all test cases. As a result, the determination of the learning parameters deploying PSO will help NN models achieve their tasks more effectively even if their topologies are fixed.

REFERENCES

[1] J. Zhao, J. Sun, C.H. Lai, W. Xu, "An improved quantum-behaved particle swarm optimization for multi-peak optimization problems", Int J Comput Math, 88(3), pp. 517-532, 2011.

[2] S. Ghosh, S. Das, D. Kundu, K. Suresh, B.K. Panigrahi, Z. Cui, "An inertia-adaptive particle swarm system with particle mobility factor for improved global optimization", Neural Comput Appl, 21(2), pp. 237-250, 2012.

[3] J. Nayak, B. Naik, H.S. Behara, "A novel nuture inspired firefly algorithm with higher order neural network: Performance analysis", Engineering Science and Technology, an International Journal, 19, pp. 197-211, 2016.

[4] J.H. Wikel, E.R. Dow, "The use of neural-networks for variable selection in QSAR", Bioorg Med Chem Soc, 3(4), pp. 645-651, 1993.

[5] E.J. Ofure, O.O. David, A.M. Oludare, A.A. M, "Artificial Neural Network model for the determination of GSM Rxlevel from atmospheric parameters", Engineering Science and Technology, an International Journal, 20, pp. 795-804, 2017.

[6] S. Haykin, Neural Networks: A Comprehensive Foundation, Upper Saddle River, Prentice Hall, New Jersey, USA, 1999.

[7] V.G. Gudise, G.K. Venayagamoorthy, "Comparison of particle swarm optimization and backpropagation as training algorithms for neural networks", In: Proceedings of the IEEE Swarm Intelligence Symposium; 24-26 April 2003; Rolla, MO, USA: IEEE. pp. 110–117.

[8] O. Üstün, İ. Yıldız, "Determination of the learning parameters in backpropagation learning algorithm by genetic algorithm", SDU Int Tech Sci, 1(2), pp. 61-73, 2009 (article in Turkish with an abstract in English).

[9] S. Sheell, T. Varshney, R. Varshney, "Accelerated learning in MLP using adaptive learning rate with momentum coefficient", In: Second International Conference on Industrial and Information Systems; 9-11 August 2007; Sri Lanka, Penadeniya: pp. 307–310.

[10] P.F. Ribeiro, W.K. Schlansker, "A hybrid particle swarm and neural network approach for reactive power control", IEEE (2006) 1–8.

[11] X.H. Yu, G.A. Chen, S.X. Cheng, "Acceleration of backpropagation learning using optimized learning rate and momentum", Electron Lett, 29(14), pp. 1288-1290, 1993.

[12] P. Jangir, S.A. Parmar, I.N. Trivedi, R.H. Bhesdadiya, "A novel hybrid Particle Swarm Optimizer with multi verse optimizer for global numerical optimization and Optimal Reactive Power Dispatch problem", Engineering Science and Technology, an International Journal 20, pp. 570-586, 2017.

[13] G.Y. Lian, K.L. Huang, J.H. Chen, F.Q. Gao, "Training algorithm for radial basis function neural network based on quantum-behaved particle swarm optimization", Int J Comput Math, 87(3), pp. 629-641, 2010.

[14] Y. Shi, R.C. Eberhart, "A modified particle swarm optimizer", In: Proceedings of the IEEE International Conference on Evolutionary Computation; May 1998; Piscataway. Anchorage, AK: IEEE. pp. 69-73.

[15] D. Chen, C. Zhao, H. Zhang, "An improved cooperative particle swarm optimization and its application", Neural Comput Appl, 20(2), pp. 171-182, 2011.

[16] S.K. Oh, H.J. Jang, W. Pedrycz, "Optimized fuzzy PD cascade controller: A comparative analysis and design", Simul Model Pract Th, 19(1), pp. 181-195, 2011.

[17] S. Tamer., C. Karakuzu, "Parçacık Sürüsü Optimizasyon Algoritması ve Benzetim Örnekleri", In: Elektrik-Elektronik-Bilgisayar Sempozyumu, Elektronik Bildirileri Kitabı; 6–10 Aralık 2006; Bursa. pp. 302–306, (article in Turkish).

[18] W.K. Wong, S.Y.S. Leung, Z.X. Guo, "Feedback controlled particle swarm optimization and its application in time-series prediction", Expert Syst Appl, 39(10), pp. 8557-8572, 2012.

Bug Localization by Using Information Retrieval and Machine Learning Algorithms

M. ERŞAHİN, S. UTKU and D.KILINÇ

[1]Computer Engineering Department, Dokuz Eylul University, Izmir/Turkey, mustafa.ersahin@gmail.com
[1]Computer Engineering Department, Dokuz Eylul University, Izmir/Turkey, semih@cs.deu.edu.tr
[1]Software Engineering Department, Celal Bayar University, Manisa/Turkey drdenizkilinc@gmail.com

***Abstract* – – In large scale software applications, bug localization is a difficult and costly process. Many issues or bugs may be reported at both development and maintenance phase of software development lifecycle. Hence, it is important for developers to discover the location of the bug. In general, source codes and bug reports are used for identifying bug location with the help of Information Retrieval (IR) techniques. In this paper, we present an IR-based bug localization approach named BugSTAiR that uses structured information of source files, source code history, bug reports and bug similarity data if exists. To do best of our knowledge it is the first system developed for JavaScript source files. The experimental results show that accuracy of the system is promising (~%30 on Top 1) on file level bug localization.**

***Keywords* – Bug Localization, Information Retrieval, Machine Learning, Bug Reports**

I. INTRODUCTION

Along with the rapid development in software industry, importance of the software maintainability and software quality have risen drastically [1]. Software quality is evaluated with many different metrics. Software development lifecycle has iterative phases from requirements analysis to maintenance. There can be different issues in each phase that threats the quality of software. Software bugs are one of the most important threats to this process since they are visible and reduce customer's confidence in software. Maintenance phase of software development lifecycle starts after software is released and sometimes maintenance cost can be more than development phases in large scale software projects. When scale of the software project getting bigger, it is difficult to find and fix coding/implementation errors. Therefore, it is important to find buggy source to reduce maintenance time and cost. There are some steps established during the bug fixing processes. The first step is bug localization. In this step, developers use bug report/information from bug/issue tracking system. After they have to overcome time consuming challenges such as reproducing the bug according to bug report, understanding the coding structure, programming logic and goal of the related flow etc. Therefore, there should be some efficient methods to automate bug localization according to the bug reports.

In general, there are two different approaches in bug localization. One of them is dynamic bug localization. Dynamic methods have some processes on the applications' execution time such as runtime traces, data monitoring, tracking execution flows etc. Researchers have proposed various spectrum-based bug localization methods by inspecting a small part of source code. Gopinath et al. [3] proposed a technique that combines spectrum-based bug localization with specification-based analysis to overcome spectrum-based bug localization method's limitations. The other approach is static bug localization. Static methods use bug reports and source code for analysis processes to locate buggy sources.

Static bug localization methods are easy to apply on any phase of software development since they have few external dependencies and have relatively low computational cost with the help of Information Retrieval (IR) algorithms. FindBug is a popular static bug localization tool that has been proposed by Hovemeyer and Pugh (2004).

Recently, many researchers have worked on IR-based bug localization techniques. IR is the science that deals with the representation, storage, organization of and access to information items [2]. IR approaches have two important concepts such as *query* and *document collection*. Each bug report represents a query and the source files to be searched indicate the document collection. IR techniques use these inputs to rank documents by relevance, according to ranked list of candidate source files that may contain the bug. Ranking process has consecutive phases that starts with bug report creation. Then user enters a bug report query into the system then IR techniques computes a numeric score for all candidate source files that match to the bug query. Finally, top-ranking candidate source files are listed for developer consideration.

The success of an IR technique is highly dependent to algorithms used in retrieval processes. Rao et al. [4] have compared the main IR techniques; Unigram, Vector Space Model (VSM), Cluster Based, Latent Dirichlet Allocation, Latent Semantic Analysis (LSA) and some various combinations. Poshyvanyk et al. [5] have developed PROMESIR which uses a probabilistic ranking method and a data acquisition method called Latent Semantic Indexing (LSI). [6] Ngyuen et al. customized the LDA approach and proposed BugScout [7]. Therefore, a different approaches and techniques

proposed to improve the efficiency of information retrieval and bug localization processes. According to these improvements, BugScout took the lead because of its good performance on some large-scale datasets. Because most of the researchers used a few bugs in the evaluation process before BugScout.

Zhou et al. [7] proposed BugLocator that used rVSM(revised Vector Space Model) and performed on some high scale open source projects. BugLocator uses text similarity between source files and bug reports. Also, it uses the information about previously fixed bugs to improve bug localization accuracy. BugLocator has better experimental results than BugScout on compared datasets. Then, another approach is introduced by Saha et al. [8] BLUIR (Bug Localization Using Information Retrieval) uses structured information analysis of source code such as class names, method names etc. BLUIR has located more bugs than BugLocator according to the experimental results on the same datasets. Thus, using structured information of source file is more efficient then simple source file as a document. Youm et al. [9] has proposed BLIA (Bug Localization using Integrated Analysis) by using some other information about bugs. In addition to the bug similarity information that BugLocator used, BLIA uses stack traces, comments in bug report and change history of the source code to have better accuracy. Also, BLIA is a bug localization algorithm that provides multi-level scoring such as method-level scoring, file-level scoring etc. In order to comparison, BLIA evaluated on the same datasets which was used to evaluate for both BugLocator and BLUIR. According to the evaluation results, BLIA was better than both BLUIR and BugLocator in those datasets.

All of these approaches evaluated on well-known open source datasets. These datasets are Eclipse, AspectJ, SWT and etc. All of these software's are developed with Java programming language. Some of implementation details of Java programming language can help to information retrieval process to have better accuracy. For example; stack trace of an exception points the buggy file and function directly according to the its own information, filename should be same with the class name that is declared as public. Across to this, JavaScript scripting language is very flexible and do not force any naming convention. Therefore, it is more difficult to have accurate results when localizing the bugs according to both structured or unstructured source file information. When we looked at the literature, there is not any experimental result on JavaScript based software and datasets.

The aim of the paper is to propose a new bug localization approach named BugSTAiR that is specifically for the software products developed using JavaScript and JavaScript-based web frameworks such as AngularJS, ReactJS etc. As far as we know there is not any proposed study that realizes bug localization for JavaScript source files. There are two different bug types related with JavaScript based web applications. Because, JavaScript can manipulate both application logic and UI components such as DOM elements. DOM elements are generally changed by Cascading Style Sheet (CSS). Therefore, it is difficult to find the exact location of UI related bug. Both JavaScript files and CSS files are considered as suspicious. According to this fact, we will focus only logic related bugs to minimize this effect on our retrieval results. We have used a commercial software to experiment our approach with hundreds of bugs.

II. BACKGROUND

In this section, we demonstrate IR-based bug localization and machine learning techniques to boost our retrieval results. The idea behind our proposed approach is finding relevant source files according to the matching words between source files and bug reports. Then, machine learning algorithms help us to find similar bugs that found before to boost our retrieval result.

A. Bug Localization Example

In literature, there are many open source software products with dataset that includes lots of bug summary. But it is difficult to find web applications developed using JavaScript-based frameworks having a bug report dataset. Figure 1 presents a real-world bug report from our commercial application developed for a bank.

Bug ID: 1433

*Summary: Account Settings – Next Button is not working after entering **security question**. Button should trigger a redirect action to the **success state***

Source Code File: securityQuestionPageHelperFactory.js

```
angular.module([     ].core").factory("securityQuestionPageHelperFactory",
    function (securityQuestionApi, smsOtpConfigFactory, smsOtp, $state) {
        "use strict";

        function securityQuestionPageHelper(securityQuestionPage) {
            var self = this;

            this.validateSecurityQuestionAnswer = function () {
                securityQuestionApi.validateSecurityQuestion({
                        "answer": securityQuestionPage.formData.answer
                    }, {
                        "skipDefaultErrorAlert": false,
                        "onSuccess": self.onValidateSecurityQuestionSuccess
                    }
                );
            };

            this.onValidateSecurityQuestionSuccess = function (resp) {
                if (securityQuestionPage.config.smsOtp) {
                    self.startSmsOtpFlow();
                } else {
                    $state.go(securityQuestionPage.config.successState.name,
                        securityQuestionPage.config.successState.params,
                        {"location": "replace"});
                }
            };
```

Figure 1: Bug summary example of a real-world web application

Bug localization process lies on finding similarity score between bug reports and source code files according to result of

common matching word algorithms. All of the source code files have a computed similarity score for each bug report. This process has some steps that should be executed in a pre-defined sequence. We will cover these pre-defined steps in next section.

B. Bug Localization Process

IR-based bug localization approaches have five general steps described as follows.

- Preprocessing: This step is related with both source code files and bug reports. All files and bug reports should be preprocessed to improve efficiency of retrieval process. This process actually removes all stop-words from source code such as language specific identifiers, punctuations etc. Also, some syntactic operations such as camel case splitting, lower case transformation and natural language processing operations like word stemming, tokenization are evaluated during the preprocessing period.
- Indexing: IR-based bug localization approaches are used to index dataset that is ready when both source file and bug reports are preprocessed and dataset is prepared. Vector space model (VSM) is one of the well-known model but there are also some other probabilistic models such as LSI.
- Querying: Query construction is one of the important part in IR processes. In general, summary and description fields of bug reports are used for retrieval.
- Similarity Computation: There are several methods to compute similarity between bug reports and source code files. Each IR approach apply one of these methods to compute the relevance.
- Retrieval: After all the steps above are realized, each IR approach apply its proposed algorithm or method to obtain better accuracy on retrieving process.

III. Proposed Approach

In this section, we give detailed information about our proposed approach including preprocessing, indexing and similarity activities. Finally, we introduce our ranking approach.

A. Architecture

Bug localization in JavaScript-based web applications are hard to implement and retrieval results cannot be as accurate as Java-based software applications. The main reason that is a user interface of application is related more than one file simultaneously. Further, a web application has a lot of same filename with different type such as featurex.html, featurex.css and featurex.js. This situation causes an extra complexity to all computations on bug localization process. Thus, we decided to design our architecture to locate non-ui related bugs such as logic, flow, etc. Figure 2 shows the general architecture of our proposed approach.

After preprocessing step, we build an index with different attributes such as class names, methods, file content and object keys. Then, the first retrieval is performed on all attributes with empirical boost parameters. After, bug similarity score for each bug report is calculated with the help of machine learning algorithms. Finally, we combine the retrieval result and similarity result according to our algorithm.

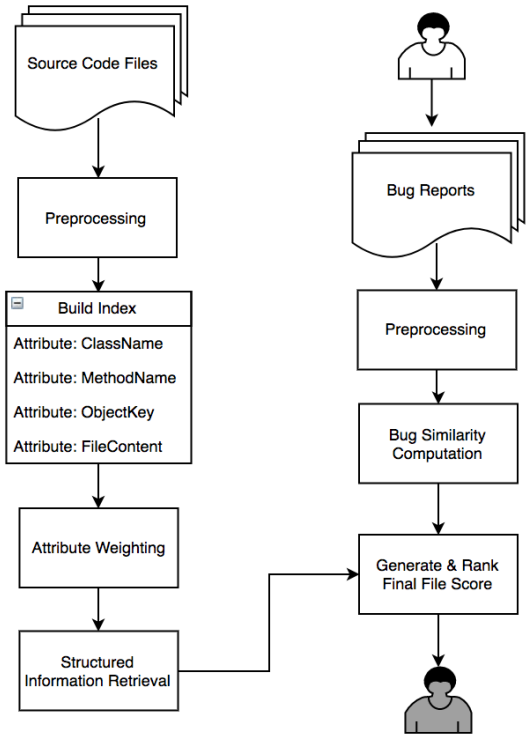


Figure 2: General view of BugSTAIR architecture.

B. Source Code & Bug Report Preprocessing

As it is mentioned in Architecture section, JavaScript files are the only accepted input to be processed. Both HTML and CSS files are excluded from source code repository. Also, all UI-related bugs are eliminated while getting the bug reports from issue tracking system. Then, all stop-words are removed from both source code files and bug reports. Stop words includes keywords such as;

- English stop words: "a", "the", "to" etc.
- Syntactic symbols/identifiers: "null", "undefined", "alert", "init" etc.
- Operators / Punctuations: "===", "!=", "<", ">" etc.

We use Camel case naming convention according to the company's software development standard. In order to increase the accuracy of the retrieval we tokenize identifiers. When we tokenize the identifiers we have individual tokens, but there can be some conflicts between bug report and source code on case

sensitivity. To solve these conflicts, we transform all the texts to lower case format.

C. Indexing

In this section, we present our source code indexing approach. We use Apache Lucene [10] to index the source code files. Lucene is one of the most-known open source IR system. In our approach, Lucene evaluates a relevance score between source code files and bug reports. There are two ways for multi parameter indexing. Some researchers prefer building a different index for each attribute and the others prefer one index file that includes all attributes in different dimension. We prefer building an index with all attributes that we need by the help of Lucene's powerful API's.

D. Query Construction and Retrieval

Source code files are called document collection and bug reports are evaluated as a query in IR process. Since the bug reports are preprocessed in the first step, query construction is done for retrieving. All words of bug report summaries are tokenized and prepared for all attributes. At this point, we give constants for each attribute according to the primitive effects/weights on retrieval result. Table 1 shows our experimental results on attribute weighting process.

Table 1: Result of Attribute Weighting Experiments

	Top 1 Score	Top 5 Score	Top 10 Score
Structured Source Code Retrieval with CLASS NAME attribute	14.23	38.79	49.46
Structured Source Code Retrieval with FUNCTION attribute	6.76	25.97	37.01
Structured Source Code Retrieval with OBJECT KEY attribute	8.18	24.91	46.26
Structured Source Code Retrieval with FILECONTENT attribute	9.6	30.96	39.85

The results indicate that CLASS NAME attribute is the most valuable attribute on our dataset.

E. Bug Similarity and Combining Results

The aim of searching a similar bug that is fixed is finding suspicious files which is modified to fix for relevant bug. There are several methods to calculate similarity. We prefer a popular machine learning approach that is name K-NN (K-Nearest Neighbor)

After we run the K-NN algorithm, we select the top three similar bugs to evaluate. The modified files of these three bugs may be better candidates for bug location. Therefore, we use these files to boost structured IR scores.

IV. CASE STUDY

In this section, we present our case study and overall experimental results. We use a commercial web application that is developed by our development team. The detailed information about application is below;

- Responsive Web Application developed using AngularJs 1.5
- Development language (Method names, variables etc.) is in English
- Bug summaries are also in English
- The number of JavaScript file related bugs is 313

A. Experimental Results

We present the experimental results using our proposed approach that is defined in previous sections. First, we prepare both source code and bug reports. Then, we perform our approach step by step. We have a base information retrieval with dataset. After, we have some improvements on retrieval results according to the steps that we execute. Table 2 shows the evaluation of our best results on different retrieval conditions.

Table 2: Result of Bug Localization

	Top 1 Score	Top 5 Score	Top 10 Score
Unstructured IR with dataset	8.18	20.64	31.67
Structured IR with dataset	5.33	28.11	44.83
Structured IR with Attribute Weighting	24.19	50.88	66.54
Structured IR with Attribute Weighting and Bug Similarity	29.71	45.04	54.67

V. CONCLUSION

In this paper, we propose BugSTAiR, a bug localization approach using both IR and machine learning algorithms for especially JavaScript based web applications.

To the best of our knowledge, BugSTAiR is the first bug localization tool that works for JavaScript based web applications by using IR and machine learning. A commercial web application and its bug report dataset is used to experiment proposed approach. Because there is not any bug report dataset of open source web application. Thus, there is not any result to compare success of the proposed approach.

In future, we would like to optimize proposed bug localization approach to localize UI-related bugs from our dataset. Localizing UI-related bugs as important as Non-UI related bugs in web application development. We also would like to generate more generic approach for software

applications that is developed with different programming languages. Finally, we would like to develop an integration plugin for common issue tracking tools.

REFERENCES

[1] Hovemeyer,D; W.Pugh, "Finding bug is easy", *ACM Sigplan Notices,39*, pp.92-106,2004

[2] Ricardo B.Y.; R.N.Berthier, "Modern Information Retrieval", *ACM Press, Newyork*, pp.189-190, 1999

[3] D. Gopinath; R.N. Zaeem; S. Khurshid, "Improving the effectiveness of spectra-based fault localization using specification", *Automated Software Engineering (ASE),2012 Proceedings of the 27th IEEE/ACM International Conference*, pp. 40-49, 2012

[4] S. Rao; A. Kak "Retrieval from software libraries for bug localization: a comparative study of generic and composite text models", *Proceedings of the 15th IEEE International Conference on Program Comprehension*, ICPC'07, pages 37-48, 2011

[5] Poshyvanyk D;Y.G. Gueheneuc; A.Marcus; G. Antoniol; V.Rajlich, "Combining probabilistic ranking and latent semantic indexing for feature identification", *Proceedings of 14th IEEE International Conference on Program Comprehension (ICPC 2006)*, Athens, Greece, 137-146, 2006

[6] Poshyvanyk D;Y.G. Gueheneuc; A.Marcus; G. Antoniol; V.Rajlich, "Feature location using probabilistic ranking of methods based on execution scenarios and information retrieval", *IEEE Transactions on Software Engineering*,33, 420-342, 2007

[7] J. Zhou , H. Zhang , D. Lo , Where should the bugs be fixed? more accurate information retrieval-based bug localization based on bug reports, in: *Software Engineering (ICSE), 2012 34th International Conference on, IEEE*, 2012, pp. 14–24 .

[8] R.K. Saha , M. Lease , S. Khurshid , D.E. Perry , Improving bug localization using structured information retrieval, in: *Automated Software Engineering (ASE), 2013 IEEE/ACM 28th International Conference on, IEEE*, 2013, pp. 345–355 .

[9] Klaus Changsun Youm,June Ahn,Eunseok Lee, "Improved bug localization based on code change histories and bug reports", *Information and Software Technology, Elsevier*, 2017

[10] Lucene, *Information about Apache Lucene*, Retrieved June 21, 2014, from http://lucene.apache.org/core/

An Overview for the National Cyber Security Strategy

C.ATAÇ[1] and S.AKLEYLEK[2]

[1] Ondokuz Mayis University, Samsun/Turkey, cihan.atac@gmail.com
[2] Ondokuz Mayis University, Samsun/Turkey, sedat.akleylek@bil.omu.edu.tr

Abstract **- Rapid developments in information and communication technologies over the last few decades have led to cyberspace becoming a part of our daily lives. Protecting cyber space from harmful activities has become a critical point of action for policy makers around the globe as societies, governments and businesses become increasingly dependent on the internet. So, it is important to prepare and develop a national cyber security strategy (NCSS) for the management of information and communication systems, ensuring citizen safety, and protecting critical infrastructures. In this study, the concepts of cyber security were mentioned and the national cyber security strategies of some of the leading were examined. In the last part, the implementation stages of a good national cyber security strategy were focused on.**

Keywords **- cyber security, NCSS, critical infrastructure, implementation stages, strategy.**

I. INTRODUCTION

The dizzying developments in information technologies are transforming social life by connecting people and providing new ways to communicate and co-operate with one another; business by making trade and manufacturing more efficient and effective; government institutions by serving online. The convenience of digitalization in every aspect of life has led to the emergence of a technology-dependent lifestyle. The continuity of the lifestyle can be achieved by ensuring that information technology is used safely and securely. Especially the cyber attacks on the critical infrastructure and economic structures of the countries have led to the serious consideration of the concept of cyber security. Thus, countries have formed national cyber security strategies with the aim of protecting cyber space.

The first steps in this area were taken by US, Japan, Russia, Germany, UK. These countries first prepared their NCSSs and then put them into practice. Thanks to these strategies, which have a vision, scope and methodology, these countries have guided at cyber security topic.

In the next sections of the study, knowledge about basic cyber security concepts were given; US, Japan, Germany and UK national cyber security strategies, which are top in the ITU 2017 report, were examined and finally the implementation steps of an effective cyber security strategy were discussed.

II. BASIC CONCEPTS AND TERMINOLOGY

There are some key concepts and terms for exploring and identifying "cyber security strategy" subject. They are listed and defined below for warming up and also creating base for the subject.

A. Cyberspace

Cyberspace, commonly referred as the virtual computing world, is a global computer network that provides and facilitate online communication [1].

B. Cyber attack

A cyber attack is a kind of maneuver that targets information systems, infrastructures, computer networks, personal computers and works through harmful actions that are usually anonymous. The attackers who employed by the states, individuals, groups, society or organizations steal, change or destroy the source by hacking a specific system[2].

C. Cyber security

Cyber security is to protect computer systems and their associated software, hardware and information from being stolen or damaged or services they provided from being disrupted or misdirected [3].

D. Cyber defense

Cyber defense is a computer network defense mechanism that includes responses to the protection of critical infrastructures and the provision of information security against attacks targeting organizations, government agencies and other potential networks [4].

E. Cyber warfare

The term cyber war includes the use of cyber space as a battlefield and the targeting of computers and networks in this war. There are threats of espionage and sabotage involving both attack and defense operations [5].

F. Cyber terrorism

Cyber terrorism is the use of the internet or cyber space to carry out acts of violence that can lead to political gains, life loss or significant bodily harm [6].

G. Critical infrastructure

The critical infrastructure is all the networks, systems and assets that need to constantly work to ensure security, economy, public health and safety with vital preventive measures for a country [7].

H. Strategy

A strategy is an overall plan or a series of plans that are specifically designed to achieve something over a long period of time [8].

III. SOME LEADING NATIONAL CYBER SECURITY STRATEGIES (NCSS)

There are lots of nations developed and published national cyber security strategy. Before mentioning the main points in preparation for implementing the national cyber security strategy, we will review national cyber security strategies of the leading countries in this regard and we will make a comparison. Taking these analyzes into consideration, we will also look at how and what should be done for preparing a correct cyber security strategy.

A. United States (US)

The United States has a settled as well as framed cyber security structure. In the country, the tasks and authorities of the institutions are completely determined and also the attempts presented to do this work are performed frequently, and the technological adaptation is guaranteed. The US routinely revisions national strategies for protected cyber space, sets innovative objectives, and shares these strategies with all the community.

The formal cyber organization of the United States has an extremely sophisticated building that is directly relevant to the decentralized management form of US federative conception. The US recognized cyber institution mainly possesses a trilateral structure: The United States Department of Defense, The Department of homeland Security and The United States Secret Service (FBI / CIA). Beyond that, various government agencies also have responsibilities and authority for their responsibility areas. Furthermore, state governments additionally would rather intentionally use these constructions by setting up a variety of components to guarantee their own cyber security, besides the national cyber security system [21].

The "*National Strategy to Secure Cyberspace*" document, released in February 2003, is the initial extensive document which describes the US cyber space region, sets out the goals as well as strategies in this amazing area, identifies the intended program of the way the national cyber space is preserved and also identifies cyber space threats. The strategic aims of this document happen to be:

- Protect US critical infrastructures against cyber attacks;
- Decrease national vulnerability to cyber strikes;
- Minimize harm and also healing time frame from cyber attacks which do appear [9].

The National Strategy to Secure Cyberspace revealed several key initiatives and actions for cyberspace security and safety to lower threats along with associated vulnerabilities and to strengthen US national security together with global cooperation. These were:

1) Improve national incident management;
2) Exercise cyber security continuity plans;
3) Improve and enrich public private information sharing mechanisms;
4) Secure the systems belonging to the Internet by bettering protocols and routing;
5) Foster the usage of reliable digital control systems;
6) Lessen and then repair software vulnerabilities;
7) Encourage national recognition program;
8) Raise enough education and training programs;
9) Improve abilities for attack attribution and also response.

The "*Cyberspace Policy Review*", prepared in 2009, has been criticized mainly for the multi pronged structure of federal government organizations included in the US cyber defense system. It has been noted in the review that some measures should be taken in order to address the problems arising from this situation to make the national cyber security systemmore effective. This document also underlined several factors, such as the White House's role as a pioneer in security politics; boosting public awareness and education; enhancing partnership between private sector and government and connecting research & development frameworks to infrastructure development [10,16].

The document "*International Strategy for Cyberspace: Prosperity, Security, and Openness in a Networked World*," which was announced in May 2011 in the United States and in the world, is a cyber security strategy document that sets out the goals and objectives of the United States in cyberspace at the international level. Some policy priorities explained in this document were [11]:

- Securing economic activities;
- Conservating US network structures;
- Applying laws about siber crime;
- Balancing of private life and safety on internet;
- Managing internet;
- International cooperation.

With the document "*The Department of Defense Cyber Strategy*" adopted on April 23, 2015, the US Armed Forces were assigned tasks like defending US network technology and systems and confidential cyber information, protecting US interests against cyber attacks, planning military and secret cyber operations, guiding operations [12].

Finally, the budget table of US for cyber security, which was separated from the year 2004 to 2017 is presented below [9,10,11,12].

Table 1: US budget for cyber security

2004	$ 18 million
2015	$ 1 billion
2016	$ 14 billion
2017	$ 19 billion

B. Japan

For Japan, which has the geography that is exposed to risks of natural disasters and also has a knowledge-based economy, provision of the security of critical information infrastructures has an important priority. This is one of the few states of global examples that first constitutes a comprehensive national strategy and institutional structuring model [16].

The Japan official cyber organization basically consists of Cybersecurity Strategic Headquarters, National Security Centre (NSC), IT Strategic HQ's, National center of Incident readiness and Strategy for Cybersecurity (NISC) which are connected to Cabinet, and Relevant Organizations [13].

The "*Basic Law on Information Technologies*", issued in 2000, laid the groundwork for institutional structuring, and in 2006 the "National Information Security Strategy" entered into force. In the first part of the strategy, while emphasizing Japan's national development goals as its main objective, the following sections focused on institutional structuring, the strategic goals within the three-year timeframe, and the legal framework required to implement the strategy.

In this document, it has been emphasized that individuals should continue their personal development in information security, raising awareness, providing R & D and technological development, encouraging international cooperation, fighting cybercrime to create a safe cyber space.

In addition to these key strategic priorities, 212 actions targeting central and local authorities, critical infrastructures, the business community and the citizen have been accomplished.

The "*Cyber Security Strategy"* document, published in 2013, was updated in 2015. In this document, the objectives were: Provide an independent, just, and safe cyberspace; and afterwards subscribe to developing economic power and sustainable growth, structuring a society where the people can live secure and trustworthy lives, and guarantee peace and robustness of the international society and national safety [15].

Some policy approaches towards achieving the the objectives are;

- Constructing secure new communication devices such as IoT;
- Providing stability and peace of the global community;
- Collaboration and cooperation with states throughout the world;
- Precautions for protection of the people and society;
- Measures for the protection of governmental bodies;
- Ensuring national security.

In addition, Japan is raising cyber safety for the 2020 Olympics. The Japanese government has decided to establish a panel to prevent any problems with cyber security during the 2020 Tokyo Olympics. A high-level bureaucrat will be found at the beginning of this work, which aims to combat cybercrime, and will coordinate with each ministry once they have identified the steps they need to make about cyber security. Apart from this, cooperation with similar institutions of other countries will be developed [22].

Lastly Cyber-related budget for Japan presented in "Overview of FY2017Budget Request" document is ¥ 12.5 billion [20].

C. Germany

Germany is quite advanced in terms of national cyber security around EU countries due to strategic perspective as well as the German Information Security Agency (BSI), a well-established institution. Arne Schönbohm, President of the Federal Office for Information Security (BSI), said: "In the 25 years since it was formed, the BSI has developed from the encryption of information, via protection of the government network, into the national cyber security authority with responsibility for the organization of information security in the digitalization process." in "*The State of IT Security in Germany 2016*" report [14].

Germany emphasizes the importance of protecting critical sub-structures and the international level of cyber security and cyber defense mechanisms in their strategies.

Germany, which published the first national policy document in the field of information security in 1989, states in this document that "Federal Government informs all relevant and responsible parties (producers, public authorities and users) in the field of information security about information technology risks and takes protective measures " [16].

In 2005, the "*National Strategy on the Protection of Information Security and Critical Infrastructures*" has been published. This strategy had three main strategic objectives. These were [16]:

- Prevention;
- Preparedness;
- Sustainability;

The "*German National Cyber Security Strategy*", which entered into force in 2011 could be summarized by items summarized as follows [16]:

- Protection of critical structures and information systems;
- Developing reliability of IT systems;
- Guarantee the security and safety of government agencies on IT;
- Being sure on controlling the cyber crime;
- Developing International cooperation to ensure cyber security;
- Trustworthy technology usage;
- Training of personnel in public institutions;
- Developing effective and secure defense vehicles.

The "*State of IT Security in Germany 2016*" published by BSI in 2016. In this 2016 Status Report, the BSI provides information about current risks for IT security in Germany as well as countermeasures [14].

In March 2015, the German government released a plan to promote IT security R&D. The plan included a budget of €180 million (~$198 million) through 2020 intended to promote specific encryption technologies and further protection of personal data and communications services. The important areas which the plan highligted were: recent technologies, safe and reliable information and communication systems,

implementation fields of IT security, confidentiality and protection of data [14].

D. UK

The UK, one of the few countries using digital technology effectively in the world, knows that prosperity and power are largely due to their good protection of technology, data and networks.

UK's National Cyber Security Strategies are presented for exhibiting its plan to build UK safe, skilled and flexible in a rapid-acting digital world.

The UK's policy on cyber security strategy obviously declares the duties belonging to the public, the private sector and the state. The Cabinet Office affiliated to the Prime Minister, the Government Communications Headquarters-GCHQ and the Ministry of Defense-MoD and also their subsidiary institutions constitute the cyber security organization in the UK [23].

The initial strategy document, "*Digital Britain*" was issued in 2009 as a necessity for reaching the UK's aim of becoming one of the world's major digital information economies [16].

In 2011, the document "*The UK Cyber Security Strategy Protecting and promoting the UK in a digital world*" was released. In this specific paper, vision for 2015 was driven as to extract enormous financial and communal worth from an energetic, secure and flexible cyberspace. Four objectives below have been listed to accomplish this vision by 2015 [24]:

1) Making the UK one of the world's safest regions to conduct business in the virtual environment;
2) The fact that UK is more resistant to cyber attacks and that citizens' interests can better protect cyber space;
3) Creating an free, stable and vibrant cyberspace;
4) Developing the concurrent data, capability and skills to support all citizens' cyber security targets.

To achieve these objectives in the document some actions were set out for what Government would make, jointly with the private sector and other countries. Some of them were:

- Cooperating with firms that possess and control UK critical infrastructure for ensuring important data and systems to remain secure and flexible;
- Set up functional partnership with the private sector on information sharing in cyberspace;
- Make a strategic summit with inspectors, including specialized business services, insurers and lawyers, for more effective management of cyber risks;
- Engage in global cooperation;
- Seek arrangement with Internet Service Providers (ISPs) on the assistance they may present to internet users;
- Promoting and developing education, vital skills and R&D at all levels;
- Establish a single point of authority for public and small businesses to stay safe online [24].

To achieve the objectives set out in this strategy paper, the UK government has allocated £ 650 million of public funding for the four-year National Cyber Security Program [24].

In 2015, "*National Cyber Security Strategy 2016-2021*" document was published. This document states that the UK has a vision of 2021 for remaining safe, resilient and self-confident in digital world. To realize this vision UK has identified the following objectives [15]:

- *Defend* : In the report it is mentioned that UK holds the methods to protect the state and nation against growing cyber threats, to take action successfully to events, to assure UK networks, data and systems.
- *Deter*: : The strategy document determined that UK is a difficult target for all kinds of aggression in cyberspace.
- *Develop* : The Strategy document states that the UK has an advanced cyber security industry, supported by world-leading scientific research and development [15].

Also below actions that is shortened, listed in strategy document:

1) More effective implementation of the cyber advocacy;
2) Make the world of the Internet more reliable;
3) To protect the government and its components;
4) Protect critical national infrastructure;
5) Contributing to the development of the public and private sector;
6) Identify cyber threats well and intervene effectively;
7) Cyber's function in deterrence;
8) Reducing cyber crime;
9) Countering aggressive foreign actors;
10) Preventing terrorism;
11) Enhancing sovereign abilities - offensive cyber;
12) Enhancing sovereign abilities – cryptography;
13) Supporting cyber security abilities;
14) Encouraging development inside the cyber security sector;
15) Actively encoraging cyber security science along with technology;
16) Working with international partners.

In the "*National Cyber Safety Strategy 2016-2021*", it is stated that a total of £ 1.9 billion will be funded in the next five years to significantly transmute the UK's cyber security.

IV. IMPLEMENTATION STAGES OF AN EFFECTIVE NCSS

Implementation stages of an effective national cyber security strategy can be examined in 4 main groups related to each other. This classification is based on our research experience on this subject. These are:

A. Government and Institutional Dimension
B. Human Dimension
C. Legal Dimension
D. Technology and Standards Dimension

A. *Government and Institutional Dimension*

1. National cyber contingency plans development

National cyber contingency plan is important and integral element of cyber security strategy. A NCP needs to have the below objectives [17,18]:

- It should have specific criteria to determine that the resulting circumstance is a crisis situation.
- The steps to be taken to triumph over the crisis has to be definitely determined within a procedure.
- Role and responsibilities of stakeholders must be identified at the time of the cyber crisis.
- Alternative plans needs to be created within the NCP based on the size of the cyber crisis.
- Work should be done to draw lessons from previous cyber crises and also to be able to review the NCP.

2. Preservation of critical infrastructure

One of the most essential elements of the national cyber security strategy is the protection of critical infrastructures (CII). Energy, health, telecommunications, water are some good examples of a CII. Here are some important things to be done [17,18]:

- Specify CII;
- Define and decrease threats about CII;
- Develop sector specific protection plans.

3. *Incidence response capability establishment*

Relevant stakeholders and the Computer Security Incident Response Team (CSIRT) play an important role in event management in coordination. Relating goals are [17,18]:

- There ought to be a central registry of national level cyber incidents.
- Roles and responsibilities of staff members must be identified.
- Team's Operational capabilities should comply with.
- Cooperation abilities of teams should be really good.

4. Cyber security exercises planning

Cyber security exercises are an important tool for assessing the preparedness of institutions and the community, such as cyber attacks, natural disasters, technology problems. Some of the things that need to be done are [17,18]:

- The issues such as the plans, infrastructure, intervention capabilities and communication that should be tested should be determined.
- A clear and clearly defined cyber exercise team should be established.
- Cyber exercises should be integrated with the life cycle of the national cyber security strategy.
- Assess the result of one or maybe the number of cyber exercises and upgrade your vision to supply the necessities of the cyber security strategy.

5. Provision of public-private partnership

Operation of critical infrastructure facilities is largely in private sector in many countries. In this respect, the good cooperation and communication between governments and the private sector is very important in terms of the protection of critical infrastructures. Providing public-private partnership is important to achieve the following objectives:

- Deterrence to prevent attackers;
- Protection from new security threats through research;
- The use of information sharing to combat new threats;
- Reaction to destroy the effect of the attacks;
- Ability to repair the devastating effects of the attacks.

6. Balance of Private Life Respect and Providing Safety

In a national cyber security strategy, the right balance must be sought between respect for private life and ensuring security [17,18].

7. Research & Development encouragement

Research and development work on cyber security is important for the development of tools against new types of attacks. For this purpose [17,18]:

- Describe the real causes of security vulnerabilities instead of correcting their effect;
- Find solutions to complex and multidimensional problems by bringing people together from different areas of expertise;
- Providing the transition from theory to practice by matching research findings and industry needs.

8. Participation in international collaboration

Collaboration and sharing of information with partners abroad is important in terms of understanding the changing threat environment and giving them the necessary response. Here are the things to be done [17,18]:

- Establishing a data repository where partners can share their experiences in cyber security;
- To create synergy between national cyber security personnel;
- Increasing the effect of struggle of international crime.

B. Human Dimension

1) *Awareness Raising*

The human is core to cyber security. At each and every step in developing and executing strategies to guarantee cyber security, people are an important component of both the process and the solution. Raising awareness of individual and institutional users about threats and weaknesses in the online world is crucial in ensuring cyber security. Common objectives are to [17,18]:

- Identify breaks of information or perhaps recognition regarding cyber security or maybe information security issues;
- Close the spaces by organize campaign or national security month, week, day.

2) *Strenghten training and educational programs*

Training and educating people on cyber security issues has very important effects for preventing attacks or lessen impacts

of these attacks to information systems. The targets of the education and training programs are as follows [17,18]:

- To improve the technical skills of employees in the information security sector by working in cooperation with universities and the private sector;
- Align cyber security training with business requirements;
- Support students to participate in workshops and training activities related to cyber security;
- Encourage strengthening of relationships between industry and academic environments on information security.

C. Legal Dimension

Cyber crime detection and determination

Preventing the increase of cyber crimes depends on the cooperation of many stakeholders on the issue and thus the determination of cybercrime. Here's what to do to handle a cybercrime:

- Establishing legal infrastructure for cyber crime;
- To increase the responsibility and power of law enforcement officers in cyber-crime fighting [17,18].

D. Technology and Standards Dimension

1) Adherence to standards

To be attached to cybersecurity related standards in production and software development is very important issue for especially in Critical Infrastructures (CI) [19].

2) Use of up-to-date and quality software and hardware

This factor identifies the use up to date hardware and the quality of software deployment and the functional requirements in public and private sectors [19].

3) *Cryptographic controls*

This factor reviews the deployment of cryptographic techniques in all sectors and users for protection of data. [19]

V. CONCLUSION

This study examined the US, Japan, Germany and UK, the leading countries in the world in creating, publishing and implementing national cyber security strategy. The visions and scopes, priorities, principles, objectives, institutional structures and approaches to implementation of the national cyber security strategies are explained.

Subsequently, the implementation sections of an effective national security strategy were assessed on four main headings. According to the assessments it has been found that the provision of cyber security in countries is equivalent to the prosperity and security of the countries, and that they are aware of the seriousness and the importance of the case. In addition, it has been noticed that countries can demonstrate strategic differences in institutional cyber security structures. One of the countries is implementing the strategy with a more distributed mechanism while the other is establishing the central structure. These countries, which constituted the national cyber emergency plans, regard the protection of their critical infrastructures as a priority. In recognition that human dimension is the weakest chain of the cyber security. They seem to pay special attention to research and development work on cyber security. They are aware of the fact that the country's cyber security can only be achieved by establishing international cooperation, not by local work. It is also well known that the legal dimension has an effective role in identifying and preventing cybercrime, and it draws attention to this issue in strategy documents. They comply with national standards and give the necessary value for the use and production of up-to-date technology and software.

ACKNOWLEDGMENTS

This research is partially supported by OMÜ under grant no. PYO.MUH.1906.17.003.

REFERENCES

[1] https://www.techopedia.com/definition/2493/cyberspace, accessed on March 10, 2018.
[2] http://www.wikizero.com/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvQ3liZXJhdHRhY2s, accessed on March 10, 2018.
[3] http://www.wikizero.com/index.php?q=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnL3dpa2kvQ29tcHV0ZXJfc2VjdXJpdHk, accessed on March 11, 2018.
[4] https://www.techopedia.com/definition/6705/cyber-defense, accessed on March 11, 2018.
[5] https://en.wikipedia.org/wiki/Cyberwarfare, accessed on March 11, 2018.
[6] https://en.wikipedia.org/wiki/Cyberterrorism, accessed on March 11, 2018.
[7] http://whatis.techtarget.com/definition/critical-infrastructure, accessed on March 11, 2018.
[8] https://www.collinsdictionary.com/dictionary/english/strategy, accessed on March 11, 2018.
[9] The White House, Washington, *National Strategy to Secure Cyberspace*, 2003. Available: https://www.energy.gov/sites/prod/files/National%20Strategy%20to%20Secure%20Cyberspace.pdf
[10] The White House, Washington, *Cyberspace Policy Review, Assuring a Trusted and Resilient Information and Communications Infrastructure*, 2009. Available: https://www.dhs.gov/sites/default/files/publications/Cyberspace_Policy_Review_final_0.pdf
[11] The White House, Washington, *International Strategy for Cyberspace, Prosperity, Security, and Openness in a Networked World*, 2011. Available: https://obamawhitehouse.archives.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf
[12] The Department of Defence, *DoD CYBER STRATEGY*, 2015. Available: https://www.hsdl.org/?abstract&did=764848
[13] The Government of Japan, *Cybersecurity Strategy*, September 2015. Available: https://www.nisc.go.jp/eng/pdf/cs-strategy-en.pdf
[14] Federal Office For Information Security, *The State of IT Security in Germany 2016*, 2016. Available: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/Publications/Securitysituation/IT-Security-Situation-in-Germany-2016.pdf?__blob=publicationFile&v=3
[15] HM Government, *National Cyber Security Strategy 2016-2021*, 2015. Available: https://www.enisa.europa.eu/topics/national-cyber-security-strategies/ncss-map/national_cyber_security_strategy_2016.pdf
[16] Güngör M., *National Information Security: Strategy And Institutional Structuring*, Thesis For Planning Expertise, T.C Ministry of Development, Information Society Directorate, March 2015
[17] ENISA, *National Cyber Security Strategies*, May 2012
[18] ENISA, *NCSS Good Practice Guide Designing and Implementing National Cyber Security Strategies*, November 2016

[19] Global Cyber Security Capacity Centre, *Cybersecurity Capacity Maturity Model For Nations*, University of Oxford, March 2016

[20] Ministry of Defence, *Defence Programs and Budget of Japan, Overview of FY2017 Budget Request*, 2017. Available: http://www.mod.go.jp/e/d_budget/pdf/281025.pdf

[21] Aytekin A., M. Sc. Thesis, *The Assessment Of Turkish National Cybersecurity Strategy And Action Plan*, Gazi University Institute Of Informatics , June 2015.

[22] https://siberbulten.com/strateji-guvenlik/japonya-2020-olimpiyatlari-icin-siber-guvenligini-arttiriyor/ , accessed on March 17, 2018.

[23] http://afyonluoglu.org/siberguvenlik/siber-guvenlik-kurumsal-yapilar/, accessed on March 18, 2018.

[24] HM Government, *The UK Cyber Security Strategy, Protecting and promoting the UK in a digital world*, November 2011. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60961/uk-cyber-security-strategy-final.pdf

Comparison of Turkey ad European Union Computer Engineering Programs

S. KILICER[1] and R. SAMLI[2*]

[1] Istanbul University, Istanbul/Turkey, sd.kilicer@gmail.com
[2]Istanbul University, Istanbul/Turkey, ruyasamli@istanbul.edu.tr
* corresponding author

Abstract **- This study aims to compare Turkey's computer engineering undergraduate programs to compare with "European Union country" programs. In this way, it is aimed to contribute to the improvement of the programs in our country. Additionally, Erasmus programs of the universities in our country are aimed at reducing the adjustment problems experienced by the students going to European Union countries. All these operations are performed with data mining algorithms.**

Keywords **– Data mining, comparison, computer engineering programs.**

I. INTRODUCTION

NUMEROUS data are generated in many areas during every day. Information is formed by the giving meaning of these data. Information is processed to aim for a purpose. Transformation of information is called "data analysis". Today, raw data conversion can be done by data mining which is the process of making predictions about the future by using the results of searching the databases with computer programs. Valuable information is tried to be reached by establishing meaningful relations between data mining and data which have no meaning. By processing the data gathered by data mining process, future usable data is generated. Data mining is now used to solve many problems and it seems to produce successful results.

Computers have become more efficient by being able to store complex data from digital machines to store data. The data that we do not want to lose is brought into the concept of data warehouse by keeping them in physical drivers. As data increases, it becomes more difficult to organize databases and data modeling is needed. Data mining is a collection of methods that are interested in discovering data that are unrelated to each other in the data stacks and making it understandable and usable for the owner of the data. It has become effective in decision-making. We are looking for links and rules using programs so that we can produce useful results for future use. Data mining is the discovery of useful information from data repositories. Data mining is the use of data analysis tools such as statistics, artificial intelligence, and machine learning to generate hidden patterns in meaningless data. Data mining also allows similar patterns of behavior and patterns of behavior among data. In this study, the similarities and differences between computer engineering departments will be tried to be obtained.

II. MATERIAL AND METHOD

In this study, computer engineering department consists of courses in Turkey and European Union. For this purpose, the characteristics of 79 university courses in computer engineering department in Turkey (mandatory/elective and ECTS information) was obtained. In this way, a total of 5635 lines of courses were created. After this process, the universities which has an Erasmus agreement (total 29 units) course list was obtained. Here, a total of 1220 rows of the course list information was obtained. This information is then classified by the J48 algorithm (a text classification algorithm).

Text mining is a formless and multi-document analyzing technology for interesting and non-trivial patterns or information extraction. Text classification is done by two steps. Selecting the appropriate algorithm and selecting the sample dataset. As the feature is reduced in text classification, the model will succeed in that area. Our purpose is to classify in text fields. In order to be able to create these models, first a file is created for training and then this file is used to learn the system. The WEKA program is used for these operations.

There are many libraries on machine learning and statistics on WEKA. In this study, the following libraries were used to perform the data mining operation.

- •Pre-processing
- •Classification
- •Clustering
- •Association Rules
- •Selecting an Attribute
- •Visualize

First, it was taught which courses belong to which class, and then tried to find out the new courses belong to which classes. Courses are classified in 9 classes such as Mathematics, Hardware, Software, Artificial Intelligence, Network and Security, Departmental Elective, Divisional Elective, Basic

Courses and Additional Courses.

III. RESULTS AND DICUSSIONS

Two training data sets were created for a learning process using WEKA program. First, the classification is made through the 238- course from the list of universities in Turkey. An average of 25 classes in each class that can form an example is specified and an education dataset is created. Secondly, it is determined which class belongs to Erasmus course list, which is Erasmus data set. Then classification with WEKA is performed (Figure 1).

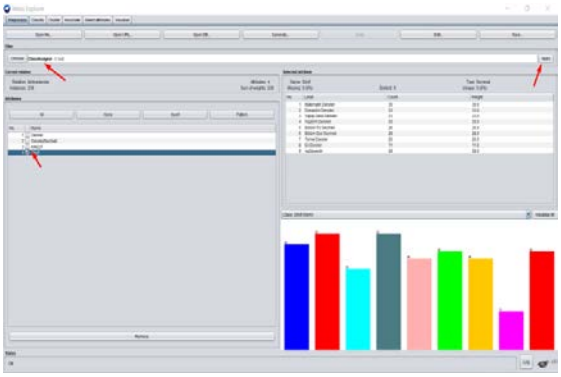


Figure 1: WEKA Classification Process

The algorithms to be used in the Classify Panel are selected. The "training set" should be selected because it is installed in the system for training purposes. The word "string to word vector" is used as a filter, and words are expressed in vector format. The "iterated lovins stemmer" was used to identify the words that are the same according to their roots (Figure 2).

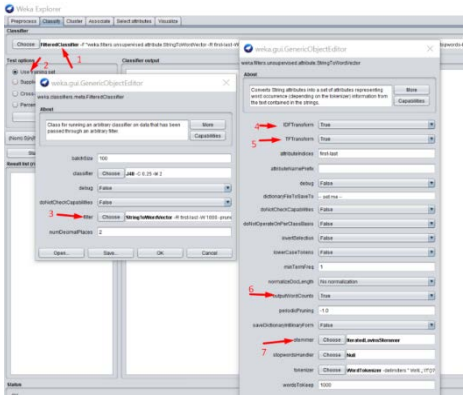


Figure 2: Text Analysis

After running this training set, the saving operation is performed (Figure 3). Then, the other data set is similarly saved (Figure 4).

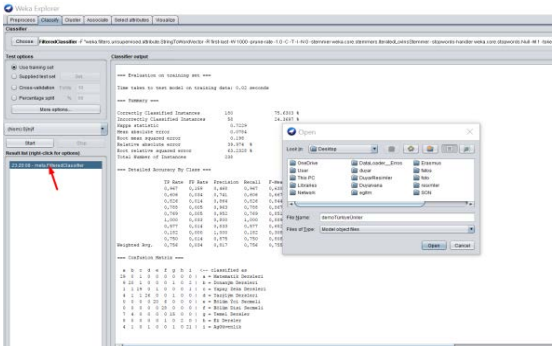


Figure 3: Data Set Saving

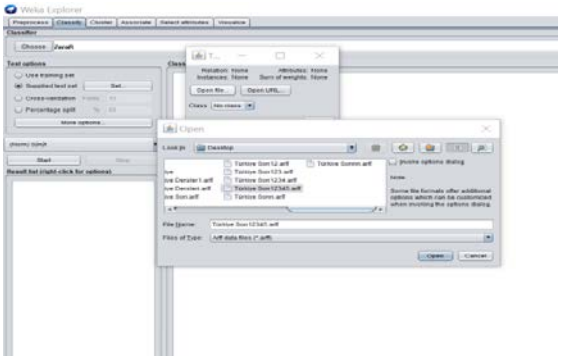


Figure 4: Training Data Set Saving

Finally, selecting the "load model" loads the resulting file in the training result (Figure 5).

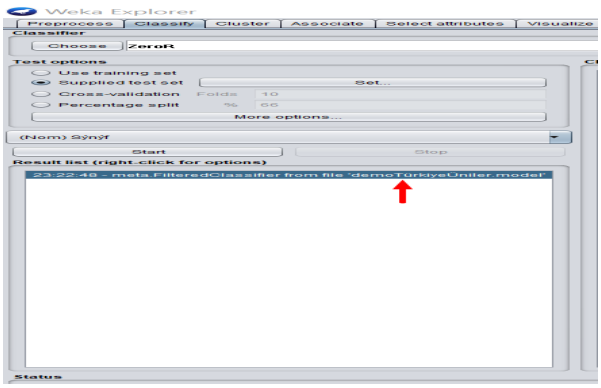


Figure 5: Attachment of the Resultant File

Then the classifications of the courses are performed (Figure 6 - Figure 7).

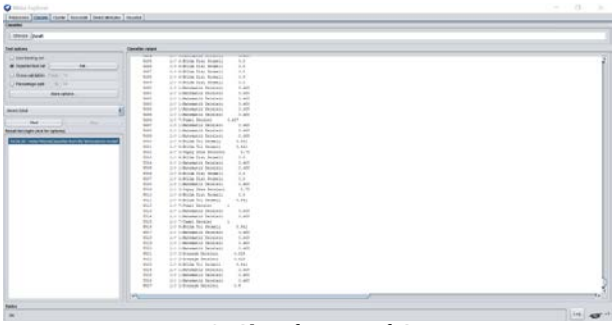


Figure 6: Classification of Courses

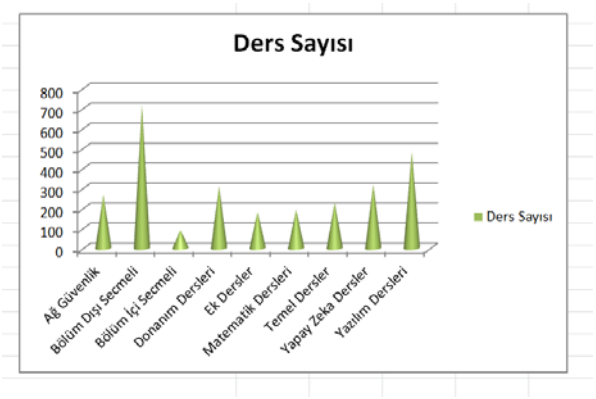


Figure 7: Classification of Courses

The same process is then carried out in the European Union data set. First, the data set for training was loaded into the system(Figure 8).

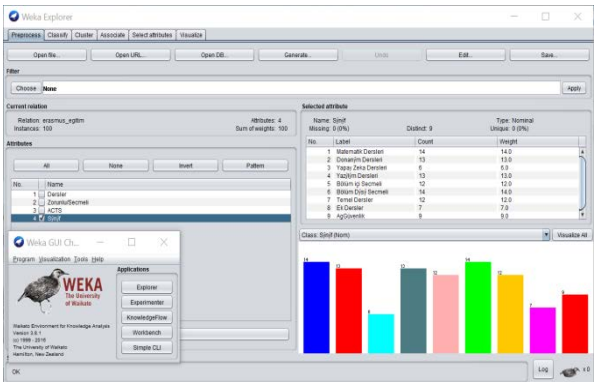


Figure 8: Data Set Saving

The data set for training is run and then the file is saved on the computer (Figure 9).

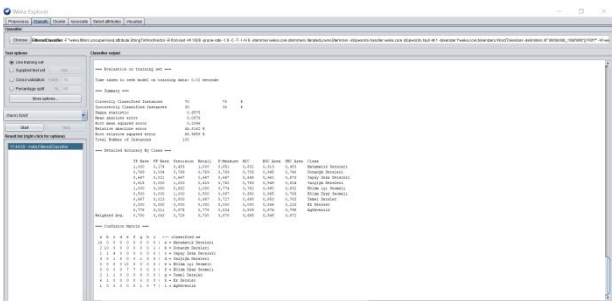


Figure 9: Training Data Set Saving

The files are loaded into the systemafter entering the system again (Figure 10).

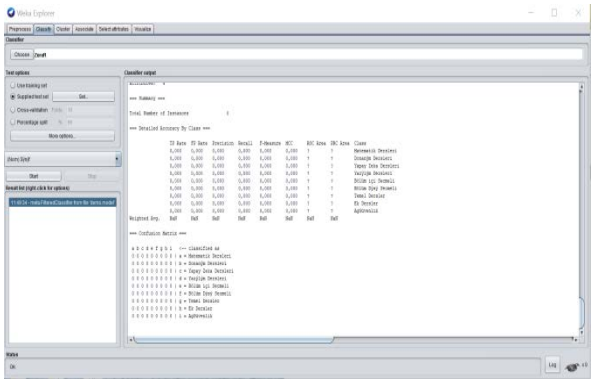


Figure 10: Training Data Set Saving

Classification of the courses is then carried out (Figure 11 - Figure 12).

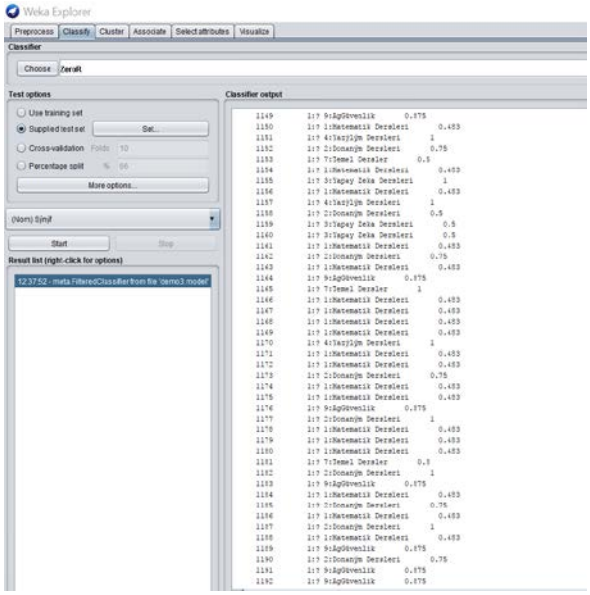


Figure 11: Classification of Courses

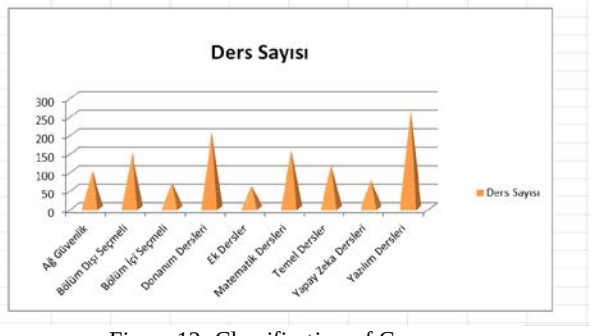


Figure 12: Classification of Courses

IV. CONCLUSIONS

Considering the values given in the results section, it is seen that Turkey universities have software courses mostly. It is important to note that the number of in-department elective courses is minimal. It is also seen that hardware and software courses are very close to each other in terms of percentage. At the highest level, students are given the opportunity to choose their own personalities, classes according to their interests, and to develop themselves in this area. When the distribution of courses in European countries is examined, it is seen that the highest rate is in software courses. It is also seen here that hardware courses have as high a level as software courses. Also, another important point that the proportion of the departmental elective courses is almost at least the ratio.

When it is looked at both graphs together, we evaluate software courses mainly in Turkey and European Union countries, the ratio is higher than 4% in Europe. Part of non-elective courses should be considered to be more in Turkey.

In this case, the student develops himself/herself by taking extracurricular courses instead of identifying the area that needs to be developed in the future. As a continuation of this situation, it can be seen more clearly when we look at the distribution of the elective courses in the department. The proportion of in-department elective courses that will allow the individual to identify himself/herself and to identify the area he/she wishes to continue in the rest of his/her life is half of that in Europe. As a result, it is not surprising that graduates should be graduated without knowing what department of individuals they want to improve themselves.

Artificial Intelligence and network security courses considering the rate in Turkey appears to be higher than in Europe. This actually means that Turkey is ready for the new technology.

Mathematical courses that the main factor in the success of the people in Turkey seem to be less than in Europe. Mathematical courses are the basic courses in the engineering field and the difficulty of the mentally thinking of the individual requires increasing the number of mathematical courses to gain engineering perspective. In addition, all computer engineers should be able to take courses from different departments by offering different departmental elective courses in order to be able to take them out of the perception that they will make software, and then they should be able to continue their education in order to improve themselves in their branches.

The outer part of the vast majority of courses learned from Turkey university, it is observed that the proportion of small courses in other classes. This shows that graduating individuals are generally forced to take similar courses. Most of the courses are compulsory and consist of the same courses. For this, the diversity of the courses belonging to the same class must be increased. This situation actually brings up a very important problem.

Erasmus students from Turkey face many grievances destined for European Union countries. It is a big problem for Erasmus students to match the courses they have taken in the other countries because of the diversity of the courses there. There are no ECTS credits of the courses there. For example, there is no match for the lesson that is taken there as software.

In this case, it is seen that the courses that are taken as one of the main courses are not matched here, so the course has to be matched with the non-departmental elective courses there, and this situation sometimes leads to the extension of the school. It is to match with the main course there, since there are no basic courses in Europe where there is a lot of diversity in courses. The individual is actually graduated without taking the courses that he should learn. These problems are thought to be solved by similar works.

V. ACKNOWLEDGEMENT

The authors wish to thank for the support of the Research Fund of Istanbul University. Project number is BEK 21443.

A quality model for evaluating maintainability of object-oriented software systems

Özlem AKALIN[1] and Feza BUZLUCA[2]

[1] Istanbul Technical University, Istanbul/Turkey, akalino@itu.edu.tr
[2] Istanbul Technical University, Istanbul/Turkey, buzluca@itu.edu.tr

Abstract **- Measuring software maintainability is of vital importance for improving software product quality. Using a software quality model in the development life cycle, the quality of the system can be continuously evaluated and improved to reduce the maintenance cost. According to ISO/IEC 25010 Software Quality Models Standard, the maintainability characteristic of software product quality is composed of five sub characteristics; modularity, modifiability, reusability, analyzability and testability. This paper proposes a quality measurement model to evaluate the maintainability of software classes in terms of their reusability and modifiability characteristics in large-scale software systems. The model is based on software properties that are strongly related to reusability and modifiability, such as size, complexity, cohesion, coupling, and inheritance. First, our method categorizes metric values of software classes in the test system as low, medium and high. This categorization is done based on the average and median values for these metrics that are obtained from reference software systems. Then, the proposed measurement method uses the levels of the metrics to calculate the reusability and modifiability scores of each class in the system. The scores fall in one of the five categories; very low, low, medium, high, and very high. The developers of the software system can examine classes with low and very low scores and then refactor them if necessary. This continuous evaluation and refactoring during the development can increase the quality of the system and reduce maintenance costs. We applied our model on two large-scale industrial mobile applications and discussed the results with the development teams of the systems. We saw that our approach could reasonably grade classes on their reusability and modifiability characteristics.**

Keywords **- Quality Model, Software Metrics, Software Maintainability, Reusability, Modifiability.**

I. INTRODUCTION

High maintenance cost is considered as an important problem for software companies, because this complex process typically consumes 50-70% of the total effort allocated to a software development project [1], [2]. Main reasons for this high cost are frequently changing customer demands and technological developments that cause the software systems to be updated constantly, and the increasing complexity of the programs. It is expected that the software systems could be developed, modified, extended, and corrected in a short time without a degradation in its performance. Therefore, many researchers study on models for measuring the maintainability characteristic of the software product quality. The objective is to help developers in detecting modules that are not properly designed and need refactoring to decrease maintenance costs.

The maintainability is defined by IEEE standard glossary of Software Engineering as "the ease with which a software system or component can be modified to correct faults, improve the performance or other attributes, or adapt to a changed environment" [3]. According to ISO/IEC 25010 Software Quality Models Standard [4], the maintainability characteristic of software product quality is composed of five sub characteristics; modularity, modifiability, reusability, analyzability and testability.

In this paper, we propose a quality model for evaluating the maintainability of large-scale object-oriented software systems. Our model is based on the reusability and modifiability, because these sub characteristics are the main factors that affect the maintenance cost. Software developers need to know which classes can be reused at different stages of the systems or in other projects. Similarly, they need to know the cost of modifying classes for adding new features, refactoring or bug fixing. In our hierarchical model, to measure reusability and modifiability we use properties of object-oriented systems such as size, cohesion, coupling, complexity, and inheritance. We assigned software code metrics for each property. The metrics can be easily from programs obtained and using their values the maintainability of a system can be measured and enhanced. We tested our model on two large-scale industrial projects and discussed our findings with the developers of the projects. The results show that our model can be used to detect software classes that have low reusability or modifiability values because they were not properly designed. Such classes can be refactored to improve the maintainability of the system

The rest of the paper starts with related work in Section 2. Section 3 gives information definition of the maintainability in ISO / IEC 25010 standard and explains the structure of the proposed model. Steps of model construction is explained in Section 4. The empirical study and validation results are shared in Section 5. Final section concludes the paper.

II. RELEATED WORK

Many studies have used different prediction models for maintainability quality. Muthanna et al. [5], investigated the use of software design metrics to statistically evaluate the maintainability quality of large software systems. They model

can be applied only to procedural programs and it is not suitable for object-oriented software systems.

Aggarwal et al. [6] developed a Fuzzy model to measure the maintainability of the software system. They defined four factors affecting maintainability, namely average number of live variable, average life span of variables, average cyclomatic complexity, and the comments ratio. Their model classifies maintainability as very good, good, average, poor and very poor. There are total 81 rules in the model for all inputs and outputs. They test the model only on the projects developed by undergraduate engineering students.

Bagheri and Gasevic [7] evaluated the maintainability of software product line feature models. They proposed different structural metrics in their study and used manually the three main sub-characteristics of the maintainability: analyzability, changeability and understandability.

Rizvi and Khan [8] investigated a maintainability model including understandability and modifiability sub-characteristics on software design phase. They proposed different design metrics for quality measurement and constructed Maintainability Estimation Model for Object-Oriented software in the design phase (MEMOOD). The model generates only project-based results and it do not give information about software classes.

Bansiya and Davis [9] built a hierarchical quality model for object-oriented designs named QMOOD, which contains structural and behavioral design properties of classes, objects, and their relationships. QMOOD has four layers which include Design Quality Attributes, Object Oriented Design Properties, Object-Oriented Design Metrics, Object Oriented Design Components.

In our study, we created a model that uses code metrics to measure the reusability and modifiability characteristics of software classes in object-oriented systems. We evaluated our model on two large-scale industrial projects and validated the results with developers of the projects.

III. Software Maintainability and Hierarchical Model

Software maintainability is defined as the necessary effort for addition of new features to the software system and elimination of defects. In the ISO / IEC 25010: 2011 "System and software quality models" standard, the maintainability quality characteristic consists of five sub-characteristics including modularity, modifiability, reusability, analyzability and testability. In our proposed model, we consider reusability and modifiability characteristics, because they are main factors that affect the maintenance cost. Reusability is defined as the degree to which a software component can be used in more than one system or in implementing other software components. Modifiability is expressed as the degree to which a software system can be efficiently modified without introducing defects or degrading existing product quality [4].

In this study, we propose a hierarchical model that uses low-level code metrics to measure the high-level quality characteristic maintainability. The general structure of the model is shown in Figure 1. The values of class-based code metrics are obtained from the software systems. These values are used to calculate values of properties such as size, complexity, cohesion, coupling and inheritance. Properties determine the values of reusability and modifiability quality characteristics for each class. Using these values, developers can find classes that increase the maintainability costs and need refactoring.

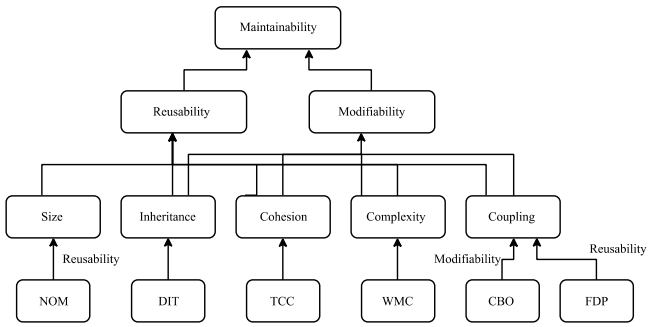


Figure 1: Hierarchical Structure of the Quality Model

IV. Construction of The Model

A. Property Selection

At the first stage of the model development, we select the main properties of object-oriented programs that can be used to determine reusability and modifiability values of classes. The selected properties and their influence on reusability and modifiability characteristics are explained below.

- Code size: The size of the source code negatively influences reusability of a class. For the simplicity of the model, effect of size on the modifiability is ignored because other factors are more dominant.
- Complexity: The complexity of the source code has a negative influence on the class' reusability and modifiability so that as complexity increases, reusability and modifiability of the class decreases.
- Cohesion: The cohesion of a class positively influences its reusability and modifiability. Therefore, when cohesion increases, the reusability and modifiability of the class increases.
- Coupling: The coupling between classes negatively influences the system's reusability and modifiability so that as the Coupling between classes increases, the reusability and modifiability of the system decreases.
- Inheritance: The degree of class inheritance has a negative influence on the reusability and modifiability of the system so that as the degree of class inheritance increases, the reusability and modifiability of the system decreases.

B. Metric selection

At the second stage of the model development, we determine metrics that should be used to measure the selected properties. Metrics are selected based on observations of authors about the metrics and on results of different research conducted in recent years [6-11]. Selected metrics are explained below.

Coupling Between Object Classes (CBO) has been proposed

by Chidamber & Kemerer [13]. If a class is coupled to more classes, its reusability and modifiability possibilities become lower.

Weighted Methods per Class (WMC) metric represents the sum of complexities of all methods in a class. Since high value of WMC metric indicates error-proneness, its reusability decreases and changes over the class take longer.

Depth of Inheritance Tree (DIT) metric calculates the depth of inheritance of given class in the class hierarchy. The deeper class have more properties and more methods, thus the possibility of its modifiability and reusability decreases.

Tight Class Cohesion (TCC) is a metric used to the relative number of methods directly connected via accesses of attributes. It is easy to refactor and understand the classes with high TCC so it has a positive effect on modifiability of a class.

Foreign Data Providers (FDP) is the number of different foreign attributes accessed by a method. since high FDP indicate that the class includes methods with different responsibilities, its reusability may become more unmanaged.

Number of Methods (NOM) metric shows the total number of methods of given class. High NOM value shows that the class has too much responsibility so it negatively affects reusability of a class.

The properties of the class, the metric value corresponding to the property, and the effect of the characteristic of the metric are shown Table 1.

C. Constructing the Prediction Model

Since values of different metrics change in very different ranges, to create an understandable model, first we categorize the values of the metrics as low (LOW), medium (MED), and high (HIGH) based on their typical values such as mean, median and standard deviation. The typical values for categorizing the metrics are obtained from open source projects that reached certain maturity. We used iPlasma tool [12] to gather metrics for each class from the reference open source software systems including WordPress-Android [16], FastHub [17], RxJava [18]. Since DIT metric values are in ordinal scale and may include asymmetric outlier values, median calculation (Med_m) is used for tagging it as shown in equation 1. Mean calculation ($Mean_m$) and standard deviation (Std) are used for other metrics, which are interval scale. Tagging rule for NOM, CBO, TCC, WMC, FDP metrics is shown in equation 2.

Our model assigns a numeric value (-1, 0, or 1) as score to each category based on the effect of the metric on the characteristic. For example, if a metric has a HIGH value and its affects the measured characteristic positively than this metric has the score +1. Table 2 shows the rules used to assign scores to metrics.

$$tag_c = \begin{cases} HIGH\ if (M_c > Med_m) \\ MED\ if (M_c = Med_m) \\ LOW\ if (M_c < Med_m) \end{cases} \quad (1)$$

Table 1: The corresponding properties of the metrics and the metric's effect on the characteristic

Property	Metric	The metric's effect on the characteristic	
		Reusability	***Modifiability***
Coupling	CBO	-	Negative
Coupling	FDP	Negative	-
Code Size	NOM	Negative	-
Inheritance	DIT	Negative	Negative
Cohesion	TCC	Positive	Positive
Complexity	WMC	Negative	Negative

Table 2: Score values corresponding tag values

	$Score_c$	
tag_c	Values for positive effect	Values for negative effect
HIGH	1	-1
MED	0	0
LOW	-1	1

$$tag_c = \begin{cases} HIGH\ if (M_c \geq Mean_m + Std) \\ MED\ if (M_c < Mean_m + Std\ or \\ \qquad M_c < Mean_m - Std) \\ LOW\ if (M_c \geq Mean_m r + Std) \end{cases} \quad (2)$$

Using the scores and the effect of metrics given in Table 1 the model calculates the output values for reusability and modifiability of each software class. The output value is the sum of all related metric scores for the quality characteristic as shown in equation 3.

$$Output_c = \sum_{k=0}^{n} Score_{c,k} \quad (3)$$

Output values change in the range between [-5,5] for reusability and [-4,4] for modifiability. In order to categorize the reusability and modifiability characteristics for each class, we also tag the output values. The output rule used in the reusability and modifiability model is shown Table 3. As a result of the output rules, we obtain a data set in the matrix form with n rows and m+1 columns for prediction where n is number of class and m is number of metric. As each column refers to a software metric, each row of the matrix refers to

Table 3: Output rules for tag values

Output	Reusability	Modifiability
VERY HIGH	$4 \leq Output_c \leq 5$	$3 \leq Output_c \leq 4$
HIGH	$2 \leq Output_c \leq 3$	$1 \leq Output_c \leq 2$
MED	$-1 \leq Output_c \leq 1$	$-1 \leq Output_c \leq 0$
LOW	$-3 \leq Output_c \leq -2$	$-3 \leq Output_c \leq -2$
VERY LOW	$-5 \leq Output_c \leq -4$	$Output_c = -4$

metric values for a software class. The last element of the row is the output tag value of software class, which is used to

categorize the reusability or modifiability characteristic.

Output intervals are not strict and can be flexible for different projects. For example, in order to find about ten percent of the worst-developed part of a project in terms of reusability, only those with an output value of "-5" can be defined as VERY LOW. Similarly, only the output value "4" can be defined as VERY HIGH to find the best developed about ten percent of a project in terms of modifiability.

V. Empirical Study and Evaluation

A. Model Results for Empirical Study

Two android mobile applications that are developed in industrial projects are examined for the evaluation of the model. They are very popular commercial projects that are extensively used in real-world. Project A has 1071 classes and Project B has 450 classes. Project A and Project B have developed in 18 months and 12 months, respectively. Since the projects were written with Java language, mean and median values in the model are obtained from open source projects written in Java. Open source project selection criteria are as follows: a) software must be written in the same language with software project which is selected to evaluate; b) open source software is selected as possible as from the similar topics to the selected project for evaluation; c) software must be at a certain maturity.

We applied our prediction model on a live release of the Project A and Project B. To tag classes according to reusability and modifiability characteristics, we first calculated mean and median values for each metrics using selected open source projects. Then we tagged metric values for Project A and Project B in accordance with the median and mean values. The metric tag values' numbers obtained for the projects are shown in Table 4 and Table 5. To create output values, we have labeled classes using the proposed model for each project. Some of the observed results are shown as:

- Project A has 173 classes labeled as VERY HIGH and 165 classes labeled as HIGH for reusability so it is considered that approximately thirty-third percent of Project A is highly reusable. Otherwise, Project A has 84 classes labeled as VERY LOW and 187 classes labeled as LOW for reusability so it can be said that about twenty-five percent of the project is quite low in terms of reusability.
- Project B has 14 classes labeled as VERY HIGH and 50 classes labeled as HIGH for reusability so it is considered that approximately fifteen percent of Project B is highly reusable. Otherwise, Project B has 18 classes labeled as VERY LOW and 166 classes labeled as LOW for reusability so it can be said that reusability of the project for about forty percent is rather difficult.
- Project A has 96 classes labeled as VERY HIGH and 229 classes labeled as HIGH for modifiability so it is considered that it is quite easy to make a modification for about thirty percent of the project. Otherwise, Project A has 67 classes labeled as VERY LOW and 270 classes labeled as LOW for modifiability so it can be said that about thirty-third percent of the project is quite low in terms of modifiability.
- Project B has 17 classes labeled as VERY HIGH and 34 classes labeled as HIGH for modifiability so it is considered that approximately ten percent of Project B is extremely easy to modify. Otherwise, Project B has 5 classes labeled as VERY LOW and 159 classes labeled as LOW for modifiability so it can be said that modifiability of the project for about thirty-third percent is rather difficult.

Table 4: Reusability results for Project A and Project B

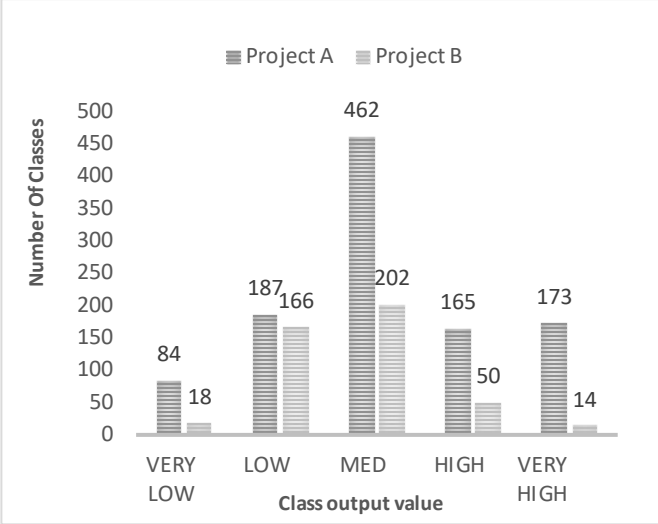
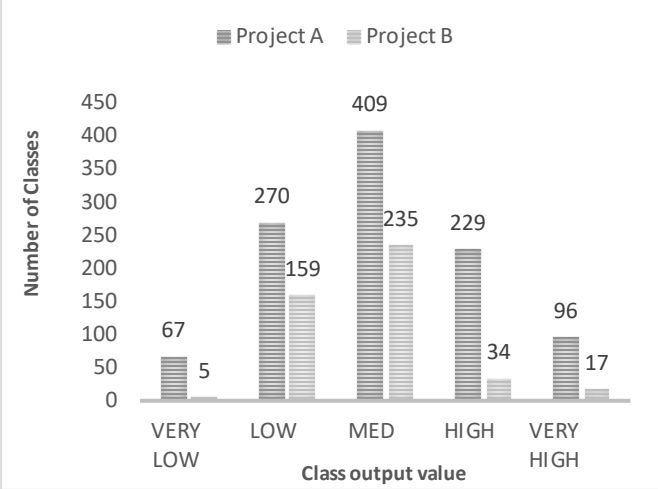


Table 5: Modifiability results for Project A and Project B



Class-based example results based on the values obtained from the model result are shown in Table 6 and Table 7. We interpreted the output values in terms of reusability and

Table 6: Class-based Reusability results

	FDP	NOM	DIT	TCC	WMC	Output
Class 1	LOW	HIGH	LOW	LOW	HIGH	MED
Class 2	LOW	HIGH	LOW	HIGH	LOW	HIGH
Class 3	HIGH	HIGH	MED	LOW	HIGH	VERY LOW

Table 7: Class-based Modifiability results

	CBO	DIT	TCC	WMC	Output
Class 1	LOW	LOW	LOW	HIGH	MED
Class 2	LOW	LOW	HIGH	LOW	VERY HIGH
Class 3	HIGH	MED	LOW	HIGH	VERY LOW

modifiability through the metrics that the classes have. The observed results for Table 6 and Table 7 are commented as:

- Class 1 has the low coupling, inheritance, cohesion and high complexity and size so its reusability is medium level. Also, since it has low coupling, inheritance, cohesion and high complexity, its modifiability is medium level. Thus, we can infer that its reusability and modifiability is medium in terms of maintainability.
- Class 2 has the low coupling, inheritance, complexity and high cohesion and size so its reusability is high level. Moreover, since it has low coupling, inheritance, complexity and high cohesion, its modifiability is very high level. As can be inferred, its reusability is high and also its modifiability is very high in terms of maintainability
- Class 3 has high coupling, size, complexity and low cohesion so its reusability is very low level. Also, since it has high coupling, complexity and low cohesion, its modifiability is very low level so that its reusability and modifiability is very low in terms of maintainability.

By examining the results obtained from the model, it becomes possible to deduce from which direction the class should be developed.

B. Model Validation and Threads to Validity

The proposed methodology is validated with the help of development teams. We use Analytical Hierarchical Process (AHP) to evaluate teams' opinions. It is a structured technique for analyzing complex multi-attribute decisions. It provides the decision maker to set priorities and make the best decision for solving a problem. Users of the AHP decompose their decision problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently [14].

We prepared a survey for each property of the reusability and modifiability characteristics to obtain metric coefficients through AHP. In our study, we aim to evaluate maintainability

	Classes				
OUTPUT	HIGH	VERY HIGH	VERY LOW	LOW	LOW
Model_Reu	4	5	1	2	2
AHP_Reu	1.23	1.22	0.94	0.9	0.97
Model_ Reu: Reusability Calculated with Model					
AHP_ Reu: Reusability Calculated with AHP					

Table 8: Reusability values for Pearson Correlation Calculation

	Classes				
OUTPUT	VERY HIGH	MED	MED	HIGH	LOW
Model_Mod	5	3	3	4	2
AHP_Mod	1.12	0.99	0.96	1.0	0.8
Model_Mod: Modifiability Calculated with Model					
AHP_ Mod: Modifiability Calculated with AHP					

Table 9: Modifiability values for Pearson Correlation Calculation

model's validation using the survey results. During the survey as size, cohesion, coupling, inheritance and complexity properties are calculated for reusability, cohesion, coupling, inheritance and complexity properties are calculated for modifiability. After the opinions of the teams were taken as a result of the survey, metric coefficients were obtained with AHP. Results are calculated for both reusability and modifiability characteristic. To identify the relationships between the model and AHP results, we analyzed Pearson correlation matrix for each value. In order to calculate the correlation between AHP and the model results, we assigned numeral values as 5, 4, 3, 2, 1 for VERY HIGH, HIGH, MED, LOW, VERY LOW model results, respectively.

Each correlation matrix has a correlation coefficient (R) for each result pair and R can range from - 1 and +1. While the value of R, +1, represents perfect positive correlation, the value of R, - 1, represents perfect negative correlation. While correlation coefficient values between 0.70 and 1.00 are accepted as a strong positive correlation in Pearson Correlation Analysis, values between - 1.00 and - 0.70 are accepted as a strong negative correlation [15].

The metrics for all the five properties of reusability and for all the four properties of modifiability were measured for both the projects. Initially, metric values were normalized and then were multiplied by their corresponding weight values according to AHP result for the projects. The results calculated with AHP and the proposed models is shown in Table 8 and Table 9.

Table 10 presents Pearson correlation coefficient between the proposed model and AHP model which is created as a result of the survey. As can be inferred, all the two coefficients depict significant positive correlation between the proposed

Table 10: Pearson Correlation Results for Project A and Project B

	Project A		Project B	
	R	R^2	R	R^2
Reusability	0.7679	0.5866	0.8819	0.7777
Modifiability	0.8282	0.681	0.7581	0.5747

model and AHP model. Therefore, it is clear that the proposed model may be able to predict a class's reusability and modifiability in terms of maintainability. The flow that occurs as a result of construction and evaluation stages has been shown Figure 2.

We need to explain some threats to validity of the proposed work for providing completeness and accuracy. The maintainability prediction model has been tested on the software projects developed in Java language. However, it is likely to be valid for the software projects which developed different object-oriented programing languages. Further research should be done to evaluate on different projects. Moreover, we used only 6 different source code metrics for development of the quality model in this study. Some of other static source code object-oriented metrics can be used for examining different properties except the properties in the proposed model.

VI. CONCLUSION

We have developed a quality model to quantify maintainability of the software classes in terms of their reusability and modifiability. We used two industrial software projects to evaluate the performance of our model in the real world. We structurally categorized software classes using code metrics and compared statistically our results to the developer teams' opinions. Consequently, our model shows that object-oriented metrics can effectively be used as predictors to evaluate maintainability of software systems.

Our model results provide different advantages in the software development phase. It generally helps to determine software class quality in terms of their maintainability level. Software developers can focus on software classes that have low quality. Moreover, the model gives information about reasons for their low or high quality levels, such as cohesion, coupling etc. Hence, developers can easily refactor classes to increase their quality level and significantly reduce the maintenance cost of the system.

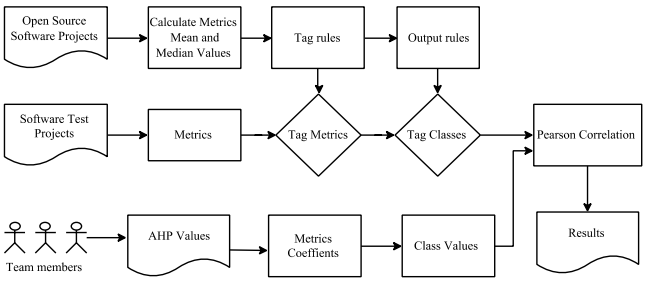


Figure 2: Construction and Evaluation of the Quality Model

REFERENCES

[1] Pigoski, Thomas M. *Practical software maintenance: best practices for managing your software investment*, Wiley Publishing, 1996.
[2] Sommerville, *Software Engineering*, 6th ed., Harlow, Addison-Wesley, 2001.S
[3] IEEE STD 610.2: *IEEE Standard Glossary of Software Engineering Terminology*, 1990.
[4] ISO/IEC 25010. ISO/IEC 25010:2011: Systems And Software Engineering – Systems And Soft-Ware Quality Requirements And Evaluation (Square) – System And Software Quality Models. Geneva: ISO, 2011.
[5] Muthanna, S., et al. "A maintainability model for industrial software systems using design level metrics." Reverse Engineering, 2000. Proceedings. Seventh Working Conference on. IEEE, 2000.
[6] Aggarwal, Krishan K., Yogesh Singh, and Jitender Kumar Chhabra. "An integrated measure of software maintainability." Reliability and maintainability symposium. Proceedings. Annual. IEEE, 2002.
[7] Bagheri, Ebrahim, and Dragan Gasevic. "Assessing the maintainability of software product line feature models using structural metrics." Software Quality Journal 19.3: 579-612,2011.
[8] Rizvi, S. W. A., and Raees A. Khan. "Maintainability estimation model for object-oriented software in design phase (memood).",2010.
[9] Bansiya, Jagdish, and Carl G. Davis. "A hierarchical model for object-oriented design quality assessment." IEEE Transactions on software engineering 28.1: 4-17, 2002.
[10] Genero, Marcela, et al. "Using metrics to predict OO information systems maintainability." International Conference on Advanced Information Systems Engineering. Springer, Berlin, Heidelberg, 2001.
[11] Singh, Charu, Amrendra Pratap, and Abhishek Singhal. "Estimation of software reusability for component based system using soft computing techniques." *Confluence The Next Generation Information Technology Summit (Confluence), 2014 5th International Conference-*. IEEE, 2014.
[12] C. Marinescu, R. Marinescu, P. Mihancea, D. Ratiu, and R. Wettel, "iPlasma:An integrated platform for quality assessment of object-oriented design," In Proceedings of the 21st IEEE International Conference on Software Maintenance, pp. 77-80, 2005.
[13] Chidamber, Shyam R., and Chris F. Kemerer. "A metrics suite for object oriented design." IEEE Transactions on software engineering 20.6 (1994): 476-493.
[14] Sharma, A., Kumar, R., Grover, P.S.: Estimation of Quality for Software Components - an Empirical Approach. ACM SIGSOFT Software Engineering Notes 33(5), 1–10 (2008)-639.
[15] Dagpinar, Melis, and Jens H. Jahnke. "Predicting maintainability with object-oriented metrics-an empirical comparison." Reverse Engineering, WCRE 2003. Proceedings. 10th Working Conference on. IEEE, 2003.
[16] https://github.com/wordpress-mobile/WordPress-Android/
[17] https://github.com/k0shk0sh/FastHub
[18] https://github.com/ReactiveX/RxJava

An approximation of the Voronoi diagram for a set of arcs

D. KOTSUR[1] and V. TERESHCHENKO[2]

[1] Taras Shevchenko National University of Kyiv, Kyiv/Ukraine, dkotsur@gmail.com
[2] Taras Shevchenko National University of Kyiv, Kyiv/Ukraine, vtereshch@gmail.com

***Abstract* – In this paper we introduce an algorithm for constructing approximate Voronoi diagram of a set of pairwise disjoint arcs on a plane. The arcs are represented by parametric curves. On the first step, we discretize curves using the proposed adaptive method. Then, we construct Voronoi diagram of the discretized objects and process the obtained Voronoi graph such that redundant edges and cells are removed. Finally, the edges of the processed graph are approximated by Bezier curves giving the parametric representation of the final Voronoi diagram. The total complexity of the described algorithm is O(n log n) in average.**

***Keywords* – Voronoi diagram, Voronoi polygon, Bezier curve, splines, approximation, discretization.**

I. INTRODUCTION

THERE is a wide variety of different algorithms for solving typical problems of computational geometry, e.g. nearest neighbor search, point localization, triangulation, convex hull construction etc.. Despite of the high efficiency, these algorithms have a very narrow application area, which is driven mainly by the complexity of the processed objects. For example, some of the well-known fast and robust algorithms for Voronoi diagram construction (e.g. "divide and conquer" [1,2] or Fortune's algorithm [3]) can be effectively applied only to a set of points or line segments. But the solution of real-world practical problems usually involves highly sophisticated geometrical objects, which normally have a very complex shape. In order to represent such shape, points and line segments are normally not sufficient. Therefore, we require a more sophisticated representation with a set parametric curves, in particular, Bezier curves, splines, circular arcs, etc.

In this paper we investigate an approach for the Voronoi diagram construction in case of the objects with a complex geometrical shape represented by parametric curve. Such Voronoi diagrams can be used in many practical applications from optimal path planning [4] and collisions avoidance by robots [5][6] to architecture and art [7].

The construction of the precise Voronoi diagram for arbitrary curves is nontrivial task due to a huge amount of specific cases, which should be considered. Therefore, the study of Voronoi diagram leads to the development of efficient approximations. However, designing the efficient algorithms for constructing the approximate Voronoi diagram is still nontrivial task, which requires further investigation.

Analysis of the recent research and publications. In a large amount of works devoted to the approximation of Voronoi diagram, authors consider the spatial discretization the 2D plane into cells using a discrete grid with a fixed step [8], [9]. In this case, the objects of interest should be also discretized and the constructed Voronoi diagram has similar grid representation. A significant drawback of these methods is that the accuracy of the Voronoi diagram depends on the density of the grid. However, by reducing the size of the cells one should expect a significant increase in the number of cells (quadratic dependence). Thus, the computational time of such algorithms can be quite large if one wants to achieve acceptable results. Another drawback is implicit representation of the results, which complicates the extraction of the topological structures.

A different way of approximating Voronoi diagram is using a Voronoi diagram for the simplest geometric objects, such as points or line segments. In particular, in [10] there is an attempt to approximate Voronoi diagram for arbitrary geometric objects using Bezier curves and Voronoi diagram for points. However, this approximation is made only for the part the Voronoi diagram (problem of finding the minimum path was solved). Approximation by means of Voronoi diagram for points has significant prospects as well as makes it possible to operate with fewer amounts of simple objects. The points on the curves can be selected based on the sampling with a fixed step or depending on the characteristics of the curve. It allows us to speed up the construction of the approximation, and at the same time, to maintain the desired accuracy in critical regions. Since this method is based on the construction of Voronoi diagram for simple geometrical objects (points), the critical regions of such approximation can be easily refined by supporting dynamic Voronoi diagram for points.

In papers [11-13] authors discuss different approaches of computing exact Voronoi diagram. For example, in [14] authors describes an algorithm for computing the precise Voronoi Diagrams of planar freeform curves. As in our case, input curves are parameterized. However, authors build the precise bisectors and junctions between the curves. This method requires solving the system of three nonlinear equations, which is not trivial task, and also applying some numerical methods. Another similar approach is described in [15]. The algorithm has complexity $O(n^2)$, where n is the number of curves. This method has similar drawback because of computing the junction points (authors use Newton-Raphson iteration scheme for solving nonlinear equations).

Novelty and ideas. The purpose of this paper is to develop an algorithm for fast and accurate approximation of Voronoi diagram. Our approach is based on the adaptive discretization of input curves and further construction of a "low-level" Voronoi diagram for discretized objects. Then this Voronoi diagram is processed in such way, that unnecessary Voronoi cells are merged together and redundant Voronoi edges are removed. The resulting Voronoi graph is approximated by parametric curves. The analysis of the algorithm complexity showed, that the average running time complexity of the poposed method is $O(N \log N)$.

II. METHODS OF FAST APPROXIMATE VORONOI DIAGRAM CONSTRUCTION

In this chapter we provide the definitions of the basic concepts related to the Voronoi diagram [16] and the proposed algorithm. We also describe and analyze the main steps of our new algorithm including discretization of the curves, Voronoi diagram construction and approximation of the resulting Voronoi edges with Bezier curves.

A. Basic concepts and statement of the problem

Definition 1. Let's suppose, that we are given a set of generator points $P = \{p_1, p_2, ..., p_n\} \subset \mathbb{R}^2$, where $n \in \mathbb{N}, n \geq 2$. We call set $\boldsymbol{P}$ the generator set of the Voronoi diagram. Denote $\boldsymbol{I_n}$ by set of generators indexes and distance between two objects $\boldsymbol{x}$ and $\boldsymbol{y}$ as $\boldsymbol{\rho(x,y)}$. In case of d-dimensional points, distance $\boldsymbol{\rho}$ can be defined as d-dimensional Euclidian distance. We call the region given by:

$$VP(p_i) = \left\{ x \in \mathbb{R}^2 \middle| \rho(x, p_i) \leq \rho(x, p_l), l \in I_n \setminus \{i\} \right\}, \quad (1)$$

where $x \in \mathbb{R}^2$, the Voronoi polygon (Voronoi cell) associated with $\boldsymbol{p_i}$. Then the Voronoi Diagram generated by $\boldsymbol{P}$ (or the Voronoi diagram of $\boldsymbol{P}$) is defined by statement:

$$VD(P) = \{VP(p_1), VP(p_2), ..., VP(p_n)\} \quad (2)$$

For any two generators $\boldsymbol{p_i}$ and $\boldsymbol{p_k}$ we define we half plane called dominance region of $\boldsymbol{p_i}$ over $\boldsymbol{p_k}$:

$$H(p_i, p_k) = \left\{ x \in \mathbb{R}^2 \middle| \rho(x, p_i) \leq \rho(x, p_k) \right\} \quad (3)$$

thus, Voronoi cell can be defined by the following statement:

$$VP(p_i) = \bigcap_{k \in I_n \setminus \{i\}} H(p_i, p_k) \quad (4)$$

The boundary (bisector) between half planes $H(p_i, p_k)$ and $H(p_k, p_i)$ denoted by $b(p_i, p_k)$ is written as:

$$b(p_i, p_k) = H(p_i, p_k) \cap H(p_k, p_i) \quad (5)$$

or

$$b(p_i, p_k) = \left\{ x \in \mathbb{R}^2 \middle| \rho(x, p_i) = \rho(x, p_j) \right\} \quad (6)$$

For given generator set $\boldsymbol{P}$ and set of indexes $\boldsymbol{I_n}$, boundary $\boldsymbol{b(p_i, p_k)}$ can be denoted in a short form as $\boldsymbol{b_{ik}}$.

Definition 2. Let's $C(t) = (x_C(t), y_C(t))$ represent a continuous parametric curve on a plane, where $t \in [0,1]$ is a parameter of the curve; $\{t_1, t_2, ..., t_{n_C}\}$ is a discrete grid such that $0 = t_1 \leq t_2 \leq ... \leq t_{n_C} = 1$. We call the set of the points $P_C = \left\{ C(t_i) \middle| i \in \overline{1, n_C} \right\}$ the discretization set of the curve $\boldsymbol{C}$, where $\boldsymbol{n_C}$ is a number of grid points.

Problem statement. Let's $\mathcal{C} = \{C_1(t), C_2(t), ..., C_m(t)\}$ represent a set of continuous pairwise disjoint parametric curves. Construct an approximate Voronoi diagram $VD(\mathcal{C})$ of a set of curves $\mathcal{C}$ such that it also has a parametric representation (for the edges of the Voronoi graph). The procedure takes into account the set of discretized curves $P_{C_1}, P_{C_2}, ..., P_{C_m}$ and it's union $\mathcal{P} = P_{C_1} \cup P_{C_2} \cup ... \cup P_{C_m}$.

B. Construction of the discretization set of a curve

The simplest curve discretization method is based on the uniform parameter gird: $|t_{i+1} - t_i| = \Delta t, \forall i = \overline{1, n_C - 1}$. However, such approach might give a non-uniform representation in terms of the distance between two neighboring points on a curve. Thus, the density of points is not controlled and might vary significantly based on the parameter $\boldsymbol{t}$.

Another popular approach is to replace an input set of curves with a new set, which is obtained by performing arc-length [17] reparametrization of the old curves:

$$l_C(t) = \int_0^t \|C'(s)\| ds, t \in [0,1] \quad (7)$$

Here $l_C(t)$ denotes the arc-length of the curve $C(t)$. The reparametrized curve $C(l)$ is obtained from $C(t)$ as follows:

$$C(l) := C\left(l_C^{-1}(l)\right), l \in [0, l_{MAX}] \quad (8)$$

where $l_{MAX} = l(1)$ and $l_C^{-1}(l)$ is an inverse function of $l_C(t)$. After discretizing the parameter $\boldsymbol{l}$ using the uniform grid we obtain the points, which are uniformly spaced on a curve. However, the application of this approach is very restricted, since we still cannot control the approximation error for the graph representing the Voronoi diagram, which is constructed based on such discretization.

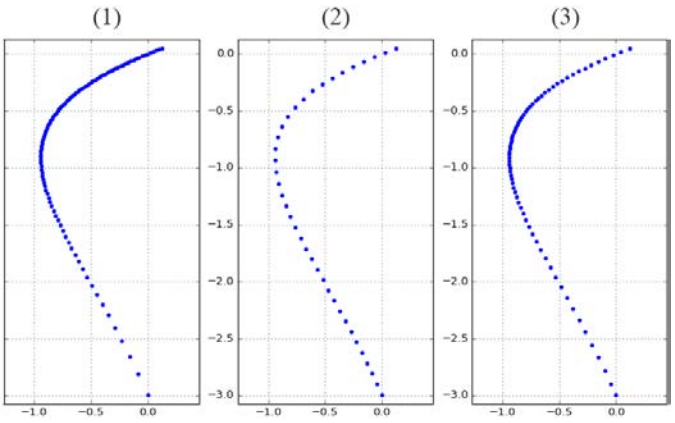


Figure 1. A visual comparison of different discretization methods. The equation of the input continuous parametric curve is following: $C(t)=\left(3\log\left(t+0.05\right)\sin(t),\log\left(t+0.05\right)\right)$, where $t\in[0,1]$: (1) uniform grid parametrization with $\Delta t=0.01$; (2) arc length parametrization with $\Delta l=0.1$; (3) adaptive discretization method proposed in this paper with $\Delta l=0.1$ and $d_{MAX}=2$;

In this paper we propose another adaptive approach based on the offset of the curve. Our new method allows us to bound the approximation error of the resulting Voronoi diagram.

An adaptive discretization method. Let's denote by $d_{\max}$ the maximal length of the Voronoi diagram edge (restricted by some field of view). The offset curve $O_d(t)$ with respect to the curve $C(t)=\left(x_C(t),y_C(t)\right)$ is defined as follows:

$$O_d(t)=C(t)+d\overrightarrow{N}(t) \tag{9}$$

where d is an offset distance and $\overrightarrow{N}(t)$ is a normal to the curve $\overrightarrow{N}(t)=\overrightarrow{w}^{\perp}(t)/\left\|\overrightarrow{w}^{\perp}(t)\right\|$ and $\overrightarrow{w}^{\perp}(t)=\left(-y_C'(t),x_C'(t)\right)$.

The curvature of $C(t)$ satisfies the following equation:

$$\kappa(t)=\frac{x_C(t)'\,y_C''(t)-x_C''(t)y_C'(t)}{\left(x_C'^{2}(t)+y_C'^{2}(t)\right)^{\frac{3}{2}}} \tag{10}$$

Discretization algorithm description:

1. Find approximate position of the points $0<s_i<1, i=\overline{1,n_\kappa}$ such that $\kappa(s_i)=0$ (using the search method on the finite grid and finite difference discretization for derivatives).
2. Split an input curve into parts between $[s_{i-1},s_i], i=\overline{1,n_\kappa}$;
3. Discretize each part of the curve using the method described below;
4. Merge discretized parts into one curve.

As it was stated above, firstly, we split the curve into the parts such that each part preserves the sign of the curvature. This is performed in order to build offset curves depending on the sign of the curvature.

The discretization of each part of the curve is based on the following procedure. Firstly, the offset curve $O_{d_{MAX}}(t)$ is constructed. Secondly, we perform an arc-length parameterization of $O_{d_{MAX}}(t)$ and obtain:

$$l_{O_{d_{MAX}}}(t)=\int_0^t\left\|O'_{d_{MAX}}(s)\right\|ds, t\in[0,1] \tag{11}$$

Then the arc-length parameter is taken as $t=l^{-1}_{O_{d_{MAX}}}(l)$. The part of the offset curve is discretized using the uniformly spaced grid with respect to the arc-length parameter l: $\left\{O_{d_{MAX}}\left(l^{-1}_{O_{d_{MAX}}}\left(l_j\right)\right)\middle| j\in I_{O_{d_{MAX}}}\right\}$. Finally, the discretized points are projected back onto the original curve using the normal vectors:

$$C\left(l_j\right)=O_{d_{MAX}}\left(l^{-1}_{O_{d_{MAX}}}\left(l_j\right)\right)-\overrightarrow{N}\left(l^{-1}_{O_{d_{MAX}}}\left(l_j\right)\right), j\in I_{O_{d_{MAX}}} \tag{12}$$

Theorem. The maximum length of the edge of the Voronoi diagram constructed based on the discretization procedure above is bounded by the discretization step of the offset curve..

Proof. This follows directly from the algorithm of construction, since we used uniform gird sampling of the arc-length parametrized offset curve at the distance $d_{\max}$.

C. The solution for arbitrary objects

At first step we build a Voronoi diagram for the union of discretization sets $\mathcal{P}$:

$$VD(\mathcal{P})=\left\{VP(p)\middle| p\in P\right\} \tag{13}$$

Let $J: P\to\mathbb{N}$ be a function, that for a given point returns index of curve, which the point of discretization belongs to:

$$J(p)=i \overset{def}{\Leftrightarrow} p\in P_{C_i} \tag{14}$$

Then the approximation of Voronoi cell associated with curve C_i is obtained by merging the Voronoi Cells for each point in the corresponding partition set:

$$\overline{VP}(C_i)\cong\bigcup_{p\in\mathcal{P}\ \&\ J(p)=i}VP(p) \tag{15}$$

Thus, the resulting approximation of Voronoi diagram for the set $\mathcal{C}$ is:

$$\overline{VD}(\mathcal{C})\cong\left\{\overline{VP}(C_i)\middle| i=\overline{1,n}\right\} \tag{16}$$

Firstly, we choose a discretized set of curve points for constructing a Voronoi diagram. The correspondence between each point of discretization and its curve is maintained. Next, we merge the Voronoi polygons, whose centers belong to the same curve, and remove the adjacent edges. Figures 2-3 show an example of Voronoi diagram construction for a set of 17 curves represented as Bezier splines.

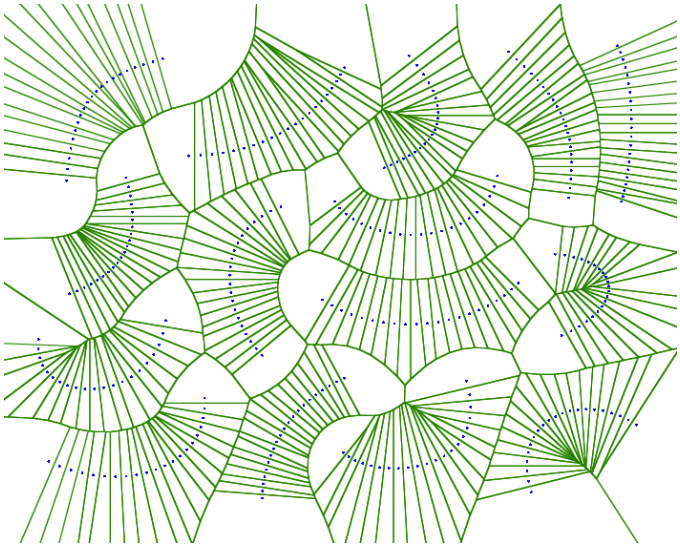


Figure 2.Example of the Voronoi diagram construction for simple objects, blue points belong to discretized curves;

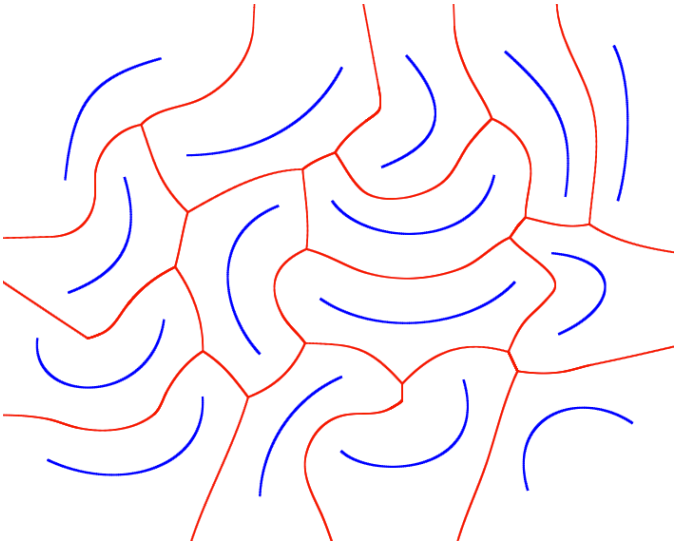


Figure 3. The final result after merging Voronoi cells and removing redundant Voronoi edges.

D. *Approximating Voronoi edge with curves*

After constructing the Voronoi diagram for discretization sets and merging cells (removing edges) of Voronoi diagram, we obtain an approximation of a Voronoi edges by polygonal chains (each chain connects two junction points). At the next step we approximate obtained chains by parametric curves.

In order to make the approximation of polyline by curves we should, firstly, choose the canonical equation of curve. Type of the approximating curve (canonical equation) depends on the desired features and parameters of the final result. In this paper, we consider the general case and make approximation by quadratic and cubic Bezier curves. Other types of curve can be considered in a similar way.

One of the most evident methods for approximation by curves is a least square fitting. We fix the first and last points (start and end point) of the Bezier curve and then use the least squares method to find best fit to the polyline.

Thus, for quadratic Bezier curves we are looking for the coordinates of point P_1:

$$B(t) = t^2 P_0 + 2(1-t)tP_1 + (1-t)^2 P_2 \tag{17}$$

In this case, least square method is reduced to the solution of a linear equation with one variable for each coordinate, which can be easily solved as follows:

$$\varphi(P_1) = \sum_{i=1}^{n} \left(P_i^* - B(t_i)\right)^2 \to \min \Rightarrow \frac{\partial \varphi}{\partial P_1} = 0 \tag{18}$$

In order to approximate the polyline with a cubic Bezier curve we should find x, y coordinates of points P_1 and P_2:

$$B(t) = t^3 P_0 + 3(1-t)t^2 P_1 + 3(1-t)^2 tP_2 + t^3 P_3 \tag{19}$$

This problem reduces to the solution of the system of linear algebraic equations for each of the coordinates. The system of equations consists of two equations:

$$\varphi(P_1, P_2) = \sum_{i=1}^{n} \left(P_i^* - B(t_i)\right)^2 \to \min \Rightarrow \begin{cases} \dfrac{\partial \varphi}{\partial P_1} = 0, \\ \dfrac{\partial \varphi}{\partial P_2} = 0. \end{cases} \tag{20}$$

Thus, the total number of equations in the system of linear equations depends on the order Bezier curve.

Implementation details. For each type of approximation curve we get an analytical solution (precise expression for each of the unknown points), which is relatively easy to implement.

III. COMPLEXITY ANALYSIS

The analysis of the algorithm's complexity is analyzed in the following theorems and lemmas.

Theorem. If m is a number of objects in the plane, which are represented by nonintersecting parametric curves, and the total number of points obtained after discretizing these m objects is *N*, then the approximation of the Voronoi diagram for the set of m arbitrary-shaped objects represented by pairwise disjoint non-intersecting parametric curves can be performed in computational time $O(N \log N)$.

Proof. In paper [1], [3], [18] authors discussed and described in detail the construction of the Voronoi diagram for a set of *N* points. The algorithm complexity is $O(N \log N)$, where *N* is a number of input points.

If Voronoi diagram is represented by doubly connected linked list, then will be correct the following lemmas:

Lemma 1. Merging two neighboring Voronoi polygons represented as doubly connected linked lists can be performed in $O(1)$ time.

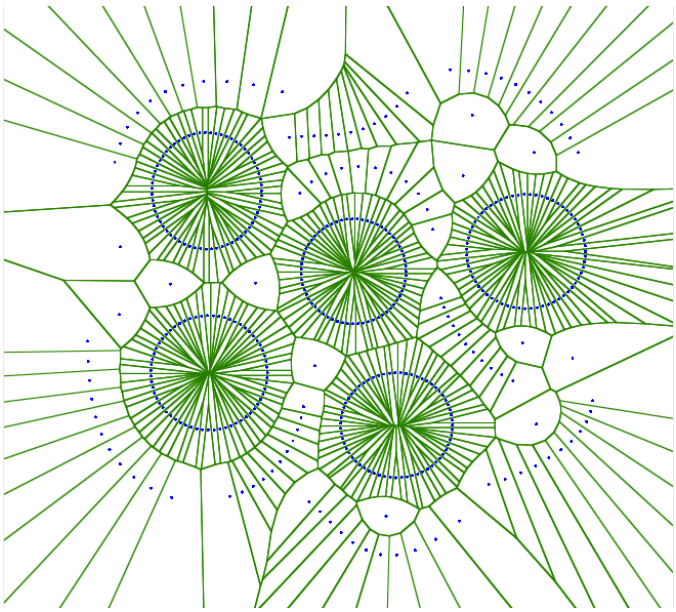


Figure 4. Voronoi diagram for discretized curves (circle, Bezier curves and points): blue points belong to discretized curves;

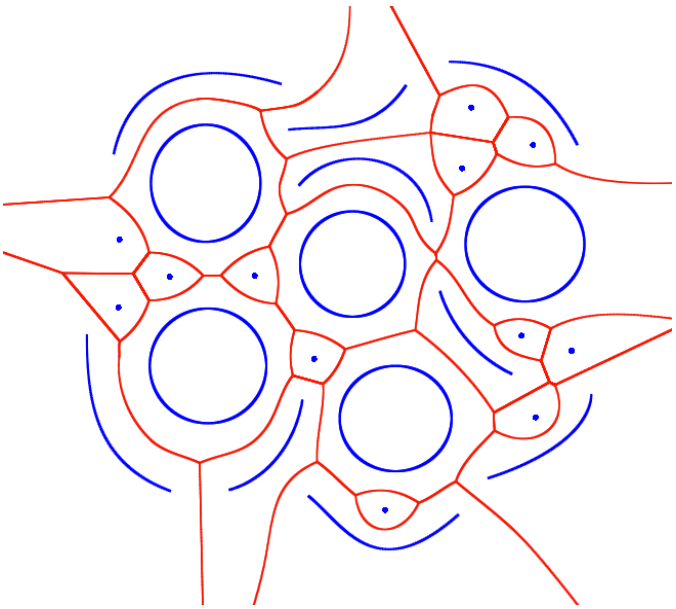


Figure 5. An approximate Voronoi diagram after postprocessing the Voronoi graph and approximation with Bezier curves;

Proof. In order to merge two neighboring Voronoi polygons we should merge corresponding doubly-connected linked lists. Merging of linked lists is a simple operation, which can be performed in $O(1)$ time by reassignment of pointers.

Lemma 2. Approximation of Voronoi diagram for arbitrary-shaped objects on a plane can be made using previously built Voronoi diagram for points in time $O(N)$.

Proof. By merging the neighboring Voronoi polygons, whose generators correspond to the same curve, we obtain an approximation of Voronoi diagram. The maximal number of Voronoi cells in this case is N. Since merging of the pair of cells takes $O(1)$, we can get approximation of Voronoi diagram in time $O(N)$ (by merging all necessary cells in pairs).

Considering lemmas 1 and 2 and next statement: quadratic or cubic Bezier curves fit polynomial chains in time $O(M)$, where M - number of points in chain; we can formulate following lemma:

Lemma 3. Approximation of the Voronoi diagram for arbitrary-shaped objects on a plane using Bezier curves can be performed in $O(N \log N)$ time.

Therefore the approximation of Voronoi diagram for set of m arbitrary-shaped objects on a plane, which are represented by non-intersecting parametric curves can be performed in time $O(N \log N)$, that concluding the proof.

IV. Implementation

During the implementation, we construct the discretized curves and build the Voronoi diagram for them using the well-known "divide and conquer" algorithm [18]. For saving the correspondence between curves indexes and points, we use a special function (which we implement as a hash-function). The resulting data structure maintains point-curve correspondence using a hash-map.

At the next step we merge the boundaries of Voronoi cells for neighbor points of the same curve to get an approximation of Voronoi cell for curve with polygonal chains. Voronoi diagram is represented by DCEL. The procedure of merging is following: we start from some point of the discretized curve p; run BFS(p) and walk through all neighbor points from the same partition. At each step we merge Voronoi cells of the neighboring points and remove unnecessary edges of Voronoi cells. During the merging procedure we also mark junction points of the resulting Voronoi graph. Figures 4-5 show the process of building the Voronoi diagram for a set of Bezier curves, ellipses and points.

The technical implementation of the proposed algorithm is preformed in programming language C++ using OpenGL and Qt framework for data visualization. The input to the algorithm can be provided either by user manually or by opening an SVG-file. The program also visualizes the main steps and intermediate results of the algorithm.

Also we tested the algorithm performance. All our experiments in this paper were carried on Intel Core i7 2.3GHz computer with 4GB RAM. Figures 6-7 illustrate the main results of the testing. In Figure 6 one can see how execution time depends on the distribution of the input data (in case of mixture of parametric curves and points).

V. Conclusion

In this paper, we propose an efficient algorithm for constructing an approximate Voronoi diagram for a set of pairwise disjoint arcs on a plane. The arcs are represented by parametric curves, which we discretize using the proposed in adaptive method. On the next step, Voronoi diagram for the discrete objects is constructed, redundant edges of the Voronoi graph are removed and corresponding Voronoi cells are merged. As a result, we obtain an approximate Voronoi

diagram, whose edges are represented by polyline chains. We also propose to approximate these chains by Bezier splines, which allows us to obtain a final representation of the Voronoi diagram by parametric curves.

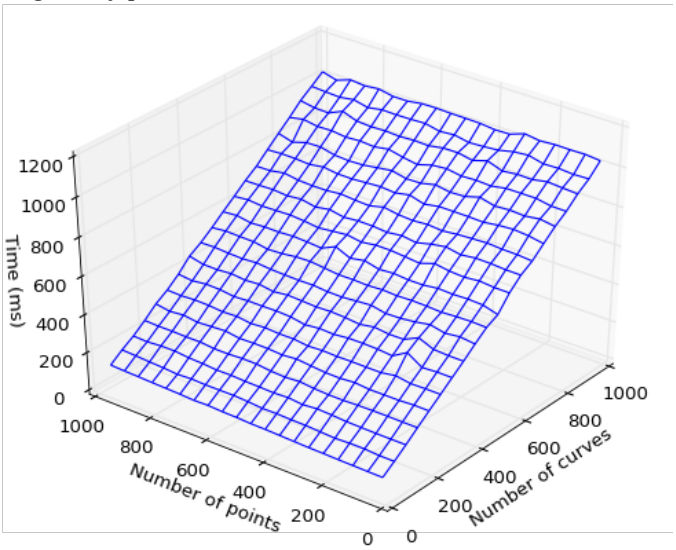


Figure 6. Execution time of the proposed approximation algorithm in case of different distributions of the input data (e.g., mixtures of points and parametric curves);

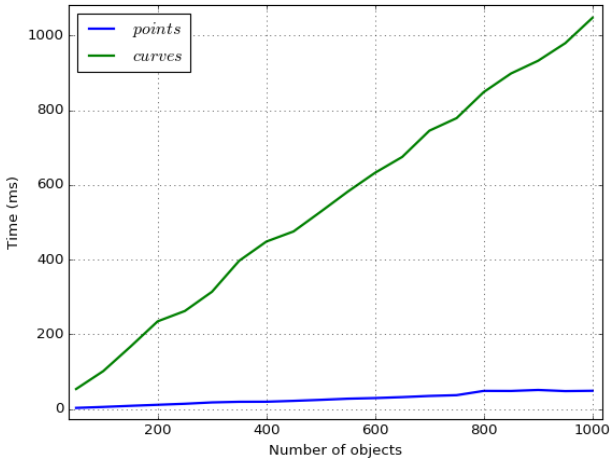


Figure 7. Comparison of the algorithm execution time in case of different input objects: parametric curves, points

In this paper we do not consider the cases, when the input curves are intersecting or sharing a common endpoint. As it was shown in [19], the application of similar to our techniques might lead to inadequacy of approximations in the mentioned above situations. Topological inconsistencies as described in [19] were handled during the merging procedure.

A significant advantage of our new approach is the ability to refine approximations for the critical areas, which could be useful for specific practical problems. We enable this by supporting dynamic structures like concatenable queues [18].

The complexity analysis shows that the computational time of the proposed algorithm is $O(N \log N)$. We have also implemented the algorithm in C++ programming language (the results are shown in Figures 2-5) and measured the running time. Plots in Figures 6-7 show the increase in computational time for curves in comparison to simple objects like points by the factor of ~20 for discretization step 0.1. The main reasons of such increase are computational overheads (curve discretization, approximation) and increase in total number of processed points.

REFERENCES

[1] F. P. Preparata and M.I. Shamos, Computational Geometry: An introduction, Springer-Verlag, Berlin, 1985, - 475 p.

[2] O. Aichholzer, W. Aigner, F. Aurenhammer, T. Hackl, B. Jüttler, E. Pilgerstorfer, M. Rabl, "Divide-and-conquer for Voronoi diagrams revisited", Computational Geometry: Theory and Applications, 2010, V 43, Is. 8, P. 688-699.

[3] S. Fortune, "A sweepline algorithm for Voronoi diagrams," Algorithmica, 1987, 2, pp. 153-174.

[4] D. Zhou, Z. Wang, S. Bandyopadhyay and M. Schwager, "Fast, On-line Collision Avoidance for Dynamic Vehicles Using Buffered Voronoi Cells," in *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1047-1054, April 2017.

[5] Masehian, E. and Amin-Naseri, M. R. (2004), A voronoi diagram-visibility graph-potential field compound algorithm for robot path planning. J. Robotic Syst., 21: pp. 275-300

[6] Rokicki, W., & Gawell, E. (2016). Voronoi diagrams – architectural and structural rod structure research model optimization. MAZOWSZE Studia Regionalne, (19), 155–164

[7] H. W. Bai, Z. L. Ye and M. Shi, "An improved approximate arc-length parameterization method for Bezier curves," *2006 7th International Conference on Computer-Aided Industrial Design and Conceptual Design*, Hangzhou, 2006, pp. 1-4.

[8] A. Sud, N. Govindaraju and D. Manocha, "Interactive computation of discrete generalized voronoi diagrams using range culling," Proc. International Symposium on Voronoi Diagrams in Science and Engineering, 2006, P. 1-10.

[9] K. E. Hoff, T. Culver, J. Keyser, M. Lin and D. Manocha, "Fast computation of generalized Voronoi diagrams using graphics hardware," Proc. of ACM SIGGRAPH Annual Conference on Computer Graphics, ACM, pp. 277–286, 1999.

[10] Y. J. Ho and J. S. Liu, "Collision-free curvature-bounded smooth path planning using composite bezier curve based on Voronoi diagram," IEEE International Symposium on Computational Intelligence in Robotics and Automation, Korea, pp. 463-468, 2009.

[11] S. Arya and T. Malamatos, "Linear-size approximate Voronoi diagrams," Proc. 13th ACM-SIAM Sympos. Discrete Algorithms, pp. 147–155, 2002.

[12] S. Arya, T. Malamatos, and D. M. Mount, "Space-efficient approximate Voronoi diagrams," Proc. of STOC, pp. 721-730, 2002.

[13] S. Har-Peled. "A replacement for Voronoi diagrams of near linear size," Proc. of FOCS, pp. 94-103, 2001.

[14] Seong et al., 2008, Voronoi diagram computations for planar NURBS curves, Proc. ACM Symp. Solid & Phys. Modeling, NY (2008), pp. 67–77.

[15] R. Ramamurthy and R. Farouki, "Voronoi diagram and medial axis algorithm for planar domains with curved boundaries: I. Theoretical foundations," J.Comput.Appl.Math. 102, pp. 119–141, 1999.

[16] F. Aurenhammer, R. Klein and D. T. Lee, "Voronoi Diagrams and Delaunay Triangulations," World Scientific Publishing Co., 2013.-p. 348.

[17] H. W. Bai, Z. L. Ye and M. Shi, "An improved approximate arc-length parameterization method for Bezier curves," *2006 7th International Conference on Computer-Aided Industrial Design and Conceptual Design*, Hangzhou, 2006, pp. 1-4.

[18] M. Shamos and D. Hoey, "Closest-point problems," Proc. 16th Annu. IEEE Sympos. Found. Comput. Sci., pp. 151-162, 1975

[19] R. Ramamurthy and R. Farouki, "Voronoi diagram and medial axis algorithm for planar domains with curved boundaries: II. detailed algorithm description," J.Comput.Appl.Math. 102, pp. 253–277, 1999

The Distance Effect in the Dosimetry Analysis of a Rat Model at GSM-900 Frequency Band: a Simulation Study

K. ATES[1], H. F. CARLAK[2] and S. OZEN[3]

[1]Akdeniz University, Antalya/Turkey, ateskayhan@gmail.com
[2]Akdeniz University, Antalya/Turkey, fezacarlak@akdeniz.edu.tr
[3]Akdeniz University, Antalya/Turkey, sukruozen@akdeniz.edu.tr

Abstract **- In this study, specific absorption rate (SAR) simulations of rat model have been carried out by using a finite integration technique (FIT). Although FIT is similar with finite difference time domain (FDTD) method in most ways, an integral form of Maxwell's equations are used in the FIT method. A monopole antenna working at 900 MHz has been designed for the electromagnetic source. Conductivity and permittivity of tissue have been selected from realistic values and implemented for the voxel based rat model.**

Simulations have been implemented with the 5 W stimulation power. Aforementioned antenna has been located at 3.5 cm and 5 cm away from the nearest point of the rat model, respectively. Total SAR values are found as 0.483 W/kg for the 3.5 cm distance and 0.315 W/kg for the 5 cm distance. Maximum SAR induced at a head region as it is expected. Furthermore, cross section of the head and body results indicate that induced SAR vary in different parts of a body because of electrical properties of each tissue. As the distance of the antenna increases, the SAR value decreases. Results show that average SAR value in 1 gr rat tissue is higher than the value in 10 gr rat tissue.

Keywords **- Specific absorption rate (SAR), dosimetry, finite integration technique (FIT), bioelectromagnetics.**

I. INTRODUCTION

DEVELOPMENT of the electricity caused that the uncountable harmful hazards either living tissue or devices using electricity which provoked the researchers investigating these cases. Therefore, many countries have been taken care of the protection rules for these dangers such as electrical shock, electromagnetic interference (EMI) [1], electromagnetic dosimetry [2], etc.

Electromagnetic power density radiated by electrical devices is absorbed by biological tissue. Therefore, operating mechanism of the tissue changes with respect to time. Electromagnetic dosimetry explains these biophysical processes, briefly. Dosimetry depends on the frequency, exposure time, electrical properties of biological tissues, electromagnetic power density and interaction mode (i.e. strong or weak axis) fundamentally.

In order to evaluate the dosimetry, numerical techniques and experimental observations have been implemented [3, 4]. In this way, researchers can analysis the biological tissues.

Nowadays, specific absorption rate is a significant indication for dosimetry analysis at the frequency range of cellular communication. Hence, epidemiological studies are carried out in order to assess the SAR. These studies include numerical calculations and experiments as mentioned [5, 6]. Furthermore, animal experiments for exposure analysis have been investigated in detailed. As a result of the various studies, it is known that temperature rises where the SAR is induced in the tissue [7].

International associations have published the SAR limits which are critical for the human health (Table 1). These SAR values have been published for a 10 gr tissue.

Table 1: SAR limits defined by international associations [8, 9]

Member	ICNIRP (W/kg)	IEEE (W/kg)
Head	2	2
Limb	4	4

In addition, according to the Federal Communications Commission (FCC), the induced SAR limit in the human head is 1.6 W/kg for a 1 gr tissue [10].

In this paper, SAR evaluations in the rat model have been investigated at the GSM-900 frequency band. Evaluations have been accomplished through finite integration technique (FIT) based on the simulation software. Following section includes underlying idea of the numerical evaluations and electrical properties of the rat model used in simulations. After that, simulation scenario is expressed and numerical dosimetry simulations are investigated.

II. MATERIALS AND METHODS

A. Specific Absorption Rate (SAR)

SAR is used for an absorbed electromagnetic energy in any biological tissue [11]. Electrical properties of the biological tissue and working frequency is defined as the fundamental parameters of SAR. Furthermore, direction and amplitude of an incident wave is significant for behavior of biological tissues. Therefore, these analyzes should be done rigorously.

SAR is calculated through equivalent tissue method as it is

shown in Equation.1. In this method, the tissue is mimicked with the electrically equivalent liquid called as a phantom. This material is prepared according to the realistic permittivity and electrical conductivity values of the related tissue.

$$SAR = \frac{dW/dt}{\rho} = \frac{Q_{ext}}{\rho} = \frac{\sigma|E|^2}{\rho} = \omega\varepsilon_0\varepsilon'' \frac{|E|^2}{\rho} = C\frac{\Delta T}{\Delta t} \quad (1)$$

Where, W is work (joule, J) and t is duration time (s). ρ represents the density of the biological tissue (kg/m^3). Q_{ext} is an external electromagnetic source (W/m^3). $|E|$ substitutes the amplitude of incident electric field (V/m). ε_0 represents the permittivity of free space which equals to 8.854×10^{-12} F/m. σ (S/m) and ε'' mimic the electrical conductivity and imaginary part of complex permittivity, respectively. ω is an angular frequency and equals to product of 2, π and f (working frequency, Hz). C is a tissue heat capacity (J/(kg.K)). T is a temperature (K).

Another method of observing the SAR is theoretical calculations including analytical calculations and numerical simulations through the mentioned formulas. In this method, electromagnetic wave, electrical properties of related tissue and density or temperature alteration, exposure time and heat capacity of the tissue should be known, basically. Apart from these methods, high precision thermal cameras, sensors and probes have been developed for observing the SAR.

When the incident wave propagates along the tissue, it must attenuate according to electromagnetic postulates [12]. Skin depth (δ) is the distance where the amplitude of the propagating wave in the material decreases on by *1/e* almost (0.368) and during that time, Poynting vector decays by $1/e^2$.

$$\delta = \left(\frac{67.52}{f}\right)\left[\sqrt{(\varepsilon')^2 + (\varepsilon'')^2} - \varepsilon'\right]^{-1/2} \quad (2)$$

where, ε' represents the real part of a complex permittivity.

B. Numerical Modeling

Numerical dosimetry analyzes have been done through the finite integration technique (FIT). Gridding process is applied as a finite integration technique like finite difference time domain (FDTD) method. However, unlike the other techniques, FIT uses integral form of Maxwell's equations [13].

In this study, the rat model which is used in simulations has been consisted of voxels. Voxel is defined as the smallest unit of the three dimensional objects. Resolution of one voxel is defined as 1.827×1.827×2.015 mm in the rat model used in studies.

Figure 1: Voxel rat model

Dielectric properties of biological structures might be obtained through complex permittivity measurement at desired frequency [14]. In this way, interaction between the incident electromagnetic wave and living tissue can be explained.

Table 2: Electrical properties of biological tissues at 900 MHz

Tissue	Epsilon	Conductivity (S/m)	Tissue	Epsilon	Conductivity (S/m)
Fat	5.462	0.051	Grey matter	52.72	0.942
Lymph node	59.684	1.0385	Ligament	45.826	0.7184
Marrow	55.043	0.0402	Gallbladder	59.142	1.2569
Bladder	18.936	0.383	Eye sclera	55.271	1.1668
Testis	60.553	1.2096	Lung	22.000	0.4567
Spleen	57.178	1.277	Pancreas	59.684	1.0385
Muscle	55.032	0.9429	Blood	61.36	1.5379
Liver	46.833	0.855	Heart	59.893	1.2298
Skin	41.405	0.8668	Eye bulb	68.902	1.6362
Gland	59.684	1.0385	Kidney	58.676	1.3922
Blood vessel	44.775	0.6961	Bone cortical	12.454	0.1433
Mucosa	46.08	0.8447	Colon wall	57.94	1.0799
Small intestine wall	59.487	2.1648	Cerebrospinal fluid	68.638	2.4126

Skin depth formula has been shown below.

$$\varepsilon^* = \varepsilon' - j\varepsilon'' \quad (3)$$

In equation (3), j represents the

$\sqrt{-1}$, ε^* indicates the complex permittivity and ε' represents the real part of the complex permittivity.

$$\sigma = \omega\varepsilon_0\varepsilon'' \quad (4)$$

Dielectric properties used in this paper is given in Table 2.

III. Simulation Results

Dosimetry simulations have been implemented using a FIT based simulation program. According to the simulation scenario (Figure 2), a monopole antenna is fed by a 5 W stimulation power due to a signal generator which injects an electric field into the environment. Designed antenna was verified by S_{11} parameters at the 900 MHz frequency range. By the way, this antenna has an electromagnetic interference with the rat model. The distance between antenna and the rat model (x) is 3.5 cm and 5 cm, respectively. Antenna is defined as a dual-systemeither converts electric field into a voltage and vice versa [7]. There are plenty of antenna types in terms of usage purposes.

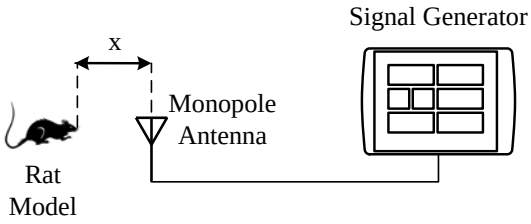
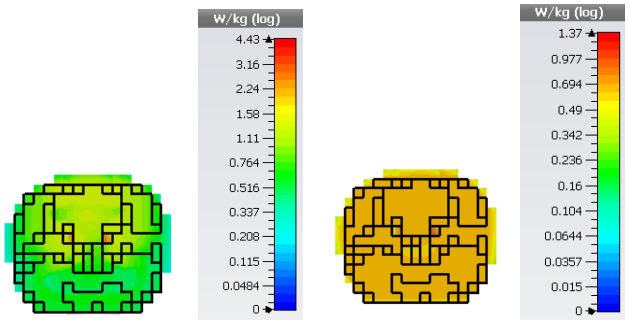


Figure 2: Simulation scenario

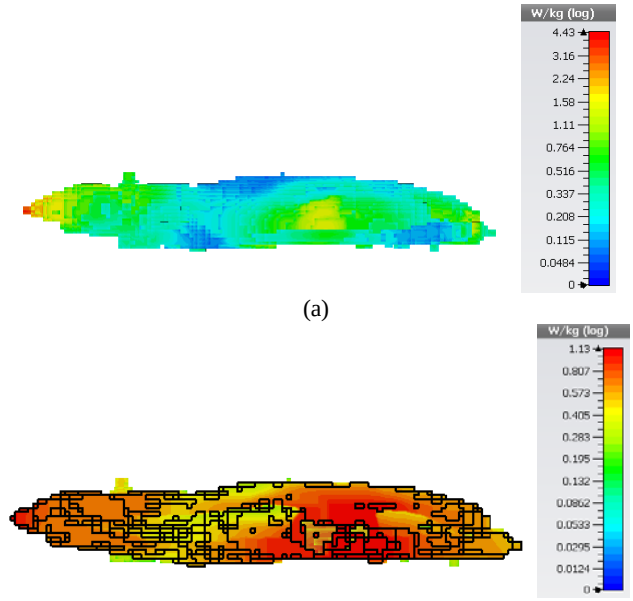
All results were assessed on the basis of their own conditions and all SAR results have been evaluated with the logarithmic scaling. Whole body average SAR values has been found as 0.483 W/kg where the x distance is adjusted as 3.5 cm and 0.315 W/kg SAR value is obtained for a 5 cm distance.



(a) (b)

Figure 3: SAR distribution of 1 gr (a) and 10 gr (b) tissue in the head region where x distance is 3.5 cm

Figure 3 represents the induced SAR in the head region where the antenna is located away 3.5 cm from the nose. Evaluations have been carried out for the 1 gr and 10 gr average tissue. Results showed that maximum values in the head region is obtained as 2.4 and 0.85 W/kg for 1 gr and 10 gr tissue, respectively.

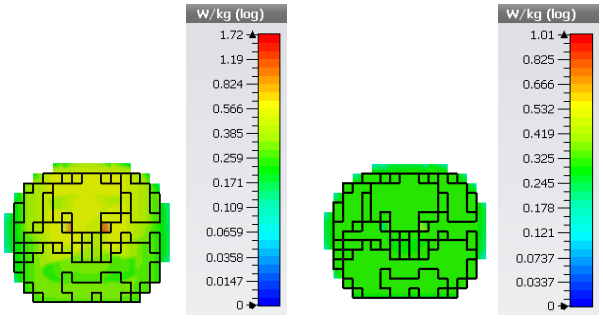


(a)

(b)

Figure 4: SAR distribution of 1 gr (a) and 10 gr (b) tissue in whole body where x distance is 3.5 cm.

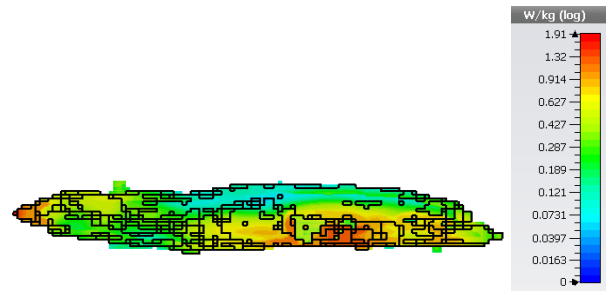
Figure 4.a denotes the sagittal view of whole body average induced SAR for a 1 gr. Tissue. In this figure, the maximum exposure is observed at the head region which is related with the distance and electrical properties of different tissues. Figure 4.b shows the induced SAR where the cross section of sagittal view of the whole body for a 10 gr. tissue. Results indicate that SAR value varies from region to region in different organs of the body such as the brain, liver and kidney. Maximum values have been observed in different organs which are the head and at middle region of the body. Also, 1.13 W/kg average SAR was obtained as the maximum value at the cross section of whole body for a 10 gr tissue.



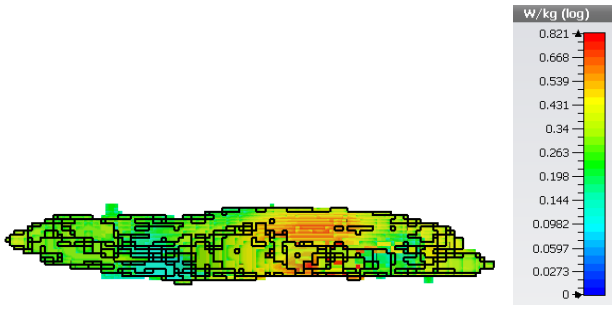
(a) (b)

Figure 5: SAR distribution of 1 gr (a) and 10 gr (b) tissue in the head region where x distance is 5 cm

Figure 5 shows the induced SAR values of the head region for 1 gr and 10 gr tissue in case of the 5 cm x distance. The maximum SAR value was calculated approximately 1.2 W/kg. The value of 0.6 W/kg was acquired for the 10 gr tissue.



(a)



(b)

Figure 6: SAR distribution of 1 gr (a) and 10 gr (b) tissue in the whole body where x distance is 5 cm.

Figure 6 indicates the average SAR values of 1 gr and 10 gr tissue at the cross section of the sagittal plane of whole body. Maximum values have been observed at the closest point of the model to the antenna and abdomen tissue of the body. Maximum values of the average SAR at cross section of whole body were obtained as 1.91 W/kg for 1 gr tissue and 0.82 W/kg for 10 gr tissue, respectively.

IV. CONCLUSION

In this study, dosimetry analysis at 900 MHz have been carried out through a FIT based numerical simulation software. Voxel based rat model was selected as the victim for the dosimetry evaluation. Electrical parameters of the rat tissue have been taken from the literature. Monopole antenna has been designed for numerical simulations and placed away from the rat model. Radiation power of the antenna, called Poynting vector, was assumed as 5 W.

Specific absorption rate (SAR) has been selected as the dosimetry evaluation criterion. As a result of the SAR simulations, head region, sagittal view of the whole body, which was evaluated only for 1 gr tissue for 3.5 cm distance, and cross section of sagittal plane of whole body have been evaluated. Furthermore, it is obviously stated that the induced SAR in 1 gr tissue is higher than the 10 gr tissue. The electromagnetic exposure decreases when x dimension increases. In other ways, the amplitude of the electromagnetic wave decays when the distance increases. Propagating wave mostly interferes with the head region of the body which is the closest part in between the body and the antenna.

Furthermore, electromagnetic exposure is related with the wavelength of the propagating wave and the physical dimensions of the victim.

REFERENCES

[1] K. Ates, S. Ozen and H. F. Carlak, "Electromagnetic radiation characteristics investigation and electromagnetic interference analysis of household appliances," in *1st International Mediterranean Science and Engineering Congress*, 2016, pp. 1-9.

[2] K. Ates, H. F. Carlak and S. Ozen, "Mitigation of the magnetic field created by the underground power cables and exposure analysis of the human body," in *1st International Mediterranean Science and Engineering Congress*, 2016, pp. 1-7.

[3] P. Gajsek, et al., "Empirical validation of SAR values predicted by FDTD modeling," *Bioelectromagnetics*, vol. 23, no. 1, pp. 37-48, 2002.

[4] K. Wake, et al., "An exposure system for long-term and large-scale animal bioassay of 1.5-GHz digital cellular phones," *IEEE Transactions on Microwave Theory and Techniques*, vol. 55, no. 2, pp. 343-350, 2007.

[5] O. Spathmann, et al. ""Head Only"-exposure of continuously growing rats to 900 MHz GSM signals," in *Electromagnetic Compatibility (EMC EUROPE), 2012 International Symposium on. IEEE*, 2012, pp. 1-5.

[6] L. Yang, et al., "SAR and temperature distribution in the rat head model exposed to electromagnetic field radiation by 900 MHz dipole antenna," *Australasian Physical & Engineering Sciences in Medicine*, vol. 36, issue. 2, pp. 251-257, 2013.

[7] K. Ates, and H. F. Carlak, "Dosimetry analysis of the human head model due to mobile phone usage at GSM-850 frequency band," in *Engineering and Technology 2017 IEEE Int. Conf. on*, pp. 1-5.

[8] *IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz*, IEEE Standard C95.1-2005, 2005.

[9] A. Ahlbom, et al., "ICNIRP (International Commission on Non-Ionizing Radiation Protection) Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic and Electromagnetic Fields (up to 300 GHz)," *Health Phys.*, vol. 74, no. 4, pp. 494-522, 1998.

[10] Federal Communications Committee (FCC) RF Safety FAQ [Online], Available: https://www.fcc.gov/engineering-technology/electromagnetic-compatibility-division/radio-frequency-safety/faq/rf-safety#block-menu-block-4, Consulted: 23 February 2018.

[11] A. V. Vorst, A. Rosen, and Y. Kotsuka, *RF/Microwave Interaction with Biological Tissues*. John Wiley & Sons, 2006.

[12] C. A. Balanis, *Advanced Engineering Electromagnetics*. John Wiley & Sons, 1999.

[13] T. Weiland, "Advances in FIT/FDTD modeling," in *Proceedings of 18th Annual Review of Progress in Applied Computational Electromagnetics*, 2002, pp. 1-14.

[14] S. Gabriel, R. W. Lau, and C. Gabriel, "The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz to 20 GHz," *Physics in Medicine & Biology*, vol. 41, no. 11, pp. 2251-2269, 1996.

[15] L. Puranen, et al., "Space efficient system for whole-body exposure of unrestrained rats to 900 MHz electromagnetic fields," *Bioelectromagnetics*, vol. 30, no. 2, pp. 120-128, 2009.

[16] C. K. Chou, et al., "Development of a rat head exposure system for simulating human exposure to RF fields from handheld wireless telephones," *Bioelectromagnetics*, vol. 20, no. S4, pp. 75-92, 1999.

Modeling and Simulation of Complex Mechanical Systems Using Electrical Circuit Analog

M.AKBABA[1], A. DALCALI[2] and M. GÖKDAĞ[3]

[1] Karabuk University, Karabuk/Turkey, mehmetakbaba@karabuk.edu.tr
[2] Karabük/Turkey, dalcali.adem@gmail.com
[3] Karabük University, Karabük/Turkey, mgokdag@karabuk.edu.tr

Abstract **- A system is a set of interacting components are connected in such a way that the vibration or response in the state of one component affects the state of others. A dynamic system is described by time differential equations; therefore, the future response of the system is determined by the present state of the system (the initial conditions) and the present input. Modeling and simulation are the prerequisite to the design of a dynamic system, as the designing and manufacturing a system and then testing it for expected performance would be unavoidably expensive. Wide spread availability of high speed and high storage capacity computers makes the simulation process even more attractive before the manufacturing stage for observing the expected performance of complex dynamic systems. However, before simulation phase of a system the mathematical model of the system must be prepared. For complex mechanical systems it is not always easy to establish the mathematical model. On the other hand, it is much easier to establish the mathematical model of an electrical system. Furthermore, simulation software is richer for electrical systems. For these reasons in this manuscript modeling of a complex mechanical systems will be realized using electrical circuit analog and then will be simulated as an electrical circuit. Advantages of this approach will be explored.**

Keywords **- Modeling and simulation, complex mechanical systems, electrical system analog.**

I. INTRODUCTION

SIMILAR to other engineering disciplines, the modeling and simulation is an art as well as science. It requires training and experience to become a competent. Scientific discovery, or understanding, involves the formulation of theory to explain observed phenomena, the design and execution of experiments to test theory and the feedback of experimental results to evolve theory. This process is known as the scientific method and plays a central role in any scientific discovery endeavor. Modeling and simulation is the prerequisite to the design of a dynamic system, as the designing and manufacturing a system and then testing it for expected performance would be unavoidably expensive. Wide spread availability of high speed and high storage capacity computers makes the simulation process even more attractive. It lives at the intersection between theory and experiment and highly valued in scientific discovery because it provides additional insights that are often impractical or impossible to discover through real-world experimental and theoretical analysis alone. It can be viewed as virtual experimentation that could be implemented in any circumstances. Benefits of modeling and simulation can be summarized as follows;

a) A simulation model executed on a computer system can compress the time frame in thousands of folds and can be used to investigate quickly the effects of a changes in a real-life situation that take place over several years.
b) Can be used to study complex systems that would be very difficult to investigate without run a simulation model on a fast-digital computer.
c) In engineering and product design the effect of various changes on a complex systems performance can be easily investigated without producing a physical prototype.
d) Can be used to investigate situation that true hands-on experiments would be dangerous in real life.
e) Optimization routines can be incorporated into the simulation models and system can be designed with optimum performance in various aspects.

There are many steps to fallow in preparation of modeling and simulation of a real life complex system that involves components from many engineering disciplines such as mechanical, electrical, electromechanical, fluid, thermal, etc. However, involving in such a wide detail is beyond the scope of this presentation. Therefore, in this study modeling and simulation of a relatively complex mechanical system using analog electrical circuit will be presented in details and the benefits accrued from such a study will be explored.

II. MODELING

A mathematical model of a dynamic system is description of the system in hand in terms of differential or integrodifferential equations governing the dynamic behavior of the system. The type of model sought will depend on both the objective of the engineer and available tools for simulation. It is the opinion of this author that analysis tools available for electrical systems are wide spread and easy to understand. Therefor the aim of this paper is to show how mechanical systems and systems from other engineering disciplines can be modelled and simulated as analogous to an electrical circuit. For this purpose, a mechanical

system is taken as an application example.

A. Analogy Between Mechanical and Electrical Systems

Basics of analogy between the mechanical and electrical systems are given in several references [1-3], but application details are not explored further. Therefor the topic did not receive the deserved attention in the existing literature and did not find wide spread use. For example, for complex systems composed of several blocks and interconnections it is not clear to the user how to translate these interconnection into the electrical circuits. This important aspect of the subject will be explained in this presentation. In a translational mechanical system there are three main system elements, which are mass (M), stiffness element (K) and viscous friction element (B). The most common stiffness element is he spring, which will be considered in this presentation. Corresponding elements in a rotational mechanical system are moment of inertia, rotational stiffness and rotational viscous friction. Similarly, in electrical systems there are well known three elements that are inductance (L), capacitance (C) and resistance. The governing equations between mechanical and electrical systems show a complete similarity as shown in the Table 1 and Table 2. This gives opportunity for a perfect analogy between mechanical and electrical systems. Analogy given below is known as direct analog or force-voltage analog in which each velocity corresponds to an electrical current and each displacement corresponds to an electrical charge.

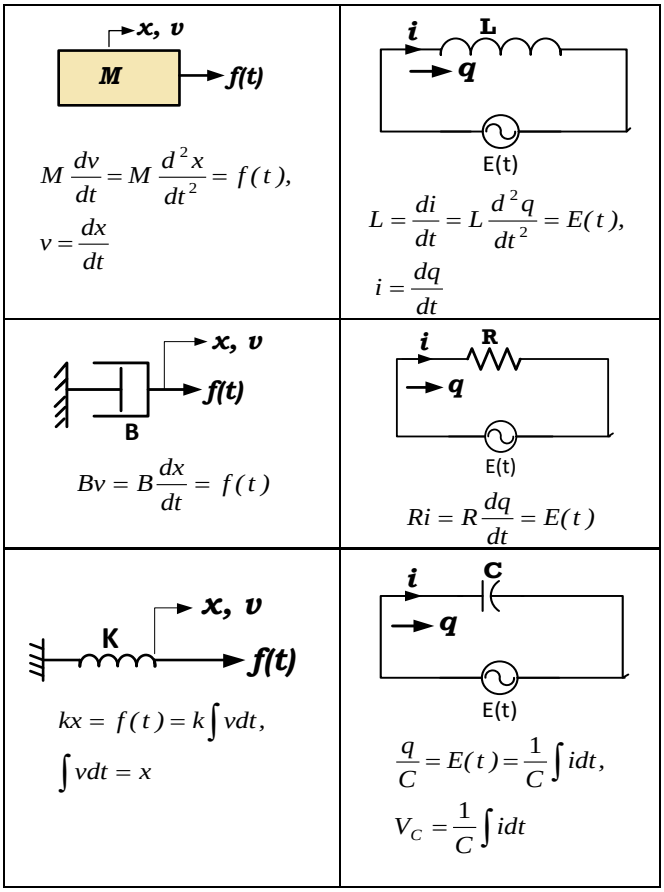
Table 1: Analogy between the electrical and mechanical quantities.

Symbol	Mechanical	Unit	Symbol	Electrical Analog	Unit
f	Force	N	E	Voltage	V
x	Displacement	m	q	Electrical charge	C
v	Speed	m/s	i	Electric current	A
M	Mass	kg	L	Inductance	H
B	Viscous friction	Ns/m	R	Electrical resistance	Ω
k	Spring	N/m	$\frac{1}{C}$	Capacitor	F

The analogy between the physical elements and related governing equations given in Tables 1 and 2 clearly show that there is a complete similarity between the two systems. Therefor the both system always can virtually represent each other. Further to this fact what remains next is the form of presentation of the interconnects between different blocks of mechanical systems in equivalent electrical circuits. This aspect can be laid down as follows:

Table 2: Analogy between governing equations of the mechanical and electrical elements.

Mechanical Elements and governing equations	Analogue Electrical elements and analog governing equations



a) Each mass in a mechanical system corresponds to a loop in the equivalent electrical circuit. Therefor the number of loops in the equivalent electrical circuit is defined by number of different masses in mechanical system.

b) The interconnect elements in each mass block in mechanical systems corresponds to the electrical elements shared between the equivalent electrical circuit loops.

c) Series elements carrying same value of force in a mechanical system corresponds to parallel elements in the equivalent electric circuit.

d) Velocity of each mass in a mechanical system corresponds to each loop current in the equivalent electrical circuit.

e) Displacement of a mass in a mechanical system correspond to the total electrical charge accumulated on each associated capacitor in equivalent electrical circuit.

f) Displacement difference between the two masses corresponds to the total charge associated on the capacitor placed on the branch of equivalent electric circuit that is common to the loops associated with each mass.

g) Series elements in mechanical systems subjected to same force are in parallel in equivalent electrical circuits or they are placed in adjacent loops when associated with the other elements.

An example mechanical system is shown in Fig. 1. A suspended system is selected in order in which the gravitational forces can be accounted for. The gravitational forces are shown as M_1g, M_2g and M_3g respectively, where g is the usual gravitational acceleration (g=9.8 m/s2). If references for

displacements x_1, x_2 and x_3 are taken as the positions of masses before application of external forces Fa(t) and Fb(t) but after system being suspended and the sprigs are lengthened, then all the gravitational forces can be dropped from the model equations. In this case the model equations of a suspended systems become like the model equations of a translational systems. Therefore, the way of modelling applied to the suspended system in hand will be equally valid for translational systems as well. Hence in the forgoing analysis gravitational forces will be omitted from the model equations and references for displacements will be taken at the equilibrium position that has taken place after system has been suspended, but before the application of the external forces.

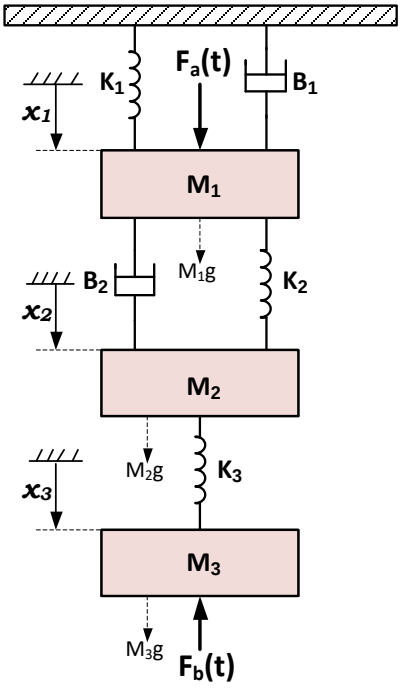


Figure 1: Selected example mechanical system.

The system parameters are selected as follows:
M_1=15 kg, M_2=12 kg, M_3=9 kg, K_1=220 N/m, K_2=350 N/m, K_3=200 N/m, B_1=20 Ns/m, B_2=15Ns/m and the applied forces are selected as:

$$F_a(t) = 60\sin(2t)e^{-0.1t} + 120\,N \quad and$$
$$F_b(t) = 30\cos(2t)e^{-0.08t} + 20\,N$$

Applying the principles of the analogy that have been explained earlier the electrical equivalent circuit of the mechanical system given in Fig. 1 will be as shown in Fig. 2;

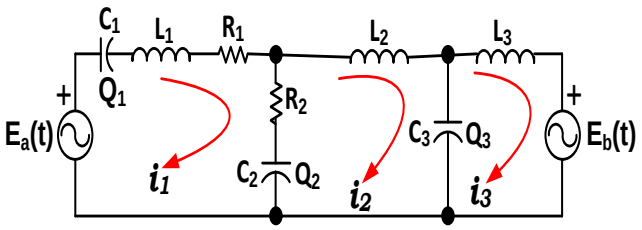


Figure 2: Equivalent Electrical circuit of the mechanical system given in Fig. 1.

The amplitudes of equivalent electrical circuits parameters naturally are L_1=M_1, L_2=M_2, L_3=M_3 C_1=1/K_1, C_2=1/K_2, C_3=1/K_3, R_1=B_1, R_2=B_2, $E_a(t)$= $F_a(t)$ and $E_b(t)$= $F_b(t)$. Equations governing the circuit given in Fig. 2 will be as follows:

$$E_a(t) = (R_1 + R_2)i_1 + L_1\frac{di_1}{dt} + \frac{Q_1}{C_1} + \frac{Q_2}{C_2} - R_2 i_2 \quad (1)$$

$$0 = -R_2 i_1 + R_2 i_2 - \frac{Q_2}{C_2} + L_2\frac{di_2}{dt} + \frac{Q_3}{C_3} \quad (2)$$

$$\frac{dQ_1}{dt} = i_1 \quad (3)$$

$$\frac{dQ_2}{dt} = i_1 - i_2 \quad (4)$$

$$\frac{dQ_3}{dt} = i_2 - i_3 \quad (5)$$

$$-E_b(t) = -\frac{Q_3}{C_3} + L_3\frac{di_3}{dt} \quad (6)$$

Organizing the six simultaneous differential equations into State-Space model one obtains:

$$\dot{X} = AX + BU$$
$$Y = CX + DU$$

Where A is the system matrix, B is the input matrix, C is the output matrix, X are state variables, Y are output variables and U are input forcing functions. For the system in hand the open form of these matrixes is given as follows:

$$\begin{bmatrix} \frac{di_1}{dt} \\ \frac{di_2}{dt} \\ \frac{di_3}{dt} \\ \frac{dQ_1}{dt} \\ \frac{dQ_2}{dt} \\ \frac{dQ_3}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R_1+R_2}{L_1} & \frac{R_2}{L_1} & 0 & -\frac{1}{L_1C_1} & -\frac{1}{L_1C_2} & 0 \\ \frac{R_2}{L_2} & -\frac{R_2}{L_2} & 0 & 0 & \frac{1}{L_2C_2} & -\frac{1}{L_2C_3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{L_3C_3} \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ Q_1 \\ Q_2 \\ Q_3 \end{bmatrix} + \begin{bmatrix} \frac{1}{L_1} & 0 \\ 0 & 0 \\ 0 & -\frac{1}{L_3} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} E_a(t) \\ E_b(t) \end{bmatrix} \quad (8)$$

The matrices A, B and U are available from the above equation. Since outputs are taken as state variables themselves then the output matrix C is a unit diagonal matrix as follows:

$$[C] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and the direct transmission matrix D will a zero matrix of 6 by 2. The system parameters are selected as: M_1=12 kg, M_2=7 kg, M_3=5 kg, K_1=180 N/m, K_2=125 N/m, K_3=100 N/m, B_1=20 Ns/m, B_2=15 Ns/m and the two applied forces are selected as:

$F_a(t) = 60\,sin(2t)e^{-0.1t} + 120\,N$ *and*

$F_b(t) = 30\,cos(2t)e^{-0.08t} + 25\,N$

The amplitudes of equivalent electrical circuits parameters naturally are L_1=M_1, L_2=M_2, L_3=M_3 C_1=1/K_1, C_2=1/K_2, C_3=1/K_3, R_1=B_1 and R_2=B_2.

The state-space equations are programmed in MATLAB environment using 4th order Runge-Kutta method (RK4) [7] and the results are shown in Figures 3 and 4 respectively for velocities and displacements of the masses.

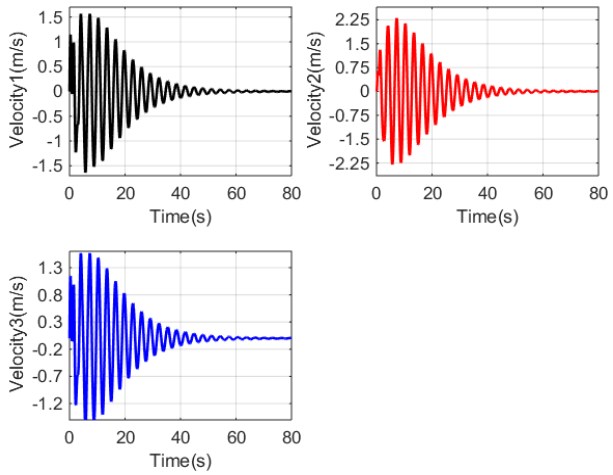


Figure 3: Variation of velocities of the three masses versus time.

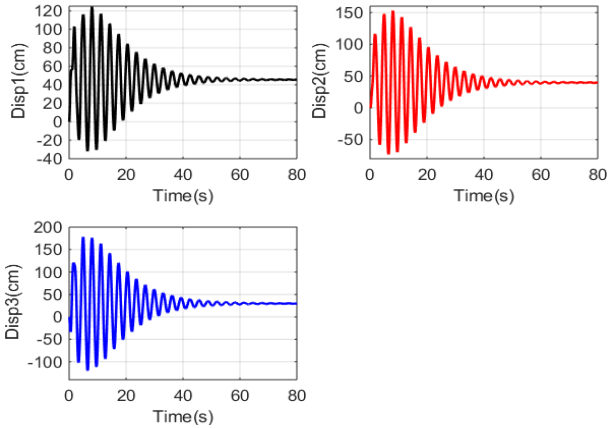


Figure 4: Variation of the displacements of the three masses versus time.

Figure 3 shows that as the time varying sine and cosine terms in force functions decay to zero and only constant terms remain then all three velocities also decay to zero. This is due to fact that after forces attain constant values then, as can be seen from Figure 4, the displacements also attain constant values and as natural result of this the velocities decay to zero. Figure 3 show that the velocity of the middle mass attains higher value then the velocities of both top and bottom masses, because the middle mass is affected both the vibration of both the top and bottom masses which are under the forcing of forces with time phase difference (sine and cosine terms are displaced by 90°).

Figure 4 shows that the displacements exhibit large excursions between positive and negative values, although the larger force on the top side has much higher constant term than the amplitude of exponentially decaying component (60 to 120 N). These large excursions in the displacements are the combined results of the effect of phase difference between the sinusoidal components of both applied forces and the natural vibrations arising from sudden application of the forces.

To verify that the effect of phase difference in the exponentially decaying sinusoidal components of the forces, the displacements are re-determined by removing the above said phase difference, i.e., converting the cosine term to sine term in Fb(t). The displacements computed under this condition is presented in Figure 5. It can be observed from this figure that when the phase difference between the Forces, the large excursions in the displacements are greatly reduced. The remaining transient components are almost the result of natural vibration arising from sudden application of the forces.

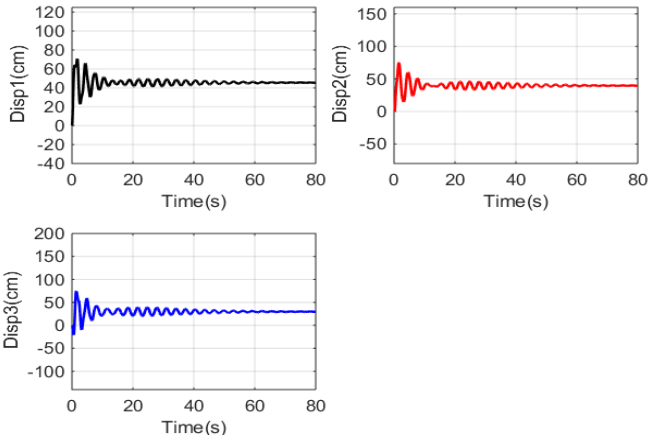


Figure 5: Variations of displacement versus time when both forces are in phase (sine and sine), i.e., $F_a(t) = 60\,sin(2t)e^{-0.1t} + 120\,N$

$$and\ F_b(t) = 30\,cos(2t)e^{-0.08t} + 25\,N$$

It can be observed from Figure 5 that when both forces are in phase, i.e. both forces are sine (or both are cosine), since they are applied in opposite direction then their effects are almost canceled. But the steady-state values are not affected. In both cases the steady-state vales of the displacements are almost 46 cm, 40 cm and 30 cm.

To verify the above argument the exponentially decaying sinusoidal components are removed from force expressions and displacements are recalculated. Results of this calculations are plotted in Figure 6 for all 3 displacements. Comparison of the Figure 5 and 6 shows that the results are almost same in both cases. There are only minor differences in the transient portions of the displacement. This clarifies the effect of phase difference between the applied forces.

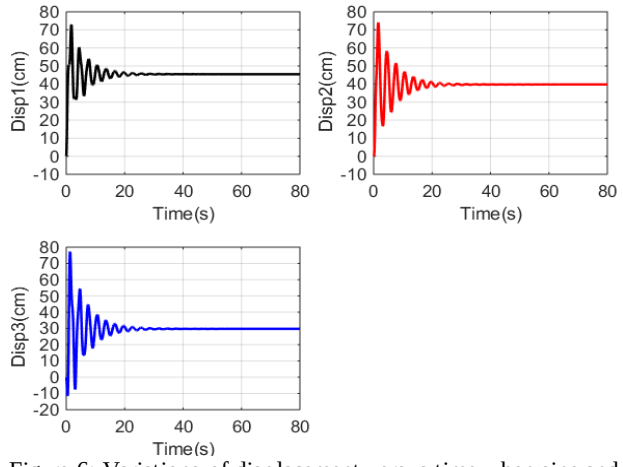


Figure 6: Variations of displacement versus time when sine and cosine terms are removed from forces, i.e., Fa(t)=120 N, Fb(t)=25 N.

The steady-state values are not affected. In both cases the steady-state vales of the displacements are almost 46 cm, 40 cm and 30 cm.

To verify the above argument the exponentially decaying sinusoidal components are removed from force expressions and displacements are recalculated. Results of this calculations are plotted in Fig. 6 for all 3 displacements. Comparison of the Figure 5 and 6 shows that the results are almost same in both cases. There are only minor differences in the transient portions of the displacement. This clarifies the effect of phase difference between the applied forces.

III. CONCLUSION

Analogous systems are the systems that can be presented by the same mathematical model but that are different physically. The concept of analogous systems is very useful in modeling and simulation or even in experimentation. The solution of the equations describing one system can be directly applied to analogous system any other field. On the other hand, when dealing with systems experimentally usually is much easier to deal with electric systems. Therefor instead of building and studying a mechanical, hydraulic, thermal, pneumatic systemor the like, we can build and study its electrical analog. In this present study full details of obtaining the equivalent electric circuit analogue of complex mechanical systems has been explained in details and it is applied to a relatively complex suspended mechanical system. Equivalent electric circuit analog has been derived through which the state-space model of the example mechanical system is obtained. Solving the obtained state-space model the velocities and displacements are computed and presented graphically without direct modeling the mechanical system in mechanical equations. The response of the mechanical system is easily obtained for various shapes of the applied force functions. In real life complex systems are usually composed of interconnections of sub-systems from different disciplines of engineering. Applying the general approach explained in this study, such complex systems can be unified into a single equivalent electric circuit and then the modeling, simulation, and analysis can be performed easily, like the example presented in this study.

REFERENCES

[1] K. Ogata, *Modern Control Engineering*. Prentice Hall, 2010.
[2] C. N. Dorny, *Understanding Dynamic Systems: Approaches to Modeling, Analysis and Design*. Prentice Hall, 1993.
[3] K. Ogata, *System Dynamics*. Pearson New Int. Edition, 2014.
[4] R. L. Woods, and K. L. Lawrence, *Modeling and Simulation of Dynamic Systems*. Prentice Hall, 1997.
[5] H. V. VU and R. S. Esfandiari, *Dynamic Systems, Modeling and Analysis*. McGraw Hill, 1997.
[6] K. Ogata, *State-Space Analysis of Control Systems*. Prentice Hall Inc., Englewood Cliffs, 1967.
[7] S. C. Chapra and R. P. Canale, *Numerical Methods for Engineers*. McGraw Hill, 2014.

3-D Modeling and Analysis of Shaded Pole Motors Using Finite Elements Method

A. DALCALI[1] and M. AKBABA[2]

[1] Karabük/Turkey, dalcali. adem@gmail.com
[2]Karabuk University, Karabuk/Turkey, mehmetakbaba@karabuk.edu.tr

Abstract **- Among the electrical machines the shaded pole motors are the most difficult ones in terms of modeling and analysis. This difficulty is arising from the non-uniform airgap and unbalanced windings on its stator. Because of these the airgap flux contains rich space harmonics, and this complicates the mathematical modeling and simulation of these machines. On the other hand, small size of these machines results in strong end effects. Hence the 3-D modeling of magnetic field analysis poses an important advantage. Therefore, in this presentation a 15 W, 4 pole shaded pole motor has been modelled in 3-D using Finite Elements (FE) method to determine some important performance parameters such as the airgap flux distribution and the saturation effects in the motor laminations. In this process magnetic field distribution have been obtained for three different case which are:**

a) Main winding is excited but shading rings and rotor cage are not excited

b) Shading rings are excited, but stator winding and rotor cage are not excited

c) Both the stator winding and shading rings are excited, but rotor cage is not excited

Finally, advantages accrued from such an analysis is discussed in detail.

Keywords **– Shaded pole motor, Finite Element Analysis, 3D model, Magnetostatics analysis, Magnetic flux density.**

I. INTRODUCTION

Shaded pole induction motors are having a simple structure in terms of design but they are the most difficult motors to analyze theoretically. The magnetic field generated by in these motors is an elliptical rotating magnetic field and this makes the analysis of these machines very complicated. There is no standard procedure in their modeling and determination of the motor performance analyses. For this reason, there are very few studies conducted on such motors [1-4].

Examining the existing literature, it is seen that researchers usually perform 2D analyzes of the SPIMs under different conditions. Sarac and Cundev have developed 4 different shaded pole induction motor (SPIM) models and performed 2D analysis in case of unloaded, rated load and locked rotor. In the analyzes the magnetic flux density, saturation region and weaknesses of the core have been determined [5]. In another study [6], Sarac energized the windings of the SPIM separately to perform the analyzes in 2D and obtained the flux waveforms on the air-gap of the motor. It takes a long time to analyze the skew effect of rotor bars by 3D analysis. For this reason, they investigated that effect in 2D analysis by dividing the rotor in n pieces displaced from each other by a very small angle. Further to this, in another study, utilizing Finite Elements Analysis (FEA) analysis they found that the motor has high flux density on the stator bridge. As a solution, they preferred to use soft magnetic material in the core and allowed for better flux distributions in critical regions [7]. Zhou and Rajanathan have worked on the optimization of starting torque using 2D FEA. They found that the factors affecting the starting torque value were the stator pole shape and the rotor slot shape [8]. Neto et al. have investigated the iron losses of a SPIMs with variable air gap as a function of skew. Although the motor structure is simple, the presence of harmonic components in the stator magneto motor force has made it difficult to analyze [9].

Since the short circuit resistance and inductance, the main winding inductance and skew are not directly taken into consideration in 2D FEA Shamlou and Mirsalim, performed 3D analyzes and found that experimental data were compatible with 3D analyzes [10]. By using genetic algorithm (GA) optimization technique, Sarac and Cvetkovski, designed and analyzed 3 different types of motor models. They obtained magnetic flux distribution, efficiency, torque, power factor and output power of each motor at nominal load. Because of the optimization studies, the efficiency and the electromagnetic torque were improved [11]. When the literature is examined, it is seen that generally 2D analyzes of the SPIM are studied. 3D modeling and analysis of electric machines is a more accurate step because of neglecting the fringing effect and leakage field areas in the 2D modeling. Also, it will be possible to design a machine closer to reality with 3D modeling. In this study, 3D magnetostatics analysis of the SPIM has been carried out in consideration of the mentioned points.

II. 3D MODEL OF SHADED POLE MOTORS

Shaded pole induction motors have salient pole stator and squirrel cage rotor. SPIM has some advantages such as being inexpensive, easiness in maintenance and direct on-line connection to a single-phase supply without any interfacing requirement. Despite these advantages, it has low starting torque, low efficiency and the direction of rotation cannot be changed in single-ring SPIMs. Because of its simplicity and low cost, it is widely used in small power applications, such as aspirators, hair dryers and toys. In Europe, as much as 10 million pieces per year of these motors are produced. Single-phase motors with capacitor are produced 700-800 thousand

pieces per year. Other alternative efficient motors are produced in Europe with around 100,000 units per year [12-15].

3D modeling offers more accurate model because of both not neglecting the fringing effect and leakage fields where mostly ignored in the 2D modeling. With 3D modeling, considering the presence of these effects and it will be possible to design a machine closer to the expected performance and analysis will be more realistic. The 3D exploded model of the SPIM is given in Fig 1. The physical and electrical properties of the SPIM are shown in Table 1.

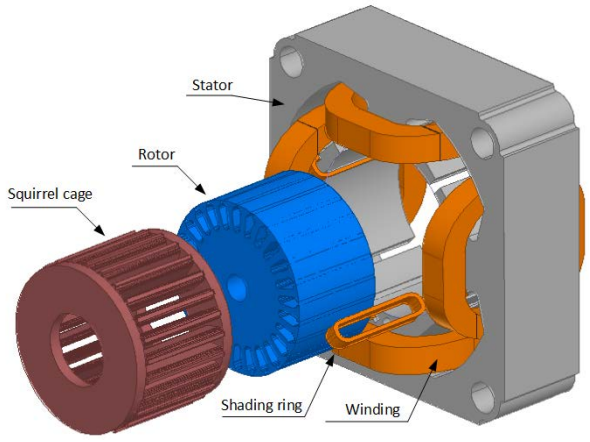


Figure 1: Exploded 3D models of SPIM

Table 1: Physical and electrical characteristics of the motor.

Properties	Unit	Value
Nominal power	W	15
Nominal voltage	V	220
Nominal current	A	0.375
Frequency	Hz	50
Pole number	-	4
Rotor speed	rpm	1305
Distance between two poles	mm	3
Core depth	mm	25
Rotor radius	mm	22
Rotor type	-	Squirrel cage
Rotor slot number	-	26
Number of main windings	-	580

III. Finite Element Analysis and Mathematical Modeling of the Motor Under Investigation

Finite element analysis is a method developed to find approximate numerical solutions on the region of magnitudes which are continuous in a certain region such as electric field, magnetic field and whose variations on this region can be expressed by partial differential equations. In this method, the region to be solved is divided into the finite number of small elements. Later, the solution is assumed to be continuous over the smallest region (mesh) with the desired magnitude and the governing partial differential equation expressing the field change is also valid for each element. To obtain a solution at any point on the region, the contributions of the elements surrounding that point are added to the account. For this reason, the sizes of the corner points of all elements in the region are linked together in a chain. As a result, linear equation with up to the number of nodes is obtained. Thus, the solution of this equation calculates the required values [3,16].

It is possible to create the desired design and behavior model of the machine with approximation methods such as finite difference and finite element used in the analysis of electric machines. By applying these methods to the machine design, it is possible to determine the electromagnetic parameters at a high accuracy [17,20]. To calculate the electromagnetic quantities, 2D and 3D models of the machine models can be created, and finite element analysis can be performed. The use of FEA provides the designer with time and economic benefits. The application of FEA to machine design, ensures that critical design parameters such as core losses, winding inductances and induced torque of the machine are determined with high accuracy. The mesh structure of the 3D modeling SPIM is shown in Fig 2. For the more precise solution, the number of elements must be higher in the close to the air gap and the rotor slot regions. When Figure 2 is examined, it can be observed that the number of elements increases especially in the regions where the fringe effect will occur. 1532583 pieces of tetrahedra were created for the solution.

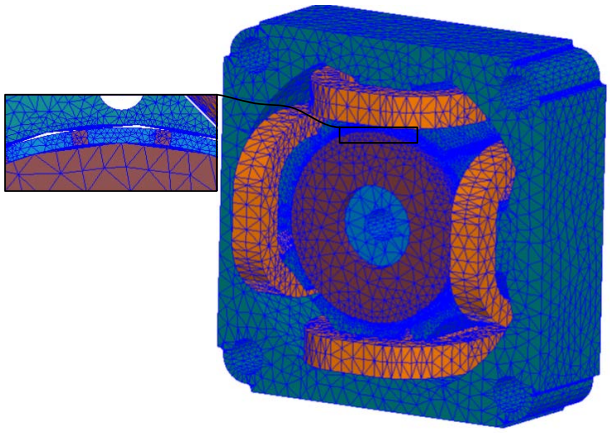


Figure 2: 3D mesh model of SPIM.

It is important to know the magnetic field distribution in AC machines since the accuracy of the calculated performance parameters are tightly related to correct calculation winding inductances. On the other hand, most of the inductance components can be determined only from the accurate magnetic field distribution in the machine magnetic circuit. The magnetic field in electric machines is expressed by Maxwell's equations. Since motor magnetic field is distributed over 3 different mediums, air, steel sheets and cupper or aluminum and steel sheets possess non-linear behavior then FEA is the more suitable method for the magnetic field analysis of the electric machines. Following Maxwell equations are used this purpose:

$$\nabla \times \vec{H} = \vec{J}$$
$$\nabla \times \vec{E} = -\frac{\partial B}{\partial t} \quad (1)$$

where, $\vec{H}$ is the magnetic field strength, $\vec{J}$ is the current density, $\vec{E}$ the electric field intensity, $\vec{B}$ is the magnetic flux density. The magnetic vector potential $\vec{A}$ is defined as in magnetic flux density in Equation 2.

$$\vec{B} = \nabla \times \vec{A} \quad (2)$$

The basic formulation of the vector potential for the magnetic field is expressed by Equation 3.

$$\nabla \times \left(\nu \nabla \times \vec{A}\right) = \vec{j} \quad (3)$$

Here ν indicates variable permeability as $B = f(H)$ characteristic of the core material is a nonlinear relation and permeability is expressed as $\nu = \frac{\partial B}{\partial H}$.

IV. Analysis and Result

.

SPIMs have a main winding and shaded pole rings on the stator and rotor cage in the rotor. In this section the magnetic field analysis is analyzed in three steps:

a) Main winding is excited but shading rings and rotor cage are not excited,

b) Shading rings are excited but main winding and rotor cage are not excited

c) Both the stator winding and shading rings are excited, but rotor cage is not excited.

When only the main winding is excited by applying 232 Ampere-turn (At), the magnetic flux density distribution and vector representation is given in Fig 3.

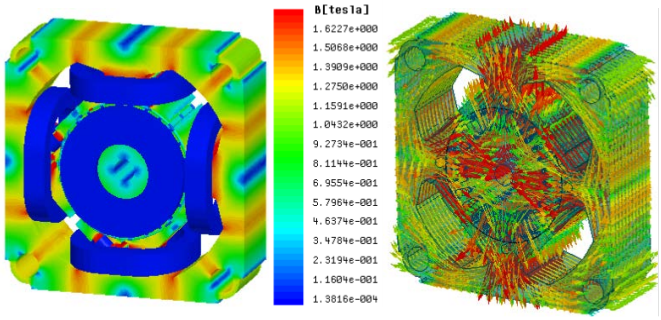


Figure 3: 3D magnetic flux distribution and vector representation when only the main winding is excited.

The distribution of the flux density obtained in Figure 3 is obtained by the following equation (3).

$$\frac{\partial}{\partial x}\left(\nu \frac{\partial A}{\partial x}\right) + \frac{\partial}{\partial y}\left(\nu \frac{\partial A}{\partial y}\right) + \frac{\partial}{\partial z}\left(\nu \frac{\partial A}{\partial z}\right) = -\vec{J} \quad (4)$$

The current density $\vec{J}$ is equal to zero in the air-gap and in the steel sheet. The magnetic flux density value is calculated from Equation 5 in 3D analyzes (4).

$$B = \sqrt{B_x^2 + B_y^2 + B_z^2} \quad (5)$$

Here, B_x, B_y and B_z are the components of magnetic flux density in the *x*, *y* and *z* axis directions respectively. When the flux distribution is examined, it can be observed that the flux density is particularly high at the corners. As expected, magnetic flux density is zero in copper which is not a magnetic material.

When only shaded pole excited by applying 40 At, the magnetic flux density distribution and vector representation is given in Fig 4.

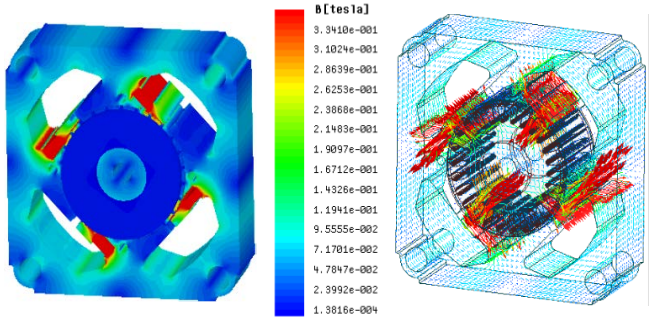


Figure 4: Flux density distribution on the machine model (only the shading rings are excited).

The flux density distributions obtained by the analysis of the excitation at the shaded pole 40 At and the main winding 232 At is shown in Fig. 5.

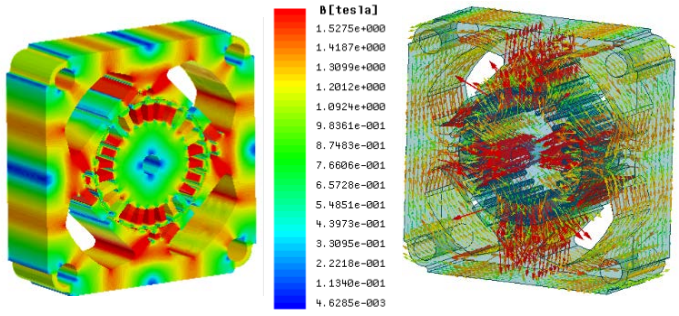


Figure 5: Flux density distribution on the machine model (both windings are excited).

Examination of the obtained flux distribution show that the flux values are relatively high especially at the pole end regions and at the regions where the shading rings are placed. The BH curve of the M43 steel used as stator and rotor material shows in

Figure 6. The BH characteristic of the steel sheet shown in Figure 6 can be modelled into closed form as in [19] and can be easily used in magnetostatics calculations to speed-up the iterations. Accordingly, the resulting inductance parameters can be modeled and used in determining the transient behavior of the motor [20-21]

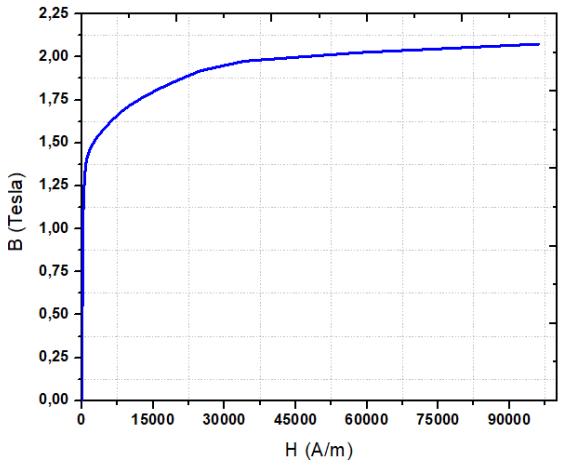


Figure 6: The BH curve of the core material (steel M43).

When all excitation conditions are examined by considering the BH curve in Fig. 6, it is seen that the flux values obtained in the core are within the targeted limits.

V. CONCLUSIONS

In this study, 3D modeling and design of a 15 W, 4 poles one phase shaded pole motor was realized. Magnetostatics analysis was performed under different excitation conditions using the Finite Elements analysis of the designed motor. Since in 2D FEM analysis end ring resistance and leakage inductances of the coils cannot be considered, to overcome these difficulties, in this study 3D modeling and analysis is performed. Thus, more accurate flux distributions are obtained, including the 3D fringing and the above said effects on the designed motor. From the obtained results it was observed that the M43 sheet, which was selected as the core material, is suitable in all three excitations conditions. A relatively higher flux density is found especially in the region of the shading rings and in the rotor slots. Therefore, in order not to drive the motor into saturation, which can cause sharp rise in the magnetizing current which in turn results in deterioration of the power factor and efficiency, in the design phase emphasis should be focused on these high flux density regions.

REFERENCES

[1] A. B. Dehkordi, "A single-phase induction machine model for real-time digital simulation," *in International Conference on Power Systems Transients*, Croatia, 2015.

[2] F. I. Kentli, "A Survey on design optimization studies of induction motors during the last decade," *Journal of Electrical & Electronics Engineering Istanbul University*, vol. 9 (2), pp. 969–975, 2009.

[3] M. Akbaba, and S.Q. Fakhro, "Field distribution and iron loss computation in reluctance augmented shaded-pole motors using finite element method," *IEEE Transactions on Energy Conversion*, vol. 7 (2), pp. 302–307, 1992.

[4] M. Akbaba and S.Q. Fakhro, "An improved computational technique of inductance parameters of shaded pole motoes using Finite Elements method," *IEEE Transactions on Energy Conversion*, vol. 7 (2), pp. 308–314, 1992.

[5] V. Sarac, and D.M. Cundev, "Electromagnetic fields calculation at single phase shaded pole motor," *Electrotechnica Electronica*, vol. 47, pp. 41–45, 2012.

[6] L. Petkovska, M. Cundev, and V. Sarac, "FEM analysis of asymmetrical magnetic field in electrical machines," *II. International Conference on Advanced Computational Methods in Engineering*, pp. 1–10, 2002.

[7] V. Sarac, G. Stefanov, and G. Cogelia, "Study of performance characteristics of single phase motors," *Automatic Control and Robotics*, vol. 15 (2), pp. 71–83, 2016.

[8] D. Zhou, C. B. Rajanathan, and A. T. Sapeluk, "Finite-element-aided design optimization of a shaded-pole induction motor for maximum starting torque," *IEEE Transactions on Magnetics*, vol. 36 (5), pp. 3551–3554, 2000.

[9] P. S. Neto, S. L. Nebeta, I. E. Chabu, and J. R. Cardoso, "Analytical modeling of shaded pole motors with non-uniform air gap," *Electric Machines and Drives Conference*, pp. 1656–1662, 2003.

[10] S. Shamlou, and M. Mirsalim, "A new restructured shaded pole induction motor- 3d finite element analysis and experimental verification," *International Journal of Smart Electrical Engineering*, vol. 4 (3), pp. 125–130, 2015.

[11] V. Sarac, and G. Cvetkovski, "Different motor models based on parameter variation using method of genetic algorithms," *Przeglad Elektrotechniczny*, vol. 3, pp. 162–165, 2011.

[12] A. Karmakar, P. K. Saha, and G. K. Panda, "D-Q Axis modelling analysis of a shaded pole induction motor and study of the non-linear behavior," *International Conference on Power, Energy and Control*, pp. 596–600, 2013.

[13] M. Andriollo, M. De. Bortoli, G. Martinelli, A. Morini, and A. "Tortella, Design improvement of a single-phase brushless permanent magnet motor for small fan appliances," *Transactions on Industrial Electronics*, vol. 57 (1), pp. 88–95, 2010.

[14] F. Parasiliti, M. Villani, and M. Castello, "PM brushless DC motor with exterior rotor for high efficiency household appliances," *International Conference on Electrical Machines*, pp. 623–628, 2014.

[15] Y. Gao, K. T. Chan, and Y, Shuang, "A novel chaotic-speed single-phase induction motor drive for cooling fans," *Fourtieth IAS Annual Meeting. Conference Record of the 2005 Industry Applications Conference*, pp. 1337–1341, 2005.

[16] S. S. Rao, "The Finite Element Method in Engineering," 4th Edition, Elsevier Science & Technology Books, 2004.

[17] S. L. Ho, and W. N. Fu, "Review and future application of finite element methodes in induction motors," *Electric Machines & Power Systems*, vol. 26 (2), pp. 111–125, 1998.

[18] K. Hameyer, and R. Belmans, "Numerical modelling and design of electrical machines and devices," *WIT Press, Southampton, Boston*, pp. 305, 1999.

[19] M. Akbaba, "A modified Froelich's type equation for accurate modeling of magnetic cores," *Electric Power Systems Components and Systems (Formerly Electric Machines and Power Systems)* Vol. 19(3), pp.303-311, 1991.

[20] M. Akbaba, S.Q.Fakhro, "Saturation Effect in in Three-Phase Induction Motors," *Electric Power Systems Components and Systems (Formerly Electric Machines and Power Systems)*, Vol. 12(3), pp.179-193, 1987.

[21] M. Akbaba, "Modeling of the Saturated Leakage Reactance of Induction Motor as a Time Varying Parameter for Transient Computations,"*lectric Power Systems Components and Systems (Formerly Electric Machines and Power Systems)*, Vol. 20(5), pp.539-548, 1992.

An Intelligent Material Placement for Electrical Installation Project

B.AKGÜL[1] and H. KUTUCU[2]

[1] Bartin University, Bartin/Turkey, bayramakgul@bartin.edu.tr
[2]Karabuk University, Karabuk/Turkey, hakankutucu@karabuk.edu.tr

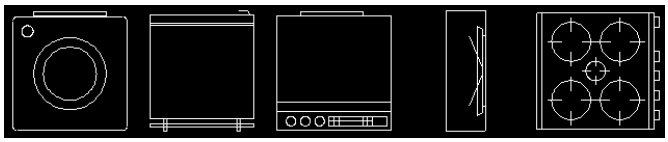
Abstract **- In this study, we consider an intelligent placement of electrical materials such as luminaire, oven, dishwasher, washing machine, refrigerator, TV-phone socket, etc. for the appropriate places for drawing electrical installation diagrams. Electrical installation project drawings of the buildings are prepared on the architectural projects which are prepared by the architects. When the electrical materials as symbols are placed on the project, with the type of building, the type and size of the inner area, the purpose of use and most importantly other descriptive symbol drawings (named as "architectural furnishing symbols") are taken into consideration. For example, if an oven symbol is found in a drawing part, then it means that this part most likely is a kitchen. Although any of one architectural furnishing has dozens of different symbols, they are basically similar to each other with small differences. In our application, "architectural furnishing symbols" are introduced to the system using artificial neural networks, then "electrical material symbols" are automatically placed according to the furnishing symbols detected at the site.**

Keywords **- Artificial neural networks, classification, computer aided design, electrical installation project**

I. INTRODUCTION

In this study, a system was developed for the installation of electrical materials depending on the furnishings for drawing the electrical installation project. This system has been integrated on the 2D vector drawing application previously developed by Akgul [1].

Furnishings in architectural projects are the name of refrigerator, television, cooker-oven, washing machine, seats, tables, beds, cabinets, etc. symbols. Some furnishings are shown in Figure 1.



(a) (b) (c) (d) (e)

Figure 1: some furnishing symbols: (a) washing machine, (b) refrigerator, (c) dishwasher, (d) television, (e) cooker-oven.

In this study, only "intelligent placement of electrical material symbols" is carried out which is a preprocess for a complete smart drawing of the electrical installation project.

The system consists of three phases; 1) selection of a closed region and discovery of the objects in the interior, 2) identification of the furnishing object found, 3) proposals are made for the electrical material suitable for the furnishing object described.

II. SELECTION OF CLOSED REGION AND DISCOVERY OF THE OBJECTS

Most of the rooms in the architectural drawing are rectangular. However, there are many rooms that are not in rectangular shape. Rooms on the drawing can be considered as closed polygons. In a CAD drawing environment, a closed polygon can be constructed via point or window selection with the cursor. The objects within a closed polygon can be easily determined according to their location and which of them can be evaluated as furnishings is determined according to the type of object. All block objects in the drawing can be evaluated as possible furnishings, but some of them should be eliminated. All "block insertion" objects in a selected closed area are listed in this phase.

III. RECOGNITION OF FURNISHING OBJECT

The problem can be defined as pattern recognition problem in general. Pattern recognition is a process of deciding for input samples using the information of some data sets [2]. More clearly pattern recognition is a way of responding to the question "what is this?".

In pattern recognition problems; artificial neural networks have been used for many problems, such as handwriting analysis [3,4,5], optical character recognition [6,7], signature recognition [8,9], fingerprint detection [10], road or building detection from satellite images [11,12,13], detection of objects in pictures or videos [14], vehicle license plate recognition [15].

There are five general steps for pattern recognition system. These are pre-processing, segmentation, display, training and testing.

For the preliminary operation of describing the symbols, the furnishing symbols are transformed into a bitmap in black and white at 128x128 or 64x64 pixel resolution as black regions 0, white regions 1. These bitmaps are converted to a matrix then represented as one-dimensional vector and normalized in double precision number to be used for feature vector in neural networks. An example of symbol segmentation is shown in Figure 2.

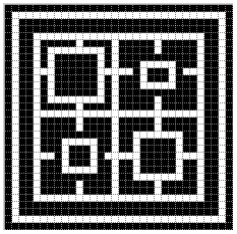


Figure 2: Symbol segmentation.

The use of the resolution initially set when converting furnishing symbols to bitmaps is an important influence. Because the line thicknesses do not change when scaling in vector drawings, but when raster images are scaled the lines become thicker or some pixels become disappear depending on the enlargement or reduction. This causes the symbol to be distorted. In the application, the initial resolution is parametrically determined, so while preparing the symbols for preprocessing, distortions are prevented.

For the artificial neural network, different scenarios were designed in Matlab environment. Table 1 shows some of these scenarios. In the table, [In Hdn Out] refers the number of neurons in Input/Hidden/Output layers. The results shown in Table 1 are visualized as the graphic in Figure 3.

In the study, 2395 128x128 pixels bitmap is used. The study was conducted on 2120 samples and the training was tested on 275 samples. Table 1 shows the accuracy of the training with different networks and resolutions. In the study, as seen in the Table 1, more successful cases have occurred in situations where there is no hidden layer. The best performance ratio is 92.36 percent with 128 input neurons and no hidden layer.

Resolution (px)	128x128		64x64		32x32	
Network [In Hdn Out]	Training time h:mm:ss	Accuracy (%)	Training time h:mm:ss	Accuracy (%)	Training time h:mm:ss	Accuracy (%)
[16 27]	0:02:42	86.55	0:00:47	88.36	0:00:12	84.73
[16 8 27]	0:11:51	76.36	0:00:53	78.99	0:01:53	69.93
[16 16 27]	0:04:22	81.82	0:01:13	82.90	0:00:22	79.35
[16 32 27]	0:04:33	74.91	0:00:24	76.09	0:00:13	70.29
[32 27]	0:01:18	86.91	0:00:27	91.27	0:00:07	87.27
[32 8 27]	0:10:36	74.18	0:02:00	80.80	0:00:53	77.54
[32 16 27]	0:02:29	85.45	0:00:56	85.87	0:00:15	81.25
[32 32 27]	0:01:38	87.27	0:00:47	87.32	0:00:13	88.04
[32 64 27]	0:02:20	86.18	0:00:42	90.22	0:00:16	52.73
[32 128 27]	0:02:12	60.00	0:01:12	57.25	0:00:17	85.82
[64 27]	0:02:08	89.45	0:00:50	91.27	0:00:11	88.00
[64 8 27]	0:20:43	78.55	0:13:04	53.99	0:01:08	77.45
[64 16 27]	0:02:33	81.45	0:03:01	73.55	0:00:18	84.73
[64 32 27]	0:02:21	82.55	0:00:53	89.49	0:00:11	85.09
[64 64 27]	0:02:25	89.09	0:00:46	90.58	0:00:16	86.18
[64 128 27]	0:04:09	36.73	0:01:49	62.68	0:00:21	87.64
[128 27]	0:04:07	88.00	0:01:18	92.36	0:00:21	88.36
[128 16 27]	0:08:15	85.82	0:03:34	88.41	0:01:06	83.27
[128 32 27]	0:05:51	87.27	0:02:29	90.55	0:00:29	88.36
[128 64 27]	0:09:29	84.36	0:01:21	90.94	0:00:48	87.64
[128 128 27]	0:09:50	60.00	0:03:07	78.16	0:00:44	87.27
[256 27]	0:14:55	90.18	0:02:40	90.58	0:00:48	89.09
[256 16 27]	0:36:02	84.36	0:06:08	81.88	0:01:37	86.18
[256 32 27]	1:26:15	86.55	0:12:41	84.78	0:01:10	88.00
[512 27]	0:42:01	85.82	0:06:28	91.67	0:01:32	88.73
[1024 27]	1:07:58	83.70	0:15:10	90.58	0:04:34	88.00
Intel Core i7-6700HQ Cpu, 2.6 Ghz, 16 GB Ram, Windows 10 x64, Matlab: R2015a (GPU and Parallel pool are used)						

Table 1: Training durations and accuracy rates of some artificial neural network scenarios using segmentation at different resolutions.

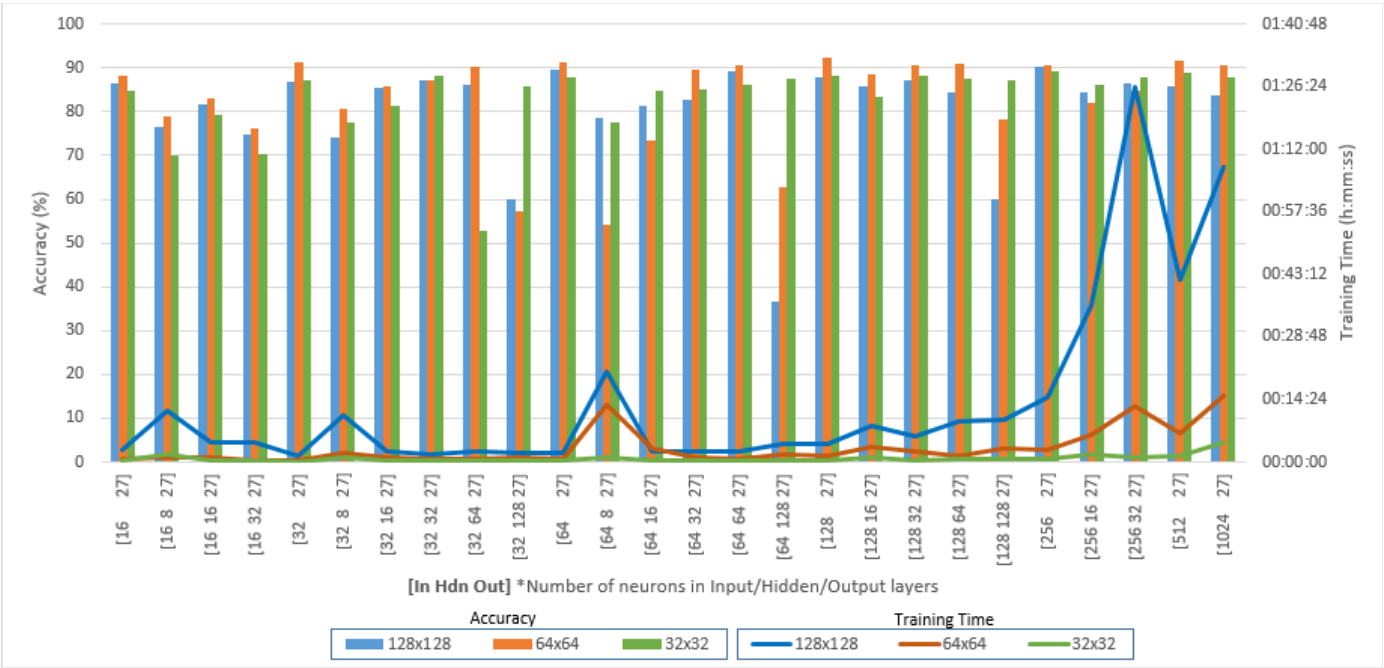


Figure 3: Training durations and accuracy rates of some artificial neural network scenarios.

For network training, *'trainscg'* function that updates weight and bias values according to the *scaled conjugate gradient method* is used. *'mse'* performance function is used to measure the network's performance according to the *mean of squared errors*. *'logsig'* transfer function which is the logarithmic sigmoid function is used to calculate a layer's output from its net input.

I. Placement of Electrical Material

In this step, a look-up table has been developed to suggest electrical materials suitable for furnishings. For example, an electrical socket for refrigerators, washing/dishwashers, and cooker furnishings have been proposed. Furthermore, the electrical socket, both antenna, and telephone jacks for the TV have been specified. In addition, according to the furnishings found, the type of room was estimated and appropriate lighting luminaire for this area also was proposed.

The electrical equipment recommended in the look-up table is located automatically, the closest wall where the furnish is fixed. For finding the nearest wall and rotating the electrical material symbol on drawing, the work in [1] is used.

II. Conclusion

In this study, only "intelligent placement of electrical material symbols" is carried out which is a preprocess for a complete smart drawing of the electrical installation project. Artificial neural networks were used to identify the furnishings in the study. It is highly probable that the proposed method recognizes an unknown symbol. In cases where it cannot be predicted, the symbol is included in the system by users, thus the success of the system is continuously improved.

The improvement of the study will continue. The study prepared in Matlab environment continues to be used in .net environment.

References

[1] İ. M. Orak and B. Akgul, "Fast and simple computer aided internal electrical wiring project design & integration of calculations," *Procedia Technol.*, vol. 1, pp. 280–285, 2012.

[2] B. D. Ripley, "Pattern Recognition and Neural Networks," *Analysis*, no. 1995, p. 403, 1996.

[3] J. Pradeep, E. Srinivasan, and S. Himavathi, "An Investigation on the Performance of Hybrid Features for Feed Forward Neural Network Based English Handwritten Character Recognition System 2 Handwritten Recognition System," vol. 10, pp. 21–29, 2014.

[4] A. Jailin Reshma, J. Jenushma James, M. Kavya, and M. Saravanan, "An overview of character recognition focused on offline handwriting," *ARPN J. Eng. Appl. Sci.*, vol. 11, no. 15, pp. 9372–9378, 2016.

[5] N. Arica and F. T. Yarman-Vural, "An overview of character recognition focused on offline handwriting," *IEEE Trans. Syst.*, vol. 31, no. 2, pp. 216–233, 2001.

[6] N. Arica and F. T. Yarman-Vural, "Optical character recognition for cursive handwriting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 801–813, 2002.

[7] K. K. Çevik and E. Dandil, "Yapay Sinir Ağları İçin Net Platformunda Görsel Bir Eğitim Yazılımının Geliştirilmesi," pp. 19–28, 2012.

[8] Z. Demir, S. Çikoğlu, F. Temurtaş, and N. Yumuşak, "Bir Yapay Sinir Ağı Modeli ile İmza Tanıma," *S.A.U Fen Bilim. Enstitüsü Derg.*, vol. 7, no. 2, pp. 44–48, 2003.

[9] T. Dash, T. Nayak, and S. Chattopadhyay, "Offline Handwritten Signature Verification using Associative Memory Net," *Int. J. Adv.*, vol. 1, no. 4, pp. 370–374, 2012.

[10] E. B. Ceyhan, Ş. Sadıroğlu, and E. Akyıl, "Parmak izi öznitelik vektörleri kullanılarak YSA tabanlı cinsiyet sınıflandırma," *J. Fac. Eng. Archit. Gazi Univ.*, vol. 29, no. 1, pp. 201–207, 2014.

[11] I. Kahraman, "An Approach for Road Network Detection from Satellite Images Using Neural Networks," no. 1, pp. 15–18, 2017.

[12] I. Kahraman, M. K. Turan, and I. R. Karas, "Road Detection from High Satellite Images Using Neural Networks," *Int. J. Model. Optim.*, vol. 5, no. 4, pp. 304–307, 2015.

[13] L. Abraham and M. Sasikumar, "Automatic building extraction from satellite images using artificial neural networks," in *Procedia Engineering*, 2012, vol. 50, pp. 893–903.

[14] J. De Vries, "Object Recognition: A Shape-Based Approach using Artificial Neural Networks," 2006.

[15] E. Tamer and B. Çizmeci, "Plaka Tanıma Sistemi için Farklı bir Yaklaşım," in *17. IEEE Sinyal İşleme ve İletişim Uygulamaları Konferansı*.

Exposure Analysis of a Human Body due to Underground Power Cables and Magnetic Field Mitigation

K. ATES[1], H. F. CARLAK[2] and S. OZEN[3]

[1]Akdeniz University, Antalya/Turkey, ateskayhan@yahoo.com
[2]Akdeniz University, Antalya/Turkey, fezacarlak@akdeniz.edu.tr
[3]Akdeniz University, Antalya/Turkey, sukruozen@akdeniz.edu.tr

Abstract **- In this paper, the magnetic field exposure analysis of a human body due to underground power cables is carried out as the simulation study. Simulations are implemented through a finite element method (FEM). Human body is modeled as a two layered cylinder. Upper layer of the cylinder is selected as an average skin tissue and the inner layer is formed as an average muscle tissue with realistic values at an extremely low frequency (ELF) region. Shielding is applied as the mitigation technique and aluminum (Al) is used as a shielding material. The thickness of 1, 2, 3 and 4 mm shielding materials are applied, respectively.**

Evaluations are implemented with respect to the magnetic flux density and the induced current density. 4 mm shielding which is the thickest material used simulations shows the best shielding results to mitigate the magnetic flux density and induced current density. The worst shielding is obtained for the 1 mm thickness of the material, as expectedly. Furthermore, as the distance of the source increases, both magnetic flux density and induced current density decrease. In other words, these parameters depend on the distance between a source and observation points. Different induced current density values of skin and muscle are observed due to the different electrical properties of tissue. Assessments have been done according to exposure limits published by the well-known organizations.

Keywords **- Magnetic field mitigation, underground power cables, finite element method, magnetic field exposure.**

I. INTRODUCTION

ELECTROMAGNETIC fields at the extremely low frequency (ELF) region, which occur from high voltage elements (transmission lines, bus bars, circuit breakers, power substations, transformer centers, etc.), seriously threaten both occupational and public safety [1, 2]. This rising danger has been thoroughly investigated in recent years. Studies have shown that the magnetic field exposure at ELF region causes negative effects such as nervousness, immune system weakening, exhaustion and headache [3]. Also, there is a nonlinear link between long term exposure of these fields and childhood leukemia [4].

International Agency for Research on Cancer (IARC) specifies these fields as 2D risk group while the National Institute of Environmental Health Sciences (NIEHS) describes 0.3-0.4 μT magnetic flux density value causing possible carcinogenic hazard. The limit values of magnetic flux density, which have been determined by international organizations have shown in Table 1.

Table 1: Exposure limits for magnetic flux density at an extremely low frequency region (μT) [5, 6]

Target	ICNIRP (2010)	IEEE (2002)
Occupational	1000	2710
Public	200	904

Many countries define their regulations based on the above values for the health of their citizens [7]. Government of Poland restricts the maximum magnetic flux density values as 160 μT and 48 μT for employees and civilian population, respectively. Limits of Switzerland are beneath the 1 μT. Russia applies 100 μT for occupational and 10 μT limits for the public health.

Another significant parameter of ELF based dosimetry is the induced current density. International Commission on Non-Ionizing Radiation Protection (ICNIRP) has published limit values for induced current density as 10 mA/m^2 for occupational and 2 mA/m^2 for public at the frequency range of 4 Hz to 1 kHz [8].

In this study, a shielding solution which is usually used for electromagnetic interference problems in order to mitigate the magnetic field and induced current density due to underground transmission lines on the human body is proposed through simulation software called COMSOL™ which utilizes a finite element method (FEM). The problem scenario is composed of a two dimensional geometry. Layered cylinder mimics the average human model and shielding material is chosen as an aluminum plate to cover the transmission lines. Simulations have been carried out for different shielding thicknesses and the results have been evaluated according to international standards.

II. MATERIALS AND METHODS

Biot-Savart Law states the magnetic field generation from the current element. This law which can be applied to many disciplines such as communication theory and electromagnetic

interference, is expressed as follows:

$$\vec{H} = \oint \frac{I.dl \times \vec{a}_R}{4\pi R^2} \quad (1)$$

where, $\vec{a}_R$ is a unit vector , R is a distance between the observation point and current source (m) and $I.dl$ is a differential current element.

External magnetic field induces electric field inside the human body according to Faraday's Law. Electric field formed in the human body can be formulated as:

$$E = \sqrt{E_x^2 + E_y^2 + E_z^2} \quad (2)$$

E_x, E_y, and E_z show the directional components of the electric field. Induced current density J inside the body is defined as:

$$\vec{J} = \sigma \vec{E} \quad (3)$$

where, σ is an electrical conductivity of interested tissue (S/m).

Since the wavelength of a power frequency is tremendously larger than the human dimension, the effects of the magnetic fields can be examined and displacement currents might be ignored according to quasi-static approximations [9].

Power frequency magnetic fields (PFMF) may pose serious risks for human health according to epidemiological studies [10]. Inadequate precautions for magnetic field cancellation around the transmission lines cause detrimental effects such as brain cancer, leukemia, miscarriage, distortion in endocrine system balance for a long-term exposure. Magnetic field radiation and electromagnetic dosimetry can be considered together. It should be noted that, electrical properties of tissue is a significant parameter for the non-ionizing dosimetry. The human body is modeled as the two-layered cylinder in this study. The electrical properties of the layers are given in Table 2 [11].

Table 2: Electrical properties of tissues at ELF frequency region

Tissue	Conductivity (σ, S/m)	Permittivity (ε)
Skin	0.43	1136
Muscle	0.35	177.19×10^5

Magnetic field mitigation process may be carried out as an internal and external ways [12]. Phase arrangement and electrical characteristics of the transmission line play a major role for an internal magnetic field mitigation process. Since the magnetic field of an each proper aligned phase cancels the magnetic field of other phases, a lower resultant magnetic field vector occurs. In order to mitigate the external magnetic field, an external device or shielding technique is preferred. Shielding technique can be applied using two types of material, either ferromagnetic or conductive [7]. If the conductive material will be selected as shielding material, then the eddy currents are induced on the shield and opposite magnetic fields occur to cancel the undesired magnetic field.

$$R = \frac{l}{\mu_0 \mu_r A} \quad (4)$$

where, R is a reluctance, l is a travelling pathway of the magnetic flux, μ_0 is a magnetic permeability of the air, μ is a magnetic permeability of the material, and A is a cross-section. Shielding through ferromagnetic materials causes more relative permeability which provides the lower reluctance. In this way, magnetic flux can be confined within the shielding material.

If the shielding is placed at the far field, the electric and magnetic fields are considered together. Otherwise, the electric and magnetic fields should be considered separately. If the electric field is the dominant field compared to the magnetic field, the wave impedance is calculated according to Equation 5. Conversely, if the magnetic field is the dominant compared to the electric field, the wave impedance is defined by Equation 6.

$$Z_{W_E} = \frac{|E|}{|H|} \approx \frac{1}{\omega \varepsilon r} \quad (5)$$

$$Z_{W_E} = \frac{|E|}{|H|} \approx \omega \mu r \quad (6)$$

In above formulas, ω is an angular frequency (rad/s) and r represents the distance between the source and the shield (m). ε is a dielectric constant and μ is a magnetic permeability of the free space. Wave impedance is an important parameter to determine the shielding efficiency at the near field shielding theory.

In this study, shielding technique was investigated to mitigate the magnetic field and aluminum (Al) material was chosen as the shielding material. Electrical conductivity of the Al is 3.77×10^7 S/m and a magnetic permeability is 1.

III. Simulation Results

Current density and magnetic flux density on the human body which are induced by underground power cables were investigated using a finite element method. In finite element method (FEM), the problem model is separated into sub spaces as small as possible. Hence, different problem subspaces which have got different physical parameters and boundary conditions can be obtained. Accuracy of solution depends on boundary conditions, initial values and the number of mesh. The more number of meshes in the problem space means using a smaller mesh structure and this reduces the error rate.

Figure 1 presents the two dimensional problem space including transmission lines, cable duct, shielding material and human model. Underground transmission lines dealt with in the problem scenario consist of copper conductors. External radius of each phase is 44.5 mm and 185 mm^2 is defined as the cross section of each phase. According to the catalogue values

published by commercial companies, the maximum current capacity is defined as 503 A.

According to the problem space (Figure 1), the distance between copper conductors of an underground transmission lines and the ground level (d) is a 1.65 meter. Each span between the copper conductors (f) is a 0.1 meter. According to electrical authority restrictions in Turkey, [13] the height of the cable duct (h) is a 0.8 meter. The length of the top of the duct (c) is a 0.6 meter and bottom part (e) is a 0.4 meter. Distance between the top of the cable duct and ground level (k) is assumed as a 1 meter. Height of the human from the ground level (b) is considered as a 1.8 m and human body was modeled as a two layered cylinder using realistic parameters. Outer layer of the cylinder mimics the skin and the thickness of the skin is a 2.5 mm. Inner layer simulates the muscle tissue whose thickness was 25 cm. In this study, proposed technique was suggested through reverse-U shape shielding material which enclose the conductors. The distance from the center of the conductor to the shield is 5 cm for an each side. Simulations were implemented for 1, 2, 3 and 4 mm thickness of shielding material and the results have been evaluated along the vertical line passing through the center point of cylinder.

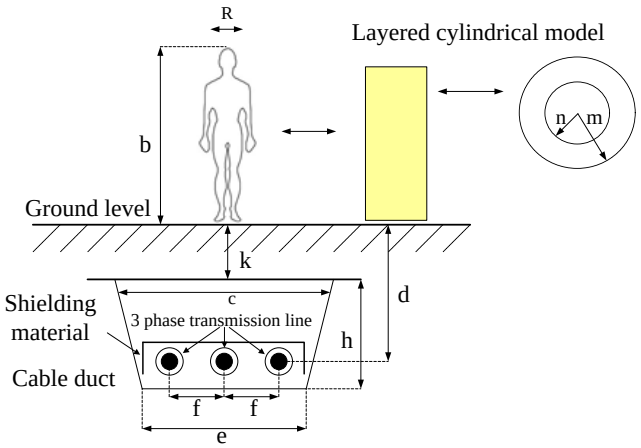


Figure 1: 2 dimensional problem scenario

Figure 2 denotes the induced current density in the body without a magnetic field mitigation. In the figure, current density norm means the amplitude of the induced current density in the human body model. Maximum current density was measured about 306.6 mA/m^2 on the closest skin layer to the conductors and 237 mA/m^2 was obtained in the muscle tissue located at the middle point of the cylinder (0.9 m above the ground level). Different induced current densities were obtained at the top and bottom tissues, because the amplitude of a magnetic field is an inversely proportional with the distance. Since the electrical conductivity of the skin is more than the muscle, higher values were calculated in the skin compared to the muscle.

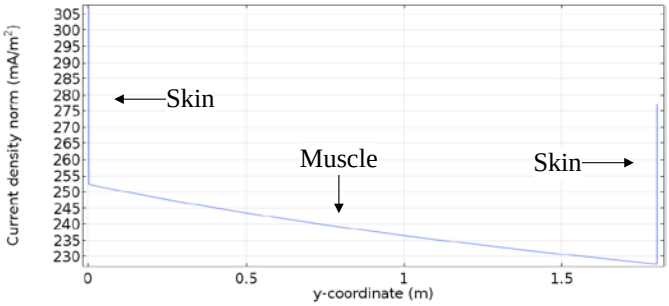


Figure 2: Induced current density in the body without shielding

Figure 3 contains the induced current densities for different thicknesses according to mitigation levels. It has been observed that the shielding efficiency was increased directly proportional with the shielding thickness. It is possible to reach the desired values which are beneath the ICNIRP limits. The most successful results were obtained with 4 mm Al shield, as it is expected.

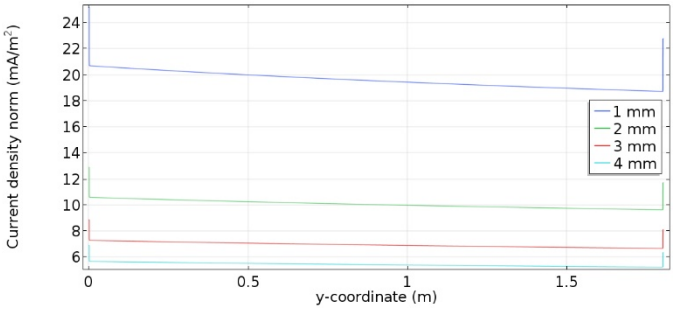


Figure 3: Induced current density in the body with Al shielding

Figure 4 shows the magnetic flux densities in the body model. Magnetic flux density was observed at the top and bottom region of the cylinder as 87.24 μT and 182.28 μT, respectively. These values are above the danger limits of some well-known international organizations but comply with the limits of ICNIRP & IEEE Organizations.

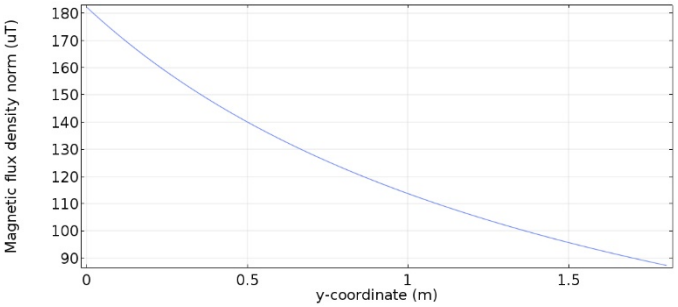


Figure 4: Magnetic flux density without shielding along the vertical axis of the body

Figure 5 depicts the results of magnetic flux density in the body depending on the shielding thicknesses. Values of the magnetic flux density at the lowest level are obtained due to the thickest shield, i.e., 4 mm Al shielding.

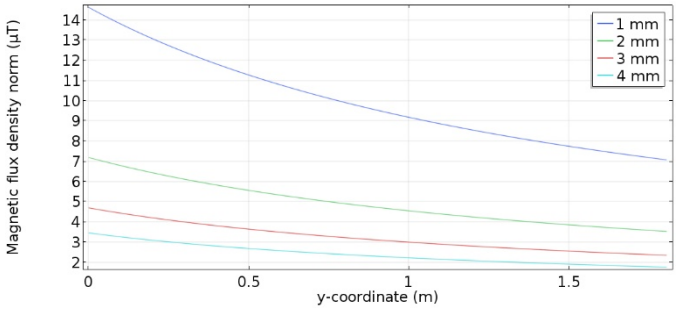


Figure 5: Magnetic flux density with Al shielding along the vertical axis of the body

IV. CONCLUSION

In this study, exposure analysis of magnetic field mitigation due to underground power cables has been investigated. Human body was represented as a double-layered cylinder. The analyzes were accomplished with commercial software, COMSOL™.

As a result of the simulations, induced current densities and magnetic flux densities due to magnetic field effects have been drastically mitigated. Magnetic flux densities in the body are ranged from 7 μT to 14.61 μT for a 1 mm thickness Al shielding material and from 3.52 μT to 7.18 μT for the 2 mm shielding thickness. According to 3 mm Al shielding, magnetic flux density was observed between 4.68 μT and 2.34 μT. The best results were obtained at 4 mm shielding in the range of 1.75 and 3.45 μT. Furthermore, different current densities were induced on each part of the tissue. Minimum induced current densities were observed in the muscle and skin tissue as 5.66 and 6.94 mA/m^2, respectively, by utilizing 4 mm Al shielding which provides the best shielding efficiency.

It has been proven in recent years that magnetic fields at the ELF region have been associated with adverse effects on a brain cancer, leukemia, miscarriage and endocrine system of the body. In order to prevent these hazards, the necessary arrangements must be performed through one of the fundamental components of EMI, source side, target side or coupling side. When the practical applications for the shielding technique are examined, it should be noted that, as the thickness increases, both the working time of the shielding material and costs increase too.

REFERENCES

[1] E. Ali and A. R. Memari, "Effects of magnetic field of power lines and household appliances on human and animals and its mitigation," in IEEE Middle East Conference on Antennas and Propagation (MECAP), pp. 1-7, 2010.

[2] M. Perić, S. S. Ilić and R. S. Aleksić, "Determination of ELF magnetic field penetrated into human body," in 7th International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology, 2007, pp. 311-314.

[3] K. Ates, H. F. Carlak and S. Ozen, "Mitigation of the magnetic field created by the underground power cables and exposure analysis of the human body," in 1st International Mediterranean Science and Engineering Congress, 2016, pp. 1-7.

[4] A. Ahlbom et al., "A pooled analysis of magnetic fields and childhood leukemia," British Journal of Cancer, vol. 83, pp. 692–698, 2000.

[5] *ICNIRP Guidelines for Limiting Exposure to Time-Varying Electric and Magnetic Fields (1 Hz- 100 kHz), Health Physics*, vol. 99, no. 6, pp. 818–836, 2010.

[6] *IEEE Standard for Safety Levels with Respect to Human Exposure to Electromagnetic Fields, 0–3 kHz*, IEEE Standard C95.6-2002, 2002.

[7] D. Bavastro, et al., "Magnetic field mitigation at power frequency: design principles and case studies," IEEE Transactions on Industry Applications, vol. 51, no. 3, pp. 2009-2016, 2015.

[8] A. Ahlbom, et al., "ICNIRP (International Commission on Non-Ionizing Radiation Protection) Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic and Electromagnetic Fields (up to 300 GHz)," *Health Phys.*, vol. 74, no. 4, pp. 494-522, 1998.

[9] W. Xi, M. A. Stuchly and O. P. Gandhi, "Induced electric currents in models of man and rodents from 60 hz magnetic fields," IEEE Transactions Biomedical Engineering, vol. 41, no. 11, pp. 1018-1023, 1994.

[10] M. A. Abd-Allah, "Interaction of ELF magnetic fields with human body organs model underneath EHV transmission lines," in 2006 IEEE PES Power Systems Conference and Exposition, pp. 1967-1970, 2006.

[11] K. Ates, H. F. Carlak, and S. Ozen, "Magnetic field exposures due to underground power cables: a simulation study," in 2nd World Congress on Electrical Engineering and Computer Systems and Science, 2016, pp. 1-7.

[12] J. C. del Pino López, P. C. Romero and P. Dular, "Parametric analysis of magnetic field mitigation shielding for underground power cables," *Renewable Energy and Power Quality Journal*, vol. 1, no. 5, pp. 519-526, 2007.

[13] "Elektrik Dağıtım Şebekeleri Enerji Kabloları Uygulama Usul ve Esasları," Turkish Electricity Distribution Authority, 2008.

[14] N. Wertheimer and E. D. Leeper, "Electrical wiring configurations and childhood cancer," American Journal of Epidemiology, vol. 109, no. 3, pp. 273-284, 1979.

[15] M. Feychting and A. Ahlbom, "Childhood leukemia and residential exposure to weak extremely low frequency magnetic fields," Environmental Health Perspectives, vol. 103, suppl. 2, pp. 59-62, 1995.

Solving The Traveling Salesman Problem Using Parallelized Artificial Bee Colony Algorithm

F. ASIL[1] and M. GOK[2]

[1] Cukurova University, Adana/Turkey, fasil@student.cu.edu.tr
[2] Cukurova University, Adana /Turkey, musgok@cu.edu.tr

Abstract - **Fast developing GPU technology increases the performance of search algorithms used to solve NP-hard problems. Travelling Salesman Problem (TSP) is a well-known NP-hard problem. In this paper, we parallelize a popular swarm algorithm, Artificial Bee Colony, to solve TSP. Proposed algorithm is tested on small scale benchmarks obtained by modifying Mandl's Swiss Road Network. Proposed implementation is tested by three experiments performed on a host PC and a GPU card. The results are compared against the results generated by the serial implementation, which is executed on the host PC. Test results for the fully connected benchmark show that the proposed parallel implementation has increased the performance of the computation up to 150 times compared to the serial implementation.**

Keywords – **Artificial Bee Colony Algorithm, GPU, Travelling Salesman Problem.**

I. INTRODUCTION

Graphics processing becomes a cheap alternative to high performance computers [1]. The ease of programming on GPU platforms contributes this popularity [2]. Nowadays, any application that performs intensive SIMD type instructions are executed on GPUs. GPUs are widely used in Computer Vision, Signal Processing, Pattern Artificial Intelligence research fields.

Swarm Algorithms are inspired by collective foraging behavior of the social animals such as ants, bees etc. Swarm intelligence is imitated to solve combinatorial search and optimization problems [3]. These algorithms can be easily adapted to run on GPU cores, since each individual of the swarm acts independently [4-6] .

A well-known NP-hard problem class is the Travelling Salesman Problem (TSP), which is often used to show the efficiency of meta-heuristic algorithms and methods [7].

In this paper, we will use the Artificial Bee Colony algorithm to solve Travelling Salesman Problem [8]. Main focus of the study is the development of the GPU implementation of the Artificial Bee Colony algorithm.

II. ARTIFICIAL BEE COLONY ALGORITHM

A bee colony's survival depends on efficient search of food resources [3, 8]. ABC algorithm is inspired by the foraging behavior of bees. There exists plenty of versions of the ABC algorithm [9]. In its simplest version the ABC algorithm consists of two types of bees. The first type are the active bees, they search the solution space, while the second type are onlookers. Onlookers join the active bees by means of a random selection procedure. The probability of selection is determined by the fitness of the active bees. The better the fitness of an active bee the more likely that onlookers join it. This strategy basically dedicates more and more search power on promising areas. There exists plenty of versions of the ABC algorithm [9]. Main steps of ABC algorithm are explained as follows:

ABC Algorithm

Initialization

Generate random scout and onlooker bees and compute the fitness of them.

Search

- *Find a new solution by modifying the current solution of the scout bee.*
- *Calculate the fitness of the solution. If the new fitness is better than the current fitness, replace the scout bee's solution with the new solution.*
- *If the number of iterations for the search expires. Go back to hive (which means go to the selection procedure)*

Selection

- *Onlookers choose to join a scout bee based on their fitness.*
- *Scout bees can be also replaced by the others if their fitness have not improved during the search.*
- *Start a new search, if the goal is not reached.*

III. TRAVELLING SALESMAN PROBLEM

Problem is basically stated as follows: Given a connected graph find the shortest path that visits all nodes and returns back to the starting node. The problem constraint is that each node must be visited only once. On a fully connected graph there exists O(n!) solutions of this problem. Optimum solution can be found by brute force i.e. testing all the node permutations. For small n this can be done but as the size

increases the number of operations become so large that it will become impractical to use brute force even with high speed computers. For example, over 120 quintillion solution permutations must be examined to solve TSP problem for 20 cities on a fully connected network.

Figure 1 shows a graph with 7 nodes, where each node represents a city. Noted that this graph is not fully connected, so some of the solutions are not feasible on this graph. Figure 2 shows two feasible solutions on this graph. ABC algorithm keeps an ordered list of nodes to represent each of these solutions. For example, the first solution can be represented as [0,1,2,3,5,6,4,0] while the latter can be represented as [2,3,1,0,6,4,5,2]. Any of the nodes in the solution list can be chosen as the starting node.

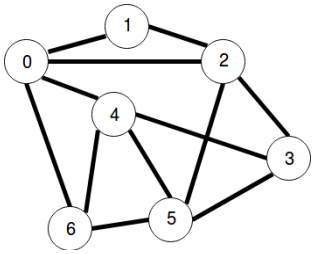


Figure 1: A graph for travelling salesman problem.

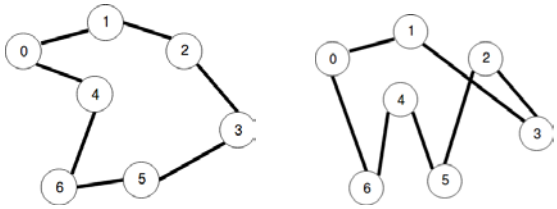


Figure 2: Two feasible solutions.

IV. PARALLEL IMPLEMENTATION OF ABC

ABC algorithm's performance is increased by using GPUs. GPUs are often used to compute independent arithmetic instructions in parallel. This method can be applied to ABC to compute the evaluation scores of all bees in parallel. Another method is running independent copies of ABC on all cores and threads. Both approaches have pros and cons. The first approach can speed up the algorithm when the evaluation function requires significant number of arithmetic operations. The second approach can help searching the solution space more effectively since many solutions can be examined at once. In this study we prefer the second approach and use GPUs to execute its own copy of the algorithm since the evaluation function for TSP is simply the total length of the solution.

The proposed parallel implementation has two parts. The host part is executed on a PC, and the kernel function is executed on the GPUs of an Nvidia card, which is connected to the host PC's main board. The host part of the program initializes the memory, launches the kernel program and orchestrates the communication between GPU cores and displays the result. Each GPU core executes a copy of the kernel function. On NVidia architecture cores are arranged as blocks, which share a common memory. In our implementation, each bee is represented by a block, all kernels in the same block checks one neighbor of that bee. By this way fast shared memory of each block is exploited. The main steps of the bee initialization and neighbor generation can be described as follows:

1. Pick a random node, set it as the current node and append it in the solution list.
2. Randomly choose a neighbor of the current node, if this node is not in the solution list, append it to the list, otherwise try another neighbor. If all neighbors are in the solution list go back to step 1.
3. Repeat step 2 until all the nodes are included in the solution list.
4. Host simply copies the most successful bee thread in each block, also keeps the record of the best bee of all blocks.
5. Go back to step 1 if the goal is not satisfied.

Our node structure implementation contains the record of the neighbors as well as the number of neighbors.

Serial implementation program is an exact copy of a kernel function of the parallel implementation.

V. RESULTS

Serial and parallel programs are written to compare the efficiency of our implementation. Tests are performed on two network models derived from Mandl's Swiss Transportation network [10]. The original network consists of 15 nodes and 21 links. The first model modifies the original network by three new edges. This model is shown if Figure 3. The second model is a fully connected network which is constructed by adding edges with random weights to Mandl's graph.

Experimental executions are performed on a host PC with an Intel® Core i7-5410U processor with speed of 2.60GHz and 8GB of memory. GPU card is Nvidia® GeForce 840M with 384 cores and 6109MB total memory. We used Visual Studio professional 2013 as a compiler with CUDA 8.0 for parallel implementations [11].

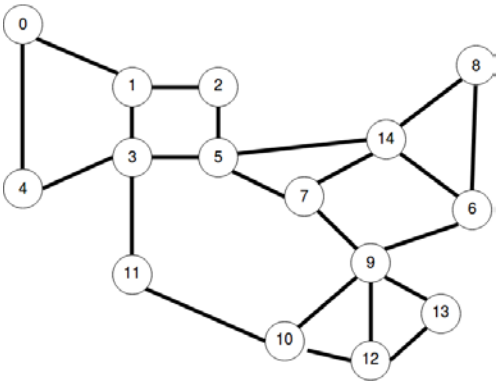


Figure 3: Modified Mandl's Network

Figure 4 shows the performance of parallel implementations on the first model. In this figure, x axis shows the number of

bees (kernels) used to solve the problem, and the y axis shows the running time in seconds. The experiments are performed for 5, 10, 20, 30, 50, 75, 100, 150, 200, 300, and 500 bees. Each experiment is repeated 5 times and their averages are displayed.

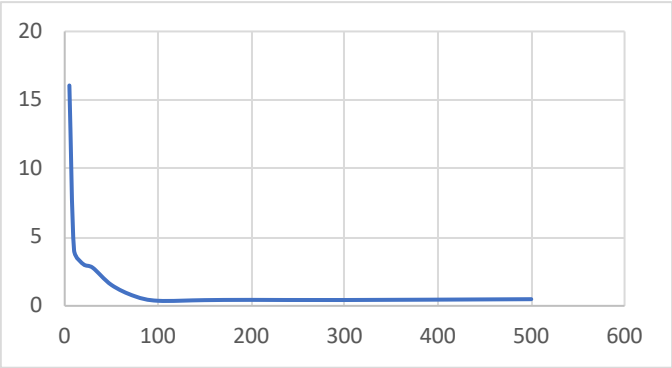


Figure 4: Execution times of parallel implementation tests.

There are only a few feasible solutions on this network, so at each test most of the execution time is spent on finding a solution. The parallel implementation speeds up until the number of bees is equal to 100. At this point the solution is found in 0.370 seconds. Increasing the number of bees more than 100 slightly decreased the speed. On the other hand, serial implementation was able to find a feasible solution in approximately 137.653 seconds with 4 bees. This means parallel implementation generates the result roughly up to 372 times faster than the serial implementation. In our experiments best timing result is found with 4 bees; increasing the number of bees slowed down the serial implementation.

To observe the performance of the parallel implementation on the fully connected graph, two types of experiments are run. In the first type of the experiments, a predetermined search score is set and we measure the time needed to reach that goal and observe the effects of increasing the number of bees.
The results for the serial and parallel experiments are shown in Figure 5. The programs were run till the cost of 50 or less had been reached. In this figure, the x axis represents the number of bees and the y axis represents the time in milliseconds. For the serial implementation, increasing number of bees was efficient at the beginning but after 40 bees, performance is degraded.
On the other hand, parallel implementation runs faster than serial implementation. There is a performance degradation after 15 bees but this decrease is not as worse as the serial implementation case. Main reason for the superiority of the parallel implementation is that the search time of each bee is not affected by others i.e. it doesn't have to wait the previous bee to work. The best result for the parallel implementation is obtained for 25 bees, at this number the solution is generated in 0,234 seconds. The best result for the serial implementation is generated by 15 bees in 35,192 seconds. This means roughly 150 times speed up is obtained on this particular program. Note that the serial implementation's performance is more sensitive to the number of bees while the parallel implementation's performance's is less.

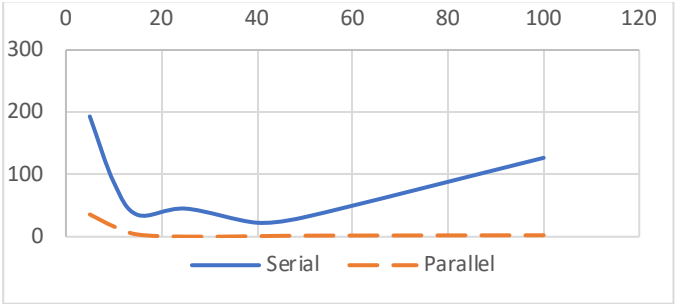


Figure 5: Comparison of serial and parallel implementations by varying number of bees.

Figure 6, shows comparison results obtained by varying the cost goal. In this figure, the x axis shows the cost goals and the y axis shows the average execution time spent to reach them. In this experiment, we set the number of bees to 15. In the previous experiment serial implementation's performance was best at this number. To test the efficiency and performance of the implementations, the goal is decreased from 60 to 48 by steps of 2. For values higher than 60, both serial and parallel implementations were able to solve the problem in one iteration. As the target cost decreases, the performance of the parallel implementation stays pretty much the same while the serial implementation's performance significantly decreases. Serial implementation reaches the target 48 in 43,369 seconds and the parallel implementation reaches this target in 1,015 second.

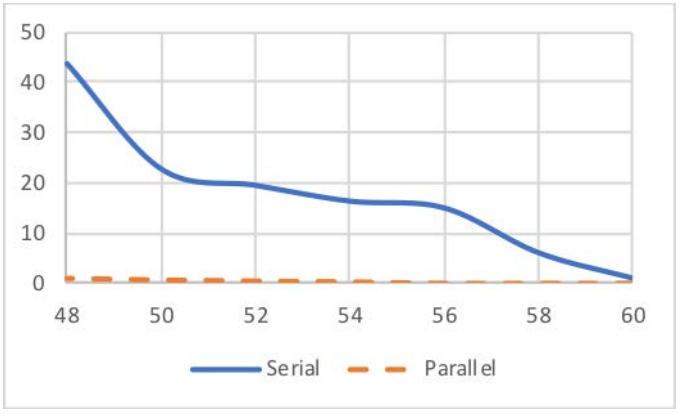


Figure 6: Comparison of serial and parallel (GPU) implementations by varying cost goal.

VI. CONCLUSION

This paper presented a TSP solution method based on ABC algorithm and its implementation on a GPU hardware. Major contributions of the study are the adaptation of ABC to the GPU implementation and the use of a new neighbor search method procedure for TSPs. Though ABC algorithm has many flavors one of the most basic implementation of the algorithm is used in the study. A more improved version of the ABC could show better results. The presented method is tested on a small-scale network. Test results for the proposed

implementation show significant improvements compared to serial tests. The future work will expand the proposed approach and search the efficiency of the proposed method on large and constrained networks.

ACKNOWLEDGMENT

Research funding was provided by the Cukurova University Research Grant FYD-2015-5454.

REFERENCES

[1] D. Patterson, *Computer Architecture: A Quantitative Approach*. Morgan Kauffman, 2017.
[2] D. Kirk, W.-M. Hwu, Programming Massively Parallel Processors: A Hands-on Approach, Nvidia, 2010
[3] J. Heaton, *Artificial Intelligence for Humans, Vol 2: Nature-Inspired Algorithms*. Heaton-Research, 2014.
[4] GH. Luo et al. "A parallel Bees Algorithm implementation on GPU." *Journal of Systems Architecture* vol. 60, no.3, pp 271-279, 2014
[5] J. Janousek, Jan and V. Snasel. "Clustering using artificial bee colony on cuda." *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pp.3803-3807, 2014.
[6] A. Delévacq et al. "Parallel ant colony optimization on graphics processing units." *Journal of Parallel and Distributed Computing* vol. 73, no. 1, pp 52-61, 2013.
[7] A. Schrijver,. "On the history of combinatorial optimization (till 1960)" *Handbooks in operations research and management science*, vol 12, pp. 1-68, 2005.
[8] D. Karaboga, "Artificial bee colony algorithm." *Scholarpedia* vol 5. No. 3 pp 6915, 20
[9] D Karaboğa, *Yapay Zeka Optimizasyon Algoritmaları*, 2014
[10] C. E. Mandl, "Evaluation and optimization of urban public transportation networks. " *European Journal of Operational Research*, vol. 5, no. 6, pp. 396-404 1980.
[11] J. Nickolls et al. "Scalable parallel programming with CUDA." In *ACM SIGGRAPH 2008 classes*, p. 16. ACM, 2008.

An Artificial Bee Colony Algorithm and Its Application to Travelling Salesman Problems: Reverse Logistics Optimisation for Accumulator Recycling Companies

O. KORKMAZ[1] and C. SEL[1]

[1]Karabuk University, Karabuk/Turkey, oguuzhaan_krkmz@hotmail.com
[1]Karabuk University, Karabuk/Turkey, cagrisel@karabuk.edu.tr

Abstract **- The reverse logistics processes, i.e. collecting the recycling materials, have a major cost for the recycling companies. Collecting the recycling materials with the shortest route, recycling companies can reduce the transportation cost. In this study, we introduce a case study on a travelling salesman problem encountered in an accumulator recycling company. We introduce a mathematical model and an artificial bee colony algorithm to reach the shortest route of the distribution network in the case. Scenario analysis for the performance evaluation of the proposed bee colony algorithm has been made on high-scale benchmark problems in the literature. (see, travelling salesman problems in OR Library, http://people.brunel.ac.uk/~mastjjb/jeb/info.html). It has been shown that the algorithm is efficient for an industrial application, it is able to reach near-optimal solutions in reasonable solution time.**

Keywords **- Accumulator recycling, Travelling salesman problem, Reverse logistics, Mathematical modelling, Artificial bee colony, A real-life case with single vehicle**

I. INTRODUCTION

The waste of the chemical accumulators is an average of 95 thousand tons per year in Turkey (please see, the data provided by "Accumulator and Recycling Industries Association", http://www.akuder.org.tr/sunum/iwes2012.pdf). An average of 57 thousand tons of lead, 11.4 thousand tons of plastic and 20.9 thousand tons of acid solution are supplied to the market per year by recycling the waste accumulators in Turkey. As an overall remark, it is necessary to collect, transport and recycle the waste accumulators.

In this work, a real case with an asymmetric traveling salesman problem is solved by using artificial bee colony algorithm from swarm based heuristic algorithms.

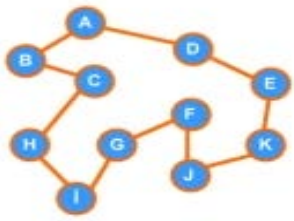
II. LITERATURE REVIEW

There are many studies on traveling salesman problems in the field. These studies are related to a number of variants of the travelling salesman problem and the solution approaches used to solve the problem. Heuristic algorithms are extensively accepted approaches to solve travelling salesman problem. Recently, Kıran et al. (2013) have used the neighborhood search algorithm to improve the initial solutions of the Artificial Bee Colony Algorithm [4]. Koçer and Akça (2014) integrate the local search algorithm, they propose an Artificial Bee Colony Algorithm improving the value of the food sources in the algorithm [2]. Venkatesh and Singh (2015) extend the travelling salesman problem. They argued that the all route would be completed by two vehicles. In their problem, only one vehicle could reach each point at the same time. Meng et al. (2016) use a neighborhood search and taboo search algorithms to improve the food sources created for initial solutions in the artificial bee colony algorithm [3]. Zhong et al. (2017) propose an algorithm that involves solving the abandoned food sources at the scout bee phase of the Artificial Bee Colony Algorithm, within the certain possibilities [5].

III. PROBLEM DEFINITION

In our problem, the collection process of the waste accumulators will be carried out by one vehicle. The vehicle starts at the recycling facility and collects the accumulators from all temporary storage points and return to the recycling facility. The problem is referred to the general form of Asymmetric Traveling Salesman Problem in the literature. The scope of the problem diagram is shown in Figure 1.

Figure 1: Symbolic Asymmetric Travelling Salesman Problem



The solution space can be calculated (N-1)! for the travelling salesman problems with the start point known and having N number of node points.

IV. ARTIFICIAL BEE COLONY ALGORITHM

Travelling salesman problems can be solved with mathematical modeling or heuristic algorithms. Mathematical programming models guarantee the optimal result instead of the heuristic algorithms. However, heuristic algorithms have some computational advantages on solving large sized real-life problems.

Heuristic methods have inspired from the nature. One of these methods, the artificial bee colony algorithm which reflects the intelligent behavior of the bees in order to find foods in real life [1]. The pseudo-code of the artificial bee colony algorithm is explained in summary in Algorithm 1.

Algorithm 1: Pseudo-Code For Artificial Bee Colony

```
1:  Initialization
2:  Evaluation
3:  iteration=1
4:  repeat
5:      Employed Bees Phase
6:      Calculate Probabilities for Onlookers
7:      Onlooker Bees Phase
8:      Scout Bees Phase
9:      Memorize the best solution so far
10:     iteration=iteration+1
11: until iteration=Maximum Iteration Number
```

V. THE CASE STUDY

In this case, an asymmetric travelling salesman problem has been considered within 17 travelling nodes. While the distances between the nodes are calculated, the real road distances are used instead of bird flight. The distances between the points are not the same for forward and backward. The recycling facility is in Ankara, Turkey. The capacity of a vehicle is assumed to be infinite.

Figure 2: Geographical Distribution of Temporary Storage Points in the Map

In this study, The Artificial Bee Colony Algorithm is coded by MATLAB software. As a stopping criterion, the limit of non-improvement is used instead of the maximum number of iterations. Parameter analysis is performed with MINITAB software and found as follows; Onlooker Bees Number: 100, Employed Bees Number: 300, Food Number: 100, Limit Value: 300 has been considered.

For comparative results, the solution value of 1807.123km is calculated by a well-known mathematical model in the literature [6]. The artificial bee colony algorithm has reached the optimal solution in every test, and the mean duration of the tests was 13.59 second.

The computational experiments are performed 5 times and mean values were taken. The features of the computer used Intel(R) Core(TM) i5-7400 CPU @ 3,00 GHz, 8gb ram, 128gb SSD.

VI. CONCLUSION

In this study, a real case study is considered, and an optimal route is determined by an artificial bee colony algorithm. In small scale problems, mathematical models provide optimal results in a very limited computational time. However, if the problem scale increases, the artificial bee colony algorithm become a successful algorithm that achieves the optimal results and overcome the resulting computational burden. The total route distance can be evaluated by increasing the number of vehicles in the next studies.

REFERENCES

[1]. D.Karaboga, B.Basturk,"A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm", *Journal Global Optimisation, Vol.39 (2007)*.

[2]. H.E.Kocer, M.R.Akca,"An Improved Artificial Bee Colony Algorithm with Local Search for Traveling Salesman Problem", *Cybernetics and Systems, Vol.45 (2014)*.

[3]. L.Meng, S.Yin, and X.Hu," A New Method Used for Traveling salesman problem Based on Discrete Artificial Bee Colony Algorithm", *Telkomnika, Vol.14* (2016).

[4]. M.S.Kıran, H.İşcan, M.Gündüz,"The analysis of discrete artificial bee colony algorithm with neighborhood operator on traveling salesman problem", *Neural Computing and Applications Vol.23 (2013)*.

[5]. Y.Zhong, J.Lin, L.Wang and H.Zhang,"Hybrid discrete artificial bee colony algorithm with threshold acceptance criterion for traveling salesman problem", *Information Sciences Vol.421 (2017)*.

[6]. Dantzig, G., Fulkerson, R., & Johnson, S. "Solution of a large-scale traveling-salesman problem", *Journal of the operations research society of America, 2(4), 393-410 (1954)*.

Selection of Facial Features using Genetic Algorithm under Different Illumination Conditions and Occlusions

O. SHARIFI[1], M. Ç. YILDIZ[2] and M. ESKANDARI[1]

[1]Toros University, Department of Computer and Software Engineering, Mersin/Turkey,
omid.sharifi@toros.edu.tr, maryam.eskandari@toros.edu.tr
[2]Milli Savunma Üniversitesi, Kara Harp Okulu Dekanlık, Ankara/Turkey, mcyildiz@kho.edu.tr

***Abstract* – This paper concentrates on a face biometric system to investigate the face biometric and problems related to face recognition under different illumination variations, pose and partial occlusion. A face recognition system is developed to recognize face images based on Principal Components Analysis (PCA). The implemented scheme applies histogram equalization and mean-and-variance normalization for image preprocessing step to reduce the effects of the illumination. In order to improve the recognition performance, we implement a feature selection method based on Genetic Algorithm (GA). The implemented method improves the recognition performance of system by selecting the optimized subset of PCA features and removing the irrelevant data. Several datasets of ORL, FERET and BANCA databases are used in order to test the robustness of the developed face recognition system.**

***Keywords* - Face Recognition, Principal Component Analysis, Genetic Algorithm, Optimization.**

I. INTRODUCTION

Face recognition, among the different biometric systems [1], has an important role in identification and verification of human beings in places with high security concern. Among several methods in face recognition, Principal Component Analysis (PCA) [2–10] is well known in case of easy implementation, reasonable performance [4,5], and effectiveness in large databases [11]. Therefore, the concentration of many researchers is on PCA improvement. Some works focus on the best eigenvectors selection [4–6, 8], to improve the performance of PCA by removing eigenvectors involving noise and also to reduce the computational time by compressing images.

On the other hand, plenty of researches emphasize on face recognition by developing robust methods to recognize faces under different illumination variations, pose, and partial occlusions [12-15]. In fact, illumination variations such as shadows, underexposure, and overexposure are considered as fundamental problems in a practical recognition system. In this respect, researchers introduce various techniques to deal with illumination changes in the past decades [16-18].

We investigate the effect of GA on a face recognition system to select the optimized subset of features in order to improve the recognition performance of the system. The system with and without optimized features are compare together in this study. In addressing the variation problems, we explore the suitability of using feature selection for recognizing face images. The effect of occlusion on face images with and without optimization of feature sets is also explored. Different experiments are performed on FERET [19], ORL [20] and BANCA [21] database in order to report the robustness of the scheme.

In fact, in this work to deal, we implement a genetic algorithm (GA) method in PCA to search the most suitable eigenvectors. By embedding GA in PCA, the optimal features can be determined. This technique enhances the performance of PCA. Since selecting eigenvectors corresponding to the largest eigenvalues always doesn't guaranty to achieve a high performance. Therefore, determining proper eigenvectors not only enhances the face recognition performance it improves the computational time too. The organization of the paper is as follow: The detail of Genetic algorithm as feature selection strategy is represented in Section II while Section III presents the explanation of face recognition system implemented in this paper. Section IV indicates experimental results and finally Section V concludes the paper.

II. GENETIC ALGORITHM

To extract the proper subset of original PCA features, GA is used in this work. According to [22], GA is a stochastic algorithm that provides an efficient method to find globally the optimal solution in large space. Generally, GA considers a population of solution which is initialized randomly in the start point. The solutions are evolved to achieve next generation of population with high quality solutions. Mostly, constant number of individuals in the population *n* is evaluated by fitness function after applying GA processes. The individuals are represented as binary code named chromosome in the population to apply most important GA operators namely selection, crossover, and mutation. The different types of crossover operator are used to generate new individuals to obtain suitable solutions, and mutation operator is used to achieve different diversity of solutions and avoid premature

convergence by keeping away from getting trapped into local maximum or local minimum.

GA attempts to remove irrelevant features and it is often performed after feature extraction. Therefore, features are extracted from the face images, and then an optimal subset of these features is selected by a feature selection method.

III. FACE RECOGNITION SYSTEM USING PCA AND GA

In the past few years, one of the most attractive areas for biometric schemes was face recognition. Mainly, the implemented method consists of four stages as shown in Figure 1. Image preprocessing is performed on face images using histogram equalization and mean variance normalization. Feature extraction module extracts a set of significant information from the raw biometric trait using PCA.

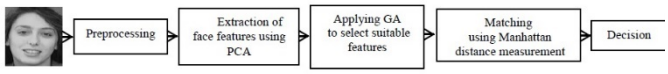


Figure 1: Structure of face recognition system.

In step three, GA algorithm was embedded into the system to extract the suitable features. Finally, the last step computes the distance scores between training and test face feature vectors using Manhattan distance measurement. Nearest Neighbor Classifier is employed to classify the individuals. The experimental results demonstrated in Section IV show that the implemented system using GA achieves an improved recognition accuracy compared to the system with only PCA.

IV. EXPERIMENTAL RESULTS

As mentioned before, in this paper PCA is used as a feature extraction method to extract the facial features. Genetic Algorithm selects an optimal subset of eigen-features (eigenvectors) to enhance the performance of face recognition. Generally, the importance of eigenvectors usually is specified based on the order of the eigenvalues, but this order is not always appropriate to define the data. In this work, binary encoding scheme is applied to represent the presence or absence of a specific eigenvector where 1 means presence of the eigenvector and 0 means absence of the eigenvector.

The performance of implemented face recognition system is experimented on several subsets of three face databases namely ORL, FERET and BANCA. The first dataset consists of ORL face database. In ORL face dataset, 10 different frontal face images for 40 different subjects are available. In our face recognition system, all 40 subjects are considered; we assigned randomly 5 images per subject for training and the rest for testing. The parameters of GA applied in this study are set and shown in Table 1. The fitness function is considered as recognition rate and the stopping condition is set to maximum number of iteration, or obtaining the highest recognition rate.

Table 1: Genetic Algorithm Parameters.

Parameter	Value
Population Size	10-15-20-25-30
Number of Iteration	80
Crossover Rate	0.95
Mutation Probability	0.05

The first set of experiments is carried out on all subjects of ORL face database as shown in Table 2. The accuracy achieved by PCA algorithm is reported based on the selection of maximum number of nonzero eigenvectors and GA-based eigenvector selection.

Table 2: Recognition Performance of ORL Face Database.

Method	Number of Selected Features	Selected Population Size	Maximum Recognition Rate
PCA	199	N/A	84.50
PCA - GA	98	20	87.00

As it is seen, the accuracy achieved by PCA algorithm with selection of maximum number of nonzero eigenvectors is 84.50%. However, a better recognition rate is achieved using PCA-GA feature selection as 87.00% with 98 eigenvectors in population size of 20. Table 3 shows the performance accuracy obtained in different population size.

Table 3: Recognition Performance of ORL Face Database in Different Population Size.

Population Size	Number of Selected Features	Maximum Recognition Rate
10	97	85.50
15	101	86.50
20	98	87.00
25	115	85.50
30	103	86.00

On the other hand, a dataset of ORL face database using 25 individuals is selected; all the selected subjects in the dataset include different facial expressions. Tables 4 and 5 report the performance accuracy, population size and number of features selected in each population.

Table 4: Recognition Performance of ORL Face Dataset in the Presence of Facial Expression and Occlusion.

Method	Number of Selected Features	Selected Population Size	Maximum Recognition Rate
PCA	124	N/A	88.00
PCA - GA	68	20	88.00

As it can be seen the best recognition performance is achieved as 88.00% in both cases with and without feature selection. We can see that when facial expressions exist there is not any improvement in recognition performance. However we can claim that the same recognition performance is obtained with small number of eigenvectors as 68 that leads to decreasing computational time.

Table 5: Recognition Performance of ORL Face Database for Different Population Size in the Presence of Facial Expression and Occlusion.

Population Size	Number of Selected Features	Maximum Recognition Rate
10	62	88.00
15	61	87.20
20	68	88.00
25	75	87.20
30	64	88.00

Another subset of ORL face database with 15 individuals is selected with different pose variations. The result of recognition performance with and without applying GA to select the optimized eigenvectors is reported in Tables 6 and the recognition performance of different population size with variation of pose is reported in Table 7.

Table 6: Recognition Performance of ORL Face Dataset in the Presence of Pose Variation.

Method	Number of Selected Features	Selected Population Size	Maximum Recognition Rate
PCA	74	N/A	89.333
PCA - GA	29	20	90.667

Maximum recognition rate without feature selection is 89.333% while with selecting the optimized subset of features we improved the recognition rate to 90.667%. We can see easily from Table 7 that this recognition rate is obtained with population size 15, 20 and 30. However, minimum number of features is selected in population size 20 that lead to decreasing the computational time and memory.

Table 7: Recognition Performance of ORL Face Database for Different Population Size in the Presence of Pose Variation.

Population Size	Number of Selected Features	Maximum Recognition Rate
10	35	88.00
15	38	90.667
20	29	90.667
25	37	89.333
30	44	90.667

A subset of FERET database with 100 individuals and 4 samples per subject is selected to perform the experiments. In the work, 2 samples are considered as train and the rest 2 for testing. The selected dataset includes all kinds of variations and occlusions. Tables 8 and 9 summarize the recognition performance, population size, and total number of selected features for FERET database.

The recognition performance of PCA feature extractor when there is not any selection and with maximum number of nonzero eigenvectors is 88.00% as reported in Table 8. On the other hand, the best recognition performance is achieved with population size 30 and selection of 95 eigenvectors out of 200 as shown in Tables 8 and 9.

It is clearly seen from Table 9 that the recognition performance is improved in all population size compared to regular PCA feature extraction method.

Table 8: Recognition Performance of FERET Face Dataset.

Method	Number of Selected Features	Selected Population Size	Maximum Recognition Rate
PCA	199	N/A	88.00
PCA - GA	95	30	91.25

Table 9: Recognition Performance of FERET Face Dataset for Different Population Size.

Population Size	Number of Selected Features	Maximum Recognition Rate
10	117	88.75
15	106	90.00
20	106	88.75
25	116	90.00
30	95	91.25

Finally, the last set of experiments considers the extracted face images from BANCA video sequences. In total, 40 subjects with 10 samples per individual are selected, 5 samples randomly are considered as training and the rest 5 as testing. The recognition performance of PCA is 97.00% while the feature selection improved the recognition performance to 100% as it is reported in Tables 10 and 11. The best recognition performance for all population size is 100 while population size of 25 selected the minimum number of features.

Table 10: Recognition Performance of BANCA Face Dataset.

Method	Number of Selected Features	Selected Population Size	Maximum Recognition Rate
PCA	199	N/A	97.00
PCA - GA	109	25	100.00

Table 11: Recognition Performance of BANCA Face Dataset for Different Population Size.

Population Size	Number of Selected Features	Maximum Recognition Rate
10	116	100.00
15	112	100.00
20	114	100.00
25	109	100.00
30	110	100.00

The PCA+GA method presented in this study is compared with only PCA method using ROC (Receiver Operator Characteristic) analysis in Figure 2 for FERET dataset.

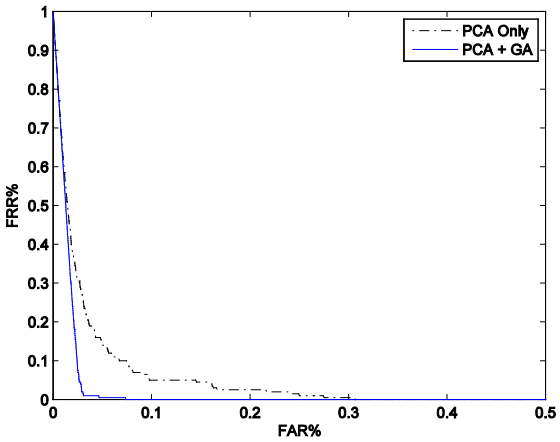


Figure 2: ROC Curves of PCA and PCA+GA for FERET Dataset.

V. CONCLUSION

In this paper, the effect of GA as a feature selection method to select the optimized sub set of features is investigated. In general, we explored face biometrics and the problem related to face recognition under different illumination variations, pose and partial occlusion. In order to improve the recognition performance, a feature selection method is used based on GA. By using GA, the optimized sub set of PCA features are selected, the optimized features try to remove irrelevant data and therefore the scheme improves the recognition performance. Analysis of the experimental results show that the feature selection strategy can improve the recognition performance by removing the eigenvectors containing noise since the selection of eigenvectors depends on the database and the available conditions and noise factors in the database. In fact, GA by selecting the optimized set of features under different factors such as pose, illumination, expression and occlusion caused to improve the recognition performance or at least we can claim the same recognition performance is obtained with small number of features compared to the system with involving all features.

REFERENCES

[1] MRTD, Machine Readable Travel Documents (MRTD). Available from: ⟨http://www.icao.int/mrtd/overview/overview.cfm⟩.

[2] M. Turk, A random walk through eigenspace, IEICE Trans. Inf. Syst. E Ser. D 84 (12) (2001) 586–1595.

[3] K. Kim, K. Jung, H. Kim, Face recognition using kernel principal component analysis, IEEE Signal Process. Lett. 9 (2) (2002) 40–42.

[4] H. Moon, P. Phillips, Computational and performance aspects of PCA-based face-recognition algorithms, PERCEPTION-LONDON 30 (3) (2001) 303–322.

[5] R. Tjahyadi, W. Liu, S. Venkatesh, Automatic parameter selection for eigen- faces, in: Proceeding of the 6th International Conference on Optimization Techniques and Applications (ICOTA), Citeseer, 2004, pp. 9–11.

[6] H. Meng, X. Ke, Further research on principal component analysis method of face recognition, in: IEEE International Conference on Mechatronics and Automation (ICMA), 2008, pp. 421–425.

[7] E. Kokiopoulou, Y. Saad, PCA without eigenvalue calculations: a case study on face recognition, in: SIAM Data Mining Conference, Newport, CA, 2005.

[8] S. Gupta, O. Sahoo, A. Goel, R. Gupta, A new optimized approach to face recognition using eigenfaces, Global J. Comput. Sci. Technol. 10 (ver 1.0) (2010) 15–17.

[9] V. Radha, M. Pushpalatha, Comparison of PCA based and 2DPCA based face recognition systems, Int. J. Eng. Sci. Technol. 2 (12) (2010) 7177–7182.

[10] Y. Zeng, D. Feng, L. Xiong, An algorithm of face recognition based on the variation of 2DPCA, J. Comput. Inf. Syst. 7 (1) (2011) 303–310.

[11] W. Zhao, R. Chellappa, Face Processing: Advanced Modeling and Methods, Inc.
Academic Press, 2006.

[12] A.M. Martinez, "Recognizing Imprecisely Localized, Partially Occluded and Expression Variant Faces from a Single Sample per Class". IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 24. (2002) 748-763, 2002.

[13] H. J. Oh, K.M Lee, S. U Lee, "Occlusion invariant face recognition using selective LNMF basis images". ACCV (Asian Conference on Computer Vision) 2006, Hyderabad, India, 120-129, January 2006.

[14] M. Ç. Yildiz, O. Sharifi, M. Eskandari , "Log-Gabor Transforms and Score Fusion to Overcome Variations in Appearance for Face Recognition"– Proceedings of International Conference on Computer Vision and Graphics, ICCVG 2016, Warsaw, Poland, September 19-21, 2016. Springer-Verlag-London, Lecture Notes in Computer Science.

[15] S. Harin, J. Sabah, "Face Recognition in the Presence of Expression and/or Illumination Variation". Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05) 144-148, 2005.

[16] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition," Pattern Recognit., vol. 46, no. 6, pp. 1691–1699, 2013.

[17] L. Qingshan, Y. Wang, L. Hanqing, M. Songde, "Occlusion Robust Face Recognition with Dynamic Similarity Features". 18th International Conference on Pattern Recognition, 2006. ICPR 2006 Volume 3, 0-0 0, 544 – 547, 2006.

[18] O. Sharifi, M. Eskandari, "Optimal Face-Iris Multimodal Fusion Scheme", Symmetry 2016, 8(6), 48; doi:10.3390/sym8060048.

[19] P.J. Philips, H. Wechsler, J. Huang, P. Rauss "The FERET database and evaluation procedure for face recognition algorithms, Image and Vision Computing, Vol.16, pp.295–306, 1998.

[20] AT&T Laboratories Cambridge, The ORL Database of Faces: http://www.cam-orl.co.uk/facedatabase.html.

[21] E. Bailly-B, S. Bengio, F. Bimbot, M. Hamouz, J., Kittler, J., Mariethoz, J., Matas, K., Messer, V., Popovici, F. Poree, B. Ruiz and J.P. Thiran, "The BANCA database and evaluation protocol", In Audio and Video-Based Biometric Person Authentication: Proc. 4th Int. Conf., AVBPA2003, LNCS 2688, pp 625–638, Germany, 2003.

[22] S. Sivanandam, S.N. Deepa, Introduction to Genetic Algorithms, Springer Verlag, 2007.

Constrained Optimization Problems Solution with Salp Swarm Algorithm and Ant Lion Optimization

P. ERDOGMUS[1], M. GOCER[1], E. DUDAK[1], N. OZKAN[1]

[1] Duzce University, Duzce/Turkey, pakizeerdogmus@duzce.edu.tr
[1]Duzce University, Duzce/Turkey, muhammedgocer20@gmail.com,
[1] Duzce University, Duzce/Turkey, ebrududak@hotmail.com
[1] Duzce University, Duzce/Turkey, nurcnozkn3@gmail.com

***Abstract* – In this study, two recent nature-inspired optimization algorithms Salp Swarm Algorithm (SSA) and Ant Lion Optimization (ALO) were introduced and a comparison with two algorithms was realized. For these comparisons, constrained optimization test problems were solved with these algorithms.**

The effects of the penalty coefficients on the solution of constrained problems were also studied. Results were submitted.

***Keywords* – Nature-inspired optimization, Salp Swarm Algorithm, Ant-Lion Optimization, Constrained optimization test problems.**

I. INTRODUCTION

The technological improvements are bringing some problems, such as energy issues and environmental pollution. With decreasing natural-resources, we must satisfy the increasing computing needs. Especially with Industry 4.0, manufacturing technologies include cyber-physical systems, the Internet of things, cloud computing and cognitive computing require intensive energy use. For the sustainable development, all the resources must be used as optimally. So, optimization is the first aim of the productions now. Devices using minimum energy and algorithms running fastest are some of the aims for the future technologies. So, optimization algorithms are the tools for realizing these aims. In recent years, optimization algorithms inspired from nature are popular.

ALO algorithm is an algorithm inspired by nature and mimics the hunting mechanisms of ant lion animals in nature. This algorithm was developed by Seyedali Mirjalili in 2015 [1]. ALO simulates hunting behavior of Ant-lion in five basic steps, such as the random walking of ants, setting traps, keeping ants in traps, catching prey and rebuilding traps. The ALO algorithm has found optimal results in the majority of classical engineering problems and has shown that it is advantageous in solving constrained problems. The ALO algorithm is preferred and recommended in order to improve the design, to achieve efficiency, to increase production, to reduce costs and to increase performance [1]. This algorithm provides a solution for how to obtain the fastest and shortest throughput. Optimal distributed generation(DG) allocation of distribution systems, radiation heat transfer, Fuzzy C-Means-ALO based hybrid technique, multi-purpose optimization and scheduling of micro-micro grids consisting of renewable energies, PID controller-controlled cruise control system, antenna array synthesis, condenser, wind integrated systems, and generator maintenance planning have been used in projects and achieved successful results.

For example, to reduce driver fatigue in cars and ensure traffic safety, the PID controller and the Bode ideal transfer function are designed and formulated as an optimization problem. The gain parameters of the designed PID controller are adjusted by the ALO algorithm, so that the proposed compensation system shows the ideal reference model of the bode [2]. This algorithm was also involved in the energy field, and the optimum VARs (Volt-ampere Reactive) capacity and Fixed Capacitors (FCs) positions were used to determine the ALO technique in order to determine the loss reduction operation of the capacitor placement based on total cost and power [3]. In addition, the ALO has been used to solve the practical hydrothermal energy production scheduling (HTPGS) problem with wind integration [4]. As a result, the applicability of the ALO algorithm has been shown in these studies. Generator Maintenance Scheduling (GMS) is an application example of ALO as a new optimization approach. In the deterministic reliability model, making preventive maintenance planning is the basic building block and the focal point of the model. It has been observed that ALO is an exceedingly promising algorithm for solving the problem of adjusting maintenance periods of power systems according to the results obtained from the studies [5]. A new image segmentation approach based on FCM and ALO was proposed on the image segmentation area [6]. An example is a radiation heat transfer problem in the electric oven. The ALO has introduced simulation and positioning of transaction parameters using an intuitive approach called modified. The measurements and results obtained in the research show that the proposed methodology is positioning the system efficiently [7]. Another example is the radial distribution networks distributed generation (DG) units, and the proposed optimization technique to determine the best size is applied to ALO. The numerical results of the proposed algorithm have been compared with those of other optimization techniques to verify the technique superiority and performance [8]. Finally, Virtual power Plant (VPP) recommends a sustainable simulation method for managing energy resources from the point of view of virtual power players working in the Smart Grid. In addition, the proposed method plans to use a suitable multi-objective frame in a suitable micro grid, and ALO algorithm is utilized [9].

The ALO algorithm also finds optimal designs for the majority of conventional engineering problems; it shows that this algorithm is advantageous in solving various constrained

problems with different search areas. This algorithm has the ability to find the most appropriate set of data.

The Salp Swarm Algorithm has been modeled the behaviors of Salp swarms that lives in the ocean, and also for ants, bears, and other living creatures, has been tried to solve the optimization problems with this algorithm. In this model, the behaviors of the leader and his followers are examined, and the behaviors of the actions of the followers and the leaders are formulated and simulated.

SSA has been tested on mathematical functions and the solution of high engineering problems such as aircraft wing design and sea propellers has been studied. According to the obtained results, SSA achieves more optimal convergence than the other algorithms. [10]

The SSA algorithm has been actively used in scientific studies for a few months. One of the areas where SSA is used is the Proton Exchange Membrane Fuel Cell (PEMFC) model. Studies on the proton exchange membrane fuel are based on the application of the SSA to find the best values of the unknown variables of the PEMFC model. During the operation, the TSD value has been reduced to the minimum to ensure compliance of the constraints with the default voltage points. The two most proton exchange membrane fuel cells, NedStack PS6 and BCS 500-W Proton Exchange Membrane has been subjected to a series of tests in accordance with the proposed working procedures. [11]

Another area where the SSA algorithm is used is chemical compound activities. This scientific work was based on the estimation of the activities of chemical compounds taking into account the herd behavior of the Salp algorithm. The Salp optimization algorithm is implemented using chemical identifiers at three levels (low, mixed and high). The nearest neighbors are defined using the fitness function of the SSA to select a minimum number of features and ensure high classification accuracy. SSA outperforms with the highest classification ability with the least number of selected features [12]

This paper is organized as follows. Section 2 presents the general constrained optimization problem solution techniques and the Nature-Inspired algorithms. Section 3 presents the solution of constrained optimization problems with Nature-Inspired Algorithms. Finally, Section 4 concludes the study.

II. NATURE-INSPIRED OPTIMIZATION ALGORITHMS

In this study, it has been studied with two different optimization algorithms inspired from nature. So, both of them are based on the similar idea. Both algorithms are originally improved for non-linear functions. So, the basic concepts are similar.

A. Constrained Optimization

A general representation of a constrained optimization problem is given below:

Design Variables: $x_1, x_2, \ldots x_d$

Objective Function: Fmin= $f(x_1, x_2, \ldots x_d)$

Constraints: $C_i = g_i(x_1, x_2, \ldots x_d) \leq 0$, i=1,2,…n

Some definitions used to solve the optimization problems with these algorithms are as follows:

N:The number of agent(particle)

d: Dimension of the non-linear function

M: Maximum iteration

P:Penalty Function

F: Fitness function

Fmin: Object function

C: Constrained Function

But as it is known, Nature-inspired optimization algorithms run based on the unconstrained optimization problem. So, the constraints are accepted as penalty function which is defined as given in Equation 1.

$$P = \sum_{i=1}^{n} K_i (\max(0, g_i(x_1, x_2, \ldots x_n)))^2 \quad (1)$$

K: Penalty coefficient for each constraint bigger than zero

If i*th* constraint satisfies the smaller than zero condition, then the maximum value will be zero. Or else P will be a positive number. This means that this solution doesn't fit for constraints. The fitness function is defined as given Equaiton 2.

$$F = Fmin + P \quad (2)$$

So the aim of the optimization algorithm is to minimize the fitness function. The optimum solution both satisfies the constraints and minimizes the objective function.

In this study, Ant-Lion Optimization and Salp Swarm Algorithm have been selected to optimize the constrained optimization test problems taken from [13].

According to our knowledge, these problems were not tested up to now, except from Welded Beam Design Problem, Pressure Vessel Design Problem, Tension-Compression String Problem given in the same page. As an example, one of the test problems (G13) is given in Equation 3.

$$\begin{aligned} &\min f(x) = e^{x_1 x_2 x_3 x_4 x_5} \\ &h_1(x) = x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 - 10 = 0 \\ &h_2(x) = x_2 x_3 - 5 x_4 x_5 = 0 \\ &h_3(x) = x_1^3 + x_2^3 + 1 = 0 \end{aligned} \quad (3)$$

B. Ant-Lion Optimization

ALO is developed by Mirjalili especially for the solution of continuous optimization problems in 2015. Antlions are accepted as insects. Their preys are ants. They pass most of their lifes as larvae. ALO is inspired from the hunting behaviour of antlions. After they build cone-shaped pits, they waits ants to fall this trap. After ant got trapped this pit, the ant lion catchs and hunts it. This situation is simulated in the algorithm, if the fitness of the ant is better than Ant-Lion, the ant is killed. In the

algorithm two position matrix is used for algorithm. One of the position matrix is for ants and the other matrix is for ant-lions. If the number of independent variable is d and if the number of agents is n, then the size of matrix is nxd.

The pseudocode of antlion is given in Figure 1.

```
Initialize the n ants' and Antlions' positions
Calculate the fitness values of them
Find elite antlion
While stopping conditions not satisfied
    for each ant
        Select an antlion using roulette wheel
        Update the position
    EndFor
    Calculate the fitness of all ants
    Replace an antlion with its corresponding ant it if becomes fitter
    Update elite if an antlion becomes fitter than the elite
EndWhile
Return Elite
```

Figure 1: The pseudocode of ALO

C. Salp Swarm Algorithm

SSA is one of the recent algorithms inspired from the Salp Swarms. SSA has only one parameter to adapt. So it is quite easy to implement like PSO.

Salps are interesting creatures. They are transparent like jellyfish. But their systems are more complex. They have complex nervous and digestive systems with a brain, heart, and intestines. SSA simulates the salps foraging behaviors. Since the algorithms is quite new, there is a few study using this algorithm. One of the salp in the swarm is selected as leader, the others are follower. So, there are two position formulas in the algorithm. The leader position is given in the equation 4.

$$x_j^1 = \begin{cases} F_j + c_1((u_j - l_j)c_2 + l_j), c_3 \geq 0 \\ F_j - c_1((u_j - l_j)c_2 + l_j), c_3 < 0 \end{cases} \quad (4)$$

In this equation x_j^1 shows the position of the leader salp, F_j is the position of jth the food source. l and u show lower and upper bounds respectively. c_2 and c_3 are random numbers uniformly generated. So, we can assume c_1 is the only parameter to adapt for each iteration adapting the exploration and exploitation ability in the algorithm. The formula of C_1 is as given in equation 5.

$$c_1 = 2e^{-(\frac{4l}{L})^2} \quad (5)$$

The pseudocode of the SSA is given in Figure 2.

```
Create initial positions for n salp considering lb,ub
While termination criteria not satisfied
    Calculate the fitness of each salp
    Find best optimal value as F
    Update c1
    For each salps do
        if salp is leader
            Update leader salp position
        else
            Update the current position of follower salp
        end
    end
  check all the salps for lb,ub
end
```

Figure 2: The pseudocode of SSA

III. SOLUTION OF CONSTRAINED OPTIMIZATION PROBLEMS

In this study, the effect of the penalty coefficient on finding the optimum value for a given problem was studied. With this aim, thirteen test problems taken from [13] were used to test the performances of the algorithms on constrained optimization problems. These problems were solved both ALO and SSA. The parameters used for the simulations were given in Table 1.

Table 1: Parameters used in the algorithms

Parameters used in the simulations	**ALO**	**SSA**
Number of agent	30	30
Maximum iteration	1000	1000

Penalty coefficients were selected as 100, 1000 and 10000 respectively. Since the first aim is to compare the optimum solutions for different penalty coefficients, solution time is not considered.

Each simulation was repeated 30 times. The best fitness, mean and the standart deviation of the simulations were recorded for each problem. These results were given in the Table 2, 3 and 4.

Table 2: The optimum values of test problems found with ALO and SSA for K=100

Function name		*Real Optimum f(x*)*	ALO	SSA
G1	F Best	-15	**11.6936**	6,72E-06
	Mean	-	11.8640696	5,2533E-07
	Std. Dev	-	0.161879	1,5883E-06
G2	F Best	0.803619	**0.003453**	3,040008
	Mean	-	4.151323	3,075078
	Std. Dev	-	2.635856	3,250941
G3	F Best	1	1.0236E+53	**0,999956**
	Mean	-	7.2589E+52	0,999994
	Std. Dev	-	1.611E+52	1,545E-05
G4	F Best	-30665.539	**-31690.3**	-33152,3
	Mean	-	-31690.3	-33137,3
	Std. Dev	-	1.1458E-06	12,657493
G5	F Best	5126.4981	354.2134	**6114,202**
	Mean	-	989.6317	17371,72
	Std. Dev	-	206.0754	9947,9112
G6	F Best	-6961.814	90827.61	**-7864,01**
	Mean	-	90827.61	-7864,01
	Std. Dev	-	7.4045E-11	1,907E-05
G7	F Best	24.306209	25.00396	**24,5658**
	Mean	-	28.58685	27,85416
	Std. Dev	-	2.605791	27,56754
G8	F Best	0.095825	**0.09583**	**0,09583**
	Mean	-	0.09583	0,09583
	Std. Dev	-	7.2835E-14	2,115E-14
	F Best	680.63005	**680.6602**	680,6904

G9	Mean	-	681.5056	681,436
	Std. Dev	-	0.50593	0,5573231
G10	F Best	7049.3307	2236.808	**4179,034**
	Mean	-	3026.249	2614,702
	Std. Dev	-	1235.49	650,56263
G11	F Best	0.75	0.7474	**0,7499**
	Mean	-	0.7474	0,750667
	Std. Dev	-	2.0719E-14	0,0019022
G12	F Best	1	**1**	**1**
	Mean	-	0.99779	1
	Std. Dev	-	0.003413	7,5034E-16
G13	F Best	0.0539498	0.257797	**0,05621**
	Mean	-	0.946218	-0,23639
	Std. Dev	-	0.478388	1,0853603

Table 3: The optimum values of test problems found with ALO and SSA for K=1000

Function name		*Real Optimum f(x*)*	ALO	SSA
G1	F Best	-15	-12.2337	**-10,9241**
	Mean	-	-7.48412	1530,472
	Std. Dev	-	1.597886	8281,059
G2	F Best	0.803619	**0.41263**	0,3735
	Mean	-	0.28837	0,28126
	Std. Dev	-	0.050264	0,036262
G3	F Best	1	1.024E+33	**1**
	Mean	-	1.024E+33	0,999991
	Std. Dev	-	1.6121E+28	4,09938E-05
G4	F Best	-30665.539	-30858.7	**-30665,5**
	Mean	-	-30858.7	-30633,1
	Std. Dev	-	7.26262E-05	65,80751
G5	F Best	5126.4981	145037.7	**1149,284**
	Mean	-	51075831	1110,652
	Std. Dev	-	77669886	96,57013
G6	F Best	-6961.814	3.004948	**-7081,47**
	Mean	-	3.184171	-7081,46983
	Std. Dev	-	0.089583414	0,00012823
G7	F Best	24.306209	25.57973	**24,44631**
	Mean	-	36.08268	28,25625
	Std. Dev	-	38.28468	1,987051
G8	F Best	0.095825	**0.09583**	**0,09583**
	Mean	-	0.09583	0,09583
	Std. Dev	-	4.88352E-14	1,49501E-14
G9	F Best	680.63005	680.7698	**680,7366**
	Mean	-	681.6468	681,2823
	Std. Dev	-	0.950806	0,587583
G10	F Best	7049.3307	4592.75	**7376,214**
	Mean	-	7665.502	4145,994
	Std. Dev	-	2164.013	2256,014
G11	F Best	0.75	**0.74965**	0,750471
	Mean	-	0.74965	0,767467
	Std. Dev	-	3.31947E-11	0,031242
G12	F Best	1	**-1**	**-1**
	Mean	-	-0.99662	-1
	Std. Dev	-	0.003421	6,33272E-16
G13	F Best	0.0539498	0.103155	**0,0601**
	Mean	-	0.989941	0,24878
	Std. Dev	-	0.800345	0,79962

Table 4: The optimum values of test problems found with ALO and SSA for K=10000

Function name		*Real Optimum f(x*)*	ALO	SSA
G1	F Best	-15	**-10.5146**	-10,1244
	Mean	-	-6.97896	-7,37078
	Std. Dev	-	1.814874	1,260880108
G2	F Best	0.803619	**0.37994**	3,071016
	Mean	-	0.29847	3,586555
	Std. Dev	-	0.038354	1,645475339
G3	F Best	1	1.024E+33	**1**
	Mean	-	1.02394E+33	0,999903
	Std. Dev	-	2.61369E+29	0,000487439
G4	F Best	-30665.539	-30858.7	**-30665,5**
	Mean	-	-30858.7	-30533,4
	Std. Dev	-	7.29262E-05	161,6072
G5	F Best	5126.4981	303.4339	**1199,777**
	Mean	-	1023.702851	1121,065
	Std. Dev	-	204.3288213	113,0436696
G6	F Best	-6961.814	2.89018	**-6961,81**
	Mean	-	3.173982	-6961,81
	Std. Dev	-	0.121496	0,005548271
G7	F Best	24.306209	**25.59076**	25,41607
	Mean	-	29.49925	28,4077
	Std. Dev	-	2.758826	2,329944984
G8	F Best	0.095825	**0.09583**	**0,09583**
	Mean	-	0.09583	0,09583
	Std. Dev	-	7.51426E-14	1,37892E-14
G9	F Best	680.63005	680.8295	**680,6848**
	Mean	-	681.6212	681,5681
	Std. Dev	-	1.032089	0,703729772
G10	F Best	7049.3307	6090.931	**7767,949**
	Mean	-	10275.51	6080,021
	Std. Dev	-	3132.618	1802,956318
G11	F Best	0.75	0.749875	**0,7499**
	Mean	-	0.750545	0,799823
	Std. Dev	-	0.003606	0,055932945
G12	F Best	1	**1**	**1**
	Mean	-	0.99526	1
	Std. Dev	-	0.007472	8,72953E-16
G13	F Best	0.0539498	0.340279	**0,052094**
	Mean	-	1.064936	-0,04819
	Std. Dev	-	0.485674	1,161516834

IV. CONCLUSION

Even if the newly developed algorithms use unimodal and multimodal unconstrained optimization test functions, real life problems are modelled as constrained optimization problem. Since the resources, workforces, materials are limited, the optimization problems have also limits too. So, in this study,

two heuristic algorithms inspired by nature have been used to solve the thirteen constrained test problems and their performances have been compared. Since these algorithms run unconstrained logic, we tested the penalty coefficients' effects on the solution too. Because if it is used small coefficients, solutions can be out of the boundaries of the constraint. But if the big coefficients are used, optimum solutions can't be found. So, when solving a constrained optimization problem, selection of the penalty coefficients is also an optimization process.

It can be listed some results from the optimum solutions given in Table 2,3 and 4.

As generally, some of the constrained problems solution are very near to best optimum solution given in the reference. These are G7, G8, G9, G11, G12. When they are examined, it can be seen that either the problems' objective function and constraints are easy or the number of variables or constraints are small.

As it can be seen there is no best penalty coefficient solving each problem best. And also, it is used a fixed coefficient among the constraints of a problem. So, this effects the optimum solution. But, according to our observation 1000 works well in these test problems generally.

So, the other results must be considered. Since the results show the fitness, even if a small deviation from boundaries can be add as a penalty. And some of the objectives are exponential or highly increasing functions. So, the small boundary exceeding effects the results.

The number of iteration and the number of agents used in the study is fixed. Both the number of iteration and the number of agents are increased, results can be improved too.

REFERENCES

[1] Seyedali Mirjalili, The Ant Lion Optimizer, Advances in Engineering Software, Volume 83, 2015, Pages 80-98.

[2] Rosy Pradhan, Santosh Kumar Majhi, Jatin Ku Pradhan, Bibhuti Bhusan Pati, Antlion optimizer tuned PID controller based on Bode ideal transfer function for automobile cruise control system, Journal of Industrial Information Integration, Volume 9, 2018, Pages 45-52,

[3] T. George, A. R. Youssef, M. Ebeed and S. Kamel, "Ant lion optimization technique for optimal capacitor placement based on total cost and power loss minimization," 2018 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, Egypt, 2018, Pages. 350-356.

[4] Hari Mohan Dubey, Manjaree Pandit, B.K. Panigrahi, Ant lion optimization for short-term wind integrated hydrothermal power generation scheduling, International Journal of Electrical Power & Energy Systems, Volume 83, 2016, Pages 158-174,

[5] E. Umamaheswari, S. Ganesan, M. Abirami and S. Subramanian, "Deterministic reliability model based preventive generator maintenance scheduling using Ant Lion Optimizer," 2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT), Nagercoil, Volume: 2016, Pages: 1-8.

[6] P. Parvathi and R. Rajeswari, "A hybrid FCM-ALO based technique for image segmentation," 2016 IEEE International Conference on Advances in Computer Applications (ICACA), Coimbatore, 2016, Pages. 342-345.

[7] K. Ksiazek, D. Polap, M. WoZniak and R. Damaševicius, "Radiation heat transfer optimization by the use of modified ant lion optimizer," 2017 IEEE Symposium Series on Computational Intelligence (SSCI), Honolulu, HI, 2017, pp. 1-7.

[8] A. H. Ali, A. R. Youssef, T. George and S. Kamel, "Optimal DG allocation in distribution systems using Ant lion optimizer," 2018 International Conference on Innovative Trends in Computer Engineering (ITCE), Aswan, Egypt, 2018, pp. 324-331.

[9] K. Hosseini, S. Araghi, M. B. Ahmadian and V. Asadian, "Multi-objective optimal scheduling of a micro-grid consisted of renewable energies using multi-objective Ant Lion Optimizer," 2017 Smart Grid Conference (SGC), Tehran, Iran, 2017, Pages. 1-8.

[10] Seyedali Mirjalili, Amir H. Gandomi, Seyedeh Zahra Mirjalili, Shahrzad Saremi, Hossam Faris, Seyed Mohammad Mirjalili, Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems, Advances in Engineering Software, Volume 114, 2017, Pages 163-191

[11] Attia A. El-Fergany, Extracting optimal parameters of PEM fuel cells using Salp Swarm Optimizer, Renewable Energy, Volume 119, 2018, Pages 641-648

[12] A. G. Hussien, A. E. Hassanien and E. H. Houssein, "Swarming behaviour of salps algorithm for predicting chemical compound activities," 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, 2017, Pages. 315-320.

[13] http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO_files/Page422.htm

Performance Evaluation of Various Binary Variants of ABC Algorithm for Solving Knapsack Problem

B. BABAYIGIT[1] and A. AYTIMUR[2]

[1] Erciyes University, Computer Engineering Department, Kayseri/Turkey, bilalb@erciyes.edu.tr
[2] Erciyes University, Graduate School of Natural and Applied Sciences, Kayseri/Turkey, asumanaytimur@gmail.com

***Abstract* - Over past two decades, many algorithms taking inspiration from natural phenomena have been proposed by the researchers. One of the famous algorithms is Artificial Bee Colony (ABC) algorithm which is inspired by the intelligent behaviors of the honey bees. Although original ABC algorithm has been proposed for solving continuous optimization problems, in order to effectively solve binary optimization problems original ABC should be modified. Using logic operator, genetic operator, and transfer function are the strategies to obtain binary solutions. Knapsack is a well-known binary optimization problem which aims to obtain a maximal knapsack packing. In this paper, various binary variants of ABC algorithm (xorABC, crossoverABC, vFunctionABC) are applied to 0–1 knapsack packing problem. The performance of three binary variants of ABC is investigated with respect to time and quality.**

***Keywords* - Binary Optimization, ABC Algorithm, Knapsack Problem, Performance Comparison.**

I. INTRODUCTION

ABC [1] which is inspired by the intelligent foraging behavior of honey bees, is a prominent optimization technique in nature inspired metaheuristics field. ABC algorithm uses three kinds of artificial bees, namely employed bees, onlooker bees, and scout bees and provides a population-based search procedure to find optimal food sources (good solutions).

In the literature, the ABC algorithm has been applied for solving various continuous optimization problems [2–6]. However, the original ABC algorithm works under the assumption that the search space is in the continuous domain. This is not suitable and effective for application to binary optimization problems which are modelled as 0–1 binary integer optimization problems. Therefore, there are several modification techniques proposed to address binary optimization problems. These are; transfer function, angle modulation, quantum-inspired bits, genetic operators, binary operators, measure of dissimilarity, and heuristic methods [7].

Knapsack problem [8] is probably the most important and widely used benchmark problem for the binary domain. This problem finds applications in the real world as Internet advertising, broadcast bandwidth, search auctions, online routing and packing problems, project selection, capital budgeting, etc. [9, 10]. Knapsack problem often occurs when a decision is required to select a limited amount of resources (a knapsack) among a given set of items where each item can be included at most once. The values of the parameters according to the knapsack capacity can be determined by using nature-inspired optimization algorithms. Because nature-inspired algorithms can find optimal or satisfactory solutions in most cases.

In this paper, three different binary versions of ABC (xorABC, crossoverABC, vFunctionABC) algorithms are utilized to solve the knapsack problem. xorABC uses exclusive OR (xor) binary operator, while crossoverABC and vFunctionABC use crossover genetic operator and V-type transfer function, respectively. An experimental comparison among these binary ABC versions are implemented using different population and iteration counts. The comparisons are performed based on two performance measures: time and the quality of the solutions.

II. KNAPSACK PROBLEM

The knapsack problem [8] is a well-known challenging combinatorial optimization problem. Given a set of n items and a knapsack capacity C. Each item has a weight and a profit. The goal is to decide a subset of items for maximum overall profit such that total weight of subset of items should not exceed the given knapsack capacity constraint. The knapsack problem can be formulated as follows:

$$Max\ f(\mathrm{x}) = \sum_{i=1}^{n} p_i.x_i \tag{1}$$

$$subject\ to: \sum_{i=1}^{n} w_i.x_i \le C \tag{2}$$

where $\mathrm{x} \in \{0,1\}$ is the binary variable which indicates whether item i is allocated to the subnet ($x_i = 1$) or not ($x_i = 0$), p_i is the profit of item i, and w_i is the weight of item i.

The knapsack problem defined in Eqs. 1 and 2 has been used to test the performance of binary metaheuristic algorithms.

III. ARTIFICIAL BEE COLONY (ABC) ALGORITHM

ABC is a swarm intelligence computation technique which

models the foraging behavior of a honeybee swarm in nature [1]. The honey bees in ABC are categorized into three groups as employed, onlooker, and scout bees. The employed bees investigate the food sources and sharing the information about the food sources to recruit the onlooker bees. The onlooker bees make decisions to select food sources depending on the information. The food source that has better quality has a higher chance of being selected by onlooker bees than lower quality one. An employed bee whose food source is abandoned as low quality becomes a scout bee and starts searching randomly for new food sources. With this mechanism, the exploitation of the best promising food sources is guaranteed by employed and onlooker bees, the exploration of the search space around the best food sources is managed by scout bee.

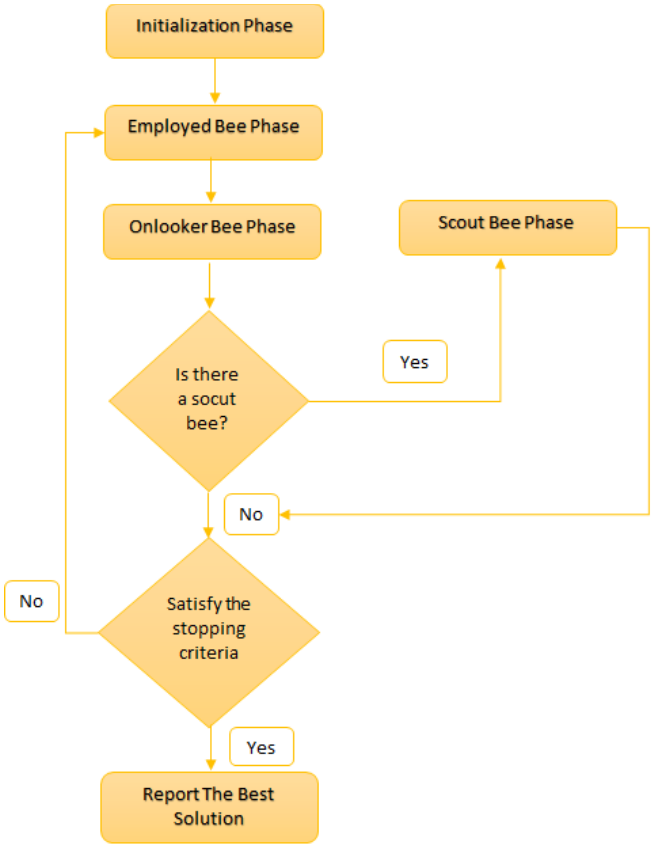


Figure 1: Flowchart of ABC algorithm.

In ABC algorithm, the position and the nectar amount of a food source are the possible solution and the quality (fitness) of associated solution, respectively. ABC algorithm has three main control parameters: the number of food sources, the value of *limit* and the maximum cycle number. The flowchart of ABC is illustrated in Figure 1.

IV. Binary Variants of ABC

Most of the nature-inspired optimization methods have been worked effectively on continuous optimization problems. However, the search space of the binary optimization problems has a binary feature which is characterized by non-differentiability and discontinuity. To address binary optimization problems, several modifications such as transfer function, angle modulation, quantum-inspired bits, genetic operators, binary operators, measure of dissimilarity, and heuristic methods [7] have been proposed to introduce binary variants of nature-inspired optimization methods.

For the initialization phase, a random binary solution is generated using Bernoulli process which is shown in Figure 2.

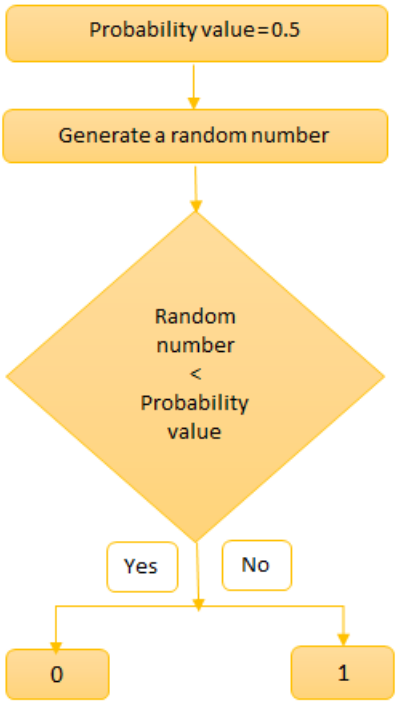


Figure 2: Generating a random binary solution.

The probability value is taken as 0.5 in this study and the random number is in the range of [0,1]. If the random number is less than the probability value, the corresponding binary solution becomes 0 otherwise 1. After initialization, positions of artificial bees are updated using binary modification techniques. In this study, of the binary modification techniques; logic operator, genetic operator, and transfer function modification techniques are selected to work in binary space for ABC algorithm and the modification procedures of each techniques are illustrated in Figure 3.

A. *Logic Operators*

For a binary solution of ABC, a new food source for an employed and onlooker bee can be produced using AND, OR, and XOR bitwise logic operators. The generating probability of '0' or '1' for AND and OR are 75%. However, when XOR logic operator is used, the changing probability of the bit is 50%. By using XOR, the new food source will be closer than old food source. For this reason, XOR logic operator is used in this study and the binary variant of ABC is named as xorABC.

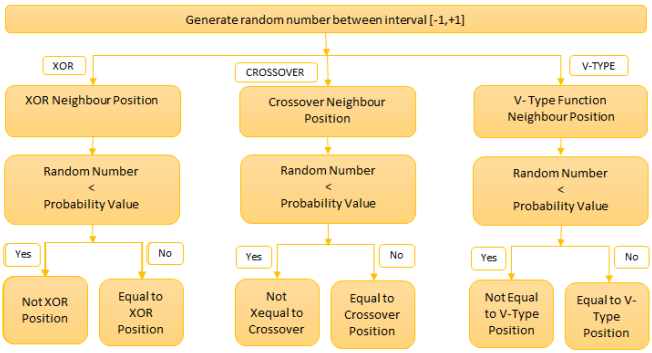


Figure 3: XOR, Crossover, and V-type modification procedures.

B. *Genetic Operator*

Genetic operator can used in a binary model of ABC algorithm. In this study, two-point crossover is used as a genetic operator. In two-point crossover, two positions on binary solutions are randomly selected and then some parts of the solutions between these positions are swapped for generating new food sources. This binary variant is named as crossoverABC.

C. *Transfer Function*

To approximate or map the continuous search space of ABC to binary, transfer functions can also be used. A transfer function first transforms all real position values to probability values in the interval of [0,1]. Then, new food sources are updated with their probability values. If a random value is less than calculated transfer function value, then the element of position vector is 0 otherwise 1. Sigmoid-type transfer function is a widely used mapping rule. However, V-type transfer function has been proposed in the recent literature [11]. With V-type transfer function, the artificial bees can stay in their current positions when their new food position values are low or switch to their complements when new food position values are high. This approach does not force to take values of 0 or 1. In this study, the ABC uses V-type transfer function is named as vFunctionABC.

V. NUMERICAL RESULTS

In this section, the experiments conducted for evaluating the three binary variants of ABC (xorABC, crossoverABC, vFunctionABC) are given. For the knapsack problem, the knapsack capacity *C* is selected as 500 kg. The profit and the weight of each item are taken p=[391 444 250 330 246 400 150 266 268 293 471 388 364 493 202 161 410 270 384 486], w=[55 52 59 24 52 46 45 34 34 59 59 28 57 21 47 66 64 42 22 23], respectively. In the experiments, the performance of three binary variants of ABC are examined using three different population values in {100, 200, 500} and five different maximum number of iterations in {50, 100, 200, 500, 1000}. The aim is to test the experimental results in terms of time and the quality of the solutions. All the tests were conducted on an Intel i7 CPU, 2.6 GHz and 16 GB of RAM.

The experimental results of every test instance of xorABC, crossoverABC, vFunctionABC are listed in Tables 1-3, respectively. The weight and time comparison for three binary ABCs are shown in Figure 4 and Figure 5.

Table 1: Solution values obtained by xorABC.

ITERATION	POPULATION	TIME(SECOND)	MAX WEIGHT	BEST POSITION RESULT 1-0
50	100	2.250278	485	1 0 0 1 1 1 0 1 0 0 1 1 1 1 1 0 0 1 0 1 1
50	200	3.188977	487	1 1 1 1 0 1 0 1 0 0 1 1 1 0 1 0 0 1 0 1 1
50	500	7.16478	500	0 1 0 1 0 1 1 1 0 1 0 1 1 1 1 1 0 0 1 1 1
100	100	4.434535	481	1 1 0 1 1 1 1 0 0 0 0 1 1 1 1 1 0 0 0 1 1
100	200	6.371734	476	1 1 1 1 0 0 0 1 0 0 1 1 1 1 1 0 0 0 1 1 1
100	500	13.543027	487	0 1 0 1 0 1 1 1 1 0 1 0 1 1 1 1 0 0 1 1 1
200	100	7.451581	478	1 0 0 1 0 1 1 1 0 0 1 1 1 1 1 0 0 1 0 1 1
200	200	11.640104	494	1 1 0 1 1 1 0 1 0 0 1 0 0 1 0 0 1 1 1 1 1
200	500	25.585254	499	0 1 0 1 0 1 1 1 1 1 0 1 1 1 0 1 1 0 1 0 1
500	100	16.737583	485	1 1 0 1 0 1 1 1 1 0 1 1 1 0 1 0 0 0 1 1 1
500	200	28.813284	491	1 1 0 1 0 1 0 0 1 1 0 1 0 1 1 1 0 0 1 0 1
500	500	62.39446	487	1 1 0 1 0 1 0 0 1 0 1 1 1 0 1 0 0 1 0 1 1
1000	100	32.460071	492	0 1 0 1 0 1 1 1 1 0 1 1 1 1 1 0 0 0 1 1 1
1000	200	57.772797	485	1 1 0 1 0 1 0 0 1 0 1 1 1 1 1 0 0 1 0 1 1
1000	500	123.815652	487	1 1 1 1 0 1 0 0 1 0 1 1 1 0 1 0 0 1 0 1 1

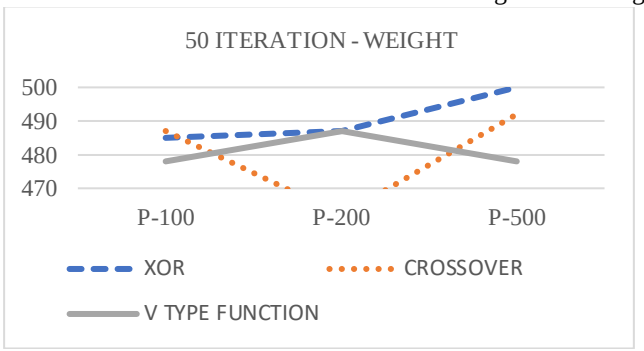
Table 2: Solution values obtained by crossoverABC.

ITERATION	POPULATION	TIME(SECOND)	MAX WEIGHT	BEST POSITION RESULT 1-0
50	100	3.861309	487	1 1 0 1 0 1 0 0 1 1 1 0 1 0 0 1 0 1 1
50	200	4.692246	462	1 1 0 1 0 1 0 1 1 0 1 0 1 0 0 1 0 1 1
50	500	9.113213	492	1 1 0 1 1 1 0 1 1 0 1 1 0 1 0 0 0 1 1 1
100	100	4.8335	498	1 1 0 1 0 0 0 1 1 0 1 1 1 1 1 0 0 1 1 1
100	200	10.172503	497	1 1 0 1 0 1 0 1 1 0 1 1 1 1 0 0 0 1 1 1
100	500	22.389429	487	1 1 0 1 0 1 0 1 1 0 1 1 0 1 1 0 0 1 1 1
200	100	10.709137	485	1 1 0 1 0 1 0 0 1 0 1 1 1 1 0 0 1 0 1 1
200	200	17.991268	499	1 1 0 1 0 1 0 0 1 0 1 0 1 0 0 1 0 1 1 1
200	500	40.609872	496	0 1 0 1 0 1 0 1 1 0 1 0 1 1 0 1 1 1 1 1
500	100	26.842518	496	0 1 0 1 1 1 0 1 1 1 1 1 1 0 0 0 1 1 1 1
500	200	42.20658	499	1 1 0 1 0 1 0 1 1 1 1 1 0 1 0 0 0 1 1 1
500	500	100.486812	497	1 1 0 1 0 1 0 1 1 1 0 1 1 1 0 0 0 1 1 1
1000	100	50.249484	485	1 1 0 1 0 1 0 0 1 0 1 1 1 1 0 0 1 0 1 1
1000	200	72.126843	499	1 1 0 1 0 1 0 1 1 1 1 1 0 1 0 0 0 1 1 1
1000	500	199.405709	492	1 1 0 1 1 1 0 1 1 0 1 1 1 0 1 0 0 0 1 1 1

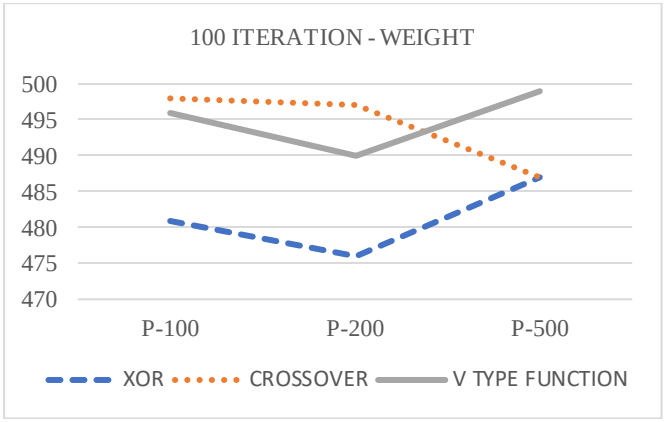
Table 3: Solution values obtained by vFunctionABC.

ITERATION	POPULATION	TIME(SECOND)	MAX WEIGHT	BEST POSITION RESULT 1-0
50	100	15.753387	478	1 0 0 1 1 0 0 0 1 1 1 1 1 0 1 1 0 1 1 1 1
50	200	60.620753	487	0 0 0 1 0 1 1 1 1 0 0 1 1 1 1 0 1 1 1 1 1
50	500	392.918376	478	0 1 0 1 0 1 0 1 1 0 1 0 1 1 0 0 1 1 1 1 1
100	100	36.56177	496	0 0 0 0 0 1 0 1 1 0 1 1 1 1 1 0 1 1 1 1 1
100	200	124.210509	490	1 1 0 1 1 1 0 1 1 0 0 1 1 1 0 0 1 0 0 1 1
100	500	822.981627	499	0 1 0 1 1 1 1 1 1 1 1 1 0 1 0 1 0 0 0 1 1
200	100	97.71337	500	1 1 0 1 1 1 0 1 0 0 1 1 0 1 0 0 1 1 0 1 1
200	200	247.205241	480	1 1 0 1 1 1 0 1 0 0 1 1 0 1 0 0 1 0 1 1
200	500	1624.497421	499	0 0 0 1 0 1 1 1 1 0 1 1 1 1 1 0 0 1 1 1 1
500	100	163.172047	496	0 1 1 1 1 1 0 1 1 1 0 1 1 0 1 0 0 0 1 1 1
500	200	590.354543	480	1 0 0 1 0 1 0 0 1 0 1 0 1 1 1 1 1 0 1 0 1 1
500	500	4130.348123	499	0 1 0 1 1 1 0 1 1 0 0 1 1 1 0 0 1 1 1 1 1
1000	100	338.10608	487	1 1 0 1 0 1 0 0 1 1 1 1 0 1 0 0 1 0 1 1 1
1000	200	1239.728107	497	0 1 0 1 0 1 0 0 0 1 1 1 1 1 0 0 1 1 1 1 1
1000	500	12203.69566	470	1 1 0 1 0 1 0 0 1 0 1 1 1 0 1 0 0 1 1 1 1

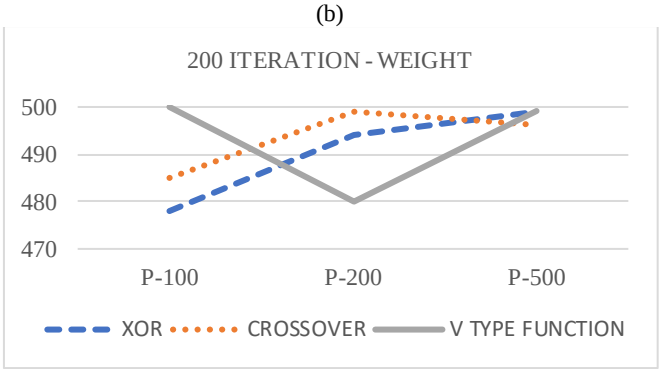
When the xor method is interpreted by itself with different populations and iteration numbers; firstly, 50 iterations were kept constant and experiments were performed with different population numbers. As the experiments were carried out with the population numbers of 100, 200 and 500; accordingly, the working time increased but the maximum weight 500 could be taken into the bag. When the number of iterations is kept constant at 100, it is seen that the number of iterations is increased to 50 and the maximum weight taken to the bag increases. The number of iterations increased at 200, as the number of populations of the bag was increased. However, increasing the number of populations at the number of iterations of 500 and 1000 did not affect the maximum weight of the bag.



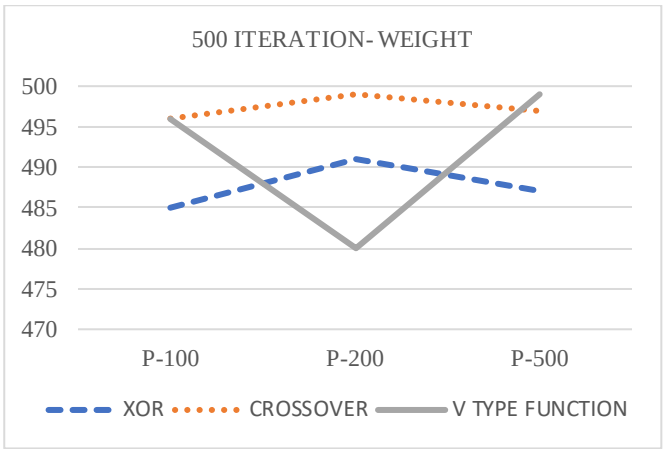
(a)



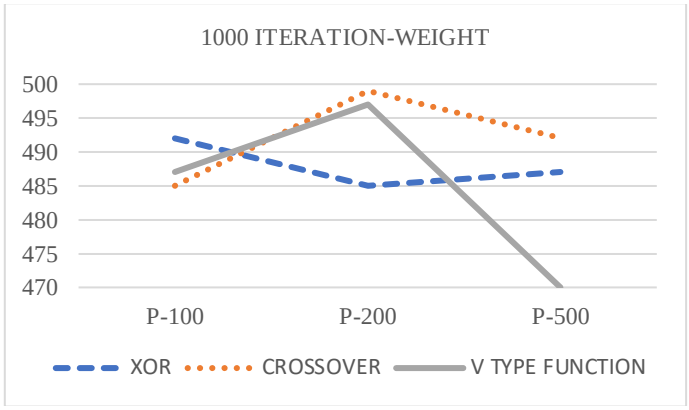
(b)



(c)

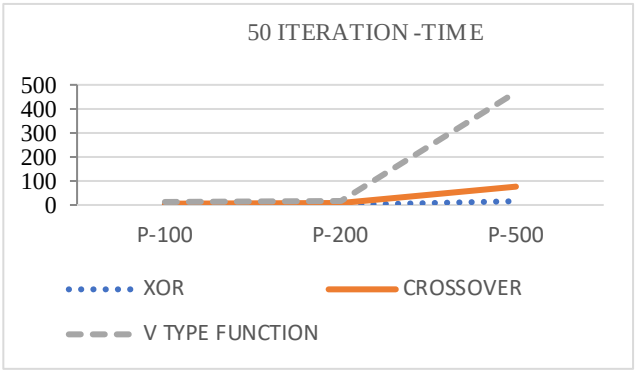


(d)

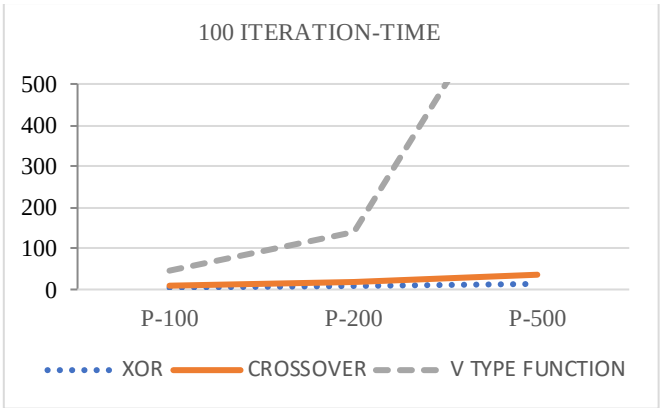


(e)

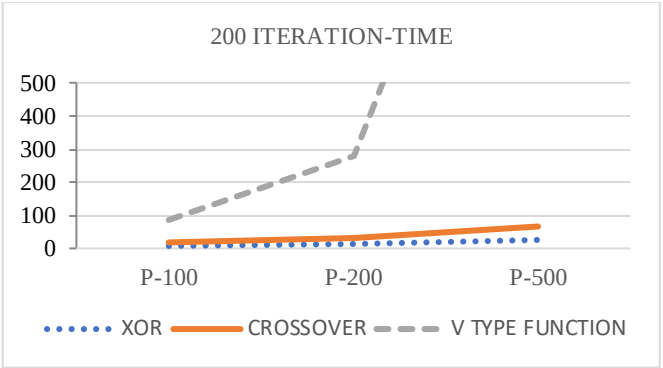
Figure 4: Weight comparison of three binary ABCs after the iteration of; (a) 50, (b) 100, (c) 200, (d) 500, (e) 1000



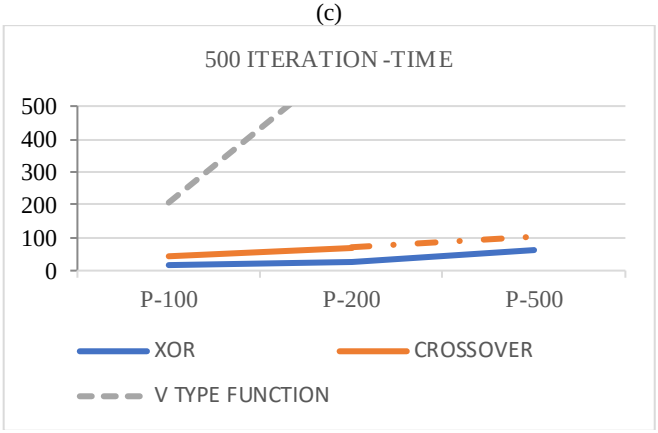
(a)



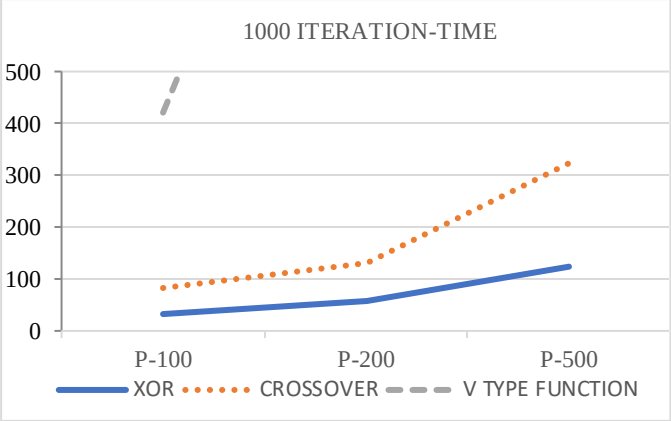
(b)



(c)



(d)



(e)

Figure 5: Time comparison of three binary ABCs after the iteration of; (a) 50, (b) 100, (c) 200, (d) 500, (e) 1000

When the crossover method is compared to different iterations and populations, as the number of iterations and the number of populations increased, the working time increased. The number of iterations is 50 and the number of populations is 500 and the maximum weight is placed on the bag. it was seen that as the number of iterations increased at 100 the maximum weight taken into the bag decreased. The number of iterations is 200, 500 and 1000, and when the number of populations is 200, the crossover is best evaluated. As the V type function is fixed at the same iteration numbers and the population numbers are increased, the working times are prolonged. The best result in V-Type is achieved in 200 iterations and 100 populations by putting 500 kg in the bag. It has been found that placing the maximum weight on the bag in the v type function is independent without increasing the number of populations and the number of iterations.

When the methods are compared with each other with the same iteration and population numbers; as shown in Figures 5 and 6; when the study periods are compared, it can be seen that the V-type result has the highest, xor has the lowest time duration, and the crossover study has time duration between xor and V-type. When the total number of item weights placed in the knapsack compared, it is possible to observe that the results obtained according to the number of iterations and the population are different.

VI. CONCLUSION

Knapsack problem is a classical challenging combinatorial optimization problem. The solution space belongs to the binary space. In this paper, three binary variants of ABC algorithm have been compared over the knapsack problem. The performance xorABC, crossoverABC, vFunctionABC are investigated in term of time and solution quality. As a result, when the run times of the same iteration and population numbers are compared, xor method works much faster than V-type and crossover method. It is concluded that the V-type function gives better results than xor and crossover when the average of the maximum weights received in the bag is taken. Future works includes adapting ABC different binary modification strategies other than this study for solving antenna thinning and real-world problems.

REFERENCES

[1] D. Karaboga, An Idea Based on Honey Bee Swarm for Numerical Optimization, Tech. Rep., Erciyes University, October 2005.

[2] D. Karaboga and B. Basturk, "On the Performance of Artificial Bee Colony (ABC) Algorithm," Appl. Soft Comput., vol. 8, no. 1, pp. 687–697, January 2008.

[3] B. Babayigit and R. Ozdemir, "Design of Non-Uniform Circular Antenna Using a Modified ABC Algorithm," *in National Electrical, Electronics and Computer Engineering Symposium (ELECO'2012)*, Bursa, Turkey, 2012, pp. 238–241.

[4] J. C. Bansal, H. Sharma, and S. S. Jadon, "Artificial bee colony algorithm:a survey," Int. J. Adv. Intell. Paradigms, vol. 5, no. 1-2, pp. 123–159, 2013.

[5] D. Karaboga, B. Gorkemli, C. Ozturk, and N. Karaboga, "A comprehensive survey: Artificial bee colony (ABC) algorithm and applications," *Artif. Intell. Rev.*, vol. 42, no. 1, pp. 21–57, 2014.

[6] A. Kumar, D. Kumar, and S. K. Jarial, "A Review on Artificial Bee Colony Algorithms and Their Applications to Data Clustering," Cybernetics and Information Technologies, vol. 17, no. 3, pp. 3–28, 2017.

[7] A. Banitalebi, M. I. A. Aziz, and Z. A. Aziz, "A self-adaptive binary differential evolution algorithm for large scale binary optimization problems," *Information Sciences*, vol. *367–368*, pp. 487–511, 2016.

[8] K. Kellerer, U. Pferschy, D. Pisinger, *Knapsack Problems*. Berlin: Springer, 2004.

[9] S. Yang and X. Yao, *Evolutionary Computation for dynamic Optimization Problems*. New York: Springer-Verlag, 2013.

[10] Y. Zhou, D. Chakrabarty, and R. Lukose, "Budget Constrained Bidding in Keyword Auctions and Online Knapsack Problems," in C. Papadimitriou and S. Zhang (Eds.), *Internet and Network Economics*, pp. 566–576. Berlin: Springer, 2008.

[11] S. Mirjalili, S. and A. Lewis, "S-shaped versus V-shaped transfer functions for binary Particle Swarm Optimization," *Swarm and Evolutionary Computation*, vol. 9, pp. 1–14, 2013.

An In-Vivo Study of Human Tibiofemoral Joint Kinematics by Using Dual Fluoroscopy System

S. UZUNER[1], M.L. RODRIGUEZ[2], L.P. LI[3], S. KUCUK[4]

[1] Duzce University, Duzce/Turkey, sabriuzuner@duzce.edu.tr
[2]University of Calgary, Calgary/Canada, mlrodrig@ucalgary.ca
[3] University of Calgary, Calgary/Canada, leping.li@ucalgary.ca
[4] Kocaeli University, Kocaeli/Turkey, skucuk@kocaeli.edu.tr

***Abstract* - A complete knowledge of Tibiofemoral (TF) joint kinematics is essential in understanding the function of healthy and pathological joint. The objective of the present study is to determine the six degrees' translations and rotations of TF joint during 10-minute in-vivo creep loading while standing using Dual Fluoroscopic (DF) images. A computational model was developed for the kinematics analysis of the right knee of a 24-year old female participant with healthy legs. Magnetic Resonance Imaging (MRI) was obtained for the unloaded joint and used for reconstruction of the knee joint model, including soft tissues. A high-resolution DF system was used to image the distal femur and proximal tibia during 10 minutes of standing. Braces were used to minimize flexions and rotations of the TF joint during the measurement. Translations and rotations of TF joint as functions of time were determined from the DF images with the Joint-Track software. Coordinate systems were established for 3D model of distal femur and proximal tibia anatomically. Rotational and translational orientations of the TF joint were calculated based on these coordinate systems. The maximum relative rotations of the distal femur with respect to the proximal tibia during 10-minute creep with approximately half body weight were 1.167 degrees in varus-valgus rotation, 4.334 degrees in internal-external rotation, and 0.541 degrees in flexion. The results showed a vertical displacement of 0.234 mm with very small rotations during 10-minute standing. Finite element modeling of the joint is in progress.**

***Keywords* - Creep, Tibiofemoral Joint Kinematics, Dual Fluoroscopy, Image-based Computer Model**

I. INTRODUCTION

IT is vital to understand TF kinematics in order to develop relevant surgical treatments, and diagnose early onset of osteoarthritis (OA)[1-3]. In addition, the results obtained from the kinematic analysis of the TF joint can be used in the analysis of TF joint finite elements to understand the deformation of the tissues in the knee joint under load. Therefore, there exist many in-vitro and in-vivo studies on TF joint kinematics in the literature. However, most previous experimental studies on TF kinematics were performed in cadaver, in-vitro, or the knee joint was not tested in physiological weight-bearing mode [4-10]. With the development of medical imaging systems such as MRI, DF, and computed tomography (CT), in-vivo TF joint kinematic analysis has drawn a lot of attention [11-16]. As compared with in-vitro studies, the results obtained from in-vivo TF joint kinematics studies have better accuracy and closeness to reality, since there are no TF contact shifts with the intact ligaments and muscles [17].

There are two main approaches in in-vivo TF joint kinematic analysis: marker-based and markerless. In marker-based approaches, markers mounted on the surface of the skin or subcutaneous bone, are used for monitoring the bone movements in the TF joint [17-20]. However, in-vivo TF joint kinematic analysis is a demanding process because the soft tissues in the TF joint complicate the in-depth examination of knee kinematics and thus it is required the errors to be mathematically corrected to improve the accuracy of the results [21-24]. On the other hand, in markerless approaches, kinematic analysis is performed without placing any markers in the TF joint. One of the most widely employed systems in this approach is the high-speed DF system. Through the noninvasive imaging, DF provides the bone positions in the TF with x-ray images [2, 15, 16, 25, 26]. When DF systems are compared with the skin marker-based systems, DF images provide translations and rotations of the bone in all 6 degrees of freedom. That's why they are not affected by soft tissue movement artifacts [25]. In addition to their high accuracy and resolution, they necessitate less time to prepare the experimental setup and analyze the data [27].

MRI and CT systems have high spatial 3D tissue geometry but low temporal resolution [28, 29]. On the other hand, DF system combines high temporal with spatial to deal with this problem [30]. Thus, a DF system allows us to capture the dynamic movement of the TF joint.

There are significant effects of variables such as gender and age on the results of TF joint kinematics. In this context, each model used for kinematic analysis is subject-specific [2]. Therefore, the results of the studies carried out in this area may not be simply generalized and still the collected data may be used to establish an important database for further biomechanical studies. The aim of the present study is to perform a kinematic analysis of the subject-specific TF joint using the advantages of the DF system and then to use the results for the analysis of TF joint mechanics with a finite element method.

II. MATERIAL AND METHODS

The procedures for the kinematic data acquisition of TF involve the following steps: (1) obtaining the 2D images of knee joint during 10 minutes standing through DF system in selected frequency (6Hz) and the 3D MRI reconstruction of the leg (TF, ankle, hip); (2) calibrating the distorted images obtained through DF; (3) identifying the coordinate system in order to determine TF tracking; and (4) matching 3D and 2D images to determine the bone displacement.

A. Acquiring MRI and DF Images

MR images of TF were obtained through 3-Tesla MRI (GE 750) from the unloaded right knee of a healthy adult female participant weighing 56.7 kg at Seaman Family MR Research Centre, Foothills Medical Centre at the University of Calgary. The MRI sequences were a high resolution MRI scan of the TF joint including 20cm above and below (Slice Thickness=1mm, FOV=24X24cm, 200 slices, Slice Spacing (SS) =0.5mm, pixel resolution 512X512 pixels). The images of hip and ankle were also obtained but at lower resolution; they were used to identify the coordinate system for femur and tibia [31]. MR images were transferred into AMIRA software (Thermo Fisher Scientific, Carlsbad, USA) for 3D image processing of knee joint. 3D knee joint can be seen in Figure 1.

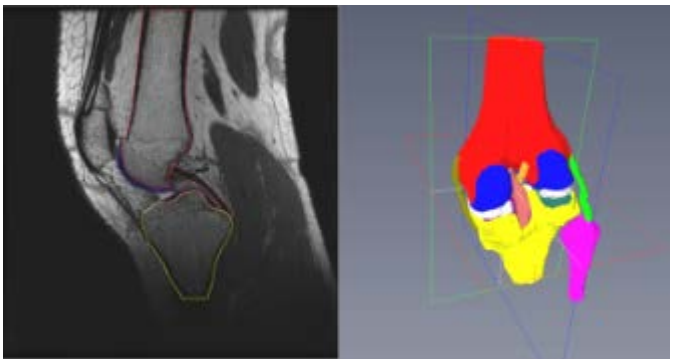


Figure 1: Segmentation and reconstruction of 3D images through MRI data of knee joint.

Dual Fluoroscopy imaging techniques were used to develop the six-degree-of-freedom kinematics model of the TF joint during loading in 10 minutes. Even though dual fluoroscopic systems have low dose X-ray technology, they have very high frame rates. The DF system consists of two high-speed X-ray generators and a high-speed solid-state video camera. This feature provides the required temporal resolution (6-250Hz) to track (translations and rotations) the bones of TF during the DF measurement [25]. DF system used in this study was shown in Figure 2. We are grateful to use the DF system at Dr. Janet Ronsky's lab at the University of Calgary. The combined MRI and DF study was approved by the Conjoint Health Research Ethics Board at the University of Calgary.

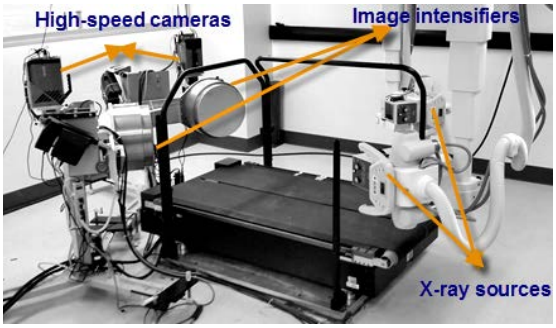
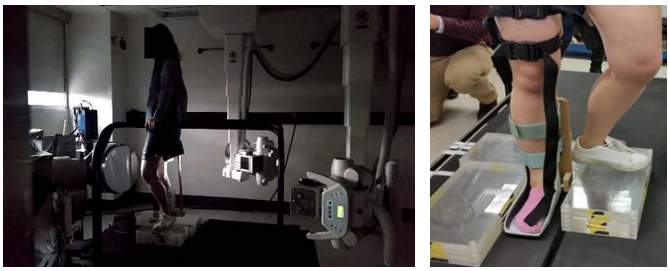


Figure 2 Dual fluoroscopic imaging system.

It was monitored in previous studies that deformation rate of soft tissues of TF was very high during early loading and it decreased rapidly after the first minute [32]. In this context, in order to track the translations and rotations of bones of TF precisely, 2D images were acquired from two cameras of the DF system at a frame rate of 6Hz, continuously for the first minute and at 6Hz for 2 s intermittently with 30-second breaks for the rest nine minutes. 563 pairs of images with high-resolution were taken by two cameras in the DF system. TF joint was tracked during loading in 10 minutes (Figure 3a). The brace was used to obtain the movement of bones of TF during the test in full extension (Figure 3b).



a) b)

Figure 3: a) DF Measurement: Right knee is loaded with approximately half body weight in full extension during 10 minutes. b) Brace was used to minimize flexion and rotation of the bones of TF.

The MRI measurement was done on unloaded knee. In order to keep the deformation of cartilages at a minimum level before the tests, the subject was taken to the hospital by a car in the early morning and she was carried to the MRI and DF Centers, which are in the same building, by a wheel chair. This effort ensured to obtain the kinematic data of TF in the right conditions to be used in future finite element analysis.

B. Image Calibrations

By nature of imaging process, the images obtained through DF system are distorted because of the structure of camera lens and the distance of the object to the camera. Therefore, these images need to be calibrated so that the TF joint kinematics can be obtained accurately from DF images. To do so, MATLAB (The MathWorks Inc., Version R2017a, Natick, MA, USA) software was employed for calibration image process with an acrylic cubic calibration frame (Figure 4).

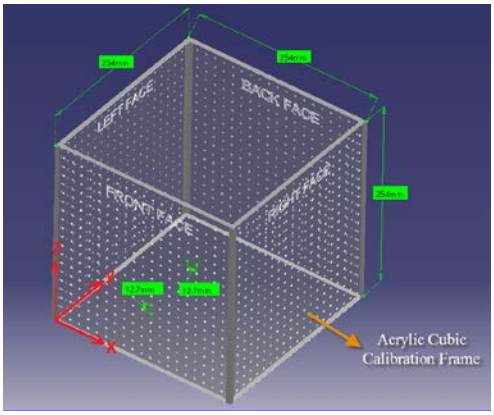


Figure 4: Acrylic cubic calibration with the frame coordinate system.

The calibration frame and a custom MATLAB R2012a procedure were developed by Derek D. Lichti et al. to correct distorted images from DF measurement [25]. Custom MATLAB R2012a software enables the acquisition of external and internal parameters of DF cameras by using acrylic cubic calibration frame features and approaches such as Direct Linear Transform, Bundle and Triangulation [14]. By using these softwares mentioned above, 2D undistorted images were obtained.

C. *Identifying the Coordinate System of Femur and Tibia*

The purpose of establishing the coordinate system is to allow us to calculate relative position between femur and tibia and to understand how the relative position changes over time. The coordinate system of femur and tibia was calculated using the anatomy of the leg [31]. In this context, the coordinate system characterizes the specific knee joint and the description of the knee motion, translations (proximal-distal (PD), anterior-posterior (AP) and medial-lateral (ML)), and rotations (Valgus-Varus (VV), Internal-External (IE), Flexion-Extension (FE)) (Figure 5).

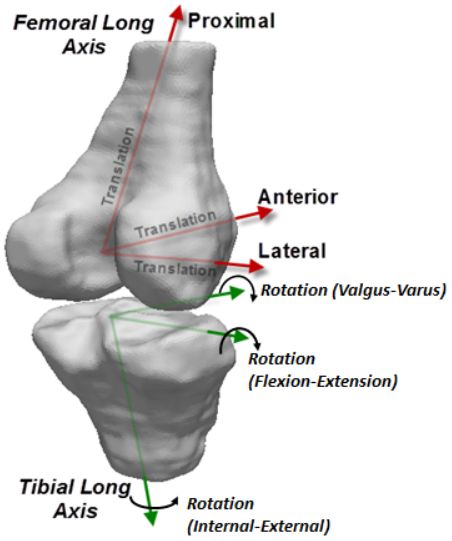


Figure 5: Coordinate systems of femur and tibia.

Because the axes to be produced based solely on knee anatomy may be less accurate, 3D models of hip and ankle were used as additional reference. The long axis for the femur and the tibia was obtained through custom MATLAB R2012a, software designed by G. B. Sharma et al [14].

This custom software creates the coordinate system for each bone as follows. For the femur, the spheres are placed in the medial and lateral and the ML axis is obtained by joining the centers of these spheres [33]. The center of the ML line is selected as the origin of the femur. As for the line of femur in the direction of AP, the ML axis is obtained by applying the cross-product method on the axis, which is the line from the origin of the femur to the center of the hip. Finally, in order to obtain the PD axis, the cross-product method is conducted on the ML axis and AP axis. Similar procedures are performed on the proximal tibia and ankle to obtain tibia long axis [14, 31].

D. *2D-3D Matching Process*

Having carried out the procedures mentioned in previous section successfully, JointTrack Biplane (University of Florida, Gainesville, FL, USA) software was then used to match the 3D model of the femur and tibia with the 2D undistorted images. The software calculates and reports the 3D motions of bones of TF through matching process until the silhouette of bones provides an optimal fit with 2D undistorted images that were initially obtained by two cameras of the DF system (Figure 6).

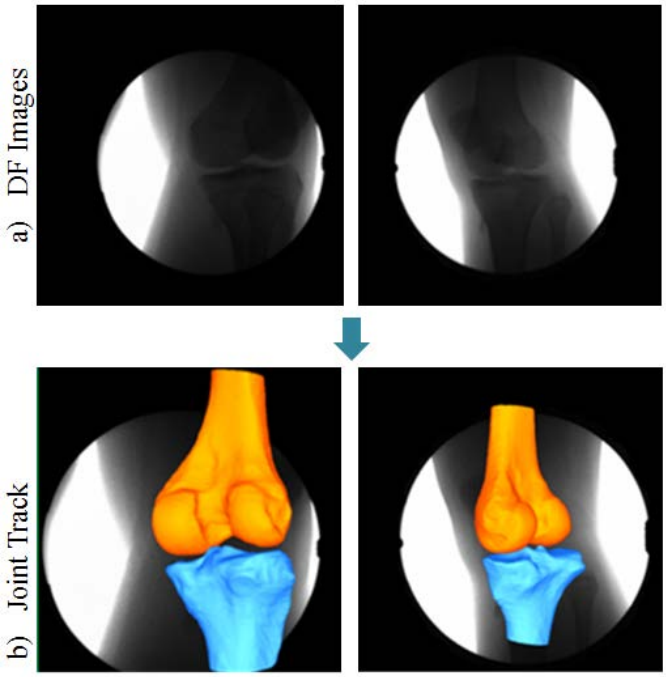


Figure 6: a) 2D images of the knee joint were obtained with DF measurement during 10 minutes standing; b) Markerless 2D–3D registration of 3D bones with pairs of DF images.

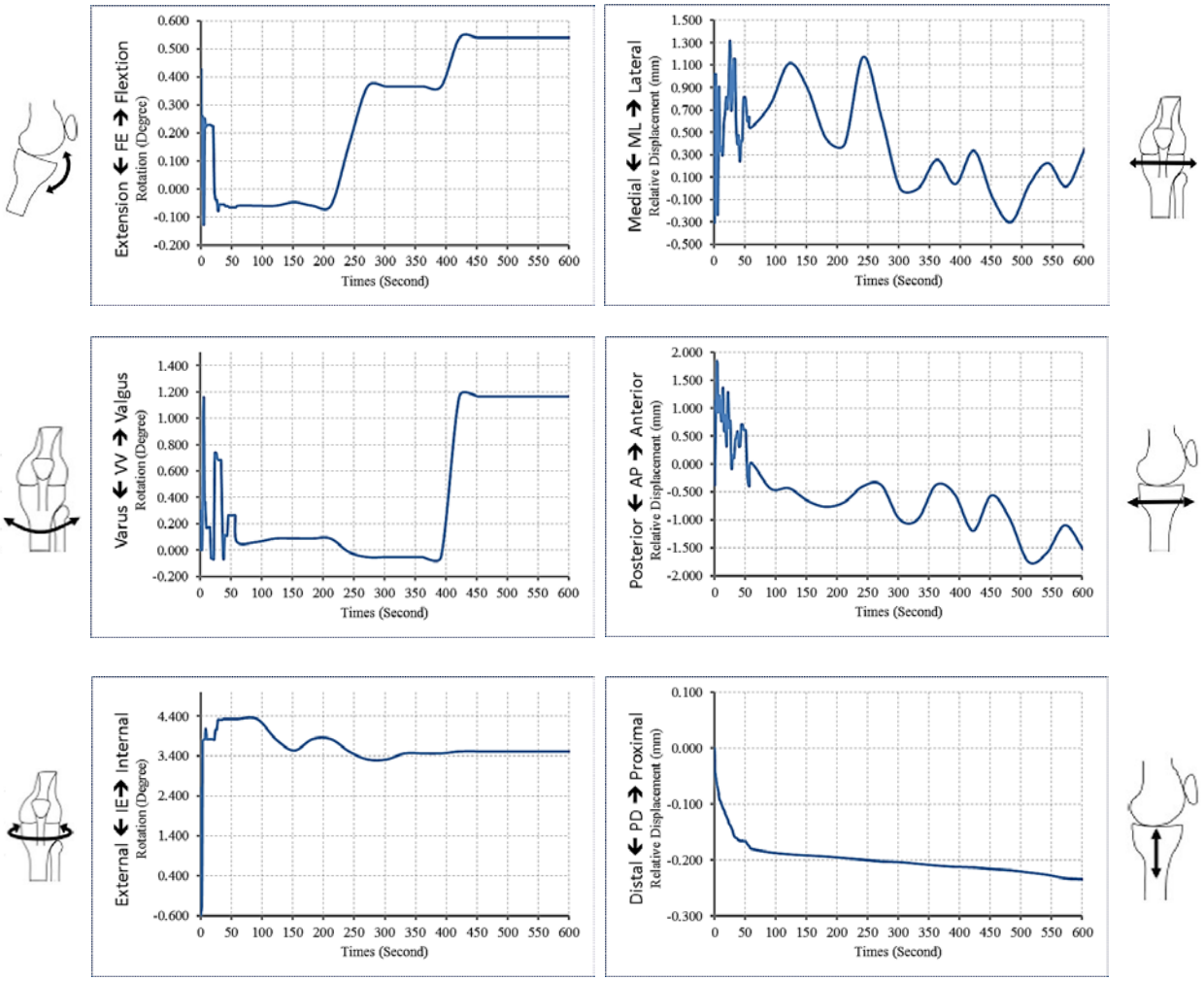


Figure 7: Relative motion for the six degrees of freedom of tibiofemoral (TF) joint during 10 minutes standing.

Cameras placed at different angles in the DF system provide images of the bones of TF at different angles in the same time through X-ray sources. In this way, data of 3D motion of the bones are acquired through these 2D images. The silhouette of the bone was overlaid manually on these two obtained images. This process was repeated meticulously, taking 56 samples from 563 pairs of images obtained during the DF measurement. The samples were taken for the first minute at 1~2-second intervals to capture the motion of bones as the maximum motion variation. For the rest nine minutes, the samples were taken at 15-second intervals. After matching process, the data of translations and rotations of the bones were obtained during 10 minutes standing and in for the selected frequency (6Hz).

III. Discussion and Conclusion

Some results of TF kinematics were presented in Figure 7. These results are relative movements of femur to tibia. The results obtained from the JointTrack software are the absolute motions of the femur and tibia with respect to their own axes. A custom MATLAB R2017a software was prepared using the Euler Angles (ZYX sequence) equations to obtain the motion of the femur relative to the tibia.

The maximum relative rotations of distal femur with respect to proximal tibia during 10-minute creep with approximately half body weight were 1.167 degrees in varus-valgus rotation, 4.334 degrees in internal-external rotation, and 0.541 degrees in flexion. Similarly, the maximum relative translations were 0.234 mm in proximal-distal direction, 1.844 mm in anterior-posterior direction, and 1.316 mm in medial-lateral direction.

The results of the present study will be used in finite element stress and strain analysis. For that study, it will be aimed to simulate creep and relaxation behaviors of the TF joint in full extension during 10 minutes. For this reason, during the DF measurement, the brace was used to guide the movement of TF joint only in full extension without substantial rotations. As it can be seen from Figure 7, the bones achieved the desired test conditions very well with very small rotations. In addition, the maximum changes in bone rotations occurred within the first few seconds of the test and rotations generally remained constant for the rest period of the test. This was the result of necessary training of the participant prior to the measurement.

The 2D-3D matching process was performed manually in

the JointTrack software. When this process was carried out, the silhouette of 3D bone models was attempted to be completely overlapped on the 2D images. Because of the human factor, it is difficult to foresee the errors that can occur in this matching process. In this context, the quantitative results may need to be verified with more data, or the results obtained from the in-vivo measurement can be compared with the results obtained from the measurement of femoral and tibial prostheses with accurately known sizes in the same condition [14]. The mechanism and the method used in this study were validated by previous studies [14, 15].

ACKNOWLEDGMENT

We greatly appreciate the help of Drs. Janet Ronsky and Gregor Kuntze and Jessica Kupper with the DF measurement at the University of Calgary. We would also like to especially thank TUBITAK (The Scientific and Technological Research Council of Turkey) for the financial and moral support of Sabri Uzuner's work on this project.

REFERENCES

[1] D. Paley, *Normal lower limb alignment and joint orientation*, in *Principles of deformity correction*. Springer. p. 1-18, 2002.

[2] K.M. Varadarajan, T.J. Gill, A.A. Freiberg, H.E. Rubash, and G. Li, *Gender differences in trochlear groove orientation and rotational kinematics of human knees*. Journal of Orthopaedic Research. **27**(7): p. 871-878, 2009.

[3] F. Liu, M. Kozanek, A. Hosseini, S.K. Van de Velde, T.J. Gill, H.E. Rubash, and G. Li, *In vivo tibiofemoral cartilage deformation during the stance phase of gait*. Journal of biomechanics. **43**(4): p. 658-665, 2010.

[4] T. Siebel and W. Kafer, *In vitro investigation of knee joint kinematics following cruciate retaining versus cruciate sacrificing total knee arthroplasty*. Acta orthopaedica belgica. **69**(5): p. 433-440, 2003.

[5] G. Li, S. Zayontz, L.E. DeFrate, E. Most, J.F. Suggs, and H.E. Rubash, *Kinematics of the knee at high flexion angles: an in vitro investigation*. Journal of Orthopaedic Research. **22**(1): p. 90-95, 2004.

[6] T. Fukubayashi, P. Torzilli, M. Sherman, and R. Warren, *An in vitro biomechanical evaluation of anterior-posterior motion of the knee. Tibial displacement, rotation, and torque*. The Journal of bone and joint surgery. American volume. **64**(2): p. 258-264, 1982.

[7] A. Garg and P. Walker, *Prediction of total knee motion using a three-dimensional computer-graphics model*. Journal of Biomechanics. **23**(1): p. 45-58, 1990.

[8] K.G. Nilsson, J. Kärrholm, and L. Ekelund, *Knee motion in total knee arthroplasty. A roentgen stereophotogrammetric analysis of the kinematics of the Tricon-M knee prosthesis*. Clinical orthopaedics and related research,(256): p. 147-161, 1990.

[9] K.G. Nilsson, J. Kärrholm, and P. Gadegaard, *Abnormal kinematics of the artificial knee: roentgen stereophotogrammetric analysis of 10 Miller-Galante and five New Jersey LCS knees*. Acta Orthopaedica Scandinavica. **62**(5): p. 440-446, 1991.

[10] A.A. Amis, W. Senavongse, and A.M. Bull, *Patellofemoral kinematics during knee flexion-extension: An in vitro study*. Journal of orthopaedic research. **24**(12): p. 2201-2211, 2006.

[11] D.A. Dennis, R.D. Komistek, W.A. Hoff, and S.M. Gabriel, *In vivo knee kinematics derived using an inverse perspective technique*. Clinical Orthopaedics and Related Research®. **331**: p. 107-117, 1996.

[12] D.K. Ramsey and P.F. Wretenberg, *Biomechanics of the knee: methodological considerations in the in vivo kinematic analysis of the tibiofemoral and patellofemoral joint*. Clinical Biomechanics. **14**(9): p. 595-611, 1999.

[13] G. Li, L.E. DeFrate, H.E. Rubash, and T.J. Gill, *In vivo kinematics of the ACL during weight-bearing knee flexion*. Journal of orthopaedic research. **23**(2): p. 340-344, 2005.

[14] G. Sharma, S. Saevarsson, S. Amiri, S. Montgomery, H. Ramm, D. Lichti, R. Lieck, S. Zachow, and C. Anglin, *Radiological method for measuring patellofemoral tracking and tibiofemoral kinematics before and after total knee replacement*. Bone and Joint Research. **1**(10): p. 263-271, 2012.

[15] G.B. Sharma, G. Kuntze, D. Kukulski, and J.L. Ronsky, *Validating Dual Fluoroscopy System Capabilities for Determining In-Vivo Knee Joint Soft Tissue Deformation: A Strategy for Registration Error Management*. Journal of biomechanics. **48**(10): p. 2181-2185, 2015.

[16] B. Ritchie, G. Kuntze, G. Sharma, J. Beveridge, J. Kupper, and J. Ronsky, *Determining In-Vivo Human Tibiofemoral Cartilage Stiffness Using Dual Fluoroscopy and Magnetic Resonance Imaging*. CMBES Proceedings. **39**(1), 2016.

[17] M. Sati, J.A. de Guise, S. Larouche, and G. Drouin, *Improving in vivo knee kinematic measurements: application to prosthetic ligament analysis*. The knee. **3**(4): p. 179-190, 1996.

[18] M.S. Andersen, D.L. Benoit, M. Damsgaard, D.K. Ramsey, and J. Rasmussen, *Do kinematic models reduce the effects of soft tissue artefacts in skin marker-based motion analysis? An in vivo study of knee kinematics*. Journal of biomechanics. **43**(2): p. 268-273, 2010.

[19] D.L. Benoit, M. Damsgaard, and M.S. Andersen, *Surface marker cluster translation, rotation, scaling and deformation: Their contribution to soft tissue artefact and impact on knee joint kinematics*. Journal of biomechanics. **48**(10): p. 2124-2129, 2015.

[20] Y. Wen, H. Huang, Y. Yu, S. Zhang, J. Yang, Y. Ao, and S. Xia, *Effect of tibia marker placement on knee joint kinematic analysis*. Gait & posture. **60**: p. 99-103, 2018.

[21] A. Cappello, A. Leardini, F. Catani, and P. La Palombara. *Selection and validation of skin array technical references based on optimal rigid model estimation*. in *Proc III Int Symp on 3-D Analysis of Hum Mov*. 1994.

[22] E. Szczerbik and M. Kalinowska, *The influence of knee marker placement error on evaluation of gait kinematic parameters*. Acta Bioeng Biomech Wroc Univ Technol. **13**: p. 43-46, 2011.

[23] T.-Y. Tsai, T.-W. Lu, M.-Y. Kuo, and C.-C. Lin, *Effects of soft tissue artifacts on the calculated kinematics and kinetics of the knee during stair-ascent*. Journal of biomechanics. **44**(6): p. 1182-1188, 2011.

[24] J. Clément, R. Dumas, N. Hagemeister, and J.A. De Guise, *Soft tissue artifact compensation in knee kinematics by multi-body optimization: Performance of subject-specific knee joint models*. Journal of biomechanics. **48**(14): p. 3796-3802, 2015.

[25] D.D. Lichti, G.B. Sharma, G. Kuntze, B. Mund, J.E. Beveridge, and J.L. Ronsky, *Rigorous geometric self-calibrating bundle adjustment for a dual fluoroscopic imaging system*. IEEE transactions on medical imaging. **34**(2): p. 589-598, 2015.

[26] J.-S. Li, T.-Y. Tsai, S. Wang, P. Li, Y.-M. Kwon, A. Freiberg, H.E. Rubash, and G. Li, *Prediction of in vivo knee joint kinematics using a combined dual fluoroscopy imaging and statistical shape modeling technique*. Journal of biomechanical engineering. **136**(12): p. 124503, 2014.

[27] M.A. Perrott, T. Pizzari, J. Cook, and J.A. McClelland, *Comparison of lower limb and trunk kinematics between markerless and marker-based motion capture systems*. Gait & posture. **52**: p. 57-61, 2017.

[28] C. Peterfy, G. Gold, F. Eckstein, F. Cicuttini, B. Dardzinski, and R. Stevens, *MRI protocols for whole-organ assessment of the knee in osteoarthritis*. Osteoarthritis and Cartilage. **14**: p. 95-111, 2006.

[29] T. Mosher, E. Walker, J. Petscavage-Thomas, and A. Guermazi, *Osteoarthritis year 2013 in review: imaging*. Osteoarthritis and cartilage. **21**(10): p. 1425-1435, 2013.

[30] E.L. Brainerd, D.B. Baier, S.M. Gatesy, T.L. Hedrick, K.A. Metzger, S.L. Gilbert, and J.J. Crisco, *X-ray reconstruction of moving morphology (XROMM): precision, accuracy and applications in comparative biomechanics research*. Journal of Experimental Zoology Part A: Ecological and Integrative Physiology. **313**(5): p. 262-279, 2010.

[31] E.S. Grood and W.J. Suntay, *A joint coordinate system for the clinical description of three-dimensional motions: application to*

the knee. Journal of biomechanical engineering. **105**(2): p. 136-144, 1983.

[32] A. Hosseini, S.K. Van de Velde, M. Kozanek, T.J. Gill, A.J. Grodzinsky, H.E. Rubash, and G. Li, *In-vivo time-dependent articular cartilage contact behavior of the tibiofemoral joint*. Osteoarthritis and cartilage. **18**(7): p. 909-916, 2010.

[33] F. Iranpour, A.M. Merican, W. Dandachli, A.A. Amis, and J.P. Cobb, *The geometry of the trochlear groove*. Clinical Orthopaedics and Related Research®. **468**(3): p. 782-788, 2010.

Detection of Light Sleep Periods Using an Accelerometer Based Alarm System

E. TURKYILMAZ, A. AKGUL, E. BOSTANCI and M. S. GUZEL

Ankara University Computer Engineering Department SAAT Lab
Ankara Turkeyegmntrkylmz@gmail.com, alperakgul1996@gmail.com, ebostanci@ankara.edu.tr, mguzel@ankara.edu.tr

***Abstract*—Light sleep is a sleeping period which occurs within each hour during the sleep. This is the period when people are closest to awakening. With this being the case people tend to move more frequently and aggressively during these periods. In this paper the most suitable moment for waking a person up will be described. The characteristics of sleeping stages, detection of light sleep periods and analysis of light sleep periods were clarified. The sleeping patterns of different subjects were analyzed. The detection of this moment and the development process of a system dedicated to this purpose will be explained, and also some experimental results that are acquired via different tests will be shared and analyzed.**

***Index Terms*—Light sleep, REM, NREM, real-time acceleration measurement, sleep stages, programmable alarm system.**

I. INTRODUCTION

In daily life most of the people needs an alarm system because they might have to wake up at a specific hour of the day. However, it may not be convenient to wake up at a random moment since sleep has a nature of continuous loops. It was observed that when people wake up in deep sleep stages, they usually have difficulties in awakening and encounter mental focus problems that endure for a certain period of time. There are 2 fundamental stages of sleep. These are Rapid Eye Movement (REM) and Non-Rapid Eye Movement (NREM) stages. All of the human-beings have these stages in their sleep. There are periods within the stages of sleep where people can wake up more easily. According to these researches, The NREM stage contains light sleep periods. As mentioned above, if the light sleep period is detected and the alarm is sounded during this period, it will be easier to wake up and feel more intense during the day. The main subject of this article is the development of an alarm system which aims to provide a solution to that issue.

The necessary literature reviews about the sleep stages and alarm systems, were made before the work on the subject started. The duration of sleep cycle segments differentiate throughout the maturation of the humans. Based on some researches, an adult human's REM sleep takes approximately 20% of total sleep time as of a newborn's 50% [1, 2]. In the time of REM sleep, high brain activity is detected. On the contrary, there is no excessive body movement. Also dreaming takes place in this stage [3, 4].

On the other hand NREM is the period that includes the light sleep. In this period people are in such a state that is between asleep and awakened and also during this interval an increase is observed on body movements [5]. Light sleep stage is indicated by these movement patterns. Fig. 1 depicts the time spent on REM and NREM stages during the sleep.

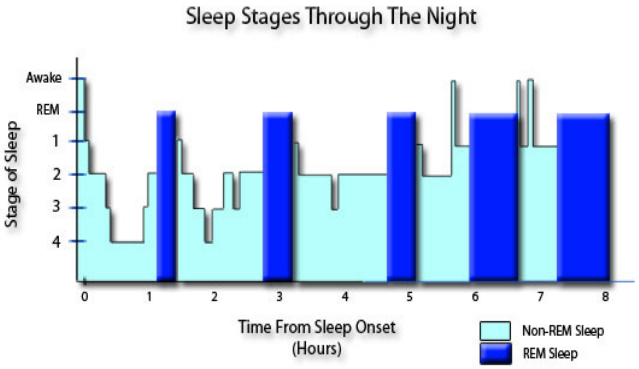


Fig. 1. Sleep stages through the night [6]

People can wake up more easily and be more energetic and strong throughout the day when they are awakened during the light sleep stage of NREM period. There are several studies and productions related to this subject. One being a mobile application called "Sleep Cycle Alarm" which uses the internal accelerometer and microphone of the device it is installed. This application works when the phone is placed somewhere nearby the sleeper. It is affected by the vibrations of the surface where the phone is placed on [7]. And also there are three patented studies that we encountered during our researches [8–10]. But these studies don't constitute an impediment to work on this topic and try to develop some other solutions.

This paper analyzes stages of the sleep and their relations with body movements and focuses on an alarm system that detects the optimal moment to wake a person up based on the change in acceleration.

The rest of the paper structured as follows. In Section II the environment of the alarm system is explained, followed by the specifications of the hardware used in the system. Section III addresses the development process of the necessary programs constructed for the system. In Section IV, the results acquired out of the experiments are shared. Finally, the paper is concluded in Section V.

II. SYSTEM ARCHITECTURE

In order to gather the necessary data, to analyze the sleep stages some hardware presence is required. In this section, the detailed features of the equipment used will be discussed. Also, the purpose of these equipments will be covered. To implement a system that meets these requirements an accelerometer and a platform to process the data is used. In this section, the hardware used will be explained in detail.

A. Raspberry Pi

Raspberry Pi is a low-cost, portable and modular computer that plugs into a display monitor. Most types of it are capable to do most of the work that a desktop computer can do. In this research as shown in the Fig. 2, a "Raspberry Pi 2 Model B" is used. This particular device has a 900 MHz quad-core ARM Cortex-A7 CPU (Central Processing Unit) and 1GB of RAM along with 4 USB 2.0 ports, 40 GPIO (General Purpose Input Output) pins, a Full HDMI port, an Ethernet port, a micro SD card slot and a VideoCore IV 3D graphics core. The IMU is plugged into one of the USB 2.0 ports, a display monitor into the Full HDMI port, the passive buzzer into the GPIO pins. The operating system is installed on an SD card and it is inserted into the SD Card slot located on the Raspberry Pi [11,12].

Fig. 2. Raspberry Pi 2 Model B

B. Accelerometer

As demonstrated in Fig. 3, a "PhidgetsSpatial 1056 3/3/3" model IMU is used to measure acceleration values. This piece of hardware also includes gyroscope and magnetometer chips. This IMU provides static and dynamic acceleration, magnetic field and angular rotation measurements in 3 axes. "PhidgetsSpatial 1056 3/3/3" makes it possible to measure real-life motion in real time. With its precise voltage supply filtering, it guarantees low noise and correct sensor operation. But this paper focuses especially on IMU's static and dynamic acceleration measurements. The accelerometer can sense a minimum 228µg (alternatively:2.2 mm/s^2).of change in acceleration. And it can measure a maximum acceleration of ±5g (alternatively: $49m/s^2$). The IMU has a minimum sampling speed of 1 samples/s (samples per second) and a maximum sampling speed of 250 samples/s [13].

Fig. 3. PhidgetSpatial 3/3/3

C. Passive Buzzer

A passive buzzer is a device that produces different types of sounds depending on the signals received by its pins. It is a low-cost device and can be replaced quite easily by another one. It requires a computer program to generate the desired tunes. Two jumper cables are needed to plug in the buzzer into a Raspberry Pi. In the system shown in Fig. 3, the buzzer is plugged into 11^{th} and 14^{th} GPIO pins of the Raspberry Pi. According to Raspberry Pi 2 Model B's GPIO pinout, 11^{th} pin is used for sending the signals from the program to the buzzer and 14^{th} pin is used for ground. A passive buzzer with two jumper cables is shown in Fig. 4.

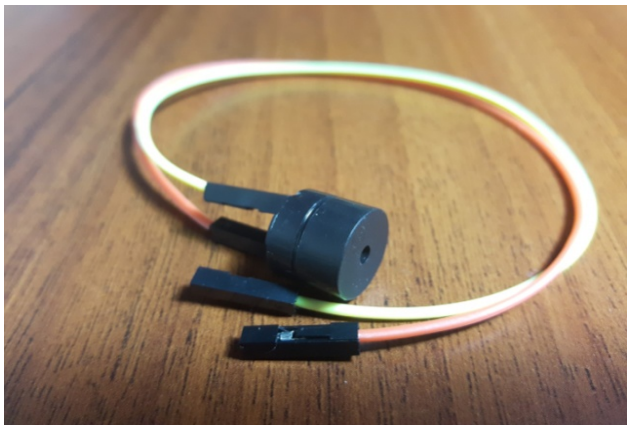


Fig. 4. Passive buzzer with two jumper cables

D. Complete System

As the system requires an accelerometer a "PhidgetSpatial 1056 3/3/3" model IMU (inertial measurement unit) is used. This hardware is operated on a "Raspberry Pi" running a "Raspbian" operating system with necessary libraries and drivers installed. The IMU is connected to a USB 2.0 port on the Raspberry Pi as displayed in Fig. 5.

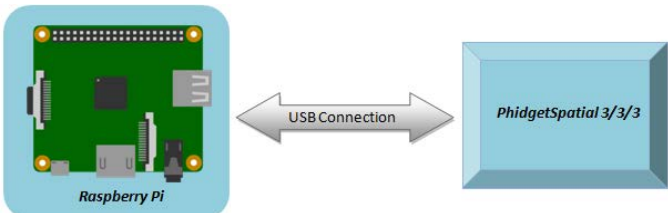


Fig. 5. Working concept of the IMU

After the installations are completed the IMU is attached to the subject's body. Then according to given sleep time as the input, the IMU starts measuring acceleration based on body movements. This process takes as long as the indicated sleep time. When the timer enters the last period of sleep time, the program seeks a specific condition to sound an alarm. This alarm is sounded by a passive buzzer plugged into the Raspberry Pi via jumper cables. As the buzzer is a passive one, it needs a dedicated program to play the desired tune. The complete system is shown in Fig. 6.

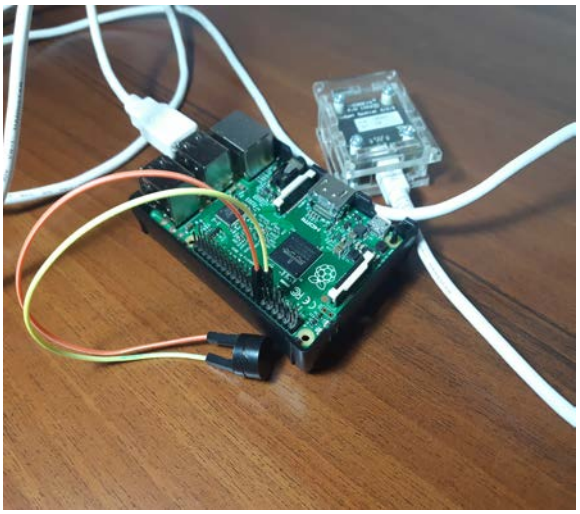


Fig. 6. The complete system

III. SOFTWARE DEVELOPMENT

During the software development process, various libraries and programming languages are used. "WiringPi" and "Phidget22" libraries have been involved and used with C programming language. This section will mainly focus on the development process of the programs which are responsible from handling the alarm system. The software developed mainly consists of three separate program files. One being the measurement program, another handling the output files and the last one generates a tune for the alarm.

A. Development Platform

All of the programming in this research is done in C programming language. This programming language is chosen because it is supported by the IMU's libraries and drivers, it is capable of running on different types of CPU architectures and C programming language provides flexibility when used on embedded systems. Also on embedded systems memory and storage capacities may often be limited. C is an efficient and proper programming language to work on such kind of systems.

B. Used Libraries

A C library called "libphidget22" is required and used to operate the IMU. This library has an accelerometer class named "PhidgetAccelerometer" which is used to gather acceleration data from accelerometer boards of the IMU.

Through "PhidgetAccelerometer_create()" function an instance of a Phidget channel is created. It takes the reference of the channel handle and returns a "PhidgetReturnCode". All the features of the accelerometer are accessed through this channel.

After creating the channel an event which is named "Phidget_setOnAttachHandler()" assigns a handler to be called when the IMU is attached to the system. "Phidget_openWaitForAttachment" function opens the Phidget channel by "Channel" PhidgetHandle and a timeout value in milliseconds. This function blocks the access to the device until the channel is opened or a timeout occurs. To measure the change in acceleration in real-time "PhidgetAccelerometer_setOnAccelerationChangeHandler()" event is used. This event is raised continuously throughout the run time and it uses a callback function to constantly detect the changes in acceleration values. This callback function will be explained in detail later in this paper. When the device is detached an event called "Phidget_setOnDetachHandler()" is raised. Then before the program is terminated the created channel is closed by "Phidget_close()" function to make the channel ready to for another use [13].

"WiringPi" is a C library which grants access to GPIO pins in all Raspberry Pi models. With it being written in C it can be easily used while working on C programming language. This particular library includes a command-line utility named "gpio" which serves as a pin programming interface. It is also useful to read and write the GPIO pins from shell scripts [14]. With the aid of "WiringPi" library, intended melodies can be played from the buzzer.

C. Creating the Alarm System

Once the alarm is set, the IMU starts measuring the acceleration of the sleeper 3 or 4 times second. This rate helps to reduce the memory space needed and still gives satisfactorily detailed results. In each time period the maximum change in acceleration at any time is calculated and stored in an array. Depending on these values, a maximum and a minimum threshold value are determined until the last period of the sleep time. While measurements are made continuously in the last period, also the measured values are compared with the predetermined threshold values. This comparison is made according to (1) [15].

$$S_{stage} = \begin{cases} S_{NREM}, \ if \ T_{min} \leq A_i \leq T_{max}, \\ S_{REM} \ , \ otherwise, \end{cases} \quad (1)$$

If a value is measured between predetermined threshold values, then the alarm is sounded. Otherwise, the alarm is sounded at the end of the sleep time. Fig. 7, shows the flowchart of the alarm system.

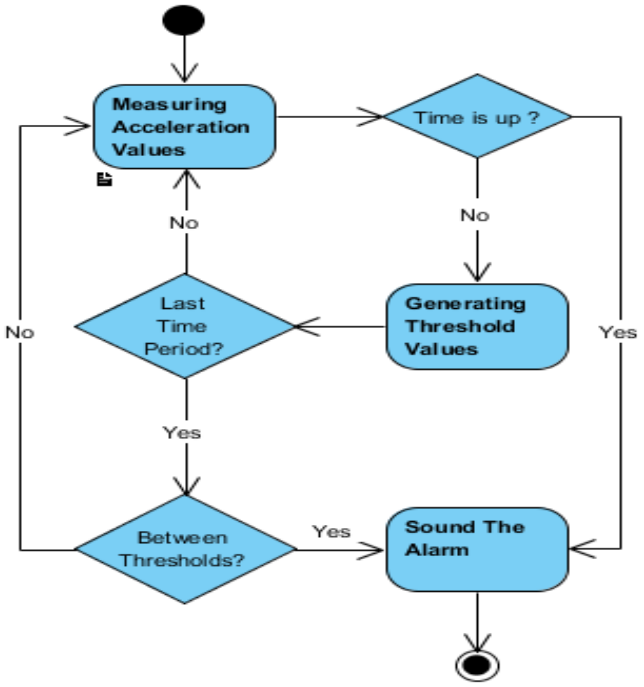


Fig. 7. Program flowchart

D. Normalizing the Acceleration Values

Normalization process is done to gather more precise acceleration values. Each acceleration value generates a 3 dimensional vector. To normalize the acceleration values, the Euclidean length of the vector is calculated and each measurement is divided by this particular value. Thanks to this operation the noise level of the measurements is reduced and all the measured values of each axes are taken into [-1 , 1] range. Calculation of the Euclidean distance in 3 dimensions is formulated in (2).

$$dist[(x,y,z),(a,b,c)] = \sqrt{(x-a)^2+(y-b)^2+(z-c)^2} \quad (2)$$

$$\begin{cases} x, & {acc_x}/{dist[(x,y,z),(a,b,c)]} \\ y, & {acc_y}/{dist[(x,y,z),(a,b,c)]} \\ z, & {acc_z}/{dist[(x,y,z),(a,b,c)]} \end{cases} \quad (3)$$

Normalization in completed as shown in (3), and based on this normalized values the Manhattan distance between back to back measurements is calculated as shown in (4). These Manhattan distance values are used as input parameters to sound the alarm.

$$\mathrm{Md(Acc_i)} = \mathrm{d(x)} + \mathrm{d(y)} + \mathrm{d(z)} \quad \begin{cases} d(x), & |x_{i+1} - x_i| \\ d(y), & |y_{i+1} - y_i| \\ d(z), & |z_{i+1} - z_i| \end{cases} \quad (4)$$

IV. Experimental Results

During the development a large amount of short-term tests are made. The aim of these short-term tests was testing whether the system worked correctly or not. Such as observing noising levels on acceleration values, controlling the timer states, checking if the buzzer worked properly and comparing the real time measurements with stored acceleration values.

After fixing the issues encountered during the short-term tests, some different long-term tests were made. Generally these long-term tests cover about 8 hours of sleep time. The outputs of these tests are written into text files and for each hour of sleep time, separate graphic charts are drawn out of them. For all the graphs drawn in this section, the x-axis represents the time in seconds and the y-axis represents the change on acceleration in g. Fig. 8 serves an example.

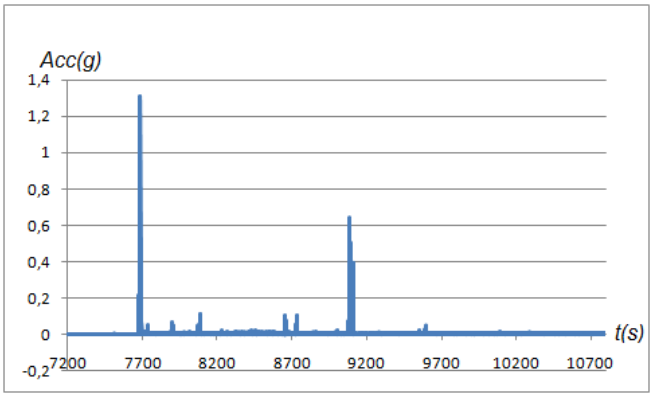


Fig. 8. Graphic chart of an hour in the sleep time

On Fig. 8, the x-axis represents time in seconds and the y-axis represents the change on acceleration in g. At the maximum peak point on the graph the subject passes to NREM period. Most of the time, the graphic charts have similar patterns when the subject is in sleep.

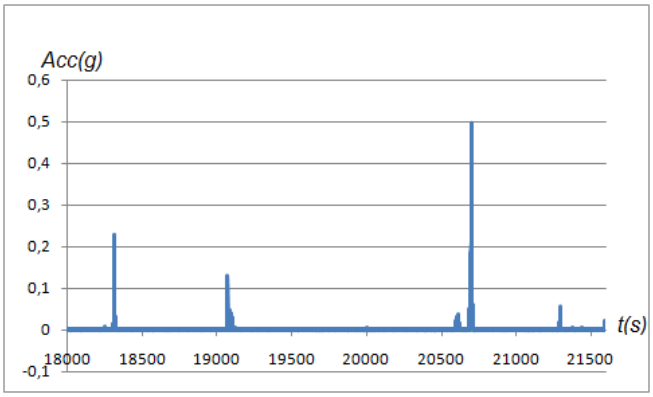


Fig. 9. Time period of minimum threshold

The minimum threshold value for this test was determined in the interval between 5th and 6th hours as 0.497g. This can be clearly seen in Fig. 9. At each hour interval the maximum change in acceleration is set as the maximum threshold value but this is updated whenever a greater value is measured. Among those maximum acceleration changes the smallest one is set as minimum threshold value.

In Fig. 10, the interval between 4th and 5th hours is graphed. The maximum threshold value was determined as 1.662g in this interval.

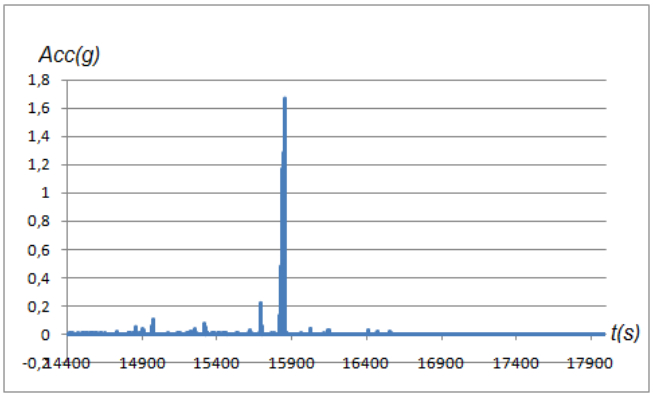


Fig. 10. Time period of maximum threshold

Due to the provision of S_{NREM} condition according to (1), in the interval between 6th and 7th hours, the measurements are stopped. Therefore the alarm sounded when 1.016g was measured. Fig. 11 depicts the awakening moment of the subject in this particular test and it can be seen that from the moment when the acceleration peaked, no more acceleration values are measured. This indicates that the alarm was sounded at this point.

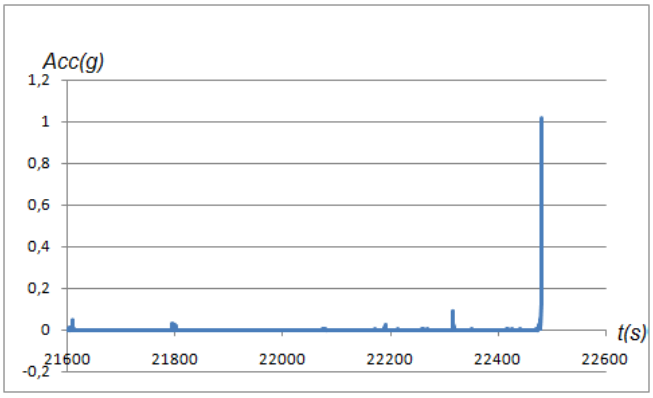


Fig. 11. Awakening moment in the last period

V. CONCLUSION

Awakening process of a human-being may prolong if they are awakened within a wrong timing. In the journey to get over this problem a good amount of experiments were conducted and many advantages of waking up within the correct timing has been observed in each particular experiment.

The tests were run on four different subjects and out of each test similar results were obtained. The result acquired belongs to one of them.

The optimal wake-up time within the last one hour varied in different tests. In the end there is a peak moment at each hour in every test. But the exact time of the NREM period can differ between the tests

As a result, a system that detects the light sleep period in the NREM stage depending on the acceleration values of the subject was built and programmed to sound the alarm when the transition to the light sleep period occurred within the desired time interval.

In order to make the system more stable, the program which operates the system can be supported by several deep learning and data mining algorithms. Long term analysis can be done according to the sleeping patterns of each particular user. Even to go further some heart-rate measuring sensors can be added on and combined with the accelerometer. In order to make the system more user-friendly, it can be integrated on a single circuit board and can be made wireless and wearable.

REFERENCES

[1] Mccarley, Robert W. Neurobiology of REM and NREM sleep. Sleep medicine, 2007, 8.4: 302-330.

[2] Roffwarg, Howard P.; Muzio, Joseph N.; Dement, William C. Ontogenetic development of the human sleep-dream cycle. Science, 1966.

[3] Dement, William C., The Stanford Sleep Book, William C. Dement, Stanford, Calif. 2006

[4] Hobson, J. Allan. REM sleep and dreaming: towards a theory of protoconsciousness. Nature Reviews Neuroscience, 2009, 10.11: 803.

[5] Gori, Sara, et al. Body movements during night sleep in healthy elderly subjects and their relationships with sleep stages. Brain research bulletin, 2004, 63.5: 393-397.

[6] Morton, Kevin, "The Five Stages Of Sleep & The Journey Through The Night", End Your Sleep Deprivation, 03 May 2017, Web, 23 Jan 2018.

[7] Virtanen, Väinö; Salmi, Aino; Salmi, Seppo. Arrangement and method to wake up a sleeping subject at an advantageous time instant associated with natural arousal. U.S. Patent No 8,493,220, 2013.

[8] Kanevsky, Dimitri; Zadrozny, Wlodek. Alarm clock system, method of operation and program product therefor. U.S. Patent No 6,928,031, 2005.

[9] Rönnholm, Valter AG. Natural alarm clock. U.S. Patent No 7,248,915, 2007.

[10] Kubey, Alan; Davidson, Andrew. REM-sleep directed visual alarm system and method. U.S. Patent No 7,956,756, 2011.

[11] PI, Raspberry. model B. URl: https://www. raspberrypi. org/products/raspberrypi-2-model-b/, Feb. 2015, Web, 18 Sep 2017.

[12] PI, Raspberry. Raspberry Pi Model B. 2015.

[13] PHIDGET SINC., "Phidget Spatial 3/3/3 – 1056_0 at Phidgets", Phidgets Inc. – Products for USB Sensing and Control, © Phidgets Inc. 2016, Web, 18 Sep 2017.

[14] Henderson, Gordon. WiringPi. Wiring Pi GPIO Interface library for the Raspberry Pi, 2013.

[15] Nam, Yunyoung; Kım, Yeesock; Lee, Jinseok. Sleep monitoring based on a tri-axial accelerometer and a pressure sensor. Sensors, 2016, 16.5: 750.

Splitting Tensile Strength of Concrete Containing Zeolite and Diatomite under the Effect of H2SO4 by ANN

YILMAZ KOÇAK[1], GIYASETTİN ÖZCAN[2], MUAMMER AKÇAY*[3] and EYYÜP GÜLBANDILAR[4]

[1] Duzce University, Duzce/Turkey, yilmazkocak@duzce.edu.tr
[2] Uludag University, Bursa/Turkey, gozcan@uludag.edu.tr
[3] Dumlupinar University, Kutahya/Turkey, makcay26@gmail.com
[4] Eskisehir Osmangazi University, Eskisehir/Turkey, egulbandilar@ogu.edu.tr

***Abstract* – In this study, it was designed to investigate with four different artificial neural network (ANN) prediction models for the concrete behavior containing zeolite and diatomite under the effect of H2SO4. The constructing purpose of this model, 3 different mixes with 27 specimens of the 28, 56 and 90 days splitting tensile strength experimental results of concrete containing zeolite and diatomite. Those experimental results used in training and testing for ANN systems were gathered from the tests. The days (age of samples), Portland cement, zeolite, diatomite, aggregate, water and hyper plasticizer as input data parameters and splitting tensile strength of concrete as an output parameter are used in the ANN models. The four different ANN models have strong potential as a feasible tool for predicting 28, 56 and 90 days the splitting tensile strength of concrete containing zeolite and diatomite in according to the training and testing results.**

***Keywords* - ANN, cement, zeolite, diatomite, splitting tensile strength.**

I. INTRODUCTION

CONCRETE, which is one of the most widely used and popular artificial construction materials, plays an important role in building technology in the world. Concrete which is a composite material, consists homogenous mixtures of cement, aggregate (fine and coarse aggregate), water, and sometime chemical admixture and mineral additives [1-2].

There are many different kinds of cements. In concrete, the most commonly used is Portland cement or blended cement. Due to economic and ecological factors like trass [3], zeolite [4], diatomite [5-6], metakaolin [7-8], pumice [9], fly ash [10-11], blast furnace slag [12], and silica fume [13-14] are intensely used in the cement and concrete technology. Some characteristics such as strength, durability and low permeability expected from good concrete are closely related not only to mix proportions but also to cement properties. Zeolite and diatomite are natural mineral material, and are abundant in our country.

Zeolite is defined as allophones that consist of alkali. Alkaline–earth cations have the crystal structure. Zeolites have water molecules in their canals. Those canals are one of the most significant properties setting apart them from other mineral groups [15-16]. Diatomite is a mineral. Diatomite consists of the fossilized siliceous shell of the microscopic single–celled alga and has possess the structural properties of amorphous silica. There are nearly fifteen thousand diatomite types in the nature. The shape of diatomite is generally a round tray or a long fish. They contain 70–90% of SiO2. They are cellular high water absorption rate materials [17].

Nowadays, many researchers have used expert systems to solve a wide variety of problems in civil engineering applications. The expert systems are Adaptive Network-based Fuzzy Inference Systems (ANFIS), fuzzy system, and so on. Artificial neural network (ANN) is one of the alternative approaches for the predicting mechanical behavior and physical properties of concrete and cement mortars [18-22]. The purpose of this study is to build an ANN system model to evaluate the effect of splitting tensile strength of concrete containing zeolite and diatomite. In the study, 7 different cements are used. Those cements PC, 5+5–10+10% diatomite and zeolite are substituted for Portland cement. For purpose of constructing this model, 3 different mixes with 27 specimens of the 28, 56 and 90 days splitting tensile strength experimental results. The splitting tensile strength experimental results of concrete containing zeolite and diatomite used in training and testing for ANN system were gathered from the concrete tests. Those 27 experimental data results were used to train the model. The ANN model had seven input parameters. Those input parameters are the age of samples (days), Portland cement, zeolite, diatomite, aggregate, water and hyper plasticizer. An output parameter is splitting tensile strength of concrete. The obtained results were compared with predicted results for splitting tensile strength of concrete.

II. EXPERIMENTAL STUDY

A. Materials

In this study, Portland cement, diatomite, zeolite, aggregate, hyper plasticizer and water are used as materials. Bolu Cement

Plant provided the cement which is CEM I 42.5 R (PC). Table 1 shows the chemical properties of PC. The physical and mechanical properties of PC are given at the Table 2. Diatomite from Kutahya region and zeolite from Balikesir–Bigadic region are used as a pozzolan. A diatomite is supplied by ASU Chemistry and Mining Firm and zeolite is supplied by a Turkish Zeolite Firm. The chemical properties of zeolite and diatomite are given at the Table 1, and physical and mechanical properties of zeolite and diatomite are given at the Table 2. Asar River aggregates are used in Duzce region as aggregate (crashed sand and crushed stone). Table 3 shows the physical properties of the aggregates. Furthermore, the type of fluid 70 produced by AYDOS Construction Chemicals Factory and new generation hyper plasticizer with solid matter content of 34.32%, intensity of 1.184 (20oC), pH value of 7.26 (20oC) are applied as concrete admixture. As a result, well water is used as mixing water from Doganli village in Duzce.

Table 1: The chemical properties of PC.

Materials	PC	Diatomite	Zeolite
Chemical composition, wt.%			
SiO_2	18.68	79.56	68.85
Al_2O_3	4.67	6.54	11.71
Fe_2O_3	3.53	2.76	1.29
CaO	64.56	2.45	3.97
MgO	0.98	0.79	1.06
SO_3	3.00	0.48	0.18
Na_2O	0.14	2.63	0.29
K_2O	0.73	0.69	2.19
S+A+F	-	88.86	81.85
Loss on ignition	3.92	3.88	10.00
Insoluble residue	0.50	75.98	37.32
Free CaO	1.74	-	-

Table 2: The physical and mechanical properties of PC.

Materials	Compressive strength, MPa		Setting time, minute		Blaine, cm^2/g	Specific gravity
	7 days	28 days	Initial	Final		
PC	29.6	52.8	118	-	4249	3.17
Diatomite	-	-	-	-	13640	2.28
Zeolite	-	-	-	-	5740	2.18

Table 3: The chemical properties of PC.

		Test Results			Standards
Unit weight, g/cm^3	Loose unit weight	1.48			TSEN 1097-3
	Dense unit weight	1.66			
Water absorption		Aggregate grading			TSEN 1097-6
		0-5, mm	**5-19, mm**	**19-30, mm**	
	Water absorption, %	0.61	1.16	1	
Moisture content, %		1.25	1.32	1.41	
Determination of organic impurities		The color of the liquid is light yellow color than colorless (Organic matter is harmless).			TSEN 1744-1

B. Methods

In this study, 3 different cements are used. Those cements are PC, 5+5–10+10% diatomite and zeolite are substituted for Portland cement. Amounts of materials to be put in the mixture are determined according to the framework of the method stated in TS 802 standards [23]. Three types of concrete are produced in according to the type and rate of mineral additive, which is substituted for the concrete. Those concretes are encoded as R, 5D5Z and 10D10Z in according to the addition rate and the used mineral additive. Consistency of fresh concrete is stated for each mixing group individually in according to TS EN 12350–2 [24]. Table 4 shows the materials amounts in concrete mixture of 1m3 sample and the characteristics of fresh concrete.

The produced concretes are poured into 15x15x15 cm cubic molds without segregation. These concretes are retained for 24 hours in the molds and harden. Then, they are cured in 23±2oC water for 28 days. After 28 days, concrete specimens extracted from the water were put into 5% H2SO4 solution and subjected to chemical effects for 56 and 90 days. The hardened concrete specimens were then subjected to splitting tensile strength tests on days 28, 56 and 90 according to TS EN 12390-6.

Table 4: Material quantity in the 1 m^3 for each concrete group.

Materials		Specific gravity	R, kg	5D5Z, kg	10D10Z, kg
Aggregate, mm	0-5	2.66	822	849	855
	5-19	2.69	586	606	611
	19-30	2.70	428	442	445
Total			1836	1897	1911
PC		3.17	400	360	320
Diatomite		2.28	-	20	40
Zeolite		2.18	-	20	40
Hyper plasticizer		1.184	4.800	4.320	3.840
Water		1	139.7	139.7	124.2

III. ARTIFICIAL NEURAL NETWORK

Artificial neural network (ANN) consists of neurons. Neurons are called an arbitrary number of simple elements. Neurons which is similar to the human brains in ANN are interconnected [26]. ANN shows simplified methods of a human brain. ANN uses new methods rather than conventional methods to solve the problems when traditional computational methods have difficult solution procedures [27]. Generally, ANN is consisted of input layer neurons, one or more hidden layer neurons and output layer neurons. The neighboring layers are fully interconnected by appropriate weights. The input layer neurons receive information from the outside environment. The input layer neurons transmit them to the hidden layer neurons without performing any calculation. Hidden layers are located between the input and output layers. Hidden layers may contain a large number of hidden processing units. All problems can be solved by a perceptron which is called one hidden layer. However, using two hidden layers is sometimes more efficient. Finally, the output layer neurons produce the network predictions to the outside world [28].

The typical neural network is composed of inputs, weights, sum function, activation function and output parts. The typical neural network is illustrated in Figure 1 [29, 30].

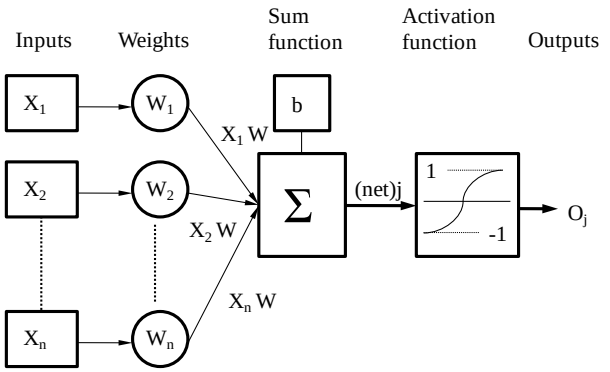


Figure 1: The artificial neuron model.

The input layer neuron comes from another neuron and it is obtained by multiplying the output of the connected neuron by the synaptic strength of the connection between them. The weighted sum of the input components (net)j are calculated by using Eq. (1) below:

$$(\text{net})_j = \sum_{i=1}^{n} w_{ij} o_i + b \qquad (1)$$

Where $(\text{net})_j$ is the weighted sum of the j^{th} neuron for the input which is received from the preceding layer with n neurons. Where w_{ij} is the weight between the j^{th} neuron in the preceding layer. And o_i is the output of the i^{th} neuron in the preceding layer and b is a fix value as an internal addition [29]. Activation function processes the net input obtained from sum function and determines the neuron output. In general, the sigmoid activation function (f $(\text{net})_j$) for multilayer feed forward models is used. The output of the j^{th} neuron $(\text{out})_j$ is computed by using Eq. (2) with a sigmoid activation function as follows [30]:

$$o_j = f(\text{net})_j = \frac{1}{1+e^{-\alpha(\text{net})_j}} \qquad (2)$$

Where α is a constant and it is used to control the slope of the semi-linear region. The sigmoid nonlinearity function activates in every layer except in the input layer. The sigmoid function gives outputs between (0, 1) represented by Eq. (2). If it desired, the outputs of this function o can be adjusted to (-1, 1) interval. Since the sigmoid processor represents a continuous function, it is particularly used in non-linear descriptions. The derivation of this function can be determined easily with regard to the parameters within $(\text{net})_j$ variable [30].

IV. Artificial Neural Network Models and Parameters

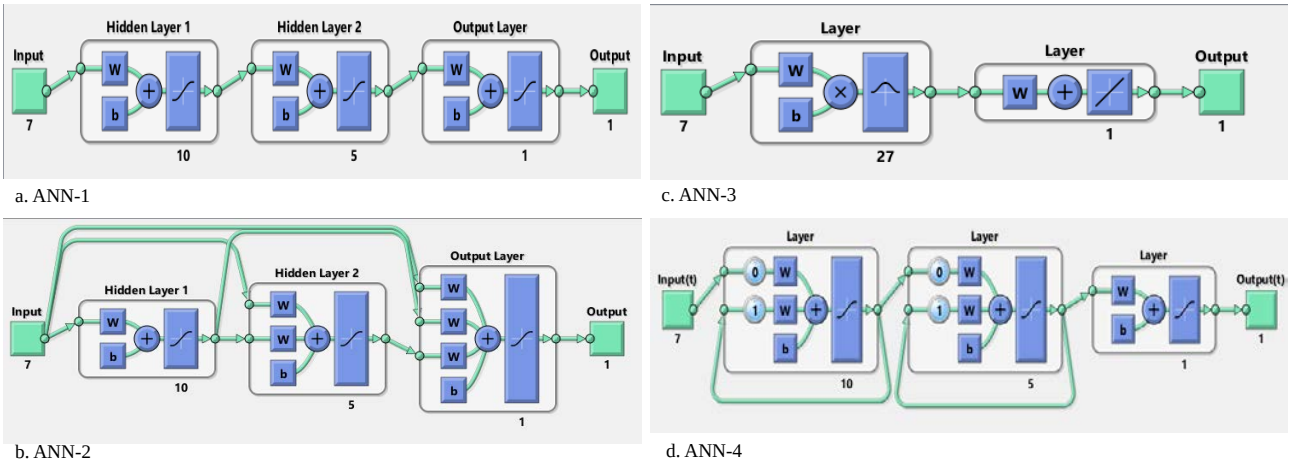
The input parameters were entered as the age of samples, PC, zeolite, diatomite, aggregate, water and hyper plasticizer in according to the Table 5. The output parameter was the splitting tensile strength of concrete in according to Table 5. Both input and output parameters were used for training and testing of the ANN model.

Table 5: The input and output quantities used in ANN Model.

		Data used in training and testing the model	
		Minimum	Maximum
Input variables	Age of samples, days	28	90
	PC, g	320	400
	Zeolite, g	0	80
	Diatomite, g	0	80
	Aggregate	1836	1912
	Water	123.2	139.7
	Hyper plasticizer	3.840	4.800
Output variable	**Splitting tensile strength, MPa**	**5**	**8.3**

To train the model, experimental 27 data were used. To test the trained model, 9 data as the average of these test results were used. The designed ANN-1 consisted of feed-forward back propagation, two hidden layers, training function (Levenberg-Marquardt), adaptation learning function (learngdm), transfer function (tansig) and performance function (MSE-mean squared error). The designed ANN-1 is demonstrated in Figure 2-a. The designed ANN-2 involved cascade-forward back propagation, two hidden layers, training function (Levenberg-Marquardt), adaptation learning function (learngdm), transfer function (tansig) and performance function (MSE-mean squared error). The designed ANN-2 is shown Figure 2-b. The designed ANN-3 model is determined generalized regression and one hidden layer. The ANN-3 is shown in Figure 2-c. The ANN-3 model does not include the training phase and it is created the regression model. The ANN-4 model consists of Elman back propagation, two hidden layers, training function (Levenberg-Marquardt), adaptation learning function (learngdm), transfer function (tansig) and performance function (MSE-mean squared error). The ANN-4 is shown in Figure 2-d. The input values of each layer in ANN-2 model is sum of multiplied by weights of the outputs of all previous layers. Elman networks are feedforward networks with the addition of layer recurrent connections with tap delays in ANN-3.

Momentum and learning rate values were determined. The models were trained through iterations. The parameter values obtained from the multilayer feed-forward neural network models. Those parameter values were given in Table 6.



a. ANN-1

c. ANN-3

b. ANN-2

d. ANN-4

Figure 2: The architectures used in the ANN for splitting tensile strength.

Table 6: The values of parameters used in models.

Parameters	ANN-1	ANN-2	ANN-3	ANN-4
Number of input layer neurons	7	7	7	7
Number of hidden layer	2	2	-	-
Number of layer	4	4	3	4
Number of first hidden layer neurons	10	10	27	10
Number of second hidden layer neurons	5	5	1	5
Number of third hidden layer neurons	1	1	-	1
Number of output layer neuron	1	1	1	1
Error after learning	$1x10^{-7}$	$1x10^{-8}$	-	$1x10^{-7}$
Learning cycle	4	5	-	4

The trained models were tested only with the input values. As a result, the predicted results were close to the experimental results.

V. RESULTS AND DISCUSSION

Two or three hidden layers were used to find more reliable solutions with multilayer feedforward network models. It is very important to accurately predict the parameters used by ANN models. The determinations of optimum number of the hidden layers' neurons in ANN model is important. To be able to find the optimum number of hidden neurons, it is started with a few numbers of neurons and then slightly increasing the number of neurons gives the best approach. For each hidden neuron number, the performance of the ANN model is monitored during this process in according to chosen performance criteria. This process will be repeated until the error becomes acceptably small or no significant improvement is observed.

In this study, it is used different neurons in the two or three hidden layers at the beginning of the process. Then, the number of neurons was increased step-by-step adding 1 neuron until no significant improvement is noted. The ANN models tried to be compared according to a root-mean squared (RMS) error criteria, the absolute fraction of variance (R2) and mean absolute percentage error (MAPE). These criteria are defined by Eqs. (3), (4) and (5) respectively [31].

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left|t_i - o_i\right|^2} \tag{3}$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{N}\left(t_i - o_i\right)^2}{\sum_{i=1}^{N}\left(o_i\right)^2} \right) \tag{4}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left| \left(\frac{t_i - o_i}{o_i} \right) \right| * 100 \tag{5}$$

Where the target value is ***t***, the network output value is ***o***, the total number of pattern is ***N***. Experimental data and average of these test results are used in the training and testing of ANN models.

In the ANN models, 27 data of experiment results were used for training. Whereas 9 data as average of these test results were used for testing. Sample number and experimental results for training and testing results which were obtained from ANN models are shown in Fig. 3 and 4, respectively. All results which were obtained from the studies and predicted ANN models by using the training and testing results for 28, 56 and 90 days splitting tensile strength of concrete were given in Figure 5 and 6, respectively.

Also, Table 7 shows input values and experimental results for testing results obtained from ANN models.

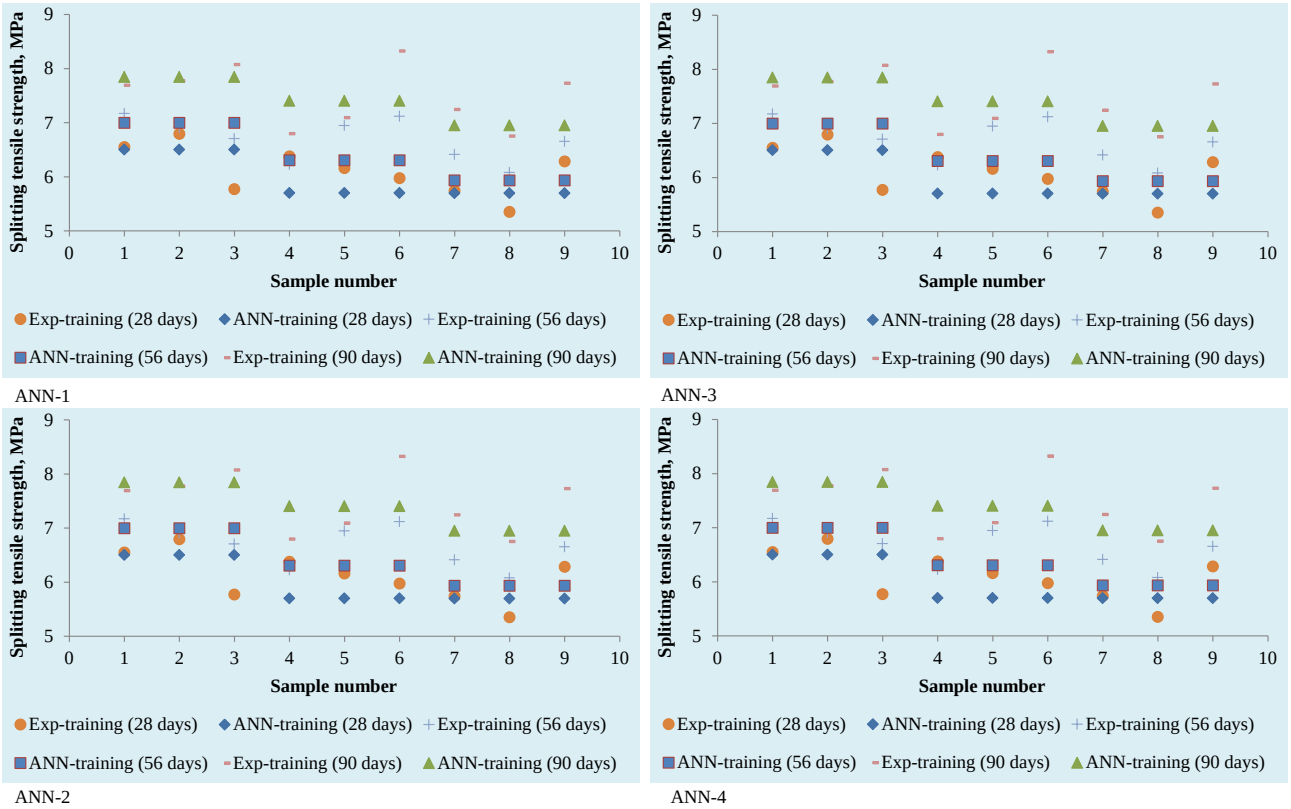


Figure 3: Comparison of splitting tensile strength experimental and training results with sample number.

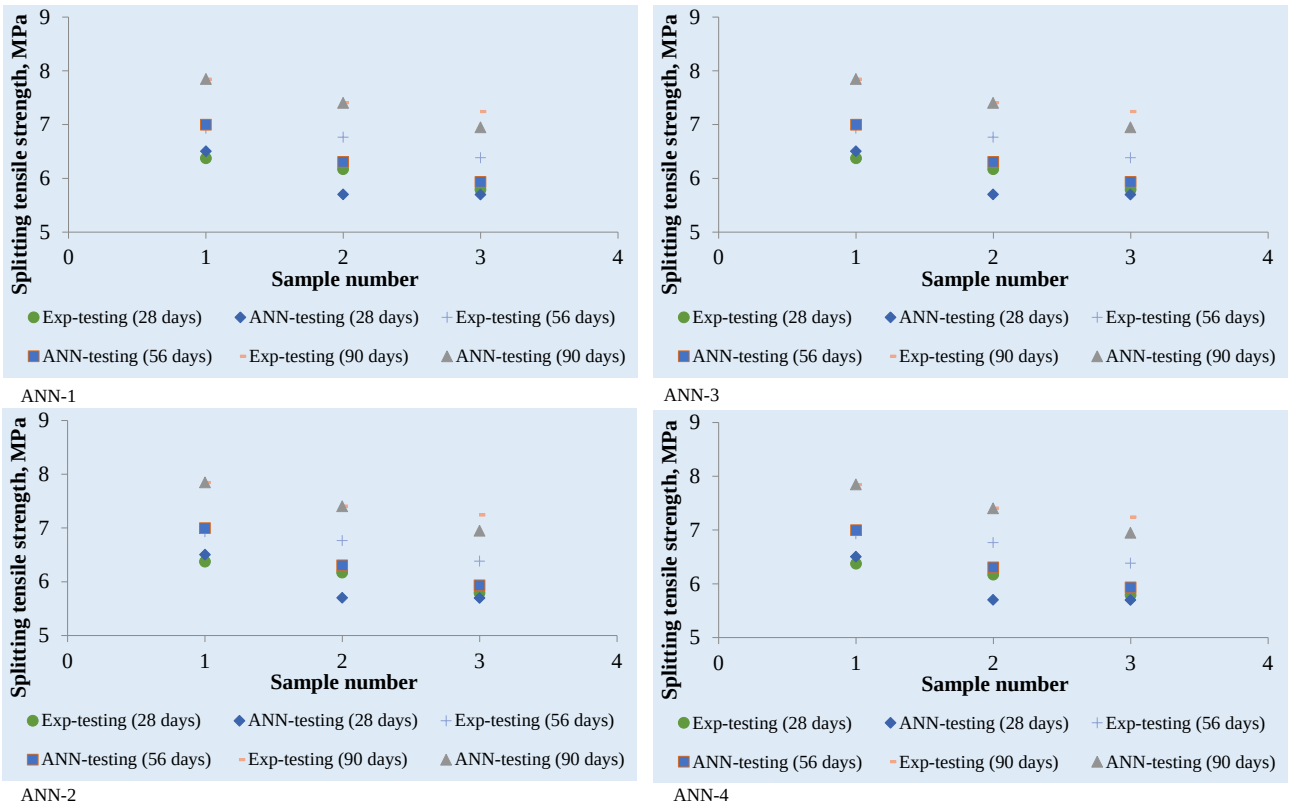
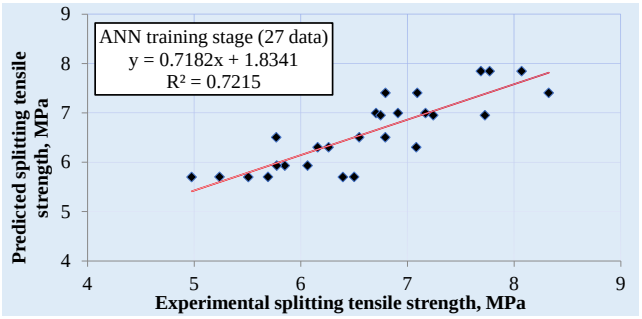
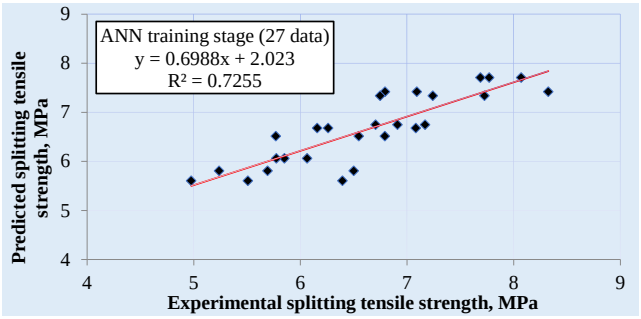


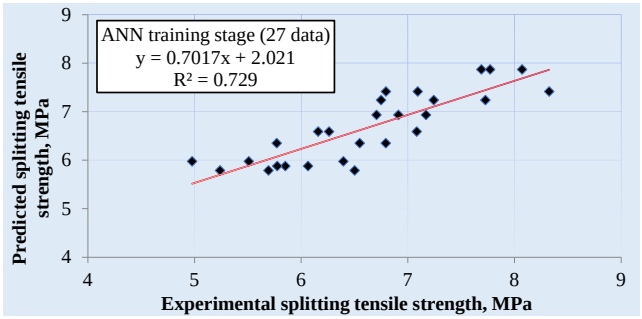
Figure 4: Comparison of splitting tensile strength average of test results and testing results with sample number.



ANN-1

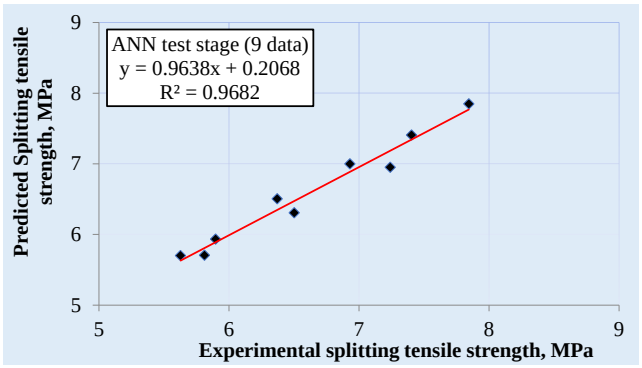


ANN-2

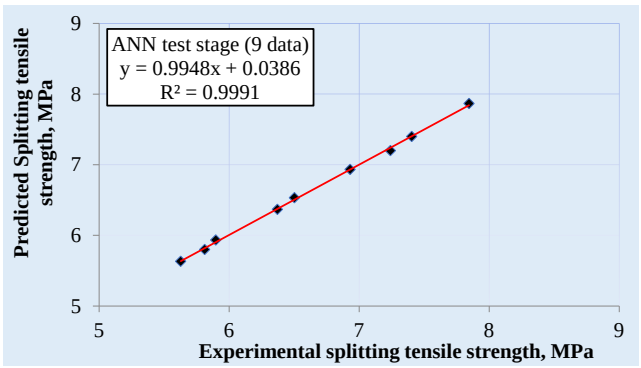


ANN-4

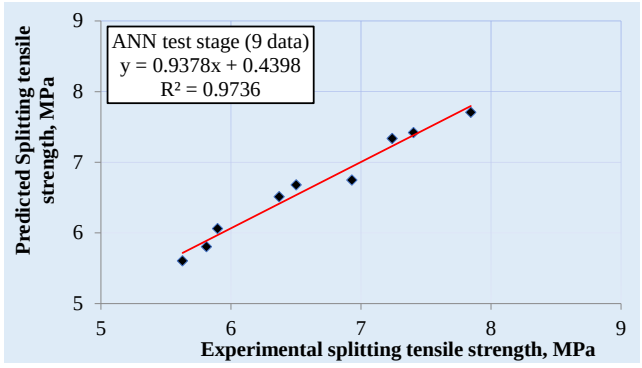
Figure 5: Comparison of splitting tensile strength experimental results with training results of model.



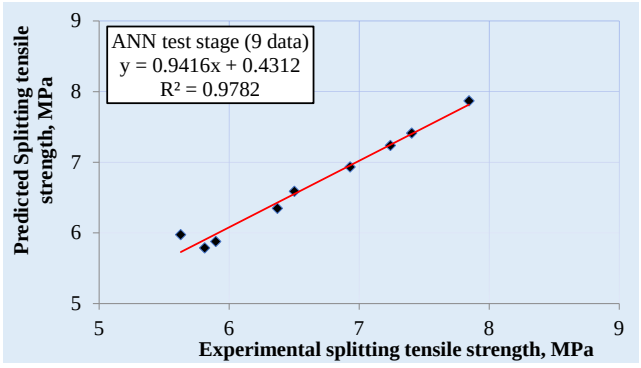
ANN-1



ANN-3



ANN-2



ANN-4

Figure 6: Comparison of splitting tensile strength average of experimental results with testing results of model.

Table 7: Comparison of splitting tensile strength average of test results with testing results obtained from ANN.

Data used in the model construction							Splitting tensile strength, MPa				
As, days	PC, kg	Z, kg	D, kg	A, kg	W, kg	HP, kg	Exp.	ANN-1	ANN-2	ANN-3	ANN-4
28	400	0	0	1836	139.7	4.800	**6.4**	6.5	6.5	6.3	6.3
28	360	20	20	1897	139.7	4.320	**5.8**	5.7	5.8	5.8	5.7
28	320	40	40	1911	124.2	3.840	**5.6**	5.6	5.6	5.6	5.9
56	400	0	0	1836	139.7	4.800	**6.9**	7.0	6.7	6.9	6.9
56	360	20	20	1897	139.7	4.320	**6.5**	6.3	6.6	6.5	6.5
56	320	40	40	1911	124.2	3.840	**5.9**	5.9	6.0	5.9	5.8
90	400	0	0	1836	139.7	4.800	**7.8**	7.8	7.7	7.8	7.8
90	360	20	20	1897	139.7	4.320	**7.4**	7.4	7.4	7.4	7.4
90	320	40	40	1911	124.2	3.840	**7.2**	6.9	7.3	7.2	7.2

For training and testing data, the linear least square fit line, its equation and the R2 values were shown in these figures. The obtained values from the training and testing in ANN model are very closer to the experimental results. This can be shown in Fig. 5 and 6. It is shown that the ANN models can generalize between input and output variables with reasonably good predictions in according to the result of testing phase in Fig. 5 and 6.

For all the station, the statistical RMS, R2 and MAPE values were given in Table 8.

Table 8: The splitting tensile strength statistical values of proposed ANN models.

Statistical parameters		R^2	RMS	MAPE
ANN-1	Training set	0.7215	0.4491	0.0530
	Testing set	0.9682	0.8177	0.0863
ANN-2	Training set	0.7255	0.4463	0.0560
	Testing set	0.9736	0.7982	0.0952
ANN-3	Training set	-	-	-
	Testing set	0.9991	0.8334	0.0911
ANN-4	Training set	0.7290	0.4440	0.0568
	Testing set	0.9782	0.8411	0.0930

The statistical values of R2, RMS and MAPE for the training ANN-1 model were found as 0.7215, 0.4491 and 0.0530, respectively. These statistical values of R2, RMS and MAPE for testing ANN-1were found as 0.9682, 0.8177 and 0.0863, respectively (Table 8). The statistical values for training of ANN-2 model were found as 0.7255, 0.4463 and 0.0560. The statistical values for testing ANN-2 model were found 0.9736, 0.7982 and 0.0952. For testing of ANN-3 model were found as 0.9991, 0.8334 and 0.0952. For training of ANN-4 model were found as 0.7290, 0.4440 and 0.0568 and these values for testing were found 0.9782, 0.8411 and 0.0930. All the statistical values show that the proposed ANN models are suitable. The proposed ANN models can predict splitting tensile strength values very close to the experimental values for the 28, 56 and 90 days.

VI. CONCLUSION

In this study, ANN models were used to predict the 28, 56 and 90 days splitting tensile strength values of concrete which contains zeolite and diatomite. In the developed models in ANN-1, 2 and 4 systems, three hidden layers were selected. There were 10 neurons in the first hidden layer. There were 5 neurons in the second hidden layer. And there were 1 neuron in the third hidden layer. The developed model in ANN-3 system, two hidden layers were selected. It is determined 27 neurons in the first hidden layer and 1 neuron in the second hidden layer. These developed models were trained with the input and output experimental data. The determined splitting tensile strength values are very closer to the experimental data which are obtained from testing for all the developed models. The statistical R2, RMS and MAPE parameter values were calculated to compare experimental data with four ANN model's results. It is found statistically significant correlations for the four developed models.

As a result, the splitting tensile strength values of concrete which contains zeolite and diatomite, can be predicted by using the ANN's. These predictions takes a quite short period of time with tiny error rates. As a conclusion, the ANN systems are practicable methods. The ANN systems can predict splitting tensile strength values of concrete containing zeolite and diatomite.

ACKNOWLEDGMENT

The authors would like to thank Duzce University Presidency of Scientific Research Projects. Duzce University Presidency of Scientific Research Projects provided financial support with the project code number 2011.03.HD.009. The authors would like to thank Duzce Yigitler Beton that enabled the tests to be carried out.

REFERENCES

[1] T.Y. Erdoğan, *Beton*, ODTÜ Geliştirme Vakfı Yayıncılık ve İletişim AŞ., ISBN / ISSN: 975–7064–67–x, Ankara, Türkiye, 2007.

[2] A. M. Neville, *Properties of Concrete, Fourth Edition*, Pearson Education Limited, ISBN-13: 978-0-582-23070-5, England, 2006.

[3] Y. Koçak, A. Dorum, B. Yılmaz and A. Uçar, "Trasın Çimento Yüzey Özelliğine, Hidratasyona ve Basınç Dayanımına Etkisi", *E–Journal Of*

New World Sciences Academy Technological Applied Sciences, vol. 5, no. 1, pp. 1–14, 2010.

[4] Y. Kocak, E. Tascı and U. Kaya, "The effect of using natural zeolite on the properties and hydration characteristics of blended cements", *Construction and Building Materials*, vol. 47, pp. 720–727, 2013.

[5] Y. Kocak and M. Savaş, "Effect of the PC, diatomite and zeolite on the performance of concrete composites", *Computers and Concrete*, vol. 17, no. 6, pp. 815–829, 2016.

[6] H. Gerengi, Y. Kocak, A. Jażdżewska, M. Kurtay and H. Durgun, "Electrochemical investigations on the corrosion behaviour of reinforcing steel in diatomite– and zeolite–containing concrete exposed to sulphuric acid", *Construction and Building Materials*, vol. 49, pp. 471–477, 2013.

[7] O. Keleştemur and B. Demirel, "Effect of metakaolin on the corrosion resistance of structural lightweight concrete", *Construction and Building Materials*, vol. 81, pp. 172–178, 2015.

[8] A. Subaşı and M. Emiroğlu, "Effect of metakaolin substitution on physical, mechanical and hydration process of White Portland cement", *Construction and Building Materials*, vol. 95, pp. 257–268, 2015.

[9] K. Yıldız, A. Dorum and Y. Koçak, "Pomza Zeolit Ve Cem I Çimentosunun Minerolojik Moleküler Elektrokinetik Ve Termal Uyumunun Yüksek Dayanımlı Betona Etkisinin Araştırılması", *Journal of The Faculty of Engineering and Architecture of Gazi University*, vol. 25, no. 4, pp. 867–879, 2010.

[10] Y. Kocak and S. Nas, "The effect of using fly ash on the strength and hydration characteristics of blended cements", *Construction and Building Materials*, vol. 73, pp. 25–32, 2014.

[11] L. Zhengqi, "Drying shrinkage prediction of paste containing meta–kaolin and ultrafine fly ash for developing ultra–high performance concrete", *Materials Today Communications*, vol. 6, pp. 74–80, 2016.

[12] H. Zhao, W. Sun, X. Wu and B. Gao, "The properties of the self–compacting concrete with fly ash and ground granulated blast furnace slag mineral admixtures", *Journal of Cleaner Production*, vol. 95, pp. 66–74, 2015.

[13] F. N. Okoye, J. Durgaprasad and N. B. Singh, "Effect of silica fume on the mechanical properties of fly ash based–geopolymer concrete", *Ceramics International*, vol. 42, pp. 3000–3006, 2016.

[14] Y. Kocak, "A Study on the effect of fly ash and silica fume substituted cement paste and mortars", *Scientific Research and Essays*, vol. 5, no. 9, pp. 990–998, 2010.

[15] F. Canpolat, "Çimento performansının geliştirilmesinde doğal zeolitin endüstriyel atıklarla birlikte çimento üretiminde kullanılması", Doktora Tezi, Sakarya Üniversitesi Fen Bilimleri Enstitüsü, Sakarya, 2002.

[16] D. Serbest, "Doğal zeolitlerin hafif yapı endüstrisinde kullanımı", Yüksek Lisans Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Eskişehir, 1999.

[17] H.Y. Aruntas and M. Tokyay, "The use of diatomite as pozzolanic material in blended cement production", *Cement and Concrete World*, vol. 1, no.4, pp. 33–41, 1996.

[18] A. Beycioglu, M. Emiroglu, Y. Kocak and S. Subasi, "Analyzing the compressive strength of clinker mortars using approximate reasoning approaches – ANN vs MLR", *Computers and Concrete*, vol. 15, no. 1, pp. 89-101, 2015.

[19] H. R. Ashrafi, M. Jalal and K. Garmsiri, "Prediction of load–displacement curve of concrete reinforced by composite fibers (steel and polymeric) using artificial neural network", *Expert Systems with Applications*, vol. 37, pp. 7663–7668, 2010.

[20] U. Atici, "Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network", *Expert Systems with Applications*, vol. 38, pp. 9609–9618, 2011.

[21] H. Yaprak, A. Karaci and I. Demir, "Prediction of the effect of varying cure conditions and w/c ratio on the compressive strength of concrete using artificial neural networks", *Neural Computing and Application*, vol. 22, pp. 133–141, 2013.

[22] S. Subasi, "Prediction of mechanical properties of cement containing class C fly ash by using artificial neural network and regression technique", *Scientific Research and Essay*, vol. 4, no. 4, pp. 289–297, 2009.

[23] Türk Standartları Enstitüsü (TSE), "Beton karışımı hesap esasları", TS 802, Ankara, Türkiye, 2009.

[24] Türk Standartları Enstitüsü (TSE), "Beton - Taze beton deneyleri - Bölüm 2: Çökme (slamp) deneyi", TS EN 12350–2, Ankara, Türkiye, 2010.

[25] Türk Standartları Enstitüsü (TSE), "Beton - Sertleşmiş beton deneyleri - Bölüm 6: Deney numunelerinin yarmada çekme dayanımının tayini", TS EN 12390–6, Ankara, Türkiye, 2010.

[26] B. B. Adhikary and H. Mutsuyoshi, "Prediction of shear strength of steel fiber RC beams using neural networks", *Construction and Building Materials*, vol. 20, no. 9, pp. 801–811, 2006.

[27] F. Demir, "Prediction of elastic modulus of normal and high strength concrete by artificial neural network", *Construction and Building Materials*, vol. 22, no. 7, pp. 1428–35, 2008.

[28] I. B. Topcu, C. Karakurt and M. Saridemir, "Predicting the strength development of cements produced with different pozzolans by neural network and fuzzy logic", *Materials and Design*, vol. 29, pp. 1986–1991, 2008.

[29] R. Parichatprecha and P. Nimityongskul, "Analysis of durability of high performance concrete using artificial neural Networks", *Construction and Building Materials*, vol. 23, pp. 910–917, 2009.

[30] F. Ozcan, C. D. Atis, O. Karahan, E. Uncuoglu and H. Tanyildizi, "Comparison of artificial neural network and fuzzy logic models for prediction of long–term compressive strength of silica fume concrete", *Advances in Engineering Software*, vol. 40, pp. 856–863, 2009.

Predicting the Compressive Strength of Concrete Containing Zeolite under the Effect of H_2SO_4 by ANFIS

GIYASETTIN ÖZCAN[1], MUAMMER AKÇAY[2], YILMAZ KOÇAK[*3] and EYYÜP GÜLBANDILAR[4]

[1] Uludag University, Bursa/Turkey, gozcan@uludag.edu.tr
[2] Dumlupinar University, Kutahya/Turkey, muammer.akcay@dpu.edu.tr
[3] Duzce University, Duzce/Turkey, yilmazkocak@duzce.edu.tr
[4] Osmangazi University, Eskisehir/Turkey, egulbandilar@ogu.edu.tr

Abstract **- This study was designed to investigate with ANFIS (Adaptive network-based fuzzy inference systems) prediction model for the behavior of concrete containing zeolite under the effect of H_2SO_4. For purpose of constructing this model, 3 different mixes with 27 specimens of the 28, 56 and 90 hydration days compressive strength experimental results of concrete containing zeolite used in training for ANFIS system was gathered from the tests. The data used in the ANFIS model are arranged in a format of six input parameters that cover the age of samples, Portland cement, zeolite, aggregate, water and hyper plasticizer and an output parameter which is compressive strength of concrete. In the model, the testing results have shown that ANFIS system has strong potential as a feasible tool for predicting 28, 56 and 90 hydration days the compressive strength of concrete containing zeolite under the effect of H_2SO_4.**

Keywords **- ANFIS, Portland cement, zeolite, compressive strength.**

I. INTRODUCTION

Concrete is one of the most commonly utilized building material in the world. As a result of technological innovations, concrete usage rate is improving more and more. In order to improve the manifold properties of fresh and hardened concrete, artificial (such as fly ash, silica fume, blast furnace slag) [1-6] or natural pozzolan (such as tuff "trass", diatomite, pumice, zeolite) are used [7-13].

There are 150 kinds of zeolite minerals known in nature. Recently, natural zeolite has been use in cement and concrete industry, and zeolite is a natural mineral found in large quantities in Turkey. In the nature, Zeolites exists in hydrated aluminosilicate minerals, which are emerged by the alteration of glass-rich volcanic rocks [14]. Natural zeolites have high amounts of reactive SiO_2 and Al_2O_3. Zeolites are minerals with crystal structure, and because of their favorable chemical properties, they are now an important raw material [15].

In today's civil engineering applications, artificial intelligence techniques are widely used to resolve manifold problems. For instance, ANFIS prediction model is used in the literature to estimate manifold properties of concrete. For instance, Kocak and Gulbandilar tried estimate the zeolite and diatomite-substituted concrete attitudes under the $MgSO_4$ in varied hydration days by using the ANFIS, and consequently stated that ANFIS model is a useful and powerful model according to the obtained results [16]. In another study, Yuan et al. tried estimate the structural and non-structural factors affecting concrete quality with ANFIS, and consequently stated that ANFIS model is a handy, appropriate, fair, quick and cheap method [17]. Additionally, modeling studies with ANFIS have recently been carried out in order to predict manifold properties of concretes prepared in varied structures and properties and it has been stated that values close to experimental results are obtained [18-22].

The objective of this study is to present a new ANFIS model that estimates compressive strength of concrete containing zeolite under the exposure to H_2SO_4. In the study, 3 varied cements, which are Portland cement, 10+10% zeolite, are used. Zeolite is substituted in Portland cement. In order to create the ANFIS model, 3 mixes are considered where each mix have distinguished properties including 27 specimens of the 28, 56 and 90 hydration days compressive strength experimental results of concrete. As mentioned, the cement is subsitite to zeolite in the meantime of ANFIS training and testing. During the ANFIS model, each set of training and testing data includes 7 parameters. Among the seven parameters, 6 of the parameters are assumed as input data. These input parameters are the age of samples (days), Portland cement, zeolite, aggregate, water and hyper plasticizer. Remaining 1 parameter is the compressive strength and it is assumed as the output data. In the ANFIS model we compare the experimental study of compressive strength results against compressive strength result predictions of ANFIS.

II. EXPERIMENTAL STUDY

A. Materials

Aggregate, Portland cement, natural zeolite, hyper plasticizer and water were used in the experimental study. We present the physical properties of the aggregates in Table 1. The aggregates are procured from the Asar River in Duzce region of Turkey, were given in Table 1. Utilized Portland cement (PC) is CEM I 42.5 R and obtained from Bolu Cement Plant. Natural zeolite of the clinoptilolite type is supplied from a Turkish Zeolite Firm. The chemical composition, physical and mechanical properties of PC and zeolite used in the study were given in Table 2. New generation hyper plasticizer as chemical admixture was used in concrete. The admixture (type of fluid 70) was acquired from AYDOS Construction Chemicals Factory. There are solid matter content of 34.32%, intensity of 1.184 (20°C), pH value of 7.26 (20°C) of the new generation hyper plasticizer. Mixing water is well water from Doganli village in Duzce.

B. Methods

In the study, total of three different cements were used with PC being the reference in concrete mixture design. The amount of PC was reduced by 0, 10, and 20% by weight being substituted by the same amount of zeolite. Quantities of material to be put into the mixture within the framework of the method stated in TS 802 standards were determined for concrete mixture design [23]. According to rate of zeolite, three types of concrete samples were produced. They were encoded as R, 10Z and 20Z according to the substitution rates of zeolite. The quantities of material of the concrete mixture samples of materials for each concrete group and specific gravity were given in the Table 3.

Table 1: The physical properties of the aggregates.

		Test result			Standards
Unit weight, g/cm^3	Loose Unit Weight	1.48			TS EN 1097-3
	Dense Unit Weight	1.66			
		Aggregate Grading			TS EN 1097-6
		0-5, mm	5-19, mm	19-30, mm	
Water absorption, %		0.61	1.16	1	
Moisture content, %		1.25	1.32	1.41	
Determination of organic impurities		The color of the liquid is light yellow color than colorless (Organic matter is harmless).			TS EN 1744-1

Table 2: The physical and mechanical properties of PC and zeolite.

Materials	PC	Zeolite			PC	Zeolite
Chemical composition, wt.%			Physical properties			
SiO_2 (S)	18.68	68.85	Blaine, cm^2/g		4249	5740
Al_2O_3 (A)	4.67	11.71	Specific gravity		3,17	2,18
Fe_2O_3 (F)	3.53	1.29	Range dimension (over sieve), %	90 µm	4.08	17.60
CaO	64.56	3.97		45 µm	-	35.80
MgO	0.98	1.06	Setting time, minute			
SO_3	3.00	0.18	Initial		118	-
Na_2O	0.14	0.29	Final		-	-
K_2O	0.73	2.19	Compressive strength (MPa)			
Loss on ignition	3.92	10.00	2 days		-	-
Free CaO	1.74	-	7 days		29.6	-
Insoluble residue	0.50	37.32	28 days		52.8	-
S+A+F	-	81.85				

Table 3: Quantities of material in the 1 m^3 for each concrete group and specific gravity of materials.

Materials		Specific gravity	PC, kg	10Z, kg	20Z, kg
Aggregate, mm	0-5	2.66	822	843	855
	5-19	2.69	586	602	611
	19-30	2.70	428	439	446
Total			1836	1884	1912
PC		3.17	400	360	320
Zeolite		2.18	-	40	80
Water		1	139.7	139.7	123,2
Hyperplasticizer		1.184	4.800	4.320	4,800

The produced concretes are poured into 150 x 150 x 150 mm cubic molds without segregation. These concretes are retained for 24 hours in the molds and harden. Hardened concrete tests have been done on three different types of concrete, for 28 hydration days, in 23±2 ^oC water. Subsequently, the concrete samples are cured in %5 H_2SO_4 and in the 56 and 90 hydration days. The compressive strengths of all the concrete samples are measured at the end of 28, 56 and 90 hydration days.

III. Adaptive Network-Based Fuzzy Inference Systems

Adaptive Neural Fuzzy Inference Systems (ANFIS) is a hybrid system which comprises both neuronal network and fuzzy logical systems. ANFIS is constructed to the neuro-fuzzy think of *"if-then"* rules that declares between input and output variables. Benefit of the ANFIS is grounded on the collaboration of neural network and fuzzy logical system. Obviously, fuzzy inference in this hybrid system is used to handle imprecision and uncertainty. On the other hand, the neuronal network handles adaptability. Takagi, Sugeno, and Kang introduced the model, which produces with fuzzy inferences rules between an input and output variables. A generally fuzzy inferences rules has the polynomial function type as follows;

$$\textit{Rule1: If } x \textit{ is } A_1 \textit{ and } y \textit{ is } B_1 \textit{ , then } f_1 = p_1 x + q_1 y + r_1$$

$$\textit{Rule 2: If } x \textit{ is } A_2 \textit{ and } y \textit{ is } B_2 \textit{ , then } f_2 = p_2 x + q_2 y + r_2 \qquad (1)$$

where, *A* and *B* are show that definition of fuzzy sets in the crisp input variables. Given that *f* is a first-order polynomial, the model becomes a first-order Sugeno fuzzy model (Eq.1). On the other hand, if *f* is a constant, the model becomes as a zero-order Sugeno fuzzy model.

In terms of ANFIS inference, output of each rule is an establishment combination of fuzzy the input variables taking part by a fixed term. The final output of the ANFIS is calculated by the weighted average ($\overline{w_i}$) with outputs on the previous layer [24,25]. The basically ANFIS architecture was shown in Figure 1.

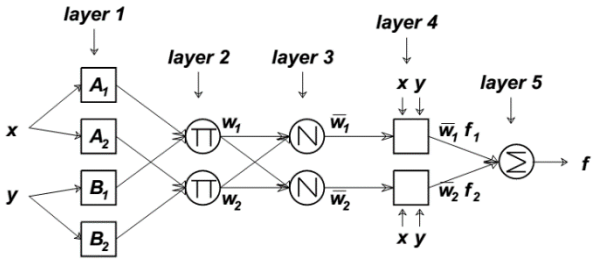


Figure 1: Equivalent ANFIS architecture.

In general, ANFIS architecture is composed of five layers. To ensure precise explanation, let assume that O_i^j demonstrated the output of the i^{th} node in j^{th} hidden layer.

The Layer 1 (Fuzzification Layer). In this layer, every node *i* in the layer is an adaptive node with following node function;

$$\begin{aligned} O_i^1 &= A_i(x), \qquad for \quad i = 1,2, or \\ O_i^1 &= B_{i-2}(y), \qquad for \quad i = 3,4 \end{aligned} \qquad (2)$$

where *x* or *y* is an input to the i^{th} node, and A_i or B_{i-2} is a membership function which is a linguistic label. On the other hand, O_i^j is show that is the membership degree of membership function such as *A* and B.

$$O_i^j = \mu A_i(x) = \frac{1}{1 + \left[\left(x - c_i\right)/a_i\right]^{2b_i}} \qquad (3)$$

where $\{a_i, b_i, c_i\}$ is demonstrated parameters of membership function and it is depend on the membership function type. The bell-shaped membership functions are widely used, and the membership degree is calculated by using Equation 3. Parameters of this layer are defined as antecedent parameters. The outputs of the layer are determined as the membership degree and are interval between 0 and 1.

The Layer 2 (Rule inference layer). In this layer, each node computes the firing strength of a rule by multiplication as follows:

$$O_i^2 = \mu A_i(x).\mu B_i(y) \qquad i = 1,2,... \tag{4}$$

The Layer 3 (Normalization layer). A member of this layer, Node *i*, computes the ratio between *i*th rule's firing strength and all firing strengths. Formally,

$$O_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2} \qquad i = 1,2,... \tag{5}$$

The Layer 4 (Consequent layer). A member of this layer, Node *i*, computes the contribution of *i*-th rule toward the overall output. Mathematically;

$$O_i^4 = \overline{w}_i.f_i = \overline{w}_i.(p_i.x + q_i y + r_i) \tag{6}$$

where w_i is defined as output of the layer 3, and $\{p_i, q_i, r_i\}$ is reckoned as the parameter set. It is referred to as the consequent parameters.

The Layer 5 (Output layer). This layer's single fixed node computes the final output as the summation of all incoming signals.

$$O_i^5 = overall\ output = \sum_i \overline{w}_i.f_i = \frac{\sum_i w_i.f_i}{\sum_i w_i} \tag{7}$$

In summary, fundamental learning strategy on the ANFIS architecture is based on backpropagation gradient descent, which decrease error signals in a bottom up approach recursively. Precisely, ANFIS starts the execution from output layer through to the nodes of input layer. This error signals are calculated by the derivative of the squared error with respect to each node's output. In fact, such learning strategy is similar to the backpropagation learning [25].

IV. Experimental Design and Model Parameters

In the ANFIS model, 27 of the experimental data were used for the training of model. On the other hand, 9 of the experimental data, which maintain the average features of all experimental data, were used for testing. According to Table 4, each set of training and testing data includes 7 parameters. Among them age of samples (days), Portland cement, zeolite, aggregate, water and hyper plasticizer are exploited as input data, and the compressive strength is utilized as output data. We present the general structure of the ANFIS model in Figure 2.

Table 4: Utilized input and output parameters and their quantities during ANFIS model.

		Utilized training and testing data of the model	
		Minimum	Maximum
Input variable	Age of samples, days	28	90
	PC, kg	320	400
	Zeolite, kg	0	80
	Aggregate, kg	1836	1912
	Water, kg	123.2	139.7
	Hyper plasticizer, kg	3.84	4.80
Output variable	Compressive strength, MPa	43.0	64.3

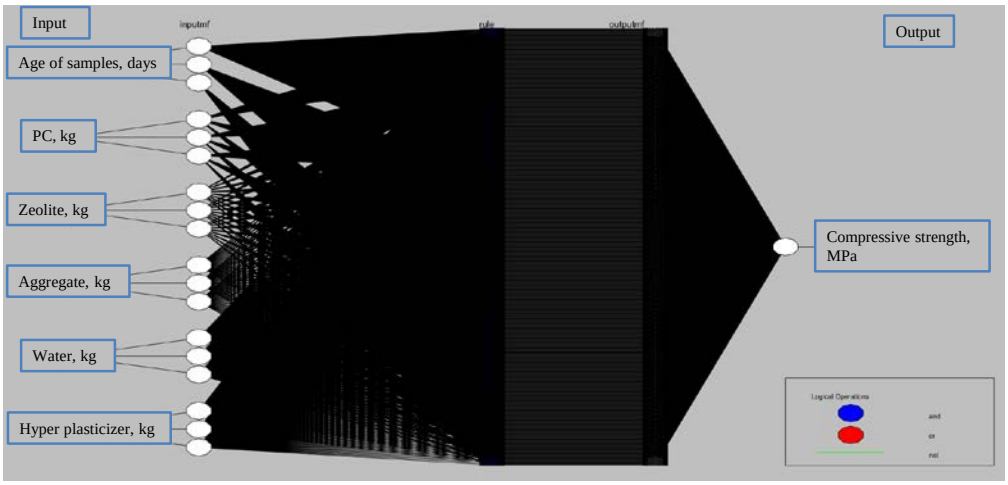


Figure 2: General structure of the model.

We have done computationally intense experimentations with various epochs. Results of the different learning algorithms with different epochs, best correlations were found through hybrid learning algorithm and 300 epochs. Furthermore, two ''gbellmf'' membership functions for each input and output variable. In our model, the bell-shaped membership function was chosen because all input variables show exponential increase and decrease in membership degree change. ANFIS model parameters were gıven Table 5. We have demonstrated membership functions of the inputs in Figure 3.

Table 5: ANFIS model parameters.

ANFIS info	Value
Number of nodes	1503
Number of linear parameters	729
Number of nonlinear parameters	54
Total number of parameters	783
Number of training data pairs	27
Number of checking data pairs	0
Number of fuzzy rules	729
Start training ANFIS	1.10744

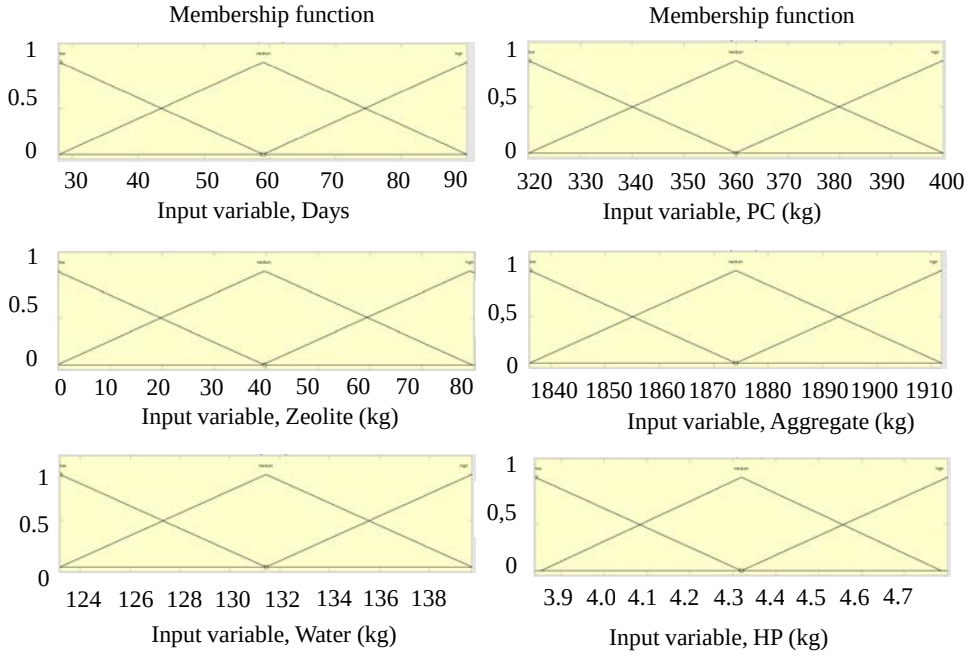


Figure 3: Membership functions of input variables.

V. Results And Discussion

The network models tried to be compared according to the R^2 (absolute fraction of variance), RMS (a root-mean squared error) and MAPE (mean absolute percentage error) criteria. These criteria are defined by equation (8), (9) and (10), respectively [26].

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{N} (t_i - o_i)^2}{\sum_{i=1}^{N} (o_i)^2} \right) \tag{8}$$

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^{N} |t_i - o_i|^2} \tag{9}$$

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \left(\frac{t_i - o_i}{o_i} \right) \right| * 100 \tag{10}$$

here N is the total number, t is the target value and o is the network output value of pattern.

Experimental results used for training were given in Figure 4, and experimental results with testing results obtained from ANFIS model were given in Figure 5. Furthermore, inputs values, experimental and testing results were given in Table 6.

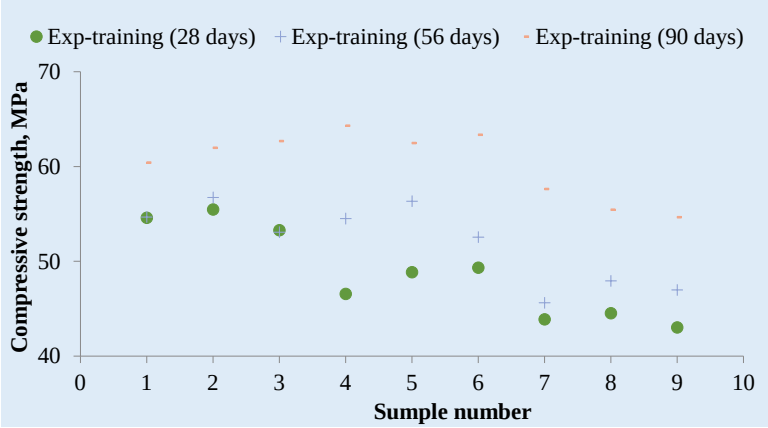


Figure 4: Experimental results used for training.

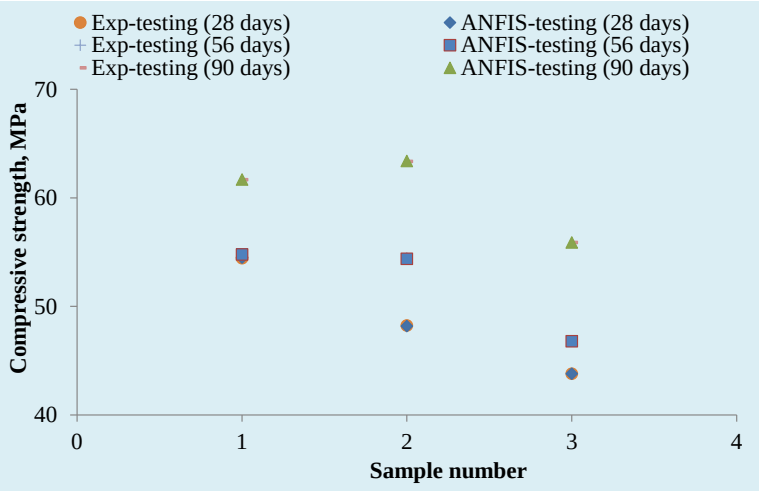


Figure 5: Experimental results with testing results obtained from ANFIS model.

Table 6: Data of input parameters, experimental results and testing results obtained from ANFIS model.

Data used in the model construction						Compressive strength, MPa	
Days	PC, kg	Zeolite, kg	Aggregate, kg	Water, kg	HA, kg	Exp., MPa	ANFIS, test
28	400	0	1836	139.7	4.800	54.4	54.5
28	360	40	1884	139.7	4.320	48.2	48.2
28	320	80	1912	123.2	4.800	43.8	43.8
56	400	0	1836	139.7	4.800	54.8	54.8
56	360	40	1884	139.7	4.320	54.5	54.4
56	320	80	1912	123.2	4.800	46.8	46.8
90	400	0	1836	139.7	4.800	61.7	61.7
90	360	40	1884	139.7	4.320	63.4	63.4
90	320	80	1912	123.2	4.800	55.9	55.9

Testing results of experimental and ANFIS models, for 28, 56 and 90 hydration days compressive strength were given in Figure 6. The statistical values for both training and testing, which are R^2, RMS and MAPE, were given in Figure 6, too.

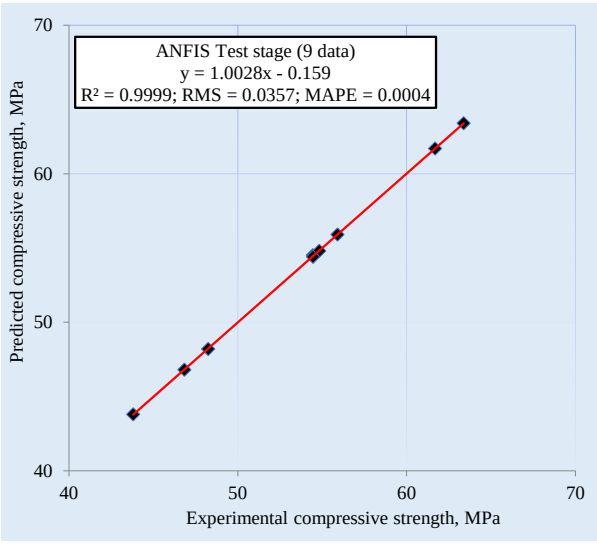


Figure 6: Comparison of experimental results with testing results of ANFIS model.

In terms of training and testing data, we present the linear least square fit line, its equation and the R^2 values in these Figures. As can be seen from Figure 6, results of the ANFIS model is very close to the experimental results. The results of testing phase in Figure 6 imply that the ANFIS model is able to present good prediction between input and output variables.

Respectively, the statistical values of R^2, RMS and MAPE of the ANFIS model were found as 0.9999, 0.0357 and 0.0004. The statistical values are denoted in Table 6. Statistical observations of Table 6, imply that proposed ANFIS model and experimental values yield close results.

VI. Conclusion

In this study, we have analyzed 27 data conducted experimental studies on 28, 56 and 90 hydration days compressive strength of concrete containing zeolite under H_2SO_4 exposure. For estimation we have exploited ANFIS model. In order to present an efficient ANFIS model, we have utilized different learning algorithms. In order to find best epoch, numerous ANFIS experiments are tested to define a model to ensure best prediction potential and is closer to experimental results. After finding the best ANFIS model, we

compared compressive strength results of ANFIS and experimental observations. R^2, RMS and MAPE statistics were used to evaluate the estimation performance of ANFIS. Results denote that prediction and the experimental values in the test stage of R^2, RMS and MAPE were found as 0.9999, 0.0357 and 0.0004 for ANFIS model, respectively.

Results mean that compressive strength of concrete containing zeolite under H_2SO_4 exposure can be estimated in the ANFIS model in a quite short time with low error rates. We conclude that ANFIS system is handy, practicable, quick, cheap and useful method for predicting the compressive strength of concrete containing zeolite under H_2SO_4 exposure.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support received from the research project of Duzce University Scientific Research Unit (Project Code Number: 2011.03.HD.09).

REFERENCES

[1] Y. Kocak and S. Nas, "The effect of using fly ash on the strength and hydration characteristics of blended cements", Construction and Building Materials, vol. 73, pp. 25–32, 2014.

[2] L. Zhengqi, "Drying shrinkage prediction of paste containing meta–kaolin and ultrafine fly ash for developing ultra–high performance concrete", Materials Today Communications, vol. 6, pp. 74–80, 2016.

[3] Y. Kocak, "A Study on the Effect of Fly Ash and Silica Fume Substituted Cement Paste and Mortars", Scientific Research and Essays, vol. 5, no. 9, pp. 990–998, 2010.

[4] A. Dorum, Y. Koçak, B. Yılmaz and A. Uçar, "Yüksek Fırın Cürufunun Çimento Yüzey Özelliklerine ve Hidratasyona Etkileri", Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi, vol. 19, pp. 47–58, 2009.

[5] H. Zhao, W. Sun, X. Wu and B. Gao, "The properties of the self–compacting concrete with fly ash and ground granulated blast furnace slag mineral admixtures", Journal of Cleaner Production, vol. 95, pp. 66–74, 2015.

[6] H. A. Alcamanda , P. H. R. Borgesa, F. A. Silvab and A. C. C. Trindade, "The effect of matrix composition and calcium content on the sulfate durability of metakaolin and metakaolin/slag alkali-activated mortars", Ceramics International vol. 44, pp. 5037–5044, 2018.

[7] Y. Koçak, A. Dorum, B. Yılmaz and A. Uçar, "Trasın Çimento Yüzey Özelliğine, Hidratasyona ve Basınç Dayanımına Etkisi", E–Journal Of New World Sciences Academy Technological Applied Sciences, vol. 5, no. 1, pp. 1–14, 2010.

[8] A. Subaşı and M. Emiroğlu, "Effect of metakaolin substitution on physical, mechanical and hydration process of White Portland cement", Construction and Building Materials, vol. 95, pp. 257–268, 2015.

[9] O. Keleştemur and B. Demirel, "Effect of metakaolin on the corrosion resistance of structural lightweight concrete", Construction and Building Materials, vol. 81, pp. 172–178, 2015.

[10] K. Yıldız, A. Dorum and Y. Koçak, "Pomza Zeolit Ve Cem I Çimentosunun Minerolojik Moleküler Elektrokinetik Ve Termal Uyumunun Yüksek Dayanımlı Betona Etkisinin Araştırılması", Journal of The Faculty of Engineering and Architecture of Gazi University, vol. 25, no. 4, pp. 867–879, 2010.

[11] H. Gerengi, Y. Kocak, A. Jażdżewska, M. Kurtay ve H. Durgun, "Electrochemical investigations on the corrosion behaviour of reinforcing steel in diatomite– and zeolite–containing concrete exposed to sulphuric acid", Construction and Building Materials, vol. 49, pp. 471–477, 2013.

[12] Y. Kocak and M. Savaş, "Effect of the PC, diatomite and zeolite on the performance of concrete composites", Computers and Concrete, vol. 17, no. 6, pp. 815–829, 2016.

[13] Y. Kocak, E. Tascı and U. Kaya, "The effect of using natural zeolite on the properties and hydration characteristics of blended cements", Construction and Building Materials, vol. 47, pp. 720–727, 2013.

[14] V. F. Rahhal, Z. Pavlık, A. Tironi, C. C. Castellano, M. A. Trezza, R. Cerny, E. F. Irassar, "Effect of cement composition on the early hydration of blended cements with natural zeolite" J Therm Anal Calorim, vol. 128, pp. 721–733, 2017.

[15] K. Yıldız, "Pomza ve zeolit katkılı yüksek dayanımlı betonların asit ve tuz etkilerine karşı dayanıklılığının araştırılması", Doktora Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2009.

[16] Y. Kocak and E. Gulbandilar, "MgSO4 Etkisindeki Betonların Basınç Dayanımının ANFIS ile Tahmini", 8. International Aggregates Syposium, Dumlupınar University, Kütahya, Turkey, pp. 251-262, 13-14 October 2016.

[17] Z. Yuan, L–N. Wang and X. Ji, "Prediction of concrete compressive strength: Research on hybrid models genetic based algorithms and ANFIS", Advances in Engineering Software, vol. 67, pp. 156–163, 2014.

[18] S. Motamedi, S. Shamshirband, D. Petković and R. Hashim, "Application of adaptive neuro–fuzzy technique to predict the unconfined compressive strength of PFA–sand–cement mixture", Powder Technology, vol. 278, pp. 278–285, 2015.

[19] A. R. Boğa, M. Öztürk and İ. B. Topçu, "Using ANN and ANFIS to predict the mechanical and chloride permeability properties of concrete containing GGBFS and CNI", Composites: Part B, vol. 45, pp. 688–696, 2013.

[20] A. Sadrmomtazi, J. Sobhani and M. A. Mirgozar, "Modeling compressive strength of EPS lightweight concrete using regression, neural network and ANFIS", Construction and Building Materials, vol. 42, pp. 205–216, 2013.

[21] M.S. Pourtahmasb, M. R. Karim and S. Shamshirband, "Resilient modulus prediction of asphalt mixtures containing Recycled Concrete Aggregate using an adaptive neuro–fuzzy methodology", Construction and Building Materials, vol. 82, pp. 257–263, 2015.

[22] Kh. Mahfuz ud Darain, S. Shamshirband, M. Z. Jumaat and M. Obaydullah, "Adaptive neuro fuzzy prediction of deflection and cracking behavior of NSM strengthened RC beams", Construction and Building Materials, vol. 98, pp. 276–285, 2015.

[23] Türk Standartları Enstitüsü (TSE), "Beton karışımı hesap esasları", TS 802, Ankara, Türkiye, 2009.

[24] K. A. Aali, M. Parsinejad and B. Rahmani, "Estimation of saturation percentage of soil using multiple regression, ANN, and ANFIS techniques", Comput. Inform. Sci., vol. 2, no. 3, pp. 127–136, 2009.

[25] J–SR. Jang, "Input selection for ANFIS learning, Fuzzy Systems", Proceedings of the Fifth IEEE International Conference, V012: pp. 1493 –1499, 1996.

[26] F. Ozcan, C. D. Atis, O. Karahan, E. Uncuoglu and H. Tanyildizi, "Comparison of artificial neural network and fuzzy logic models for prediction of long–term compressive strength of silica fume concrete", Advances in Engineering Software, vol. 40, pp. 856–863, 2009.

Investigation of the Effect of Mesh Density and Element Type on Behavior of Biphasic Soft Tissues in Finite Element Analysis

S.UZUNER[1], E. ZURNACI[2], M.L. RODRIGUEZ[3], S.KUCUK[4]

[1]University of Duzce, Cumayeri Vocational School, Duzce, Turkey, sabriuzuner@duzce.edu.tr
[2]University of Duzce, Cumayeri Vocational School, Duzce, Turkey, ermanzurnaci@duzce.edu.tr
[3]University of Calgary, Department of Mechanical and Manufacturing Engineering, Calgary, Canada, mlrodrig@ucalgary.ca
[4]University of Kocaeli, Faculty of Technology, Kocaeli, Turkey, skucuk@kocaeli.edu.tr

Abstract **- The finite element method (FEM) is a computational technique that is often used to solve biomedical engineering problems. The biphasic cartilage model plays important role in representing the mechanical behaviour of the articular cartilage. In order to obtain accurate results in finite element analysis of articular cartilage, it is necessary to determine appropriate FEM parameters such as mesh density and finite element type. Models with small element sizes in the FEM allow more accurate results to be obtained however it requires longer calculation time. In contrast, large element size can lead to non-precision results while shortening the calculation time. The type of the elements may also change the results of FEM analysis for biomechanical problems. The purpose of this study is; to evaluate the effect of the mesh size and type of the finite element on the results of the numerical biphasic tissues. In this study, in order to achieve this goal a series of compression analyzes were performed on the 3D biomedical models with different mesh density and element types using ABAQUS 6.13 software and the results were compared. The analysis results showed that mesh density element type and element type had little effect on the maximum reaction force. On the contrary, the mesh density had greatly increased the computational time.**

Keywords **- Biphasic tissues, FEM, mesh density, element type, biomechanical analysis.**

I. INTRODUCTION

Articular cartilage is a special connective tissue and its most important feature is to provide a lubricated surface and facilitate the delivery of the load [1]. Articular cartilage is a hydrated tissue which covers the end of femur and tibia, giving support to the knee joint in order to overcome loads due to body weight in daily activities such as standing, walking and sitting [2]. Articular cartilage composed by a porous matrix saturated with water. Most of it weight is water (68%-85%). The solid matrix in composed of chondrocytes, negatively charged proteoglycans and collagen fibres. Contact among tissues has low friction and usually low wear [3].

Biphasic cartilage model [4] has shown to be effective to explain the mechanical behaviour of articular cartilage. This model has a deformable homogeneous solid embedded in an incompressible fluid phase. The momentum equation for the biphasic model will be:

$$div(\sigma) = \mathrm{div}(\sigma^{s}) + \mathrm{div}(\sigma^{f}) = 0 \quad (1)$$

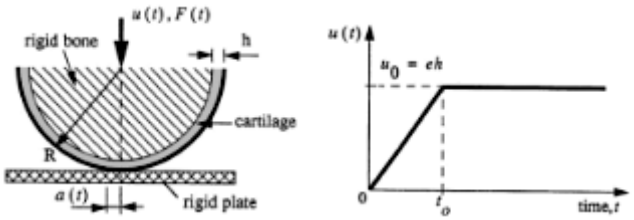
where S and f represent the solid an fluid phase of the model. This theory coincides with the consolidation procedure if fluid is assumed inviscid [5]. Similar procedures to the model which will be present the computational simulation on cartilage disks under unconfined compression [6] and spherical indenter on a disk [7].

Computational models to explain cartilage behaviour has been developed over the last decades [8-13]. Especially finite element (FE) models are commonly used for stress analysis of Articular Cartilage [14-18]. The selection of element types and mesh density plays an important role in the analysis about the behaviour of biphasic soft tissue using the finite element method. In addition this selection affects the accuracy of analysis results and the time of analysis [19].

In this study, the effects of mesh density and element type on the simulation results were investigated in the solution of contact problems of a simple geometry containing biphasic tissues using the finite element method. The role of fluid load support has been included in the computational model.

II. MODEL DESCRIPTION

In the analysis the geometry to be used was shown in Figure 1.a [20]. It consists in a spherical cartilage surface (R=399 mm) in contact with a rigid plate. Cartilage thickness was h=1 mm and was attached to a rigid bone. Rigid plate was fixed while rigid bone was allowed to move vertically against the plate. The movement was controlled by displacement history shown in Figure 1.b, involving a ramp compression during t_0=100 s and then a relaxation period of 900 s.



a) b)

Figure 1: Geometry of the model used (a) ramp function of the movement of the rigid bone (b).

For validation purposes, three displacements (u_0 in Figure 1.b) were used: 0.05, 0.08 and 0.10 mm which led to compression ratios of 5%, 8% and 10% respectively. Cartilage was assumed to be sealed in order to be compared with theory solution. Material properties of the cartilage were given in Table 1.

Table 1: Cartilage material properties

Material Type	Elastic isotropic
Young's Module	0.50 MPa
Poisson's ratio	0.3
Permeability (Isotropic)	0.002 mm^4/Ns
Void ratio	4

III. Development of Numerical Model

As the bone and attached cartilage had a spherical shape and displacement was restricted to be in the direction normal to the rigid plate, axisymmetric model was used. Rigid bone and plate were modelled as axisymmetric analytical rigid with a rigid point attached to them. Cartilage was created as axisymmetric deformable. In order to compare the theory solution cartilage surfaces had to be sealed, which is the default boundary condition for a poroelastic material in ABAQUS.

Figure 2 shows the three parts after assembly. As seen in Figure 2, the contacting area was in the vicinity of the symmetry axis. Horizontal length of the three parts was set to be 35 mm and was enough to catch meaningful results. Thickness of 1 mm was chosen for cartilage. Initially, cartilage touched the rigid plate just in one point located in the axis of symmetry.

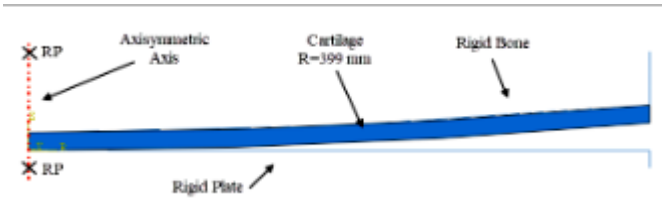


Figure 2: Axisymmetric model of it consists in a spherical cartilage surface in contact with a rigid plate and was attached to a rigid bone.

Firstly, in order to determine closeness of numerical result to analytic results, in the computational model 8-node axisymmetric quadrilateral, biquadratic displacement, bilinear pore pressures (CAX8P) were used. The model contained 8 layers of elements in the radial direction and 100 elements in the circumferential direction, leading to a total of 800 elements (Figure 3).

Figure 3: Mesh density with 800 elements.

Surface to surface discretization method was chosen for interaction. Small formulation was used for sliding formulation. Soils type step was used, to take into account the poroelastic behaviour of the material. Total time was 1000 s, with a transient consolidation fluid response. The displacement of rigid bone consists in a ramp between 0 and 100 s and was held constant until 1000. Maximum value of the ramp was set to be 1. As mentioned before, simulation result was compared with those theory results obtained from Herzog et al. (1998).

IV. Results

Finite element analysis was carried out using the developed numerical model. Relationship between reaction force and time was shown in Figure 4 with different strain (e=0.05, 0.08 and 0.10). The results of simulation and theory were agreed during the compression period until reaching maximum compression where reaction force incremented when compression was greater. Then, reaction force tended to be higher in all cases (including peak reaction force) for simulation results. Then, reaction force tended to be higher in simulation results with respect to theoretical results.

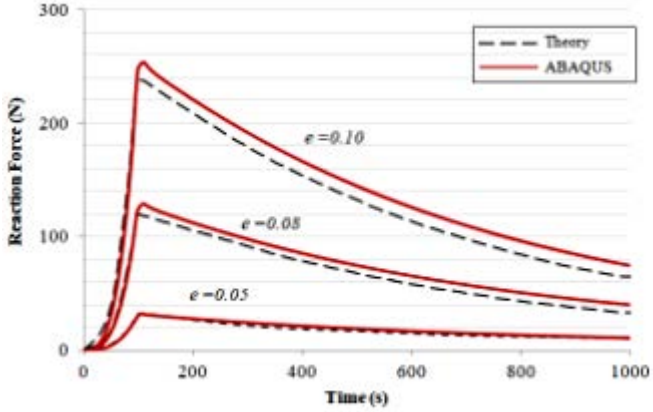


Figure 4: Reaction force change for different strains. Black lines represent theory results obtained by Herzog et al. (1998). Red lines where obtained using ABAQUS.

The relative difference went from less than 1% for the smallest compression up to less than 10% for the highest compression when comparing peak reaction forces (Table 2).

Table 2. Theoretical and numerical analysis results

	Peak Force (N)		
Displacement	Teorical Value	Numerical Value	% Error
0.05 mm	31.643	31.634	-0.02

0.08 mm	118.798	128.662	8.30
0.10 mm	237.636	254.241	6.98

When the displacement was double, the reaction force increased approximate 7 times for Theory and as for ABAQUS solutions, the reaction force increased approximately 8 times. A monotonically decrease in force was observed during relaxation period without reaching equilibrium condition. Simulation and theory had a better agreement when displacement had smaller values. At any case the behaviour of both graphics turned to be similar: forces increased with displacement until maximum compression was reached and then started to decrease with time. This is the typical behaviour of a viscoelastic material [21].

A. *Influence of Mesh Density on the Result of Simulation*

In finite element analysis, the mesh density is a critical issue closely related to the accuracy of the finite element analysis and the computational time [22]. At this section of the study; the effect of mesh density on the accuracy of numerical analysis results in biomechanical analysis applications was investigated by studying on finite element analysis of the biphasic cartilage model.

In order to determine the effect of remeshing process in results, theory results for peak force when e = 0.08 strain was used since the difference between the theoretical and the simulation results was the highest. Five different cases including different mesh density were determined. The cases were given by Table 3.

Table 3: The parameters specified for the cases.

Case	#of elements in radial direction	#of elements in circumferential direction	Total Numbers
1	2	25	50
2	4	50	200
3	8	100	800
4	16	200	3200
5	32	400	12800

Case 1 represents the mesh with fewer elements and case 5 is the mesh with the biggest number of elements. Each case has 4 times more elements than the previous one.

The finite element analysis of biphasic tissue was performed for different mesh density (Table 3). Results were shown in Figure 5. Notice the scale used for force in this plot, meaning that peak force do not significantly change when remeshing process was applied. Maximum difference was between case 1 and 3 and was 0.88%.

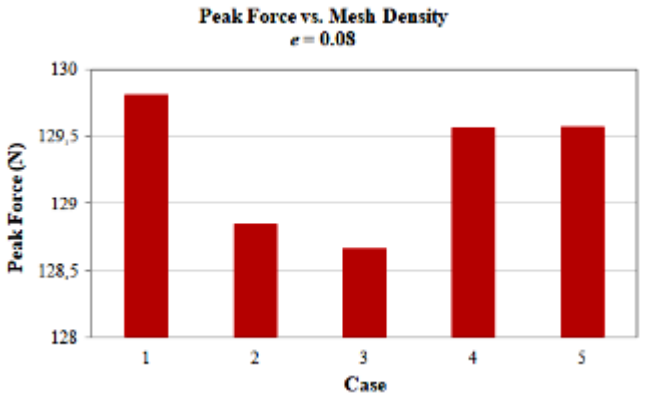


Figure 5: The peak force values for different mesh density.

The modelling solution with too many elements in finite element analysis the longer the computational time that was required. Computational times which increased about 56 times between the aforementioned cases (20.5 vs. 1158.3 seconds of CPU time). Results were shown in Figure 6. The finite element analysis were carried out on quad-core Intel i5-2450M CPU 2.5 GHz computers with 4 GB RAM.

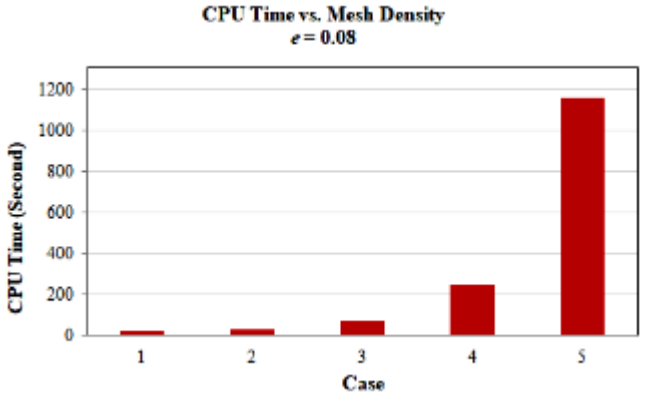


Figure 6: Computational times for different mesh density.

B. *Influence of element type on the result of simulation*

The provision of faster contact convergence and better fluid pressure distribution compatibility is directly related to the selection of element type in the finite element analysis of biphasic textures [23]. In order to determine the influence of element type on results simulation was rerun using some of the available axisymmetric elements. The types of elements used were given in Table 4. Case 1 to 3 has quadrilateral elements while 4 and 5 have triangular elements. Results for peak force when e = 0.08 were shown in Figure 7. When compared the cases with each other, highest difference happened while changing from 8 to 4 nodes quadrilateral elements (Cases 1 and 3), being less than 1%.

As seen in Figure 7, elements with 8 node for fluid pressure gives numerically more accurate results however significantly reduces contact-convergence [23, 24]. For this reason, in the knee joint analysis, the finite elements should be selected as the high node elements in the soft tissues interested.

Table 4: Elements type and description used to rerun the simulation. Case 1 represents the base case.

Case	Element Name	Element Description
1	CAX8P	An 8-node axisymmetric quadrilateral, biquadratic displacement, bilinear pore pressure.
2	CAX8PH	An 8-node axisymmetric quadrilateral, biquadratic displacement, bilinear pore pressure, hybrid, linear pressure.
3	CAX4P	A 4-node axisymmetric quadrilateral, bilinear displacement, bilinear pore pressure.
4	CAX6MP	A 6-node modified quadratic axisymmetric triangle, pore pressure, hourglass control.
5	CAX6MPH	A 6-node modified quadratic axisymmetric triangle, pore pressure, hybrid, linear pressure, hourglass control.

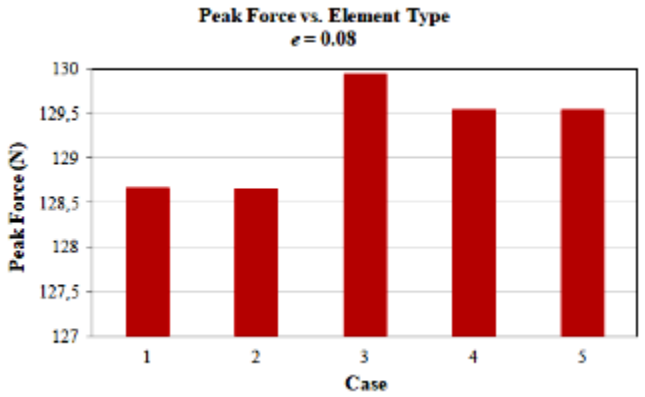


Figure 7: Peak Reaction Force for different element types.

The use of Hybrid elements cases (2 and 5 with an "H" on the element name) implies that the behaviour of the material is incompressible (Poisson's Ratio = 0.5) or close to incompressible (Poisson's Ratio > 0.475). As it was mentioned before Poisson's Ratio was used 0.3 (not even close to incompressible). This shows that sometimes consistent results with certain output variables can be obtained even when using the wrong element type. One reason for this consistency might be that volume of cartilage does not change too much due to the small strain and geometry.

V. Discussion

ABAQUS simulation was compared to the theoretical results of a spherical cartilage contacting a rigid plate. Some assumptions like Young's Module and Poisson's Ration needed to be made in order to obtain a solution. The main findings and conclusions from the study can be outlined as follows.

1) The change in element type was caused a maximum of 1% difference between the theoretical and numerical analysis results.
2) The effect of mesh density was played an important role for the computational time. The increase in mesh density increased the time of the analysis. The difference of the analysis times between the lowest and highest mesh density was determined to be about 56 times.
3) As a result of the change in mesh density, the maximum difference for the theoretical and numerical analysis results was 0.88%.
4) When the displacement goes up to two times; the difference between the theoretical and numerical analysis results was not more than 7%, although the reaction force increased about 8 times.

Results for this particular case was shown little dependence on mesh density and element type of finite element model. While computational time depends on element density of biphasic tissue numerical model.

References

[1] A. J. Sophia Fox, A. Bedi and S. A. Rodeo, "The Basic Science of Articular Cartilage: Structure, Composition, and Function," Sports Health, vol. 1(6), pp. 461-468, 2009.

[2] A. M. Liphardt, A. Mündermann, S. Koo, N. Bäcker, T. P. Andriacchi, J. Zange, J. Mester and M. Heer, "Vibration training intervention to maintain cartilage thickness and serum concentrations of cartilage oligometric matrix protein (COMP) during immobilization," *Osteoarthritis Cartilage*, vol. 17(12), pp. 1598-1603, 2009.

[3] S. R. Oungoulian, S. Chang, O. Bortz, K. E Hehir, K. Zhu, C. E Willis, C. T. Hung and G. A. "Ateshian. Articular cartilage wear characterization with a particle sizing and counting analyzer," *Journal of Biomechanical Engineering*, vol. 135(2), pp. 0245011–0245014, 2013.

[4] V. C. Mow, S. C. Kuei, W. M. Lai and C. G. Armstrong, "Biphasic creep and stress relaxation of articular cartilage in compression? Theory and experiments," *Journal of Biomechanical Engineering*, vol. 102, pp. 73–84, 1980.

[5] B. R. Simon, "Multiphase Poroelastic Finite Element Models for Soft Tissue Structures," *Mechanics Reviews*, vol. 45(6), pp. 191, 1992.

[6] L. P. Li, M. D. Buschmann and A. Shirazi-Adl, "A fibril reinforced nonhomogeneous poroelastic model for articular cartilage: inhomogeneous response in unconfined compression," *Journal of Biomechanics*, vol. 33(12), pp. 1533–1541, 2000.

[7] Q. Meng, Z. Jin, J. Fisher and R. Wilcox, "Comparison between FE Bio and Abaqus for biphasic contact problems," *Proceedings of the Institution of Mechanical Engineers. Part H, Journal of Engineering in Medicine*, vol. 227(9), pp. 1009–1019, 2013.

[8] N. Hosoda, N. Sakai, Y. Sawae and T. Murakami, "Finite Element Analysis of Articular Cartilage Model Considering the Configuration and Biphasic Property of the Tissue," *In: Lim C.T., Goh J.C.H. (eds)* in in *Conf. Rec. 2009 13th International Conference on Biomedical Engineering. IFMBE Proceedings*, 23, 1883-1887, Springer, Berlin, Heidelberg.

[9] H. Guo and R. L. Spilker, "Biphasic finite element modeling of hydrated soft tissue contact using an augmented Lagrangian method," *Journal of Biomechanical Engineering*", vol. 133(11), pp. 111001-111007, 2011.

[10] M. J. A. Latif, N. H. Hashim, R. Ramlan, J. Mahmud, A. Jumahat and M. R. A. Kadir, "The Effects of Surface Curvature on Cartilage Behaviour in Indentation Test: A Finite Element Study," *Procedia Engineering*, vol. 68(1), pp. 109-115, 2013.

[11] *Q. Zhang, L. Gao, P. Xiao, C. Zhang and J. Ye,* "Finite element analysis on the mechanical behavior of articular cartilage under rolling load," in *Conf. Rec. 2014 IEEE International Conference on Mechatronics and Automation, Tianjin*, pp. 936-940.

[12] N. H. Hashim, M. J. A. Latif, Y. L. Jaafar, R. Ramlan and J. Mahmud, "Computational and Experimental Study of Articular Cartilage Thickness on Biomechanical Behavior," *International Journal of Applied Engineering Research*", vol. 12(16), pp. 5849-5856, 2017.

[13] M. H. Doweidar and M. Doblare, "Finite Element Modeling and Simulation of the Multiphysic Behavior of Articular Cartilage, "*Numerical Methods and Advanced Simulation in Biomechanics and Biological Processes*", pp. 37-53, 2018.

[14] R. L. Spilker, J. K. Suh and V. C. Mow, "A finite element analysis of the indentation stress-relaxation response of linear biphasic articular cartilage," *Transactions of the ASME Journal of Biomechanical Engineering*, vol. 114(2), pp. 191–201, 1992.

[15] V. C. Mow, S. C. Kuei, W. M. Lai and C. G. Armstrong, "Biphasic creep and stress relaxation of articular cartilage in compression: Theory and experiments," *Journal of Biomechanical Engineering*, vol. 102(1), pp. 73–84, 1980.

[16] J. K. Suh and B. Shai, "Finite element formulation of biphasic poroviscoelastic model for articular cartilage," *Transactions of the ASME Journal of Biomechanical Engineering*, vol. 120(2), pp. 195–201, 1998.

[17] M. R. DiSilvestro, Q. Zhu, and J. K. Suh, "Biphasic poroviscoelastic simulation of the unconfined compression of articular cartilage: Ii–effect of variable strain rates," *Journal of Biomechanical Engineering*, vol. 123(2), pp. 198–200, 2001.

[18] M. R. DiSilvestro and J. K. Suh. "A cross-validation of the biphasic poroviscoelastic model of articular cartilage in unconfined compression, indentation, and confined compression," *Journal of Biomechanics*, vol. 34(4), pp. 519–525, 2001.

[19] Y. Liu and G. Glass, "Effects of Mesh Density on Finite Element Analysis," *SAE Technical Paper*, pp. 1375, 2013

[20] J. Z. Wu, W. Herzog and M. Epstein, "Evaluation of the finite element software ABAQUS forbiomechanical modelling of biphasic tissues," *Journal of Biomechanics*, vol. 31(2), pp. 165–169, 1998b.

[21] R. J. Dewall, T. Warghese, M. A. Kliewer, J. M. Harter and E. M. Hartenbach, "Compression-Dependent Viscoelastic Behavior of Human Cervix Tissue," *Ultrason Imaging*, vol. 32(4), pp. 214-228, 2010.

[22] T. M. Shashikant and R. S. Bindu, "Effect of Mesh Size on Finite Element Analysis of Plate Structure," in *Conf. Rec. 2015 International Journal of Engineering Science and Innovative Technology (IJESIT)*, pp. 181–185.

[23] Y. Dabiri and L. P. Li, "Altered Knee Joint Mechanics in Simple Compression Associated with Early Cartilage Degeneration," *Computational and Mathematical Methods in Medicine*, vol. 2013, pp. 1-11, 2013

[24] M. K. Miraki, "Finite Element Study of the Healthy and Meniscectomized Knee Joints Considering Fibril Reinforced Poromechanical Behaviour for Cartilages and Menisci," Ph.D. dissertation, Dept. Mech. and Manuf. Eng., Harvard Univ., Alberta, CA, 2013.

Application of Artificial Intelligence Methods in Software Testing

Y.Kozina[1], N. Volkova[2] , O. Osadchiy[3]

[1] Odessa National Polytechnic University, Odessa/Ukraine yuliyakc21@@gmail.com
[2] Odessa National Polytechnic University, Odessa/Ukraine, volkovanp30@gmail.com
[3] Odessa National Polytechnic University, Odessa/Ukraine, oleosadchiy@gmail.com

***Abstract* - Today, the role of software testing in the life cycle of software development has significantly increased. The process of software testing implementation include: planning, designing, creating, executing and supporting of the tests. At the same time, there was a transition from direct testing to quality assurance, covering the software development cycle in general.**
In this paper, we proposed the directions for solving problems in software testing. Solving of this problems requires knowledge, logical reasoning, the experience of an engineer for quality control software. Using an artificial intelligence methods in a software testing allowed us to found the ways for executing of tests classification and optimization.

***Keywords* – software testing, artificial intelligence, quality assurance, test-design.**

I. INTRODUCTION

Currently, software testing is one of the most established ways to ensure the quality of software development. As one of the main phases of the software development process, which includes: design, code development, testing. The last point characterized by a large contribution to the overall complexity of product development [1]. For solving of problems in software testing are requires to make decisions in the intellectual area [2]. Therefore, the goal of this thesis is to perform analysis of problems in software testing and to found ways of solving them by applying of Artificial Intelligence methods.

II. DESCRIPTION PROBLEM

The main problem that was detected in software testing is limited time. Usually, it is not possible to develop and execute tests that will cover 100% of the required functionality and test all possible input and output values within timeframe set. The next problem is the correct classification of bugs according to their criticality [3]. It significantly affects their further lifecycle and system effectiveness in general. The one more problem is software quality assessment. For software quality assessment different methods are used: statistical methods, standardization and certification method, software testing, expert evaluation method, computer method (metrics), formal methods of quality assurance (such as automation software testing, models building etc.) and others. However none of the methods gives ability to predict quality level of the released product.

III. SOLVING PROBLEM

To reduce tests amount keeping low quality risk level the following test-design techniques are used.

Approach 1 – tests classification and optimization:
– equivalent classes detection (input data is combined into classes, if some test detects some bug then we can assume that all other tests of this class will also detect this bug and vice versa);
– boundary values (values located on boundaries of the equivalent classes incoming data and around them);
– cause-effect relationships analysis allows us to select the most effective tests that cover more functionality with less cases amount (the method uses logic algebra and operates with concepts of "cause" and "effect"). By cause in this case we understand separate input condition or equivalent class. By effect – output condition or system transformation. As example – we fill some input form. Here fields are cause and button press is effect.

Approach 2 – Bayesian networks usage:
The method consists in assigning all the consequences to the causes, meaning clarification of cause and effect relationships (CER), by building CER tables. In the beginning because CER tables are not convenient for huge specifications they are divided into smaller "working" sections, trying to allocate independent CER groups into separate tables. Further based on semantic content of the specification so called truth table is built, where for each possible consequences combination matching cause is set. The tables also contain constraints for CER combinations. The same table can be built for equivalent classes. Then each table row is transformed to the test.

Bayesian network – graph based probability model that consists of multiple variables and their probabilistic dependencies according to Bayes [4].

Approach 3 – bug severity definition. Correctly define bug severity is very important task for QA engineer. Correct classification of bugs according to their criticality significantly affects their further lifecycle and system effectiveness in general. To resolve this task we need appropriate knowledge, logical reasoning, QA experience etc. Meaning that it belongs to plane of human intellectual work. So, it is possible to solve

this task by applying Artificial Intelligence (AI) methods. Among them we can find classification tasks. According to what was presented above, the task of bugs classification methodology development in software testing is absolutely relevant. The solution for this task can be implemented based on AI methods to increase classification reliability and human factor reduction. We propose to use here mathematical methods of fuzzy logic, as one of the AI area directions that allows to find solutions for hard-formable tasks, such as classification tasks. Fuzzy logic is mathematics section generalizing the concepts of classical logic and set theory, based on fuzzy set concept as an object with the belonging function of element to a set taking any values from interval [0 to 1] and not only 0 and 1 (true and false) [5]. So, how the given theory is applied for bugs classification. Bug severity plays the role of linguistic variable. Rules database is formed according to the properties that describe bug severity for each class. Then one of the fuzzy inference system algorithms is applied. In this case we use Mamdani algorithm. For the classification quality assessment we should estimate its reliability by defining percentage of incorrect classification. Errors of first and second kind are also taken into account. The conducted experiments showed that high reliability of the classification is achieved, that is about 90% when the described algorithm is applied. (metrics of source code, processes etc.) and detect program modules that can contain bugs with high probability. This allows QA engineers to correctly prioritize tests, find bugs earlier and developers – to fix those defects as early as possible.

Approach 4 – software quality assessment. For software quality assessment different methods are used: statistical methods, standardization and certification method, software testing, expert evaluation method, computer method (metrics), formal methods of quality assurance (such as automation software testing, models building etc.) and others. However none of the methods gives ability to predict quality level of the released product. Therefore it is impossible to define what actions will be needed from technical support side after the product is released. So today we cannot say that software quality assessment methodologies are mature. It is especially difficult to identify factors that affect the software quality and also these factors causes of action. We think that the most suitable for these tasks resolving is recently developed fuzzy logic mechanism. Here the fuzzy logic methods can also be used. For linguistic evaluation of input variables the following characteristics can be applied: resource-saving, scalability, integration, performance, support, interface characteristic.

Approach 5 – prediction of possible bugs number. Modern software development approaches often use different iterative methods. Project planning, coding and testing either are done in parallel or alternate cyclically for each new release. Testing process can be done continuously because changes are constantly introduced into the source code. The bigger source code volume is the harder to keep the project's high quality. For earlier bugs detection models of bugs prediction can be used. These models use different metrics of previous releases (metrics of source code, processes etc.) and detect program modules that can contain bugs with high probability. This allows QA engineers to correctly prioritize tests, find bugs earlier and developers – to fix those defects as early as possible. Using a Neural Network technologies for bugs prediction is effective. Neural Networks is one of Artificial Intelligence directions. As we can see thanks to Neural Networks learning ability and using our previous experience we can estimate test coverage as well as define the most vulnerable code parts and predict bugs.

IV. CONCLUSION

The executed investigation are showed, that the application of AI methods for software testing is effective. Also the AI directions – such as Machine Learning and Data Science currently are actively used in various fields.

REFERENCES

[1] M. Amroune, P.J. Charrel, N. Zarour and J.M. Inglebert, "A Model-Driven Engineering Approach To Develop A Cooperative Information System", International Journal of Software Engineering & Applications (IJSEA), Vol.4, No.3, pp. 29-43, 2013.

[2] Kozina Y., Kozin A., "Artificial intelligence system for technical diagnostics of photomasks", Technology and construction in electronic equipment, vol. 1, pp.7-9, 2012.(in Russian)

[3] Kanner K, "Software Testing". – Kiev:DianaSoft, 2011.- 544 p. (in Russian)

[4] Bayesian network. [Electronic resource].- Access mode: https://en.wikipedia.org/wiki/Bayesian_network.

[5] Fuzzy Logic Toolbox User Guide. [Electronic resource].- Access mode:http://www.mathworks.com/help/pdf_doc/fuzzy.pdf.

An Interactive Learning Method Based on Agent Systems

G.ATALI[1*], D.KARAYEL[2], S.S.OZKAN[3]

[1] Sakarya University, Sakarya/Turkey, gatali@sakarya.edu.tr
[2] Sakarya University, Sakarya/Turkey, dkarayel@sakarya.edu.tr
[3] Sakarya University, Sakarya/Turkey, sozkan@sakarya.edu.tr

***Abstract* – When the evolution of human being is analyzed, it is seen that almost all activities are developed by trial and error method. However, it is also known that these methods are not productive and cause much time loss. Therefore, various methods have been developed to solve the problems encountered in different situations. As to educational activities, they mainly have quite different qualities than others. In general, education and training is done collectively by bringing people at the same level together. Although we accept individuals who are brought together of the same nature, this is not a very correct approach in many cases. In fact, people have different levels of knowledge and experience. Scientific studies have even shown that each individual's learning method differs. Therefore, whatever method is used, it is difficult to obtain the desired efficiency in collective learning. In this study, a new learning method based on interactive learning and knowledge based learning systems has been developed with reference to the personal differences that emerged during learning. With this method, first, the strong and weak aspects of the person related to the subject to be learned are determined and then the interactive learning is provided with the designed interface. In this learning method where learning is performed at different periods with asynchronous learning techniques, creating permanent learning by taking individual differences into account is the ultimate goal.**

***Keywords* - Interactive Learning, Individual Learning, Agent systems**

I. INTRODUCTION

Today, with the increasing need of learning, it is seen that the learning activities have increased correspondingly. When effective and qualified theories of learning are examined, it is known that learning is differentiated from the individual point of view. Asynchronous learning is very important, especially for individuals with learning disabilities. Thanks to asynchronous learning, a type of learning that occurs at the speed of individual learning independent of location and time, the individual can learn the knowledge to be learned at the level of his own speed and perception. By this means, the persistence and continuity of the information is ensured. The applicability of the asynchronous learning method knowledge based learning systems can be provided with internet-based distance learning modules or desktop software based on a number of knowledge-based learning systems . In such systems, access to a knowledge base is provided to the learner with the designed software, so that one is able to learn the information in the direction of his/her own speed and ability independently of time and place. While the literature summary was being edited, it was tried to choose studies that are close to the subjects mentioned. M.Günel et al. studied on analyzing the Argument Based Science Inquiry (ABSI) approach implementations with respect to questioning in the classrooms and tried to explore the relationship between levels of questioning and negotiation of ideas. As a result of their studies, it was seen that the teacher's questioning strategies and the level of practice were effective in the formation and continuation of the negotiation process in the class [1]. In addition, F. Yeşildağ et al. investigated the effect of the Argument Based Science Inquiry (ABSI) approach used during research interrogation based activities on the achievement of primary school students in chemistry subjects. In this quasi-experimental designed research, pre-post tests and ABI reports were used as data collection tools. The sample of the study consists of 8th grade students from a primary school in a disadvantaged social economic area of Erzurum [2]. As to Ö. Aydın et al., they aimed to determine the opinions about argumentation by examining the effects of different applying methods of argumentation on metacognition and logical thinking abilities of science-technology teacher candidates [3]. D. Öztürk et al. examined the effects of web-based learning on nursing students. In web based education where urinary catheterization skills are measured, it is observed that the knowledge and abilities of the people increase with this method [4]. Likewise, J. Peng and colleagues chose a web-based educational platform at Tsinghua University. They provided a knowledge-based training network through a web-based platform to many users in the campus [5].

Thanks to this software created in this study, an interactive learning platform is provided to the person depending on his/her own learning speed. It is provided to add questions to the database independently of the teacher and re-examine questions that he/she recorded in learning weaknesses. Until full learning occurs, questions that are difficult to learn are asked again and again through a certain algorithm and the learning is achieved. Depending on the learning rate, the frequency of encountering the questions is changing and the questions that the person is weaker are presented more frequently. The questions that are difficult to learn are asked frequently while the learned question is carried over to a lower step. In contrast to conventional

interactive learning methods, a more innovative algorithm is being operated at this stage. Thanks to this new generation algorithm, it can be possible to present questions to the students more effectively. Even, these questions may be original or already learned questions previously.

II. INTERACTIVE LEARNING

Educational systems implemented on a virtual platform independent of time and place are called interactive training. Like many countries in the world, our country has adopted and implemented the interactive education model for many years. Through interactive learning, people can focus on the subject to learn with their own exploration and research skills being independent of the teacher and away from negative adverse effects of classroom motivation. These and similar advantages of the interactive training model as well as the learning methods realized with the direct participation of the teacher have weakened interactive learning and therefore the search for alternative learning methods has been started.

By means of an alternative interactive learning method developed by this study, people are tested with questions according to their learning competence in a data repository that they can update independently of the teacher. It is aimed to accelerate the active learning process by means of presenting the information to the learner with variance calculations on the questions in the data repository of the learning proficiency levels.

III. DATA REPOSITORY

Interactive learning programs require an effective database because of the differences in learning subjects. Instead of a separate database for each type of question, a flexible and highly effective database is crucial to the unity of the subjects to be learned. The same is true for presenting questions to the user. Designed in this context, MS SQL database structure provides a flexible and consistent structure to learner. In addition, the information repository, which is formed by separately grading the topics and questions to be learned, determines the learning sequence and realizes more effective learning.

In this study, parallel to the development of imaging technologies, it is agreed that the questions constituting the data repository are added to the data base in the form of images. As shown in Fig. 1, it is possible to add the images that are displayed in any picture format such as Jpeg, Png, Gif, Bmp to the data repository easily. In this way, the l can easily add, delete or update the images he/she has received with mobile phone, camera or computer to the data repository.

Figure 1: Visual recording for data repository

IV. KNOWLEDGE BASED LEARNING

The response rate is set to 0 (zero) for the first encountering with each question added to the data repository. Whenever the same question is asked to the learner, a rating process is performed at the rate of response according to the given answer. This rate is set separately for each question and the difficulty level of the problem varies depending on this rating rate. All questions in the data repository are recalculated according to their rating rates just before being asked to the learner. Therefore, if all the questions are considered as a series, the median, variance, mode values and also minimum and maximum response rates will vary in the response rates of this sequence. According to the answers given to the questions, the response rate of the related question in the database is updated whenever the question is answered. These updates are used to identify topics that are difficult to learn and that have been learned before. Thanks to a developed algorithm, subjects that are difficult to learn and subjects that have already been learned are distinguished and students are presented them independently of the teacher. According to personal preferences, the learner is taught interactively at his/her own learning speed.

V. ALGORITHM AND FLOWCHART

The flowchart is a diagram that is formed by describing the events that will take place in the process in blocks related to each other in order to increase functionality and validity in the field of software and process management. It is important that the algorithm of the coding to be done is shown in a schematic way in advance in order to eliminate possible operational errors. The flow chart of the interactive learning software is given in Fig. 2.

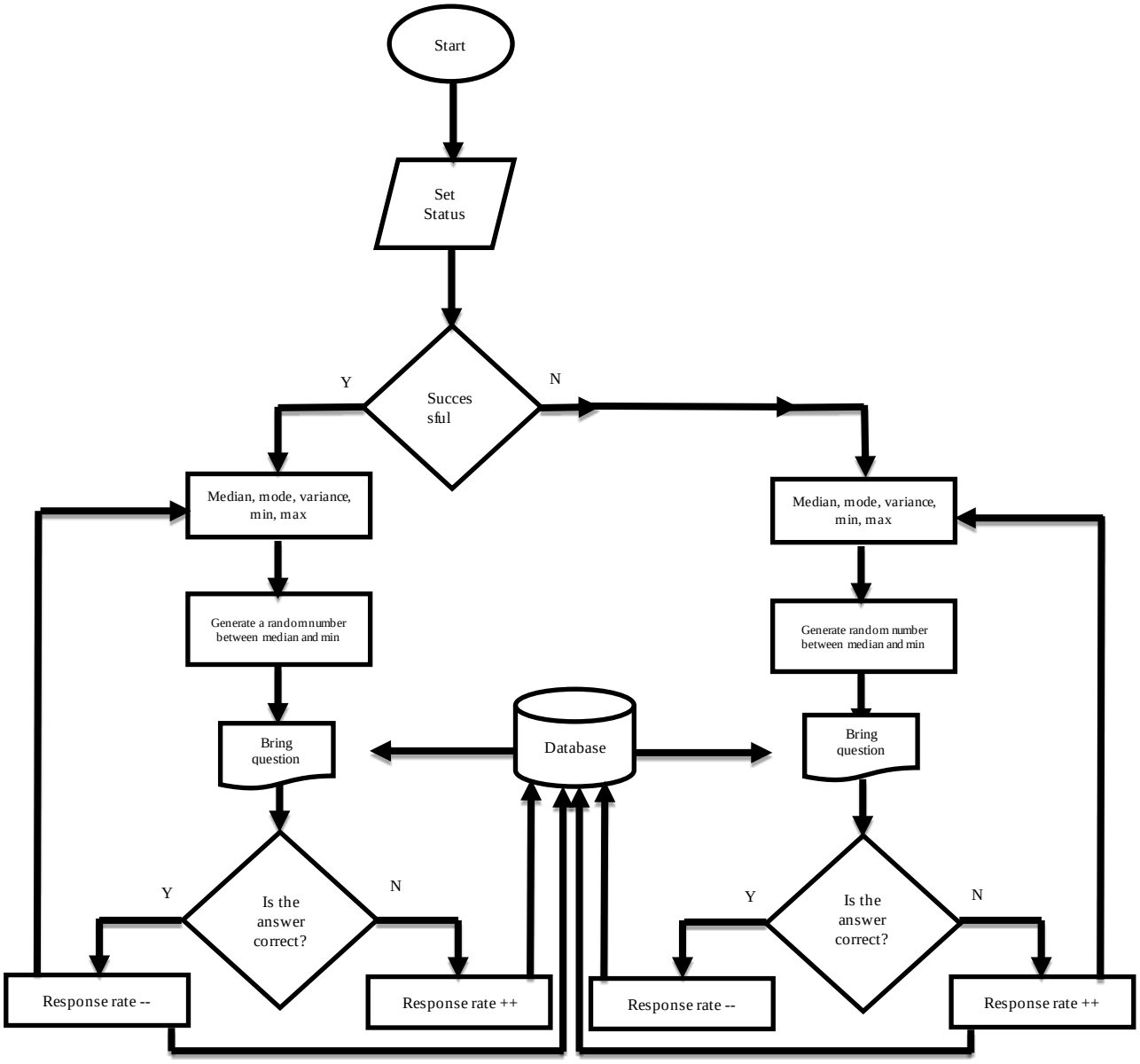


Figure 2: Algorithm of the software

As it is seen in Fig. 2, first of all there is a need for an information repository consisting of the questions to be learned in order to provide learning. Questions can be added to the database by the user in any visual format without any restrictions on the number of questions. On this account, the user can use the software for whatever he wants to learn.

During the instruction, the interface given in Figure 3 is used. User-friendly design techniques have been utilized to design this interface which is created with the C# encoding language in the Microsoft Visual Studio environment. Thus, it is aimed to motivate learners to learn by using aesthetics and ergonomics in the use of software. The algorithm running in the background

With the response to the question on the screen, the answer rating of the relevant question in the database is updated and therefore the median, mode, variance, maximum and minimum value calculation in the designed variable array are calculated

of the interface shown in Fig.3 is presented in detail in Fig.2. In terms of functioning, this algorithm takes place in the following way; first of all the preference is taken about practicing on the subjects that are well known or encountering questions or topics which were poorly learned before. Following this preference, an array is created in the database taking the response ratings with variable rates ranging from 0 to 100 into account. The median, mode, variance, largest and smallest sequence values of the sequential array are calculated after each element of this sequence is ranked using the Quicksort Algorithm. With these calculated values, the questions related to the learning rating are brought to the person from the database randomly.

again. Thanks to the array showing variable values for each answer, the learned questions reach low ratings while the questions that are not learned reach high ratings. The questions

whose answers are learned reach low ratings while the ones whose answers are not learned reach high ratings.

At the beginning of the algorithm, the user is offered two options: to practice on well-known topics or to practice on topics that have been poorly learned. These options are presented depending on the specified and recorded response rates.

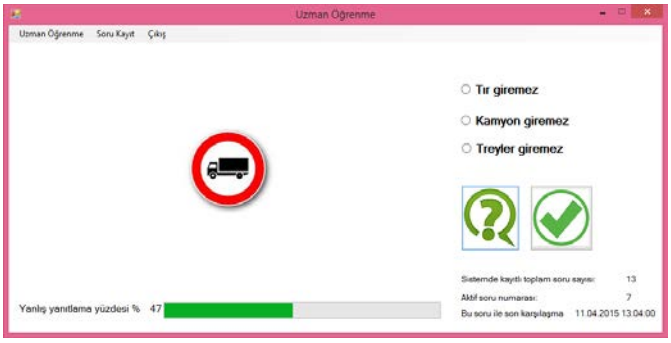


Figure 3: Knowledge-based learning interface

VI. RESULTS AND SUGGESTIONS

Thanks to the technologies developed for widespread access to information, nowadays knowledge based education systems have been quite developed. By passing the created data repositories through an effective training process, a variety of topics and information can be taught more easily and permanently. The knowledge-based interactive learning method which is created in this study consists essentially of two phases: In the first stage, the person is more intensely focused on the issues that he / she has difficulty in learning. Thus, the person can access the information in a form based on learning, not memorizing. As the second stage, the permanence of the information which is important in terms of the continuity of the information acquisition is presented to the person effectively. It is possible to obtain both the learning outcomes gained by the person experiencing difficulties in learning as well as the continuity of previously learned information. In both stages, information is presented to the person through interactive learning method and thanks to this, the acquisition of knowledge is achieved asynchronously, independent of the teacher.

The created software can be put into practice in mobile applications in parallel with the developing technology, and information that is already easy to access can be presented to the people more effectively.

REFERENCES

[1] M. Günel, S. Kıngır, Ö. Geban. Argümantasyon Tabanlı Bilim Öğrenme (ATBÖ) Yaklaşımının Kullanıldığı Sınıflarda Argümantasyon ve Soru Yapılarının İncelenmesi. Eğitim ve Bilim 2012, Cilt 37, Sayı 164

[2] F. Yeşildağ-Hasançebi, M. Günel, Argümantasyon Tabanlı Bilim Öğrenme Yaklaşımının Dezavantajlı Öğrencilerin Fen Bilgisi Başarılarına Etkisi, İlköğretim Online, 12(4), 1056-1073, 2013

[3] Ö. Aydın, F. Kaptan, Fen-Teknoloji Öğretmen Adaylarının Eğitiminde Argümantasyonun Biliş Üstü ve Mantıksal Düşünme Becerilerine Etkisi ve Argümantasyona İlişkin Görüşler, Eğitim Bilimleri Araştırmaları Dergisi, Uluslararası E-Dergi, Cilt:4 Sayı:2, Ekim 2014

[4] D. Öztürk, L. Dinç, Effect of web-based education on nursing students' urinary catheterization knowledge and skills, Nurse Education Today 34 (2014) 802–808, August 2013

[5] J. Peng, D. Jiang, X. Zhang, Design and implement a knowledge management system to support web-based learning in higher education, 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems, Procedia Computer Science 22 (2013) 95 – 103, 2013

Occupational Health And Safety In The Chemical Industry; Modeling and Software Developed On The Basis Of Fuzzy Logic To Prevent Job Accidents

M.ZİLE[1]

[1] Mersin University, Mersin/Turkey, mehmetzile@yahoo.com

Abstract **- Methods: In the provinces of Ankara, Adana and Mersin, 9 chemical working job security survey studies and work safety experts were carried out in this study. Computer software is developed by creating a fuzzy logic based risk assessment analysis model considering various hazards in occupational health and safety in the chemical industry. Results: It is very difficult to model risk in a workplace environment with many specific and ambiguous hazards and model a system to simulate these hazards. Software was developed by creating a fuzzy logic based risk assessment analysis model in consideration of many hazards in occupational health and safety in the chemical industry. An alternative approach to risk assessment has been proposed using fuzzy decision making approach and matrix method. With this approach, occupational health and safety in the Chemical Industry specialists are provided with blurred linguistic assessments before calculations are made, and the inconsistencies in decision making are reduced by taking arithmetic averages of these assessments. It has been seen that the three most important risks in the metal work done by creating the fuzzy logic based modeling of work safety risk analysis model and software using the fuzzy decision making approach and the matrix method to increase the job security in the chemical industry. Conclusion: It is very difficult to assess risk in the chemical industry which are many specific and indeterminate hazards, and to model a system to simulate these hazards. In this study, a computer software and hardware was developed by creating a fuzzy logic based risk assessment analysis model considering many hazards in occupational health and safety in the chemical industry.**

Keywords **– Fuzzy logic, Safe and Health, Job Accidents, Chemical Industry**

I. INTRODUCTION

The chemical industry is a sector that provides resources for materials produced in various fields of the industry. The area covered by the raw materials needed for many sectors is called the chemical industry [1]. New substances are produced by chemical reactions in factories. The main chemical industry is products, paint, cleaning products, cosmetics, pharmaceuticals, fertilizers, agrochemicals, thermoplastics, varnishes, synthetic fibers, petrochemical products [2]. The Turkish chemical industry consists mainly of facilities where production of various chemical raw materials and consumer products such as petrochemical, soap, detergent, fertilizer, medicine, paint varnish, synthetic fiber, soda are carried out. Companies operating in the sector differ in terms of scale and capital resources [3]. A large number of companies operating in the sector are small and medium-sized enterprises, as well as large-scale firms and multinational corporations. The main input of plastics production is provided by the petrochemical sector at 90%. The petrochemical sector is a large scale, capital and technology intensive sector [4]. The plastic and rubber sector is an import-dependent sector with over 90%. About 2600 chemical substances and preparations are produced in the chemical sector, and some of the methods and technologies used in these products are capable of meeting global competition. The chemical industry, which is the most basic stone of the industry, includes risks due to the fact that it is a sector where the business accidents are also experienced intensively with the contribution of the economy [5]. In the chemical industry, where insurance companies have the highest risk category in their policies, it is possible to ignore any of the necessary precautions that can lead to irreversible life-threatening risks. The chemical industry, one of the indispensable industries due to the products it provides to the production enterprises, faces as an area in which necessary precautions and consciousness must be provided in the headings of occupational health and safety [6]. The chemical sector is a sector that needs to be developed not only in production and manufacturing but also in occupational safety and occupational health areas [7]. The chemical sector has become a branch of institutional firms, whose procedures are monitored and developed. Today, business is divided into a wide variety of segments within the chemical industry of the accident. Costs and losses paid by companies in the chemical sector make these firms more sensitive to workers' health [8].

II. DANGEROUS CAUSES OF BUSINESS INJURIES AND CAUSES OF OCCUPATIONAL DISEASES IN CHEMICAL INDUSTRY

In the chemical industry, where dangerous substances are used in business, the first thing to do to minimize the likelihood of an accident is feasibility, measurement and control [9]. The risk capacity of the product, the suitability of the ambient conditions, the appropriate conditions for it, the necessary

clothing-dress regulations, the disciplined and organized work processes, the routine environment controls, etc., which are used in the enterprises which are exposed to combustible, explosive, Essential actions are life-saving arguments. Accident prevention is not only a matter that can be imposed on employers and legislators; workers' consciousness, trade union informants, internal audit professionals and their attachment to the standard in industry chambers [10]. Knowledgeable about technical developments, followed by changing processes, specialized in their work, competent in the workforce, is vital for both the employer and the worker. It is a fact that the largest share in the establishment of safe working in the sector falls to the state [11]. Accidents such as acid and gas burns, phosphor burns, digestive and respiratory deformations, cut and wound interactions are the most common accidents in the industry. Occupational diseases in the chemical industry are seen as lung diseases, cancer, liver and kidney disorders, anemia and skin diseases. This type of accidents and interactions that are likely to be experienced are minimized in companies that have specialized in occupational health and safety [12].

III. FUZZY LOGIC CONCEPT

The fundamental difference between fuzzy logic and classical logic is that classical logic proposals use only extreme values. Something like Aristotle's conjecture is or is not a member of a cluster [13]. In other words, it is either black or white. In the real world you can not find full black or white. Fuzzy logic works with the grids as soon as possible in accordance with the real life. In extreme cases there are black or white. Modeling and controlling complex systems using classical logic methods is thus difficult because the data must be clear and precise. Fuzzy logic frees the person from this requirement and allows for a more qualitative description. Saying only middle age means 41.5 years for one person is enough for many applications. Thus, the reduction of information to a considerable extent will be the subject, and a qualitative description that can be more easily understood rather than a mathematical description can be made [14]. The central concept of fuzzy theory is fuzzy clusters. The concept of the cluster may come a little mathematically, but it is easy to understand. For example, if we examine the concept of "middle age", we can see that the boundaries of this concept vary from person to person. We can not easily formulate the concept mathematically because the exact boundaries are not the subject. But in general, aged between 40 and 55 can be considered as middle aged borders. When we want to express this concept graphically, a curve like the anticipation bell curve will emerge. This curve is called the "curriculum of ownership" and shows which value in the concept is which weight. A fuzzy set can be clearly represented by its ownership function [15]. The property function can take any value between 0 and 1. It is possible to make the desired fine adjustment between such an ownership function and "absolutely belonging" or "absolutely not belonging". Another important concept in fuzzy logic is the concept of linguistic variable [16]. The linguistic variable is a variable that can be defined as words such as "hot" or "cold". The values of a linguistic variable are expressed in fuzzy sets. For example, "hot", "cold" and "very hot" expressions can be used for the room temperature linguistic variant. Each of these three terms is modeled as separate fuzzy sets [17]. The greatest benefit of fuzzy logic is that it can be easily modeled with "human-specific experience" and allows even indefinite concepts to be expressed mathematically. For this reason it is particularly suitable for approaching nonlinear systems. A rule base is created for this. The rule base should include all the observations, experiences and mathematical correlates, ie all knowledge, of the human being concerned with the subject studied up to that point. The better and wider the rule base is, the more precise and accurate the results are obtained [18].

IV. RISK ASSESSMENT ANALYSIS IN CHEMICAL INDUSTRY

t: hazard (up to 1 hazard number T),

d: danger value,
dtmax: max value of danger,

Nd: non-hazardous value (membership level $u_{violence}=0$),
dd: very low dangerous value (membership level $u_{violence}=0.2$),
Dd: low dangerous value (membership level $u_{violence}=0.4$),
Od: moderate danger value (membership level $u_{violence}=0.6$),
yd: high dangerous value (membership level $u_{violence}=0.8$),
Yd: very high dangerous value (membership level $u_{violence}=1.0$)

t(1): workplace ambient temperature (when it drops below 15 ^{0}C and exceeds 25 ^{0}C), $d_1^{min}=10^0$C and $d_1^{max}=35^0$C,
t(2): the air flow rate (when the speed drops below 0,5 m/sec and rises above 1 m/sec), d_2^{min}=0,4 m/sec and d_2^{max} = 1,1m/sec,
t(3): relative humidity (over 60%), d_3^{max} = 60% [19]

Simple suffocating gases
t (4): carbon dioxide, d_4^{max} = 5000 ppm.
t (5) : lead, d_5^{max} = 0,15 mg/m^3 ,
t (6) : mercury, d_6^{max} = 0,075 mg/m^3 ,
t (9) : arsenic, d_7^{max} = 0,5 mg/m^3 ,
t (8): beryllium, d_8^{max} = 2 mg/m^3
t (9) : ammonia, d_9^{max} = 25 ppm,
t (10) : chlorine, d_{10}^{max} = 1 ppm,
t (11) : nitrogen dioxide, d_{11}^{max} = 5 ppm,
t (12) : sulphur dioxide, d_{12}^{max} = 0,1 ppm,
t (13) : ozone, d_{13}^{max} = 0,1 ppm
t (14) : arsine, d_{14}^{max} = 0,05ppm,
t (15) : phosphine, d_{15}^{max} = 0,3 ppm,
t (16) : stibine, d_{16}^{max} = 0,1 ppm,
t (17):ethane
t (18):propane and butane , d_{17}^{max} =1000 ppm
t (19):acetylene
t (20):hydrogen
t (21):nitrogen,
t (22):argon,
t (23):neon,
t (24):helium,
t (25):ethylene
t (26):propylene

Chemical suffocating gases
t (27):carbon monoxide , d_{27}^{max} =50 ppm
t (28):hydrogen cyanide , d_{28}^{max} =10 ppm
t (29):hydrogen sulphide , d_{29}^{max} =10 ppm

Irritating gases
t (30):ammonia , d_{30}^{max} =25 ppm
t (31):chlorine , d_{31}^{max} =1 ppm
t (32):sulfur dioxide , d_{32}^{max} =2 ppm
t (33):phosgene , d_{33}^{max} =0,1 ppm
t (34):nitrogen oxides , d_{34}^{max} =3 ppm
t (35):ozone , d_{35}^{max} =0.1 ppm
t (36):formaldehyde , d_{36}^{max} =0.75 ppm

Gases with systemic poisoning effects
t (37):asrin , d_{37}^{max} =0.05 ppm
t (38):stibine , d_{38}^{max} =0.1 ppm
t (39):phosphin , d_{39}^{max} =0.3 ppm
t (40):nickel carbony, d_{40}^{max} =l 0.05 ppm
t (41):carbon sulphide , d_{41}^{max} =10 ppm

Fibrogenic powders
t (42):quartz , d_{42}^{max} =0.1 mg / m3
t (43):silica, d_{43}^{max} = 0.1 mg / m3
t (44):ridmit , d_{44}^{max} =0.05 mg / m3
t (45):tropolis , d_{45}^{max} =0.1 mg / m3
t (46):amosite , d_{46}^{max} =0.5 fiber / cm3
t (47):chrysotile , d_{47}^{max} =2 fibers / cm3
t (48):crosidolite , d_{48}^{max} =0.2 fiber / cm3
t (49):talc , d_{49}^{max} =2 mg / m3
t (50):mica , d_{50}^{max} =3 mg / m3
t (51):coal dust , d_{51}^{max} =2 mg / m3

Toxic dusts
t (52):lead , d_{52}^{max} =0.15 mg / m3
t (53):chromium , d_{53}^{max} =0,5 mg / m3
t (54):cadmium , d_{54}^{max} =0.05 mg / m3
t (55):vanadium , d_{55}^{max} =0.05 mg / m3
t (56):trinitrotoluene , d_{56}^{max} =0,5 mg / m3
t (57):compounds arsenic, d_{57}^{max} = 0.2 mg / m3
t (58):dinitrophenol, d_{58}^{max} = 0,2 mg / m3

Allergic dusts
t (59):cotton
t (60):linen
t (61):hemp
t (62):sisal hemp
t (63):jute
t (64):platinum compounds
t (65):wood dust
Boring (inert) powders
t (66):Esd , d_{66}^{max} =10 mg / m3
t (67):calcium carbonate
t (68):cellulose-paper fiber
t (69):emery
t (70):glycerin misti
t (71):gypsum
t (72):kaolin
t (73):limestone
t (74):magnesite
t (75):marble
t (76):glass wool
t (77):pentaerythritol
t (78):paris plaster
t (79):portland cement
t (80):lipstick
t (81):silicon,
t (82):silicon carbide
t (83):starch
t (84):sakoroz
t (85):titanium dioxide
t (86):zinc stearate
t (87):zinc oxide dust

Industrial solvers
t (88):gasoline
t (89):nail polish remover
t (90):toluene
t (91):trichlorethylene
t (92):ethylacetate
t (93):benzol
t (94):halogenated hydrocarbons
t (95):alcohols
t (96):ethers
t (97):glycol derivatives
t (98):esters
t (99):ketones

Hydrocarbons
t (100):hexane , d_{100}^{max} =50 ppm
t (101):hexane isomers , d_{101}^{max} =500 ppm
t (102):octane, d_{102}^{max} = 300 ppm
t (103):cyclohexane , d_{103}^{max} =300 ppm
t (104):turpentine , d_{104}^{max} =100 ppm

Nitro-hydrocarbons
t (105):nitroethane , d_{105}^{max} =100 ppm
t (106):benzene , d_{106}^{max} =10 ppm,
t (107):toluene , d_{107}^{max} =50 ppm
t (108):xylene , d_{108}^{max} =100 ppm
t (109):tetrachloromethane , d_{109}^{max} =5 ppm
t (110):trichloroethane , d_{110}^{max} =350 ppm
t (111):trichlorotrifluoroethane d_{111}^{max} =1000 ppm
t (112):methyl alcohol , d_{112}^{max} =200 ppm
t (113):ethyl alcohol , d_{113}^{max} =1000 ppm
t (114):propyl alcohol , d_{114}^{max} =200 ppm
t (115):ethylether , d_{115}^{max} =400 ppm
t (116):isopropyl ether , d_{116}^{max} =250 ppm

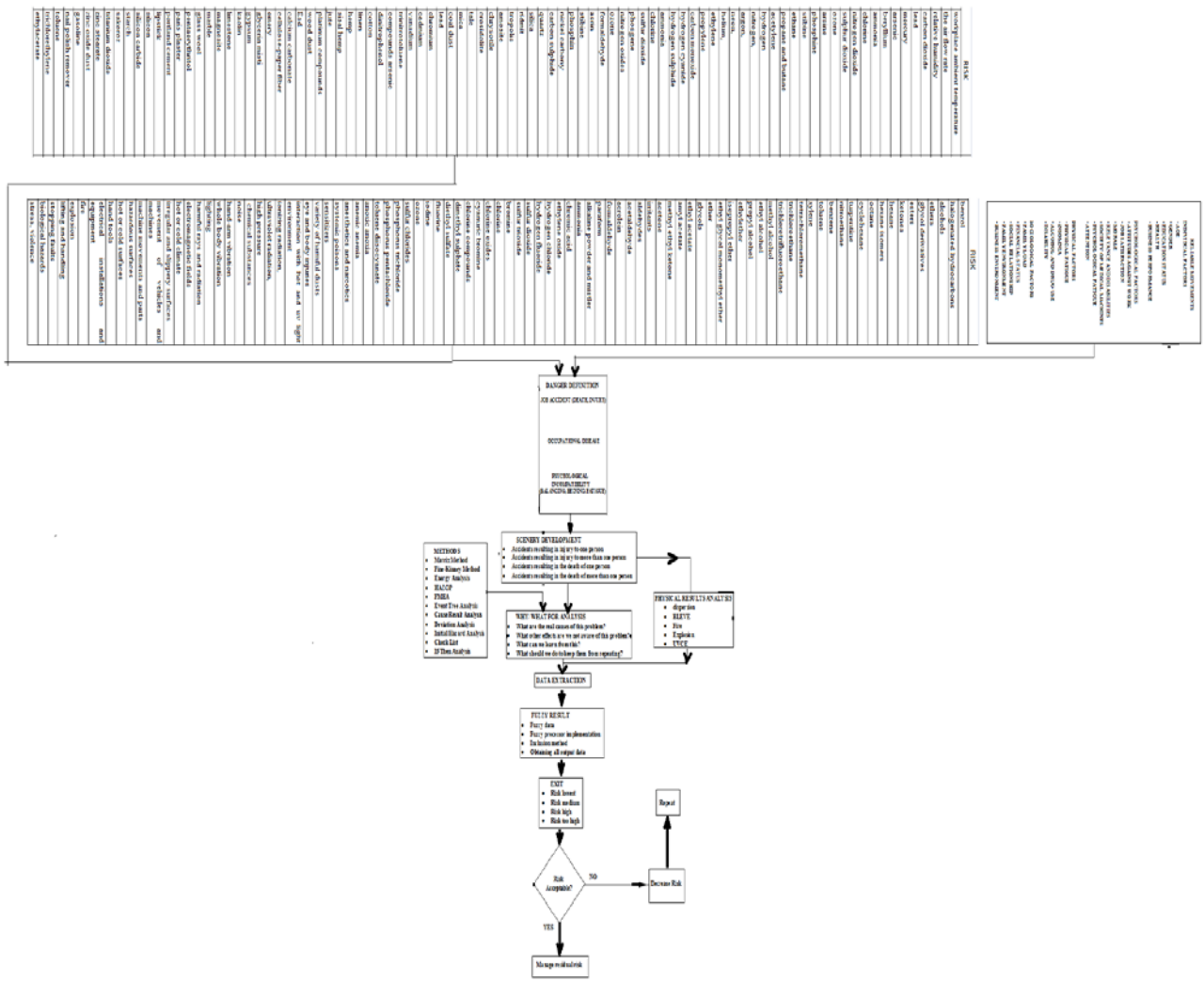


Figure 1: Creating risk analysis modelling in chemical industry.

t (117):ethyl glycol monomethyl ether , d_{117}^{max} =5 ppm
t (118):ethyl glycol monoethyl ether , d_{118}^{max} =50 ppm
t (119):glycols
t (120):ethyl acetate, d_{120}^{max} = 400 ppm
t (121):amyl acetate , d_{121}^{max} =100 ppm
t (122):methyl ethyl ketone , d_{122}^{max} =200 ppm
t (123):acetone , d_{123}^{max} =750 ppm
t (124):irritants
t (125):aldehydes
t (126):acetaldehyde
t (127):acrolein
t (128):formaldehyde
t (129):paraform
t (130):alkaline powder and mistler
t (131):ammonia
t (132):chromic acid
t (133):ethylene oxide
t (134):hydrogen chloride
t (135):hydrogen fluoride
t (136):sulfur dioxide
t (137):sulfur trioxide
t (138):bromine
t (139):chlorine
t (140):chlorine oxides
t (141):cyanuric bromine
t (142):chlorine compounds
t (143):dimethyl sulphate
t (144):diethyl sulfate
t (145):fluorine
t (146):iodine
t (147):ozone
t (148):sulfur chlorides
t (149):phosphorus trichloride
t (150):phosphorus pentachloride
t (151):toluene diisocyanate
t (152):anoxic anoxia
t (153):anemic anemia
t (154):anesthetics and narcotics
t (155):systemic poisons
t (156):sensitizers

t (157) :variety of harmful dusts,

t (158) :eye and body injuries
t (159) :interaction with hot and uv light environment,
t (160) :ionizing radiation,
t (161) :ultraviolet radiation,
t (162) : high pressure
t (163) : chemical substances
t (164 : noise
t (165) : hand-arm vibration
t (166) : whole body vibration
t (167) : lighting
t (168) : harmful rays and radiation,
t (169) : electromagnetic fields,
t (170) : hot or cold climate,
t (171) : irregular and slippery surfaces,
t (172) : movement of vehicles and machines,
t (173) : machine movements and parts,
t (174) : hazardous surfaces,
t (175) : hot or cold surfaces,
t (176) : hand tools,
t (177) : electrical installations and equipment,
t (178) : fire,
t (179) : explosion,
t (180 : lifting and handling,
t (181) : stopping faults,
t (182) : biological hazards,
t (183) : stress, violence

The dangers in the chemical industry are defined as above given by making researches and examinations in 9 different factories in Ankara, Adana and Mersin. The risk assessment analysis in the chemical industry is modeled as shown in Figure 1. The hazards created by insecure environments and the unsafe motions of human origin are entered into the hazard identification block as data [20]. These hazards are classified as hazardous and injured chemical work accidents, chemical occupational diseases and psychological disorders in the hazard identification block. The data from here was entered into the scenery development block and the result grouping was done. The reasons for the receipt are determined by the cause analysis block. We have received a risk decomposition. In order to reduce the unacceptable risks, repetition cycle has been done and according to this cycle energy, ventilation, pressure, humidity, temperature, etc. the system is automatically adjusted to reduce the risks of operation and at the same time warning alerts were sent to the business manager automatically [21]. The probability of emergence of danger is given in Table 1.

Table 1. Probability chart of emergence of hazard.

PROBABLE VALUE	POSSIBILITY OF LEAVING
Very small (1)	Hardly ever
Small (2)	Very few (once a year), only in abnormal situations.
Middle (3)	Little (once every three months)
High (4)	Frequently (once a week)
Very high (5)	Very often (every day), under normal working conditions

Id=5 'emergence of danger in the day'(membership level $u_{probability}$ =1.0)
Id=4 'dangerous emergence weekly' (membership level $u_{probability}$ =0.8)
Id=3 'emergence of danger in the moon' (membership level $u_{probability}$ =0.6)
Id=2 'emergence of danger in three months' (membership level $u_{probability}$ =0.4)
Id=1 'the emergence of danger in the year' (membership level $u_{probability}$ =0.2)
Id=0 'never appear dangerous' (membership level $u_{probability}$ =0.0)
are determined. In Table 2, the severity rating is determined.

Table 2. Violence rating table.

PROBABLE VALUE	POSSIBILITY OF LEAVING
Very small (1)	Hardly ever
Small (2)	Very few (once a year), only in abnormal situations.
Middle (3)	Little (once every three months)
High (4)	Frequently (once a week)
Very high (5)	Very often (every day), under normal working conditions

'Death, incapacity to work permanently' (membership level $u_{violence}$=1.0)
'Serious injury, long-term treatment, occupational disease' (membership level $u_{violence}$=0.8)
'Light injury, inpatient treatment / injury' (membership level $u_{violence}$=0.6)
'There is no loss of work day, outpatient treatment without permanent effect' (membership level $u_{violence}$ = 0.4)
'No business hours lost, need immediate relief, first aid' (membership level $u_{violence}$ = 0.2)
'The danger does not emerge at all' (membership level $u_{violence}$ = 0.0)
are determined.

PROBABILITY / VIOLENCE	1 VERY LOW	2 LOW	3 MIDDLE	4 HIGH	5 VERY HIGH
1 VERY LOW	NEGLIGIBLE 1	NEGLIGIBLE 2	LOW 3	LOW 4	LOW 5
2 LOW	NEGLIGIBLE 2	LOW 4	LOW 6	MIDDLE 8	MIDDLE 10
3 MIDDLE	LOW 3	LOW 6	MIDDLE 9	MIDDLE 12	HIGH 15
4 HIGH	LOW 4	MIDDLE 8	MIDDLE 12	HIGH 16	HIGH 20
5 VERY HIGH	LOW 5	MIDDLE 10	HIGH 15	HIGH 20	EXTREMELY HIGH 25

Figure 2: Score matrix.

As shown in Fig. 2, the score matrix is determined. In the linguistic variable membership function, the score table in Table 3 are determined. Risk assessment score level are determined.
Risk = Violence x Probability [19].
The risks with a score of 25 are very high and the membership rate is $u_{RİSK}$ = 1,0.
Risks between 15 and 25 are high risk and membership level is $u_{RİSK}$ = 0,8.
Risks between 8 and 12 are medium risk and membership level is $u_{RİSK}$ = 0,6.
The risks ranging from 3 to 6 are low, and the membership grade is $u_{RİSK}$ = 0,4.

Risks with a score of 1 and 2 are very low, and the membership level is $u_{RİSK} = 0{,}2$.
If the score is 0, there is no risk and the membership grade is $u_{RİSK} = 0{,}0$ [20].

Table 3. Score table.

NON-TOLLABLE RISKS (25)	The business should not be started until the risk is reduced to an acceptable level, and if there is an ongoing activity, it should be stopped immediately. If it is still not possible to reduce the risk despite the precautionary measures, the activity should be canceled.
IMPORTANT RISKS (15,16,20)	The business should not be started until the specified risk is reduced, if there is an ongoing activity, it should be stopped immediately. Emergency measures should be taken if the risk is to continue, and as a result of these measures, the continuation of the activity should be decided.
MEDIUM LEVEL RISKS (8,9,10,12)	Activities should be initiated to reduce the identified risks.
INCURED RISKS (2,3,4,5,6)	Additional control processes may not be needed to remove the identified risks. However, existing controls should be maintained.
INCREDIBLE RISKS (1)	It may not be necessary to plan the control processes and to keep records of the activities to be carried out in order to eliminate the identified risks.

Firstly, perunit values of the defined hazards are calculated as $d_t{}^n = d_t / d_t{}^{max}$, t=1, ……T', then perunit values of the undefined hazards are calculated as $\underline{d}_t{}^n = \underline{d}_t / \underline{d}_t{}^{max}$, t=T'+1,…..T, a fuzzy number scale is defined by the maximum method, $\underline{d}_t{}^n, r_t$, t=T'+1,……,T, the fuzzy definition of each defined hazard being $\underline{c}_t = \underline{W}_t . \underline{d}_t{}^n$ is calculated and all linguistic hazards are $c_t = \underline{W}_t . r_t$ for each t=T'+1,…..T is calculated on the basis of the comparison method of the blurred numbers in the order of the hazards [21].

Table 4. Linguistic variable table.

LANGUAGE VARIABLES	FUZZY VALUE
Nd: non-hazardous value	(0,0,0)
dd: verylow dangerous value	(1,1,2)
Dd: low dangerous value	(1,2,3)
Od: moderate danger value	(2,3,4)
yd: high dangerous value	(3,4,5)
Yd: very high dangerous value	(4,5,5)

Table 4 shows the linguistic variable table, the fuzzy decision-making approach in Table 5, and the probability-risk assessments determined by three different job security experts using the matrix method and the risk membership ratings of these assessments in Table 6, conversion of the risks obtained into linguistic variables of probability-severity membership ratios is given in Table 7. Once the blur sets are defined and their membership functions are assigned, the rules are written for each combination of control variables. These rules relate input variables to output variables by using If-Then expressions in decision-making. The condition 'If' is a prelude to the result of each rule. In general, each rule is shown as 'If', and then 33489 combined rules are created when identical result expressions are issued. The software was generated as shown in Figure 3 and the priority order of the risks was obtained as in Table 8.

Table 5. Fuzzy decision making approach in an applied chemical industry and probability-severity evaluations of risk using matrix method.

	RISK	ODSS	VIOLENCE
1	workplace ambient temperature	0,2-0,4-0,6	0,0-0,2-0,4
2	the air flow rate	0,2-0,4-0,6	0,0-0,2-0,4
3	relative humidity	0,2-0,4-0,6	0,0-0,2-0,4
4	carbon dioxide	0,2-0,2-0,4	0,6-0,8-1,0
5	lead	0,2-0,2-0,4	0,6-0,8-1,0
6	mercury	0,2-0,2-0,4	0,6-0,8-1,0
7	arsenic	0,2-0,2-0,4	0,6-0,8-1,0
8	beryllium	0,2-0,2-0,4	0,6-0,8-1,0
9	ammonia	0,2-0,2-0,4	0,6-0,8-1,0
10	chlorine	0,2-0,2-0,4	0,6-0,8-1,0
11	nitrogen dioxide	0,2-0,2-0,4	0,6-0,8-1,0
12	sulphur dioxide	0,2-0,2-0,4	0,6-0,8-1,0
13	ozone	0,2-0,2-0,4	0,6-0,8-1,0
14	arsine	0,2-0,2-0,4	0,6-0,8-1,0
15	phosphine	0,2-0,2-0,4	0,6-0,8-1,0
16	stibine	0,2-0,2-0,4	0,6-0,8-1,0
17	ethane	0,2-0,2-0,4	0,6-0,8-1,0
18	propane and butane	0,2-0,2-0,4	0,6-0,8-1,0
19	acetylene	0,2-0,2-0,4	0,6-0,8-1,0
20	hydrogen	0,2-0,2-0,4	0,6-0,8-1,0
21	nitrogen,	0,2-0,2-0,4	0,6-0,8-1,0
22	argon,	0,2-0,2-0,4	0,6-0,8-1,0
23	neon,	0,2-0,2-0,4	0,6-0,8-1,0
24	helium,	0,2-0,2-0,4	0,6-0,8-1,0
25	ethylene	0,2-0,2-0,4	0,6-0,8-1,0
26	propylene	0,2-0,2-0,4	0,6-0,8-1,0
27	carbon monoxide	0,2-0,2-0,4	0,6-0,8-1,0
28	hydrogen cyanide	0,2-0,2-0,4	0,6-0,8-1,0
29	hydrogen sulphide	0,2-0,2-0,4	0,6-0,8-1,0
30	ammonia	0,2-0,4-0,4	0,2-0,4-0,6
31	chlorine	0,2-0,4-0,4	0,2-0,4-0,6
32	sulfur dioxide	0,2-0,4-0,4	0,2-0,4-0,6
33	phosgene	0,2-0,4-0,4	0,2-0,4-0,6
34	nitrogen oxides	0,2-0,4-0,4	0,2-0,4-0,6
35	ozone	0,2-0,4-0,4	0,2-0,4-0,6
36	formaldehyde	0,2-0,4-0,4	0,2-0,4-0,6
37	asrin	0,2-0,4-0,4	0,4-0,6-0,8
38	stibine	0,2-0,4-0,4	0,4-0,6-0,8
39	phosphin	0,2-0,4-0,4	0,4-0,6-0,8
40	nickel carbony	0,2-0,4-0,4	0,4-0,6-0,8
41	carbon sulphide	0,2-0,4-0,4	0,4-0,6-0,8
42	quartz	0,4-0,4-0,4	0,4-0,6-0,6
43	silica	0,4-0,4-0,4	0,4-0,6-0,6
44	ridmit	0,4-0,4-0,4	0,4-0,6-0,6
45	tropolis	0,4-0,4-0,4	0,4-0,6-0,6
46	amosite	0,4-0,4-0,4	0,4-0,6-0,6
47	chrysotile	0,4-0,4-0,4	0,4-0,6-0,6
48	crosidolite	0,4-0,4-0,4	0,4-0,6-0,6
49	talc	0,4-0,4-0,4	0,4-0,6-0,6
50	mica	0,4-0,4-0,4	0,4-0,6-0,6
51	coal dust	0,4-0,4-0,4	0,4-0,6-0,6
52	lead	0,4-0,4-0,4	0,4-0,6-0,6
53	chromium	0,4-0,4-0,4	0,4-0,6-0,6
54	cadmium	0,4-0,4-0,4	0,4-0,6-0,6
55	vanadium	0,4-0,4-0,4	0,4-0,6-0,6
56	trinitrotoluene	0,4-0,4-0,4	0,4-0,6-0,6
57	compounds arsenic	0,4-0,4-0,6	0,2-0,4-0,4
58	dinitrophenol	0,4-0,4-0,6	0,2-0,4-0,4
59	cotton	0,4-0,4-0,6	0,2-0,4-0,4
60	linen	0,4-0,4-0,6	0,2-0,4-0,4
61	hemp	0,4-0,4-0,6	0,2-0,4-0,4
62	sisal hemp	0,4-0,4-0,6	0,2-0,4-0,4
63	jute	0,4-0,4-0,6	0,2-0,4-0,4
64	platinum compounds	0,4-0,4-0,6	0,2-0,4-0,4
65	wood dust	0,4-0,4-0,6	0,2-0,4-0,4
66	Esd	0,4-0,4-0,6	0,2-0,4-0,4
67	calcium carbonate	0,4-0,4-0,6	0,2-0,4-0,4
68	cellulose-paper fiber	0,4-0,4-0,6	0,2-0,4-0,4
69	emery	0,4-0,4-0,6	0,2-0,4-0,4
70	glycerin misti	0,4-0,4-0,6	0,2-0,4-0,4
71	gypsum	0,4-0,4-0,6	0,2-0,4-0,4
72	kaolin	0,4-0,4-0,6	0,2-0,4-0,4
73	limestone	0,4-0,4-0,6	0,2-0,4-0,4
74	magnesite	0,4-0,4-0,6	0,2-0,4-0,4
75	marble	0,4-0,4-0,6	0,2-0,4-0,4
76	glass wool	0,4-0,4-0,6	0,2-0,4-0,4
77	pentaerythritol	0,4-0,4-0,6	0,2-0,4-0,4
78	paris plaster	0,4-0,4-0,6	0,2-0,4-0,4
79	portland cement	0,4-0,4-0,6	0,2-0,4-0,4
80	lipstick	0,4-0,4-0,6	0,2-0,4-0,4
81	silicon	0,4-0,4-0,6	0,2-0,4-0,4
82	silicon carbide	0,4-0,4-0,6	0,2-0,4-0,4
83	starch	0,4-0,4-0,6	0,2-0,4-0,4
84	sakoroz	0,4-0,4-0,6	0,2-0,4-0,4
85	titanium dioxide	0,4-0,4-0,6	0,2-0,4-0,4
86	zinc stearate	0,4-0,4-0,6	0,2-0,4-0,4
87	zinc oxide dust	0,4-0,4-0,6	0,2-0,4-0,4
88	gasoline	0,4-0,4-0,6	0,2-0,4-0,4
89	nail polish remover	0,4-0,4-0,6	0,2-0,4-0,4
90	toluene	0,4-0,4-0,6	0,2-0,4-0,4
91	trichlorethylene	0,4-0,4-0,6	0,2-0,4-0,4
92	ethylacetate	0,4-0,4-0,6	0,2-0,4-0,4

	RISK	ODSS	VIOLENCE
93	benzol	0,4-0,4-0,6	0,2-0,4-0,4
94	halogenated hydrocarbons	0,4-0,4-0,6	0,2-0,4-0,4
95	alcohols	0,4-0,4-0,6	0,2-0,4-0,4
96	ethers	0,4-0,4-0,6	0,2-0,4-0,4
97	glycol derivatives	0,4-0,4-0,6	0,2-0,4-0,4
98	esters	0,4-0,4-0,6	0,2-0,4-0,4
99	ketones	0,4-0,4-0,6	0,2-0,4-0,4
100	hexane	0,4-0,4-0,6	0,2-0,4-0,4
101	hexane isomers	0,4-0,4-0,6	0,2-0,4-0,4
102	octane	0,4-0,4-0,6	0,2-0,4-0,4
103	cyclohexane	0,4-0,4-0,6	0,2-0,4-0,4
104	turpentine	0,4-0,4-0,6	0,2-0,4-0,4
105	nitroethane	0,4-0,4-0,6	0,2-0,4-0,4
106	benzene	0,4-0,4-0,6	0,2-0,4-0,4
107	toluene	0,4-0,4-0,6	0,2-0,4-0,4
108	xylene	0,4-0,6-0,6	0,4-0,6-0,6
109	tetrachloromethane	0,4-0,6-0,6	0,4-0,6-0,6
110	trichloroethane	0,4-0,6-0,6	0,4-0,6-0,6
111	trichlorotrifluoroethane	0,4-0,6-0,6	0,4-0,6-0,6
112	methyl alcohol	0,6-0,6-0,6	0,2-0,4-0,4
113	ethyl alcohol	0,6-0,6-0,6	0,2-0,4-0,4
114	propyl alcohol	0,6-0,6-0,6	0,2-0,4-0,4
115	ethylether	0,6-0,6-0,6	0,2-0,4-0,4
116	isopropyl ether	0,6-0,6-0,6	0,2-0,4-0,4
117	ethyl glycol monomethyl ether	0,6-0,6-0,6	0,2-0,4-0,4
118	ether	0,6-0,6-0,6	0,2-0,4-0,4
119	glycols	0,6-0,6-0,6	0,2-0,4-0,4
120	ethyl acetate	0,6-0,6-0,6	0,2-0,4-0,4
121	amyl acetate	0,6-0,6-0,6	0,2-0,4-0,4
122	methyl ethyl ketone	0,6-0,6-0,6	0,2-0,4-0,4
123	acetone	0,6-0,6-0,6	0,2-0,4-0,4
124	irritants	0,6-0,6-0,6	0,2-0,4-0,4
125	aldehydes	0,6-0,6-0,6	0,2-0,4-0,4
126	acetaldehyde	0,6-0,6-0,6	0,2-0,4-0,4
127	acrolein	0,6-0,6-0,6	0,2-0,4-0,4
128	formaldehyde	0,6-0,6-0,6	0,2-0,4-0,4
129	paraform	0,6-0,6-0,6	0,2-0,4-0,4
130	alkaline powder and mistler	0,6-0,6-0,6	0,2-0,4-0,4
131	ammonia	0,6-0,6-0,6	0,2-0,4-0,4
132	chromic acid	0,6-0,6-0,6	0,2-0,4-0,4
133	ethylene oxide	0,6-0,6-0,6	0,2-0,4-0,4
134	hydrogen chloride	0,6-0,6-0,6	0,2-0,4-0,4
135	hydrogen fluoride	0,6-0,6-0,6	0,2-0,4-0,4
136	sulfur dioxide	0,6-0,6-0,6	0,2-0,4-0,4
137	sulfur trioxide	0,6-0,6-0,6	0,2-0,4-0,4
138	bromine	0,6-0,6-0,6	0,2-0,4-0,4
139	chlorine	0,6-0,6-0,6	0,2-0,4-0,4
140	chlorine oxides	0,6-0,6-0,6	0,2-0,4-0,4
141	cyanuric bromine	0,6-0,6-0,6	0,2-0,4-0,4
142	chlorine compounds	0,6-0,6-0,6	0,2-0,4-0,4
143	dimethyl sulphate	0,6-0,6-0,6	0,2-0,4-0,4
144	diethyl sulfate	0,6-0,6-0,6	0,2-0,4-0,4
145	fluorine	0,6-0,6-0,6	0,2-0,4-0,4
146	iodine	0,6-0,6-0,6	0,2-0,4-0,4
147	ozone	0,6-0,6-0,6	0,2-0,4-0,4
148	sulfur chlorides	0,6-0,6-0,6	0,2-0,4-0,4
149	phosphorus trichloride	0,6-0,6-0,6	0,2-0,4-0,4
150	phosphorus pentachloride	0,6-0,6-0,6	0,2-0,4-0,4
151	toluene diisocyanate	0,6-0,6-0,6	0,2-0,4-0,4
152	anoxic anoxia	0,6-0,6-0,6	0,2-0,4-0,4
153	anemic anemia	0,6-0,6-0,6	0,2-0,4-0,4
154	anesthetics and narcotics	0,6-0,6-0,6	0,2-0,4-0,4
155	systemic poisons	0,6-0,6-0,6	0,2-0,4-0,4
156	sensitizers	0,6-0,6-0,6	0,2-0,4-0,4
157	variety of harmful dusts	0,6-0,6-0,6	0,2-0,4-0,4
158	eye and body injuries	0,6-0,6-0,6	0,4-0,6-0,8
159	interaction with hot and uv light environment	0,6-0,6-0,8	0,2-0,2-0,4
160	ionizing radiation,	0,6-0,6-0,8	0,2-0,2-0,4
161	ultraviolet radiation,	0,6-0,6-0,8	0,2-0,2-0,4
162	high pressure	0,6-0,6-0,8	0,2-0,2-0,4
163	chemical substances	0,6-0,6-0,8	0,2-0,2-0,4
164	noise	0,6-0,6-0,8	0,2-0,2-0,4
165	hand-arm vibration	0,6-0,6-0,8	0,2-0,2-0,4
166	whole body vibration	0,6-0,6-0,8	0,2-0,2-0,4
167	lighting	0,6-0,6-0,8	0,2-0,2-0,4
168	harmful rays and radiation	0,6-0,6-0,8	0,2-0,2-0,4
169	electromagnetic fields	0,6-0,6-0,8	0,2-0,2-0,4
170	hot or cold climate	0,6-0,6-0,8	0,2-0,2-0,4
171	irregular and slippery surfaces	0,6-0,6-0,8	0,2-0,2-0,4
172	movement of vehicles and machines	0,6-0,6-0,8	0,2-0,2-0,4
173	machine movements and parts	0,6-0,6-0,8	0,2-0,2-0,4
174	hazardous surfaces	0,6-0,6-0,8	0,2-0,2-0,4
175	hot or cold surfaces	0,6-0,6-0,8	0,2-0,2-0,4
176	hand tools	0,6-0,6-0,8	0,2-0,2-0,4
177	electrical installations and equipment	0,2-0,4-0,4	0,8-0,8-1,0
178	fire	0,2-0,4-0,4	0,8-1,0-1,0
179	explosion	0,2-0,4-0,4	1,0-1,0-1,0
180	lifting and handling	0,4-0,6-0,8	0,2-0,2-0,4
181	stopping faults	0,4-0,6-0,8	0,2-0,2-0,4
182	biological hazards	0,4-0,6-0,8	0,0-0,2-0,2
183	stress, violence	0,4-0,6-0,8	0,0-0,2-0,2

Table 6. Converting risks into probability-severity membership removal linguistic variables using a fuzzy decision-making approach and matrix method in an applied chemical industry

	RISK	ODSS	VIOLENCE
1	workplace ambient temperature	dd,Dd,Od	Nd,dd,Dd
2	the air flow rate	dd,Dd,Od	Nd,dd,Dd
3	relative humidity	dd,Dd,Od	Nd,dd,Dd
4	carbon dioxide	dd,dd,Dd	Od,yd,Yd
5	lead	dd,dd,Dd	Od,yd,Yd
6	mercury	dd,dd,Dd	Od,yd,Yd
7	arsenic	dd,dd,Dd	Od,yd,Yd
8	beryllium	dd,dd,Dd	Od,yd,Yd
9	ammonia	dd,dd,Dd	Od,yd,Yd
10	chlorine	dd,dd,Dd	Od,yd,Yd
11	nitrogen dioxide	dd,dd,Dd	Od,yd,Yd
12	sulphur dioxide	dd,dd,Dd	Od,yd,Yd
13	ozone	dd,dd,Dd	Od,yd,Yd
14	arsine	dd,dd,Dd	Od,yd,Yd
15	phosphine	dd,dd,Dd	Od,yd,Yd
16	stibine	dd,dd,Dd	Od,yd,Yd
17	ethane	dd,dd,Dd	Od,yd,Yd
18	propane and butane	dd,dd,Dd	Od,yd,Yd
19	acetylene	dd,dd,Dd	Od,yd,Yd
20	hydrogen	dd,dd,Dd	Od,yd,Yd
21	nitrogen,	dd,dd,Dd	Od,yd,Yd
22	argon,	dd,dd,Dd	Od,yd,Yd
23	neon,	dd,dd,Dd	Od,yd,Yd
24	helium,	dd,dd,Dd	Od,yd,Yd
25	ethylene	dd,dd,Dd	Od,yd,Yd
26	propylene	dd,dd,Dd	Od,yd,Yd
27	carbon monoxide	dd,dd,Dd	Od,yd,Yd
28	hydrogen cyanide	dd,dd,Dd	Od,yd,Yd
29	hydrogen sulphide	dd,dd,Dd	Od,yd,Yd
30	ammonia	dd,dd,Dd	dd,Dd,Od
31	chlorine	dd,Dd,Dd	dd,Dd,Od
32	sulfur dioxide	dd,Dd,Dd	dd,Dd,Od
33	phosgene	dd,Dd,Dd	dd,Dd,Od
34	nitrogen oxides	dd,Dd,Dd	dd,Dd,Od
35	ozone	dd,Dd,Dd	dd,Dd,Od
36	formaldehyde	dd,Dd,Dd	dd,Dd,Od
37	asnin	dd,Dd,Dd	Dd,Od,yd
38	stibine	dd,Dd,Dd	Dd,Od,yd
39	phosphin	dd,Dd,Dd	Dd,Od,yd
40	nickel carbony	dd,Dd,Dd	Dd,Od,yd
41	carbon sulphide	dd,Dd,Dd	Dd,Od,yd
42	quartz	Dd,Dd,Dd	Dd,Od,Od
43	silica	Dd,Dd,Dd	Dd,Od,Od
44	ridmit	Dd,Dd,Dd	Dd,Od,Od
45	tropolis	Dd,Dd,Dd	Dd,Od,Od
46	amosite	Dd,Dd,Dd	Dd,Od,Od
47	chrysotile	Dd,Dd,Dd	Dd,Od,Od
48	crosidolite	Dd,Dd,Dd	Dd,Od,Od
49	talc	Dd,Dd,Dd	Dd,Od,Od
50	mica	Dd,Dd,Dd	Dd,Od,Od
51	coal dust	Dd,Dd,Dd	Dd,Od,Od
52	lead	Dd,Dd,Dd	Dd,Od,Od
53	chromium	Dd,Dd,Dd	Dd,Od,Od
54	cadmium	Dd,Dd,Dd	Dd,Od,Od
55	vanadium	Dd,Dd,Dd	Dd,Od,Od
56	trinitrotoluene	Dd,Dd,Dd	Dd,Od,Od
57	compounds arsenic	Dd,Dd,Od	dd,Dd,Dd
58	dinitrophenol	Dd,Dd,Od	dd,Dd,Dd
59	cotton	Dd,Dd,Od	dd,Dd,Dd
60	linen	Dd,Dd,Od	dd,Dd,Dd
61	hemp	Dd,Dd,Od	dd,Dd,Dd
62	sisal hemp	Dd,Dd,Od	dd,Dd,Dd
63	jute	Dd,Dd,Od	dd,Dd,Dd
64	platinum compounds	Dd,Dd,Od	dd,Dd,Dd
65	wood dust	Dd,Dd,Od	dd,Dd,Dd
66	Esd	Dd,Dd,Od	dd,Dd,Dd
67	calcium carbonate	Dd,Dd,Od	dd,Dd,Dd
68	cellulose-paper fiber	Dd,Dd,Od	dd,Dd,Dd
69	emery	Dd,Dd,Od	dd,Dd,Dd
70	glycerin misti	Dd,Dd,Od	dd,Dd,Dd
71	gypsum	Dd,Dd,Od	dd,Dd,Dd
72	kaolin	Dd,Dd,Od	dd,Dd,Dd
73	limestone	Dd,Dd,Od	dd,Dd,Dd
74	magnesite	Dd,Dd,Od	dd,Dd,Dd
75	marble	Dd,Dd,Od	dd,Dd,Dd
76	glass wool	Dd,Dd,Od	dd,Dd,Dd
77	pentaerythritol	Dd,Dd,Od	dd,Dd,Dd
78	paris plaster	Dd,Dd,Od	dd,Dd,Dd
79	portland cement	Dd,Dd,Od	dd,Dd,Dd
80	lipstick	Dd,Dd,Od	dd,Dd,Dd
81	silicon	Dd,Dd,Od	dd,Dd,Dd
82	silicon carbide	Dd,Dd,Od	dd,Dd,Dd
83	starch	Dd,Dd,Od	dd,Dd,Dd
84	sakoroz	Dd,Dd,Od	dd,Dd,Dd
85	titanium dioxide	Dd,Dd,Od	dd,Dd,Dd
86	zinc stearate	Dd,Dd,Od	dd,Dd,Dd
87	zinc oxide dust	Dd,Dd,Od	dd,Dd,Dd
88	gasoline	Dd,Dd,Od	dd,Dd,Dd
89	nail polish remover	Dd,Dd,Od	dd,Dd,Dd
90	toluene	Dd,Dd,Od	dd,Dd,Dd
91	trichlorethylene	Dd,Dd,Od	dd,Dd,Dd
92	ethylacetate	Dd,Dd,Od	dd,Dd,Dd

	RISK	ODSS	VIOLENCE
93	benzol	Dd,Dd,Od	dd,Dd,Dd
94	halogenated hydrocarbons	Dd,Dd,Od	dd,Dd,Dd
95	alcohols	Dd,Dd,Od	dd,Dd,Dd
96	ethers	Dd,Dd,Od	dd,Dd,Dd
97	glycol derivatives	Dd,Dd,Od	dd,Dd,Dd
98	esters	Dd,Dd,Od	dd,Dd,Dd
99	ketones	Dd,Dd,Od	dd,Dd,Dd
100	hexane	Dd,Dd,Od	dd,Dd,Dd
101	hexane isomers	Dd,Dd,Od	dd,Dd,Dd
102	octane	Dd,Dd,Od	dd,Dd,Dd
103	cyclohexane	Dd,Dd,Od	dd,Dd,Dd
104	turpentine	Dd,Dd,Od	dd,Dd,Dd
105	nitroethane	Dd,Dd,Od	dd,Dd,Dd
106	benzene	Dd,Dd,Od	dd,Dd,Dd
107	toluene	Dd,Dd,Od	dd,Dd,Dd
108	xylene	Dd,Od,Od	Dd,Od,Od
109	tetrachloromethane	Dd,Od,Od	Dd,Od,Od
110	trichloroethane	Dd,Od,Od	Dd,Od,Od
111	trichlorotrifluoroethane	Dd,Od,Od	Dd,Od,Od
112	methyl alcohol	Od,Od,Od	dd,Dd,Dd
113	ethyl alcohol	Od,Od,Od	dd,Dd,Dd
114	propyl alcohol	Od,Od,Od	dd,Dd,Dd
115	ethylether	Od,Od,Od	dd,Dd,Dd
116	isopropyl ether	Od,Od,Od	dd,Dd,Dd
117	ethyl glycol monomethyl ether	Od,Od,Od	dd,Dd,Dd
118	ether	Od,Od,Od	dd,Dd,Dd
119	glycols	Od,Od,Od	dd,Dd,Dd
120	ethyl acetate	Od,Od,Od	dd,Dd,Dd
121	amyl acetate	Od,Od,Od	dd,Dd,Dd
122	methyl ethyl ketone	Od,Od,Od	dd,Dd,Dd
123	acetone	Od,Od,Od	dd,Dd,Dd
124	irritants	Od,Od,Od	dd,Dd,Dd
125	aldehydes	Od,Od,Od	dd,Dd,Dd
126	acetaldehyde	Od,Od,Od	dd,Dd,Dd
127	acrolein	Od,Od,Od	dd,Dd,Dd
128	formaldehyde	Od,Od,Od	dd,Dd,Dd
129	paraform	Od,Od,Od	dd,Dd,Dd
130	alkaline powder and mistler	Od,Od,Od	dd,Dd,Dd
131	ammonia	Od,Od,Od	dd,Dd,Dd
132	chromic acid	Od,Od,Od	dd,Dd,Dd
133	ethylene oxide	Od,Od,Od	dd,Dd,Dd
134	hydrogen chloride	Od,Od,Od	dd,Dd,Dd
135	hydrogen fluoride	Od,Od,Od	dd,Dd,Dd
136	sulfur dioxide	Od,Od,Od	dd,Dd,Dd
137	sulfur trioxide	Od,Od,Od	dd,Dd,Dd
138	bromine	Od,Od,Od	dd,Dd,Dd
139	chlorine	Od,Od,Od	dd,Dd,Dd
140	chlorine oxides	Od,Od,Od	dd,Dd,Dd
141	cyanuric bromine	Od,Od,Od	dd,Dd,Dd
142	chlorine compounds	Od,Od,Od	dd,Dd,Dd
143	dimethyl sulphate	Od,Od,Od	dd,Dd,Dd
144	diethyl sulfate	Od,Od,Od	dd,Dd,Dd
145	fluorine	Od,Od,Od	dd,Dd,Dd
146	iodine	Od,Od,Od	dd,Dd,Dd
147	ozone	Od,Od,Od	dd,Dd,Dd
148	sulfur chlorides	Od,Od,Od	dd,Dd,Dd
149	phosphorus trichloride	Od,Od,Od	dd,Dd,Dd
150	phosphorus pentachloride	Od,Od,Od	dd,Dd,Dd
151	toluene diisocyanate	Od,Od,Od	dd,Dd,Dd
152	anoxic anoxia	Od,Od,Od	dd,Dd,Dd
153	anemic anemia	Od,Od,Od	dd,Dd,Dd
154	anesthetics and narcotics	Od,Od,Od	dd,Dd,Dd
155	systemic poisons	Od,Od,Od	dd,Dd,Dd
156	sensitizers	Od,Od,Od	dd,Dd,Dd
157	variety of harmful dusts	Od,Od,Od	dd,Dd,Dd
158	eye and body injuries	Od,Od,Od	Dd,Od,yd
159	interaction with hot and uv light environment	Od,Od,yd	dd,dd,Dd
160	ionizing radiation,	Od,Od,yd	dd,dd,Dd
161	ultraviolet radiation,	Od,Od,yd	dd,dd,Dd
162	high pressure	Od,Od,yd	dd,dd,Dd
163	chemical substances	Od,Od,yd	dd,dd,Dd
164	noise	Od,Od,yd	dd,dd,Dd
165	hand-arm vibration	Od,Od,yd	dd,dd,Dd
166	whole body vibration	Od,Od,yd	dd,dd,Dd
167	lighting	Od,Od,yd	dd,dd,Dd
168	harmful rays and radiation	Od,Od,yd	dd,dd,Dd
169	electromagnetic fields	Od,Od,yd	dd,dd,Dd
170	hot or cold climate	Od,Od,yd	dd,dd,Dd
171	irregular and slippery surfaces	Od,Od,yd	dd,dd,Dd
172	movement of vehicles and machines	Od,Od,yd	dd,dd,Dd
173	machine movements and parts	Od,Od,yd	dd,dd,Dd
174	hazardous surfaces	Od,Od,yd	dd,dd,Dd
175	hot or cold surfaces	Od,Od,yd	dd,dd,Dd
176	hand tools	Od,Od,yd	dd,dd,Dd
177	electrical installations and equipment	dd,Dd,Dd	yd,yd,Yd
178	fire	dd,Dd,Dd	yd,Yd,Yd
179	explosion	dd,Dd,Dd	yd,Yd,Yd
180	lifting and handling	Dd,Od,yd	dd,dd,Dd
181	stopping faults	Dd,Od,yd	dd,dd,Dd
182	biological hazards	Dd,Od,yd	Nd,dd,dd
183	stress, violence	Dd,Od,yd	Nd,dd,dd

Table 7. Obtaining risk membership ratings using a fuzzy decision-making approach and matrix method in an applied chemical industry.

	RISK	RISK MEMBERSHIP LEVEL
1	workplace ambient temperature	0,08
2	the air flow rate	0,08
3	relative humidity	0,08
4	carbon dioxide	0,208
5	lead	0,208
6	mercury	0,208
7	arsenic	0,208
8	beryllium	0,208
9	ammonia	0,208
10	chlorine	0,208
11	nitrogen dioxide	0,208
12	sulphur dioxide	0,208
13	ozone	0,208
14	arsine	0,208
15	phosphine	0,208
16	stibine	0,208
17	ethane	0,208
18	propane and butane	0,208
19	acetylene	0,208
20	hydrogen	0,208
21	nitrogen,	0,208
22	argon,	0,208
23	neon,	0,208
24	helium,	0,208
25	ethylene	0,208
26	propylene	0,208
27	carbon monoxide	0,208
28	hydrogen cyanide	0,208
29	hydrogen sulphide	0,208
30	ammonia	0,132
31	chlorine	0,132
32	sulfur dioxide	0,132
33	phosgene	0,132
34	nitrogen oxides	0,132
35	ozone	0,132
36	formaldehyde	0,132
37	asrin	0,198
38	stibine	0,198
39	phosphin	0,198
40	nickel carbony	0,198
41	carbon sulphide	0,198
42	quartz	0,212
43	silica	0,212
44	ridmit	0,212
45	tropolis	0,212
46	amosite	0,212
47	chrysotile	0,212
48	crosidolite	0,212
49	talc	0,212
50	mica	0,212
51	coal dust	0,212
52	lead	0,212
53	chromium	0,212
54	cadmium	0,212
55	vanadium	0,212
56	trinitrotoluene	0,212
57	compounds arsenic	0,1518
58	dinitrophenol	0,1518
59	cotton	0,1518
60	linen	0,1518
61	hemp	0,1518
62	sisal hemp	0,1518
63	jute	0,1518
64	platinum compounds	0,1518
65	wood dust	0,1518
66	Esd	0,1518
67	calcium carbonate	0,1518
68	cellulose-paper fiber	0,1518
69	emery	0,1518
70	glycerin misti	0,1518
71	gypsum	0,1518
72	kaolin	0,1518
73	limestone	0,1518
74	magnesite	0,1518
75	marble	0,1518
76	glass wool	0,1518
77	pentaerythritol	0,1518
78	paris plaster	0,1518
79	portland cement	0,1518
80	lipstick	0,1518
81	silicon	0,1518
82	silicon carbide	0,1518
83	starch	0,1518
84	sakoroz	0,1518
85	titanium dioxide	0,1518
86	zinc stearate	0,1518
87	zinc oxide dust	0,1518
88	gasoline	0,1518
89	nail polish remover	0,1518
90	toluene	0,1518
91	trichlorethylene	0,1518
92	ethylacetate	0,1518

	RISK	RISK MEMBERSHIP LEVEL
93	benzol	0,1518
94	halogenated hydrocarbons	0,1518
95	alcohols	0,1518
96	ethers	0,1518
97	glycol derivatives	0,1518
98	esters	0,1518
99	ketones	0,1518
100	hexane	0,1518
101	hexane isomers	0,1518
102	octane	0,1518
103	cyclohexane	0,1518
104	turpentine	0,1518
105	nitroethane	0,1518
106	benzene	0,1518
107	toluene	0,1518
108	xylene	0,2809
109	tetrachloromethane	0,2809
110	trichloroethane	0,2809
111	trichlorotrifluoroethane	0,2809
112	methyl alcohol	0,198
113	ethyl alcohol	0,198
114	propyl alcohol	0,198
115	ethylether	0,198
116	isopropyl ether	0,198
117	ethyl glycol monomethyl ether	0,198
118	ether	0,198
119	glycols	0,198
120	ethyl acetate	0,198
121	amyl acetate	0,198
122	methyl ethyl ketone	0,198
123	acetone	0,198
124	irritants	0,198
125	aldehydes	0,198
126	acetaldehyde	0,198
127	acrolein	0,198
128	formaldehyde	0,198
129	paraform	0,198
130	alkaline powder and mistler	0,198
131	ammonia	0,198
132	chromic acid	0,198
133	ethylene oxide	0,198
134	hydrogen chloride	0,198
135	hydrogen fluoride	0,198
136	sulfur dioxide	0,198
137	sulfur trioxide	0,198
138	bromine	0,198
139	chlorine	0,198
140	chlorine oxides	0,198
141	cyanuric bromine	0,198
142	chlorine compounds	0,198
143	dimethyl sulphate	0,198
144	diethyl sulfate	0,198
145	fluorine	0,198
146	iodine	0,198
147	ozone	0,198
148	sulfur chlorides	0,198
149	phosphorus trichloride	0,198
150	phosphorus pentachloride	0,198
151	toluene diisocyanate	0,198
152	anoxic anoxia	0,198
153	anemic anemia	0,198
154	anesthetics and narcotics	0,198
155	systemic poisons	0,198
156	sensitizers	0,198
157	variety of harmful dusts	0,198
158	eye and body injuries	0,36
159	interaction with hot and uv light environment	0,1716
160	ionizing radiation,	0,1716
161	ultraviolet radiation,	0,1716
162	high pressure	0,1716
163	chemical substances	0,1716
164	noise	0,1716
165	hand-arm vibration	0,1716
166	whole body vibration	0,1716
167	lighting	0,1716
168	harmful rays and radiation	0,1716
169	electromagnetic fields	0,1716
170	hot or cold climate	0,1716
171	irregular and slippery surfaces	0,1716
172	movement of vehicles and machines	0,1716
173	machine movements and parts	0,1716
174	hazardous surfaces	0,1716
175	hot or cold surfaces	0,1716
176	hand tools	0,1716
177	electrical installations and equipment	0,2838
178	fire	0,3069
179	explosion	0,33
180	lifting and handling	0,156
181	stopping faults	0,156
182	biological hazards	0,078
183	stress, violence	0,078

Figure 3: Creating software on a risk assessment analysis in chemical industry.

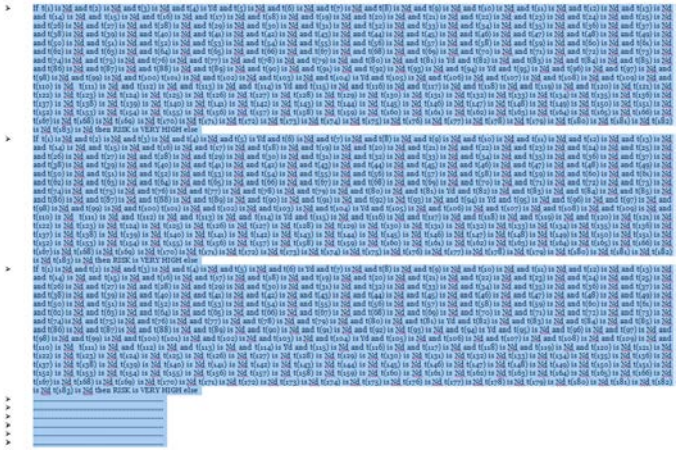
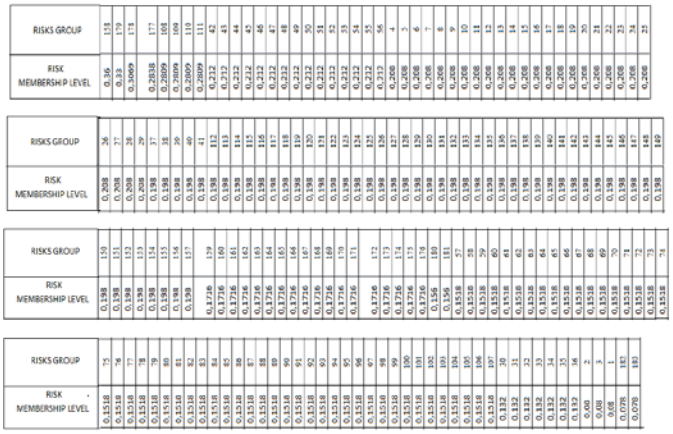


Table 8. Fuzzy decision-making approach in an applied chemical industry and sequence of risks using the matrix method.



I. CONCLUSIONS

The location of the research has been 9 chemical-working factories in the provinces of Ankara, Adana and Mersin. The dangers in the chemical industry are defined as above given by making researches and examinations in 9 different factories in Ankara, Adana and Mersin. When the reasons for these accidents were asked, the majority of the workers seemed to be involved in the chemical business accidents due to the lack of employees, but the employers did not explain the underlying reasons for the absurdity. In this study, software was developed by creating a fuzzy logic based risk assessment analysis model in consideration of many hazards in occupational health and safety in chemical industry. An alternative approach to risk assessment has been proposed using fuzzy decision making approach and matrix method. With this approach, occupational health and safety specialists are provided with blurred linguistic assessments before calculations are made. Thermal comfort (temperature, humidity, air flow rate) and clean ventilation in the chemical industry environment, particulate matter-heavy

metal in the ambient air, internal noise, vibration, lighting, gasses in the environment, a centralized control system was developed. It has been observed that the risks are reduced in these 9 chemical working factories. The industrial hygiene control method depends on many factors such as the shape, the structure of the harmful substance or agent, the entry route to the body, the exposure time, concentration in the atmosphere of the working environment. In the establishment phase of the workplace, hazard prevention and engineering controls are carried out by methods such as appropriate projecting or replacement of the used hazardous materials, application of separation or ventilation methods, taking into consideration the chemical factors. Administrative controls should be made where the exposure to workers is controlled by reducing the working hours of persons working in hazardous chemical environments and by applying other working rules. Where engineering technical controls are not sufficient to reduce the exposure limit to an acceptable level, the last resort is to use personal protective equipment. Personal protective equipment should be used in conjunction with engineering and technical controls and other methods. In order to protect workers from harmful chemical agents or minimize exposure to harmful substances, protective measures must be taken first in the source of the agent, then in the environment, and as a last resort, if necessary to take precautions in the receiver.

REFERENCES

[1] Gens, A., Hurley, J. F., Tuomisto, J. T. and Friedrich, R., Health impacts due to personal exposure to fine particles caused by insulation of residential buildings in Europe. Atmospheric Environment, 2014, 84, 213-221.

[2] Van Orden, D. R., Lee, R. J., Allison, K. A. and Addison, J., Width Distributions of Asbestos and Non-Asbestos Amphibole Minerals. Indoor and Built Environment, 2009, 18 (6), 531-540.

[3] Youn, Y., Comparison of Asbestos Exposure and Risk Assessment by Asbestos Mining Type in Korea. Ronald, F. and Samuel, P., Asbestos: Risk Assessment, Epidemiology, and Health Effects International journal of toxicology, 2011, 25, 139-141.

[4] Mannan, S., Lees' Loss Prevention İn The Process Industries: Hazard Identification, Assessment And Control, 4th Ed. Butterworth-Heinemann, 2012.

[5] Jones, R., Lehr, W., Simecek - Beatty, D., Areal Locatıons Of Hazardous Atmospheres, Technıcal Documentatıon, Noaa Technical Memorandum Nos Or&R, 2013, Pp 43, 96.

[6] Occupational Injuries and Illnesses in the United States by Industry; United States Bureau of Labor Statistics: Washington, DC, USA, 2012.

[7] Van Den Bosch, C.J.H., Weterings, R.A.P.M., Methods For The Calculation Of Physical Effects Cpr14e,Yellow Book, Comimittee For The Prevention Of Disasters, 3rd Ed., Nl, 2005.

[8] Oginska-Bulik, N. Occupational Stress And Its Consequences In Healthcare Professionals: The Role Of Type D Personality. Int J Occup Med Environ Health 2006;19:113–22.

[9] Amyotte, P. Dust Explosions: Understanding the Myths and Realities of Dust Explosions for a Safer Workplace; Butterworth-Heinemann: Oxford, UK, 2013.

[10] Rudel, RA; Perovich, LJ. Endocrine disrupting chemicals in indoor and outdoor air. Atmos. Environ 2009, 43, 170–181.

[11] WHO Guidelines for Indoor Air Quality: Dampness and Mould; World Health Organization Regional Office for Europe: Copenhagen, Denmark, 2009.

[12] Reducing Risks, Promoting Healthy Life; World Health Organization: Geneva, Switzerland, 2002.

[13] C.C. Lee, Fuzzy Logic In Control Systems: Fuzzy Logic Controller, IEEE Transactions, 1990, Vol20, No. 2.

[14] L. A. Zadeh, Fuzzy Logic Computing With Words, Pieee Transactions On Fuzzy Systems, 1996, Vol.4, No.2.

[15] K. M. Passino, S. Yurkovich, Fuzzy Control, 1998, Addison Wesley.

[16] M. Du, T. Fan, W. Su, H. Li, Design Of A New Practical Expert Fuzzy Controller In Central Air Conditioning Control System, Ieee Pacific-Asia Workshop On Computational Intelligence And Industrial Application, 2008.

[17] M. S. I. Md., S. Z. Sarker, K. A. A. Rafi, M. Othman, Development Of A Fuzzy Logic Controller Algorithm For Air Conditioning System, ICSE Proceedings, 2006.

[18] Gerla G., Fuzzy Logic: Mathematical Tools for Approximate Reasoning Kluwer Academic Publishers Dordrecht, 2001.

[19] Abuswer, M.; Amyotte, P.; Khan, F.; Morrison, L, An optimal level of dust explosion risk management : Framework and application. Journal of Loss Prevention in the Process Industries 2013, 26, 1530–1541.

[20] M. Zile, Analysis Of Job Accidents Transferred To The Judiciary And Creation Project Of Causality Modeling, University Of Cukurova, 2013, Publication No: E / 2013/600, Isbn: 978-605-01-0469-1, Adana.

[21] M. Zile, Forensic Cases Based On Maintenance And Repair Projects For Causing Approaches In Work Accidents And Measures To Be Taken, University Of Cukurova, 2015, No: E/2015/635, Isbn: 978-605-01-0710-4, 626, Adana.

Occupational Health And Safety In The Metal Industry; Modeling and Software Developed On The Basis Of Fuzzy Logic To Prevent Job Accidents

M.ZİLE[1]

[1] Mersin University, Mersin/Turkey, mehmetzile@yahoo.com

Abstract **- Methods: In the provinces of Adana and Mersin, 10 metal working job security survey studies and work safety experts were carried out in this study. Computer software is developed by creating a fuzzy logic based risk assessment analysis model considering various hazards in occupational health and safety in the Metal Industry. Results: It is very difficult to model risk in a workplace environment with many specific and ambiguous hazards and model a system to simulate these hazards. Software was developed by creating a fuzzy logic based risk assessment analysis model in consideration of many hazards in occupational health and safety in the Metal Industry. An alternative approach to risk assessment has been proposed using fuzzy decision making approach and matrix method. With this approach, occupational health and safety in the Metal Industry specialists are provided with blurred linguistic assessments before calculations are made, and the inconsistencies in decision making are reduced by taking arithmetic averages of these assessments. It has been seen that the three most important risks in the metal work done by creating the fuzzy logic based modeling of work safety risk analysis model and software using the fuzzy decision making approach and the matrix method to increase the job security in the metal industry. Conclusion: It is very difficult to assess risk in the metal industry which are many specific and indeterminate hazards, and to model a system to simulate these hazards. In this study, a computer software and hardware was developed by creating a fuzzy logic based risk assessment analysis model considering many hazards in occupational health and safety in the metal industry.**

Keywords **– Fuzzy logic, Safe and Health, Job Accidents, Metal Industry**

I. INTRODUCTION

The metal processing industry in Turkey, is located in a prominent position in the world with high quality products . In today's world where technology has developed rapidly, the sector continues its investments in order to keep up with the change [1]. Turkey, the metal processing industry, is located in a prominent position in the world with high quality products. In today's world where technology has developed rapidly, the sector continues its investments in order to keep up with the change [2]. The metalworking sector, which attaches great importance to R&D work, continues its search for new markets without interruption. Due to high technology engineering and R&D power, machine manufacturing industry has taken an important place in the world [3]. In particular, a rapid growth sector captures momentum in recent years, rising star in the position of Turkey's economy and exports[4]. Turkey average 15 percent increase in the machinery sector exports in the last five years, the rise in exports ranks third in the world [5]. It is noteworthy that the sector carries out a significant part of its exports to EU countries. Delivering considerable quantities of products to the EU market, where quality and technology are of great importance, constitute one of the most important evidences of the qualifications possessed by the machine manufacturing industry [6]. It requires knowledge, experience and expertise in the area of occupational health and safety due to the risks involved in the metal industry structure [7]. With the fulfillment of the occupational health and safety processes, it is important to take precautions to prevent occupational accidents and occupational diseases [8]. In the metal sector, mostly working with machines and benches, a wide range of cutting, drilling and rigging tools are used in the production process [9]. In the area of occupational health and safety in the sector, mechanical hazards (from machinery and equipment), physical hazards (noise, inadequate ventilation, insufficient or excessive lighting), dangerous methods and hazards arising from the process are the foregrounds [10]. Businesses operating in the metal sector also have business-specific hazards other than common hazards and are specially made to operate their detection, analysis and measures to be taken [11]. Because of the different production methods and forms used in the workplaces where the same product is produced in the sector, risks to be met due to structural differences with machinery and equipment may also vary [12].

II. DANGEROUS CAUSES OF BUSINESS INJURIES AND CAUSES OF OCCUPATIONAL DISEASES IN METAL INDUSTRY

In the provinces of Adana and Mersin, 10 metal working job security survey studies and work safety experts were carried out. Those working in the metal industry may use copper, lead, arsenic, antimony, zinc, sulfur dioxide, sulfuric acid, carbon monoxide, cadmium containing particles, bauxite, silica,

caustic soda, aluminum oxide, cryolite and hydrofluoric acid fumes, fluorite and aluminum fluoride powders during the melting of metals, mercury, cyanide powders, hydrogen cyanide, arsenic gas, noise, hot stress, electrical hazards. Chemical hazards that are detected by employees are exposed to a variety of harmful dusts, gases, fumes and other chemicals; silica, toxic metals are sulfur dioxide, carbon monoxide, sulfuric acid, nickel carbonyl, fluorides, arsenic, mercury and cyanide, while other hazards are glare and infrared radiation, noise, electrical hazards, back and upper limb damage caused by furnaces and molten metal. Formaldehyde, nitrosamines, microbiological contaminants, laceration of hands, musculoskeletal injuries, hot stress, molds, molds, molds, waxes, alkanolamines, petroleum sulfonates, borates, cellulose derivative thickeners, corrosion inhibitors and biocides petroleum products, sodium nitrate, sodium nitrite, sodium carbonate, sodium silicate, silicone oils, and biocides have been identified. It has been determined that the hazards encountered by the workers in the rolling mills are injuries, burns, eye injuries, heat stress, noise, vibration, metalworking fluids, surface cleaners, acids, ionizing radiation, harmful gases and steam. The hazards of workers exposed to welding and hot cutting processes include fire, burns, inhalation of radiant heat metal fumes and other contaminants, electrical hazards, noise, ultraviolet radiation, ozone, nitrogen dioxide, carbon monoxide, fluorides, pressurized gas cylinders, explosions, physical strain, gases, fumes, irritants, disintegration products, sources such as fatigue, musculoskeletal injuries, hot stress, ultraviolet radiation, metal fillings or various components of electrode steel such as nickel or chromium, welding wires, paints, greases, metal surface fouling, carbon monoxide, carbon dioxide It has been determined that the shells are in contact with hydrocarbons, interaction with hot and uv light environment, inert gas usage as protective, metal smoke fire, noise, ionizing radiation. It has been determined that workers in the lathe are exposed to hazards such as noise, vibration, accidents and particularly hand injuries, eye and body injuries due to splashes of metal shavings. Workers in sanding and polishing have been found to be affected by exposure to the lungs, depending on the surface to be sanded in the case of dangerous eye damage, vibration, intensive use without respiratory protection. Workers in electroplating have been found to be exposed to chromium and nickel exposure risks of cancer, cyanides, caustic and corrosive chemicals-related burns and irritation, electric shock, wet and dry sliding and falling hazards, dust explosions, ergonomic hazards. It has been determined that physical hazards, burn hazards, dust exposure hazards resulting from hazards grinders and carriers exposed to glass and enamel workers are dangerous. Workers in engraving processes are exposed to hydrofluoric acid exposure, caustic and corrosive chemicals, burns and irritation, and danger of burns. It has been determined that workers exposed to galvanization are at risk of burns, burns and irritation due to caustic and corrosive chemistry, metal smoke fire, lead exposure. It has been determined that heat workers are exposed to hazardous hazards, burns and irritation due to caustic and corrosive chemicals, hydrogen explosion risk, carbon monoxide exposure, cyanide exposure, fire risk. The risk of exposure to metalworking workers is found to be burn hazard, risk of dust explosion, acetylene, zinc metal fume. It has been determined that workers exposed to phosphatizing agents are exposed to caustic and corrosive chemicals and burns and irritation. workers exposed to linings have been found to be toxic, flammable, sensitizing, carcinogenic (chromium) exposure to various chemicals.

III. FUZZY LOGIC CONCEPT

Fuzzy logic is based on fuzzy set and sub set. In a classical approach, an entity is a member of the coop or is not. When expressed mathematically, "1" is the value when the element is a member of the conglomerate, and "0" when the element is not a member of the conglomerate [13]. Fuzzy logic is the extension of classical cluster representation. Each entity in the fuzzy entity set has a membership level. Contrary to classical clusters, the membership grades of fuzzy clusters can change in infinite number of intervals (0,1). These are a whole bunch of grades of membership that are continuous and unbroken [14]. Binary variables such as cold-warm, fast-slow, light-dark in sharp clusters are likened to the real world by being softened by flexible qualifiers such as a little cold, a little warm, a little darkness in the fuzzy logic. Once fuzzy variables are defined and membership functions are assigned to them. There are some rules that are used in defining fuzzy clusters [15]. In principle, the number of fuzzy sets assigned to each variable is usually a single number. This provides the presence of a centre point to prevent numerical oscillation between adjacent values. Second, the number of fuzzy sets is usually between 3 and 9. To describe causal relations, we must be able to distinguish one subset from the other by the use of linguistic variables [16]. The greater the number of sub-clusters, the more difficult it becomes. It is easy to distinguish between short, medium and long variable values. But with a lot of data, this situation becomes more difficult. The linguistic descriptions of the subclasses can be interpreted. At the same time, each fuzzy set must sum up the compound sets [17]. This overlay provides a continuous control area for the fuzzy controller. Once the blur sets have been defined and assigned their membership functions, the rules must be written for each combination of the control variable. These rules will relate input variables to output variables using 'If-Then' expressions in decision-making. The condition 'If' is a prelude to the result of each rule. In general, each rule is shown in 'If' (prefix) 'Then' (result) style [18].

IV. RISK ASSESSMENT ANALYSIS IN METAL INDUSTRY

t: hazard (up to 1 hazard number T),

d: danger value,

dtmax: max value of danger,

Nd: non-hazardous value (membership level $u_{violence}=0$),
dd: very low dangerous value (membership level $u_{violence}=0.2$),
Dd: low dangerous value (membership level $u_{violence}=0.4$),
Od: moderate danger value (membership level $u_{violence}=0.6$),
yd: high dangerous value (membership level $u_{violence}=0.8$),
Yd: very high dangerous value (membership level $u_{violence}=1.0$)

t(1): workplace ambient temperature (when it drops below 15 ^{0}C and exceeds 25 ^{0}C)
d_1^{min} =10^{0}C and d_1^{max} = 35^{0}C,
d_2^{min} = 0,4 m/sec and d_2^{max} = 1,1 m/sec,
t(2): the air flow rate (when the speed drops below 0,5 m/sec and rises above 1 m/sec)
t(3): relative humidity (over 60%), d_3^{max} = 60%

t(4): carbon monoxide, d4max = 50 ppm,
t(5): carbon dioxide, d4max = 1000 ppm,
t(6): hydrogen cyanide, d5max = 10 ppm,
t (7) : lead, d_7^{max} = 0,15 mg/m^3 ,
t (8) : mercury, d_8^{max} = 0,075 mg/m^3 ,
t (9) : arsenic, d_9^{max} = 0,5 mg/m^3 ,
t (10) : hydrogen sulphide, d_{10}^{max} = 20 ppm ,
t (11) : beryllium, d_{11}^{max} = 2 mg/m^3
t (12) : ammonia, d_{12}^{max} = 25 ppm,
t (13) : chlorine, d_{13}^{max} = 1 ppm,
t (14) : nitrogen dioxide, d_{14}^{max} = 5 ppm,
t (15) : sulphur dioxide, d_{15}^{max} = 0,1 ppm,
t (16) : ozone, d_{16}^{max} = 0,1 ppm
t (17) : arsine, d_{17}^{max} = 0,05ppm,
t (18) : phosphine, d_{18}^{max} = 0,3 ppm,
t (19) : stibine, d_{19}^{max} = 0,1 ppm,
t (20) :alkanolamines
t (21) :aluminum oxide
t (22) :biocides petroleum
t (23) :antimony
t (24) :bauxite
t (25) :borates
t (26) :cadmium containing particles
t (27) :caustic soda
t (28) :cellulose derivative thickeners
t (29) :copper
t (30) :corrosion inhibitors
t (31) :cryolite
t (32) :cyanide powders
t (33) :fluorite and aluminum fluoride powders
t (34) :hydrocarbons
t (35) :hydrofluoric acid fumes,
t (36) :fluorides
t (37) :greases
t (38) :Formaldehyde
t (39) :fumes and other chemicals
t (40) :disintegration products,
t (41) :irritants,
t (42) :laceration of hands,
t (43) :lead,
t (44) :metal smoke fire,
t (45) :metal surface fouling,
t (46) :metalworking fluids,
t (47) :microbiological contaminants,
t (48) :molds,
t (49) :musculoskeletal injuries,
t (50) :musculoskeletal injuries,
t (51) :nickel carbonyl,
t (52) :nitrogen dioxide,
t (53) :nitrosamines,
t (54) :petroleum sulfonates,
t (55) :silica,
t (56) :sodium carbonate,
t (57) :sodium nitrate,
t (58) :sodium nitrite,
t (59) :sodium silicate,
t (60) :zinc,
t (61) :pressurized gas cylinders,
t (62) :physical strain
t (63) :waxes,
t (64) :welding wires,
t (65) :variety of harmful dusts,
t (66) :eye and body injuries
t (67) :interaction with hot and uv light environment,
t (68) :ionizing radiation,
t (69) :ultraviolet radiation,
t (70) : high pressure
t (71) : chemical substances
t (72) : noise
t (73) : hand-arm vibration
t (74) : whole body vibration
t (75) : lighting
t (76) : harmful rays and radiation,
t (77) : electromagnetic fields,
t (78) : hot or cold climate,
t (79) : irregular and slippery surfaces,
t (80) : movement of vehicles and machines,
t (81) : machine movements and parts,
t (82) : hazardous surfaces,
t (83) : hot or cold surfaces,
t (84) : hand tools,
t (85) : electrical installations and equipment,
t (86) : fire,
t (87) : explosion,
t (88) : lifting and handling,
t (89) : stopping faults,
t (90) : biological hazards,
t (91) : stress, violence

The dangers in the metal industry are defined as above given by making researches and examinations in 10 different factories in Adana and Mersin. The risk assessment analysis in the metal industry is modeled as shown in Figure 1. The hazards created by insecure environments and the unsafe motions of human origin are entered into the hazard identification block as data. These hazards are classified as hazardous and injured work accidents, occupational diseases and psychological disorders in the hazard identification block. The data from here was entered into the scenery development

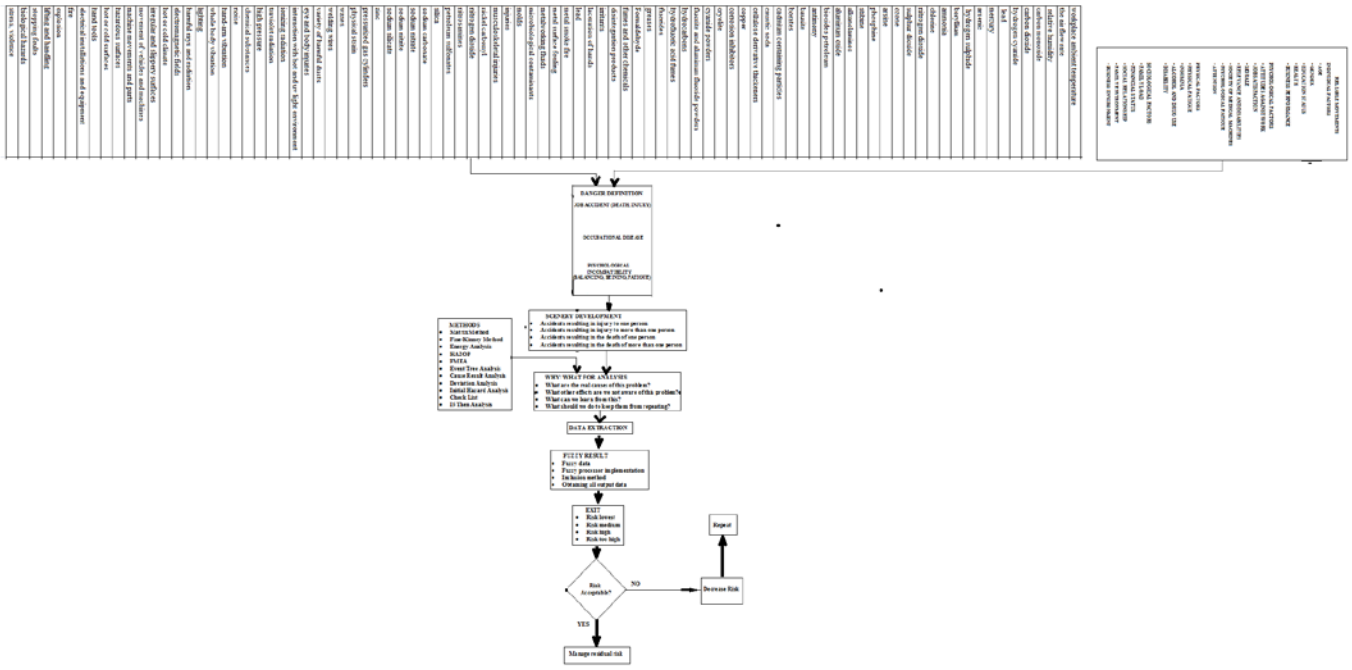


Figure 1: Creating risk analysis modelling in metal industry.

block and the result grouping was done. The reasons for the receipt are determined by the cause analysis block. We have received a risk decomposition. In order to reduce the unacceptable risks, repetition cycle has been done and according to this cycle energy, ventilation, pressure, humidity, temperature, etc. the system is automatically adjusted to reduce the risks of operation and at the same time warning alerts were sent to the business manager automatically. The probability of emergence of danger is given in Table 1.

Table 1. Probability chart of emergence of hazard.

PROBABLE VALUE	POSSIBILITY OF LEAVING
Very small (1)	Hardly ever
Small (2)	Very few (once a year), only in abnormal situations,
Middle (3)	Little (once every three months)
High (4)	Frequently (once a week)
Very high (5)	Very often (every day), under normal working conditions

Id=5 'emergence of danger in the day'(membership level $u_{probability}$ =1.0)
Id=4 'dangerous emergence weekly' (membership level $u_{probability}$ =0.8)
Id=3 'emergence of danger in the moon' (membership level $u_{probability}$ =0.6)
Id=2 'emergence of danger in three months' (membership level $u_{probability}$ =0.4)
Id=1 'the emergence of danger in the year' (membership level $u_{probability}$ =0.2)
Id=0 'never appear dangerous' (membership level $u_{probability}$ =0.0)
are determined. In Table 2, the severity rating is determined.

Table 2. Violence rating table.

PROBABLE VALUE	POSSIBILITY OF LEAVING
Very small (1)	Hardly ever
Small (2)	Very few (once a year), only in abnormal situations,
Middle (3)	Little (once every three months)
High (4)	Frequently (once a week)
Very high (5)	Very often (every day), under normal working conditions

'Death, incapacity to work permanently' (membership level $u_{violence}$=1.0)
'Serious injury, long-term treatment, occupational disease' (membership level $u_{violence}$=0.8)
'Light injury, inpatient treatment / injury' (membership level $u_{violence}$=0.6)
'There is no loss of work day, outpatient treatment without permanent effect' (membership level $u_{violence}$ = 0.4)
'No business hours lost, need immediate relief, first aid' (membership level $u_{violence}$ = 0.2)
'The danger does not emerge at all' (membership level $u_{violence}$ = 0.0)
are determined.

PROBABILITY / VIOLENCE	1 VERY LOW	2 LOW	3 MIDDLE	4 HIGH	5 VERY HIGH
1 VERY LOW	NEGLIGIBLE 1	NEGLIGIBLE 2	LOW 3	LOW 4	LOW 5
2 LOW	NEGLIGIBLE 2	LOW 4	LOW 6	MIDDLE 8	MIDDLE 10
3 MIDDLE	LOW 3	LOW 6	MIDDLE 9	MIDDLE 12	HIGH 15
4 HIGH	LOW 4	MIDDLE 8	MIDDLE 12	HIGH 16	HIGH 20
5 VERY HIGH	LOW 5	MIDDLE 10	HIGH 15	HIGH 20	EXTREMELY HIGH 25

Figure 2: Score matrix.

As shown in Fig. 2, the score matrix is determined. In the linguistic variable membership function, the score table in

Table 3 are determined. Risk assessment score level are determined. Risk = Violence x Probability [19]. The risks with a score of 25 are very high and the membership rate is $u_{RİSK}$= 1,0. Risks between 15 and 25 are high risk and membership level is $u_{RİSK}$ = 0,8. Risks between 8 and 12 are medium risk and membership level is $u_{RİSK}$ = 0,6. The risks ranging from 3 to 6 are low, and the membership grade is $u_{RİSK}$ = 0,4. Risks with a score of 1 and 2 are very low, and the membership level is $u_{RİSK}$ = 0,2. If the score is 0, there is no risk and the membership grade is $u_{RİSK}$ = 0,0 [20].

Table 3. Score table.

NON-TOLLABLE RISKS (25)	The business should not be started until the risk is reduced to an acceptable level, and if there is an ongoing activity, it should be stopped immediately. If it is still not possible to reduce the risk despite the precautionary measures, the activity should be canceled.
IMPORTANT RISKS (15,16,20)	The business should not be started until the specified risk is reduced. If there is an ongoing activity, it should be stopped immediately. Emergency measures should be taken if the risk is to continue, and as a result of these measures, the continuation of the activity should be decided.
MEDIUM LEVEL RISKS (8,9,10,12)	Activities should be initiated to reduce the identified risks.
INCURED RISKS (2,3,4,5,6)	Additional control processes may not be needed to remove the identified risks. However, existing controls should be maintained.
INCREDIBLE RISKS (1)	It may not be necessary to plan the control processes and to keep records of the activities to be carried out in order to eliminate the identified risks.

Firstly, perunit values of the defined hazards are calculated as $d_t^n = d_t / d_t^{max}$, t=1, ……T', then perunit values of the undefined hazards are calculated as $\underline{d}_t^n = \underline{d}_t / \underline{d}_t^{max}$, t=T'+1,…..T, a fuzzy number scale is defined by the maximum method, $\underline{d}_t^n, r_t$, t=T'+1,……,T, the fuzzy definition of each defined hazard being $\underline{c}_t = \underline{W}_t . \underline{d}_t^n$ is calculated and all linguistic hazards are $c_t = \underline{W}_t . r_t$ for each t=T'+1,…..T is calculated on the basis of the comparison method of the blurred numbers in the order of the hazards [21].

Table 4. Linguistic variable table.

LANGUAGE VARIABLES	FUZZY VALUE
Nd: non-hazardous value	(0,0,0)
dd: verylow dangerous value	(1,1,2)
Dd: low dangerous value	(1,2,3)
Od: moderate danger value	(2,3,4)
yd: high dangerous value	(3,4,5)
Yd: very high dangerous value	(4,5,5)

Table 4 shows the linguistic variable table, the fuzzy decision-making approach in Table 5, and the probability-risk assessments determined by three different job security experts using the matrix method and the risk membership ratings of these assessments in Table 6, conversion of the risks obtained in Table 7 into linguistic variables of probability-severity membership ratios is given in Table 8. Once the blur sets are defined and their membership functions are assigned, the rules are written for each combination of control variables. These rules relate input variables to output variables by using If-Then expressions in decision-making. The condition 'If' is a prelude to the result of each rule. In general, each rule is shown as 'If', and then 8281 combined rules are created when identical result expressions are issued. The software was generated as shown in Figure 3 and the priority order of the risks was obtained as in Table 9.

Table 5. Fuzzy decision making approach in an applied metal industry and probability-severity evaluations of risk using matrix method.

	RISK	ODSS	VIOLENCE
1	workplace ambient temperature	0,2-0,4-0,6	0,0-0,2-0,4
2	the air flow rate	0,2-0,4-0,6	0,0-0,2-0,4
3	relative humidity	0,2-0,4-0,6	0,0-0,2-0,4
4	carbon monoxide	0,4-0,6-0,8	0,8-1,0-1,0
5	carbon dioxide	0,0-0,2-0,4	0,6-0,8-1,0
6	hydrogen cyanide	0,0-0,2-0,4	0,6-0,8-1,0
7	lead	0,0-0,2-0,4	0,6-0,8-1,0
8	mercury	0,0-0,2-0,4	0,6-0,8-1,0
9	arsenic	0,0-0,2-0,4	0,6-0,8-1,0
10	hydrogen sulphide	0,0-0,2-0,4	0,6-0,8-1,0
11	beryllium	0,0-0,2-0,4	0,6-0,8-1,0
12	ammonia	0,0-0,2-0,4	0,2-0,4-0,6
13	chlorine	0,0-0,2-0,4	0,2-0,4-0,6
14	nitrogen dioxide	0,0-0,2-0,4	0,6-0,8-1,0
15	sulphur dioxide	0,0-0,2-0,4	0,6-0,8-1,0
16	ozone	0,2-0,4-0,6	0,2-0,4-0,6
17	arsine	0,0-0,2-0,4	0,6-0,8-1,0
18	phosphine	0,2-0,4-0,6	0,6-0,8-1,0
19	stibine	0,2-0,4-0,6	0,6-0,8-1,0
20	alkanolamines	0,2-0,4-0,6	0,6-0,8-1,0
21	aluminum oxide	0,0-0,2-0,4	0,6-0,8-1,0
22	biocides petroleum	0,0-0,2-0,4	0,2-0,4-0,6
23	antimony	0,4-0,6-0,8	0,6-0,8-1,0
24	bauxite	0,4-0,6-0,8	0,6-0,8-1,0
25	borates	0,4-0,6-0,8	0,6-0,8-1,0
26	cadmium containing particles	0,0-0,2-0,4	0,6-0,8-1,0
27	caustic soda	0,2-0,4-0,6	0,2-0,4-0,6
28	cellulose derivative thickeners	0,4-0,6-0,8	0,2-0,4-0,6
29	copper	0,2-0,4-0,6	0,6-0,8-1,0
30	corrosion inhibitors	0,0-0,2-0,4	0,6-0,8-1,0
31	cryolite	0,2-0,4-0,6	0,6-0,8-1,0
32	cyanide powders	0,4-0,6-0,8	0,6-0,8-1,0
33	fluorite and aluminum fluoride powders	0,2-0,4-0,6	0,6-0,8-1,0
34	hydrocarbons	0,6-0,8-1,0	0,6-0,8-1,0
35	hydrofluoric acid fumes	0,6-0,8-1,0	0,6-0,8-1,0
36	fluorides	0,4-0,6-0,8	0,6-0,8-1,0
37	greases	0,4-0,6-0,8	0,6-0,8-1,0
38	Formaldehyde	0,6-0,8-1,0	0,6-0,8-1,0
39	fumes and other chemicals	0,0-0,2-0,4	0,0-0,2-0,4
40	disintegration products	0,0-0,2-0,4	0,0-0,2-0,4
41	irritants	0,4-0,6-0,8	0,6-0,8-1,0
42	laceration of hands	0,4-0,6-0,8	0,6-0,8-1,0
43	lead	0,4-0,6-0,8	0,6-0,8-1,0
44	metal smoke fire	0,2-0,4-0,6	0,6-0,8-1,0
45	metal surface fouling	0,2-0,4-0,6	0,2-0,4-0,6
46	metalworking fluids	0,2-0,4-0,6	0,6-0,8-1,0
47	microbiological contaminants	0,2-0,4-0,6	0,0-0,2-0,4
48	molds	0,0-0,2-0,4	0,2-0,4-0,6
49	injuries	0,0-0,2-0,4	0,6-0,8-1,0
50	musculoskeletal injuries	0,0-0,2-0,4	0,6-0,8-1,0
51	nickel carbonyl	0,0-0,2-0,4	0,6-0,8-1,0
52	nitrogen dioxide	0,0-0,2-0,4	0,6-0,8-1,0
53	nitrosamines	0,0-0,2-0,4	0,6-0,8-1,0
54	petroleum sulfonates	0,0-0,2-0,4	0,2-0,4-0,6
55	silica	0,0-0,2-0,4	0,6-0,8-1,0
56	sodium carbonate	0,0-0,2-0,4	0,2-0,4-0,6
57	sodium nitrate	0,0-0,2-0,4	0,2-0,4-0,6
58	sodium nitrite	0,0-0,2-0,4	0,2-0,4-0,6
59	sodium silicate	0,0-0,2-0,4	0,2-0,4-0,6
60	zinc	0,0-0,2-0,4	0,4-0,6-0,8
61	pressurized gas cylinders	0,0-0,2-0,4	0,4-0,6-0,8
62	physical strain	0,0-0,2-0,4	0,4-0,6-0,8
63	waxes	0,0-0,2-0,4	0,2-0,4-0,6
64	welding wires	0,0-0,2-0,4	0,2-0,4-0,6
65	variety of harmful dusts	0,0-0,2-0,4	0,2-0,4-0,6
66	eye and body injuries	0,2-0,4-0,6	0,2-0,4-0,6
67	interaction with hot and uv light environment	0,4-0,6-0,8	0,2-0,4-0,6
68	ionizing radiation	0,2-0,4-0,6	0,2-0,4-0,6
69	traviolet radiation	0,2-0,4-0,6	0,2-0,4-0,6
70	high pressure	0,2-0,4-0,6	0,2-0,4-0,6
71	chemical substances	0,4-0,6-0,8	0,2-0,4-0,6
72	noise	0,2-0,4-0,6	0,2-0,4-0,6
73	hand-arm vibration	0,0-0,2-0,4	0,2-0,4-0,6
74	whole body vibration	0,4-0,6-0,8	0,2-0,4-0,6
75	lighting	0,2-0,4-0,6	0,0-0,2-0,4
76	harmful rays and radiation	0,2-0,4-0,6	0,0-0,2-0,4
77	electromagnetic fields	0,0-0,2-0,4	0,0-0,2-0,4
78	hot or cold climate	0,2-0,4-0,6	0,0-0,2-0,4
79	irregular and slippery surfaces	0,2-0,4-0,6	0,0-0,2-0,4
80	movement of vehicles and machines	0,2-0,4-0,6	0,2-0,4-0,6
81	machine movements and parts	0,4-0,6-0,8	0,2-0,4-0,6
82	hazardous surfaces	0,0-0,2-0,4	0,0-0,2-0,4
83	hot or cold surfaces	0,0-0,2-0,4	0,0-0,2-0,4
84	hand tools	0,2-0,4-0,6	0,0-0,2-0,4
85	electrical installations and equipment	0,2-0,4-0,6	0,8-1,0-1,0
86	fire	0,2-0,4-0,6	0,8-1,0-1,0
87	explosion	0,0-0,2-0,4	0,8-1,0-1,0
88	lifting and handling	0,0-0,2-0,4	0,0-0,2-0,4
89	stopping faults	0,0-0,2-0,4	0,0-0,2-0,4
90	biological hazards	0,0-0,2-0,4	0,0-0,2-0,4
91	stress, violence	0,0-0,2-0,4	0,0-0,2-0,4

Table 6. Fuzzy decision making approach in an applied metal industry and risk rating of risk assessments using the matrix method.

	RISK	RISK MEMBERSHIP LEVEL
1	workplace ambient temperature	0,00-0,08-0,24
2	the air flow rate	0,00-0,08-0,24
3	relative humidity	0,00-0,08-0,24
4	carbon monoxide	0,32-0,60-0,80
5	carbon dioxide	0,00-0,16-0,40
6	hydrogen cyanide	0,00-0,16-0,40
7	lead	0,00-0,16-0,40
8	mercury	0,00-0,16-0,40
9	arsenic	0,00-0,16-0,40
10	hydrogen sulphide	0,00-0,16-0,40
11	beryllium	0,00-0,16-0,40
12	ammonia	0,00-0,08-0,24
13	chlorine	0,00-0,08-0,24
14	nitrogen dioxide	0,00-0,16-0,40
15	sulphur dioxide	0,00-0,16-0,40
16	ozone	0,04-0,16-0,36
17	arsine	0,00-0,16-0,40
18	phosphine	0,16-0,32-0,60
19	stibine	0,16-0,32-0,60
20	alkanolamines	0,12-0,32-0,60
21	aluminum oxide	0,00-0,16-0,40
22	biocides petroleum	0,00-0,08-0,24
23	antimony	0,24-0,48-0,80
24	bauxite	0,24-0,48-0,80
25	borates	0,24-0,48-0,80
26	cadmium containing particles	0,00-0,16-0,40
27	caustic soda	0,04-0,16-0,36
28	cellulose derivative thickeners	0,08-0,24-0,48
29	copper	0,12-0,32-0,60
30	corrosion inhibitors	0,00-0,16-0,40
31	cryolite	0,12-0,32-0,60
32	cyanide powders	0,24-0,48-0,80
33	fluorite and aluminum fluoride powders	0,12-0,32-0,60
34	hydrocarbons	0,36-0,64-1,0
35	hydrofluoric acid fumes	0,36-0,64-1,0
36	fluorides	0,24-0,48-0,80
37	greases	0,24-0,48-0,80
38	Formaldehyde	0,36-0,64-1,0
39	fumes and other chemicals	0,00-0,04-0,16
40	disintegration products	0,00-0,04-0,16
41	irritants	0,24-0,48-0,80
42	laceration of hands	0,24-0,48-0,80
43	lead	0,24-0,48-0,80
44	metal smoke fire	0,24-0,32-0,60
45	metal surface fouling	0,04-0,16-0,36
46	metalworking fluids	0,24-0,32-0,60
47	microbiological contaminants	0,00-0,08-0,24
48	molds	0,00-0,08-0,24
49	injuries	0,00-0,16-0,40
50	musculoskeletal injuries	0,00-0,16-0,40
51	nickel carbonyl	0,00-0,16-0,40
52	nitrogen dioxide	0,00-0,16-0,40
53	nitrosamines	0,00-0,16-0,40
54	petroleum sulfonates	0,0-0,2-0,4
55	silica	0,0-0,2-0,4
56	sodium carbonate	0,00-0,08-0,24
57	sodium nitrate	0,00-0,08-0,24
58	sodium nitrite	0,00-0,08-0,24
59	sodium silicate	0,00-0,08-0,24
60	zinc	0,00-0,12-0,32
61	pressurized gas cylinders	0,00-0,12-0,32
62	physical strain	0,00-0,12-0,32
63	waxes	0,00-0,08-0,24
64	welding wires	0,00-0,08-0,24
65	variety of harmful dusts	0,00-0,08-0,24
66	eye and body injuries	0,04-0,16-0,36
67	interaction with hot and uv light environment	0,08-0,24-0,48
68	ionizing radiation	0,04-0,16-0,36
69	traviolet radiation	0,04-0,16-0,36
70	high pressure	0,04-0,16-0,36
71	chemical substances	0,08-0,24-0,48
72	noise	0,04-0,16-0,36
73	hand-arm vibration	0,00-0,08-0,24
74	whole body vibration	0,08-0,24-0,48
75	lighting	0,00-0,08-0,24
76	harmful rays and radiation	0,00-0,08-0,24
77	electromagnetic fields	0,00-0,04-0,16
78	hot or cold climate	0,00-0,08-0,24
79	irregular and slippery surfaces	0,00-0,08-0,24
80	movement of vehicles and machines	0,04-0,16-0,36
81	machine movements and parts	0,08-0,24-0,48
82	hazardous surfaces	0,00-0,04-0,16
83	hot or cold surfaces	0,00-0,04-0,16
84	hand tools	0,00-0,08-0,24
85	electrical installations and equipment	0,16-0,40-0,60
86	fire	0,16-0,40-0,60
87	explosion	0,00-0,20-0,40
88	lifting and handling	0,00-0,04-0,16
89	stopping faults	0,00-0,04-0,16
90	biological hazards	0,00-0,04-0,16
91	stress, violence	0,00-0,04-0,16

Table 7. Converting risks into probability-severity membership removal linguistic variables using a fuzzy decision-making approach and matrix method in an applied metal industry

	RISK	ODSS	VIOLENCE
1	workplace ambient temperature	dd,Dd,Od	Nd,dd,Dd
2	the air flow rate	dd,Dd,Od	Nd,dd,Dd
3	relative humidity	dd,Dd,Od	Nd,dd,Dd
4	carbon monoxide	Dd,Od,yd	yd,Yd,Yd
5	carbon dioxide	Nd,dd,Dd	Od,yd,Yd
6	hydrogen cyanide	Nd,dd,Dd	Od,yd,Yd
7	lead	Nd,dd,Dd	Od,yd,Yd
8	mercury	Nd,dd,Dd	Od,yd,Yd
9	arsenic	Nd,dd,Dd	Od,yd,Yd
10	hydrogen sulphide	Nd,dd,Dd	Od,yd,Yd
11	beryllium	Nd,dd,Dd	Od,yd,Yd
12	ammonia	Nd,dd,Dd	dd,Dd,Od
13	chlorine	Nd,dd,Dd	dd,Dd,Od
14	nitrogen dioxide	Nd,dd,Dd	Od,yd,Yd
15	sulphur dioxide	Nd,dd,Dd	Od,yd,Yd
16	ozone	dd,Dd,Od	dd,Dd,Od
17	arsine	Nd,dd,Dd	Od,yd,Yd
18	phosphine	dd,Dd,Od	Od,yd,Yd
19	stibine	dd,Dd,Od	Od,yd,Yd
20	alkanolamines	dd,Dd,Od	Od,yd,Yd
21	aluminum oxide	Nd,dd,Dd	Od,yd,Yd
22	biocides petroleum	Nd,dd,Dd	dd,Dd,Od
23	antimony	Dd,Od,yd	Od,yd,Yd
24	bauxite	Dd,Od,yd	Od,yd,Yd
25	borates	Dd,Od,yd	Od,yd,Yd
26	cadmium containing particles	Nd,dd,Dd	Od,yd,Yd
27	caustic soda	dd,Dd,Od	dd,Dd,Od
28	cellulose derivative thickeners	Dd,Od,yd	dd,Dd,Od
29	copper	dd,Dd,Od	Od,yd,Yd
30	corrosion inhibitors	Nd,dd,Dd	Od,yd,Yd
31	cryolite	dd,Dd,Od	Od,yd,Yd
32	cyanide powders	Dd,Od,yd	Od,yd,Yd
33	fluorite and aluminum fluoride powders	dd,Dd,Od	Od,yd,Yd
34	hydrocarbons	Od,yd,Yd	Od,yd,Yd
35	hydrofluoric acid fumes	Od,yd,Yd	Od,yd,Yd
36	fluorides	Dd,Od,yd	Od,yd,Yd
37	greases	Dd,Od,yd	Od,yd,Yd
38	Formaldehyde	Od,yd,Yd	Od,yd,Yd
39	fumes and other chemicals	Nd,dd,Dd	Nd,dd,Dd
40	disintegration products	Nd,dd,Dd	Nd,dd,Dd
41	irritants	Dd,Od,yd	Od,yd,Yd
42	laceration of hands	Dd,Od,yd	Od,yd,Yd
43	lead	Dd,Od,yd	Od,yd,Yd
44	metal smoke fire	dd,Dd,Od	Od,yd,Yd
45	metal surface fouling	dd,Dd,Od	dd,Dd,Od
46	metalworking fluids	dd,Dd,Od	Od,yd,Yd
47	microbiological contaminants	dd,Dd,Od	Nd,dd,Dd
48	molds	Nd,dd,Dd	dd,Dd,Od
49	injuries	Nd,dd,Dd	Od,yd,Yd
50	musculoskeletal injuries	Nd,dd,Dd	Od,yd,Yd
51	nickel carbonyl	Nd,dd,Dd	Od,yd,Yd
52	nitrogen dioxide	Nd,dd,Dd	Od,yd,Yd
53	nitrosamines	Nd,dd,Dd	Od,yd,Yd
54	petroleum sulfonates	Nd,dd,Dd	dd,Dd,Od
55	silica	Nd,dd,Dd	Od,yd,Yd
56	sodium carbonate	Nd,dd,Dd	dd,Dd,Od
57	sodium nitrate	Nd,dd,Dd	dd,Dd,Od
58	sodium nitrite	Nd,dd,Dd	dd,Dd,Od
59	sodium silicate	Nd,dd,Dd	dd,Dd,Od
60	zinc	Nd,dd,Dd	Dd,Od,yd
61	pressurized gas cylinders	Nd,dd,Dd	Dd,Od,yd
62	physical strain	Nd,dd,Dd	Dd,Od,yd
63	waxes	Nd,dd,Dd	dd,Dd,Od
64	welding wires	Nd,dd,Dd	dd,Dd,Od
65	variety of harmful dusts	Nd,dd,Dd	dd,Dd,Od
66	eye and body injuries	dd,Dd,Od	dd,Dd,Od
67	interaction with hot and uv light environment	Dd,Od,yd	dd,Dd,Od
68	ionizing radiation	dd,Dd,Od	dd,Dd,Od
69	traviolet radiation	dd,Dd,Od	dd,Dd,Od
70	high pressure	dd,Dd,Od	dd,Dd,Od
71	chemical substances	Dd,Od,yd	dd,Dd,Od
72	noise	dd,Dd,Od	dd,Dd,Od
73	hand-arm vibration	Nd,dd,Dd	dd,Dd,Od
74	whole body vibration	Dd,Od,yd	dd,Dd,Od
75	lighting	dd,Dd,Od	Nd,dd,Dd
76	harmful rays and radiation	dd,Dd,Od	Nd,dd,Dd
77	electromagnetic fields	Nd,dd,Dd	Nd,dd,Dd
78	hot or cold climate	dd,Dd,Od	Nd,dd,Dd
79	irregular and slippery surfaces	dd,Dd,Od	Nd,dd,Dd
80	movement of vehicles and machines	dd,Dd,Od	dd,Dd,Od
81	machine movements and parts	Dd,Od,yd	dd,Dd,Od
82	hazardous surfaces	Nd,dd,Dd	Nd,dd,Dd
83	hot or cold surfaces	Nd,dd,Dd	Nd,dd,Dd
84	hand tools	dd,Dd,Od	Nd,dd,Dd
85	electrical installations and equipment	dd,Dd,Od	yd,Yd,Yd
86	fire	dd,Dd,Od	yd,Yd,Yd
87	explosion	Nd,dd,Dd	yd,Yd,Yd
88	lifting and handling	Nd,dd,Dd	Nd,dd,Dd
89	stopping faults	Nd,dd,Dd	Nd,dd,Dd
90	biological hazards	Nd,dd,Dd	Nd,dd,Dd
91	stress, violence	Nd,dd,Dd	Nd,dd,Dd

Table 8. Obtaining risk membership ratings using a fuzzy decision-making approach and matrix method in an applied metal industry.

	RISK	RISK MEMBERSHIP LEVEL
1	workplace ambient temperature	0,106
2	the air flow rate	0,106
3	relative humidity	0,106
4	carbon monoxide	0,573
5	carbon dioxide	0,186
6	hydrogen cyanide	0,186
7	lead	0,186
8	mercury	0,186
9	arsenic	0,186
10	hydrogen sulphide	0,186
11	beryllium	0,186
12	ammonia	0,106
13	chlorine	0,106
14	nitrogen dioxide	0,186
15	sulphur dioxide	0,186
16	ozone	0,186
17	arsine	0,186
18	phosphine	0,36
19	stibine	0,36
20	alkanolamines	0,346
21	aluminum oxide	0,186
22	biocides petroleum	0,106
23	antimony	0,506
24	bauxite	0,506
25	borates	0,506
26	cadmium containing particles	0,186
27	caustic soda	0,186
28	cellulose derivative thickeners	0,266
29	copper	0,346
30	corrosion inhibitors	0,186
31	cryolite	0,346
32	cyanide powders	0,506
33	fluorite and aluminum fluoride powders	0,36
34	hydrocarbons	0,666
35	hydrofluoric acid fumes	0,666
36	fluorides	0,506
37	greases	0,506
38	Formaldehyde	0,666
39	fumes and other chemicals	0,066
40	disintegration products	0,066
41	irritants	0,506
42	laceration of hands	0,506
43	lead	0,506
44	metal smoke fire	0,386
45	metal surface fouling	0,186
46	metalworking fluids	0,386
47	microbiological contaminants	0,106
48	molds	0,106
49	injuries	0,186
50	musculoskeletal injuries	0,186
51	nickel carbonyl	0,186
52	nitrogen dioxide	0,186
53	nitrosamines	0,186
54	petroleum sulfonates	0,2
55	silica	0,2
56	sodium carbonate	0,106
57	sodium nitrate	0,106
58	sodium nitrite	0,106
59	sodium silicate	0,106
60	zinc	0,146
61	pressurized gas cylinders	0,146
62	physical strain	0,146
63	waxes	0,106
64	welding wires	0,106
65	variety of harmful dusts	0,106
66	eye and body injuries	0,186
67	interaction with hot and uv light environment	0,266
68	ionizing radiation	0,186
69	traviolet radiation	0,186
70	high pressure	0,186
71	chemical substances	0,266
72	noise	0,186
73	hand-arm vibration	0,106
74	whole body vibration	0,266
75	lighting	0,106
76	harmful rays and radiation	0,106
77	electromagnetic fields	0,066
78	hot or cold climate	0,106
79	irregular and slippery surfaces	0,106
80	movement of vehicles and machines	0,186
81	machine movements and parts	0,266
82	hazardous surfaces	0,066
83	hot or cold surfaces	0,066
84	hand tools	0,106
85	electrical installations and equipment	0,386
86	fire	0,386
87	explosion	0,2
88	lifting and handling	0,066
89	stopping faults	0,066
90	biological hazards	0,066
91	stress, violence	0,066

Figure 3: Creating software on a risk assessment analysis in mines.

Table 9. Fuzzy decision-making approach in an applied mine and sequence of risks using the matrix method.

RISKS GROUP	34	35	38	4	23	25	32	36	37	41	42	43	44	46	66	18	19	33	20	29	31	28	67	71	74	81	54	55	87	5	6	7	8	9	10	11	14	15	16	17	21	26	27	30
RISK MEMBERSHIP LEVEL	0,666	0,666	0,666	0,573	0,506	0,506	0,506	0,506	0,506	0,506	0,506	0,506	0,386	0,386	0,386	0,386	0,36	0,36	0,36	0,346	0,346	0,346	0,266	0,266	0,266	0,266	0,2	0,2	0,2	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186	0,186

RISKS GROUP	45	49	50	51	53	68	69	70	72	80	60	61	62	1	3	12	13	22	47	48	56	57	58	59	63	64	65	73	75	76	78	79	84	39	40	77	82	83	88	89	90	91
RISK MEMBERSHIP LEVEL	0,185	0,185	0,185	0,185	0,185	0,185	0,185	0,185	0,185	0,185	0,146	0,146	0,146	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,106	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066	0,066

V. CONCLUSIONS

In the metal sector, mostly working with machines and benches, a wide range of cutting, drilling and rigging tools are used in the production process. In the area of occupational health and safety in the sector, mechanical, physical hazards, dangerous methods and hazards arising from the process are the foregrounds. Businesses operating in the metal sector also have business-specific hazards other than common hazards and are specially made to operate their detection, analysis and measures to be taken. Because of the different production methods and forms used in the workplaces where the same product is produced in the sector, risks to be met due to structural differences with machinery and equipment may also vary. The location of the research has been ten metal-working factories in the provinces of Adana and Mersin. The dangers in the metal industry are defined as above given by making researches and examinations in 10 different factories in Adana and Mersin. When the reasons for these accidents were asked, the majority of the workers seemed to be involved in the business accidents due to the lack of employees, but the employers did not explain the underlying reasons for the absurdity. In this study, software was developed by creating a fuzzy logic based risk assessment analysis model in consideration of many hazards in occupational health and safety in metal industry. An alternative

approach to risk assessment has been proposed using fuzzy decision making approach and matrix method. With this approach, occupational health and safety specialists are provided with blurred linguistic assessments before calculations are made. Thermal comfort (temperature, humidity, air flow rate) and clean ventilation in the metal industry environment, particulate matter-heavy metal in the ambient air, internal noise, vibration, lighting, gasses in the environment, a centralized control system was developed. It has been observed that the risks are reduced in these 10 metalworking factories.

REFERENCES

[1] Gens, A., Hurley, J. F., Tuomisto, J. T. and Friedrich, R., Health impacts due to personal exposure to fine particles caused by insulation of residential buildings in Europe. Atmospheric Environment, 2014, 84, 213-221.

[2] Van Orden, D. R., Lee, R. J., Allison, K. A. and Addison, J., Width Distributions of Asbestos and Non-Asbestos Amphibole Minerals. Indoor and Built Environment, 2009, 18 (6), 531-540.

[3] Youn, Y., Comparison of Asbestos Exposure and Risk Assessment by Asbestos Mining Type in Korea. Ronald, F. and Samuel, P., Asbestos: Risk Assessment, Epidemiology, and Health Effects International journal of toxicology, 2011, 25, 139-141.

[4] Mannan, S., Lees' Loss Prevention İn The Process Industries: Hazard Identification, Assessment And Control, 4th Ed. Butterworth-Heinemann, 2012.

[5] Jones, R., Lehr, W., Simecek - Beatty, D., Areal Locatıons Of Hazardous Atmospheres, Technıcal Documentatıon, Noaa Technical Memorandum Nos Or&R, 2013, Pp 43, 96.

[6] Occupational Injuries and Illnesses in the United States by Industry; United States Bureau of Labor Statistics: Washington, DC, USA, 2012.

[7] Van Den Bosch, C.J.H., Weterings, R.A.P.M., Methods For The Calculation Of Physical Effects Cpr14e,Yellow Book, Comimittee For The Prevention Of Disasters, 3rd Ed., Nl, 2005.

[8] Oginska-Bulik, N. Occupational Stress And Its Consequences In Healthcare Professionals: The Role Of Type D Personality. Int J Occup Med Environ Health 2006;19:113–22.

[9] Amyotte, P. Dust Explosions: Understanding the Myths and Realities of Dust Explosions for a Safer Workplace; Butterworth-Heinemann: Oxford, UK, 2013.

[10] Rudel, RA; Perovich, LJ. Endocrine disrupting chemicals in indoor and outdoor air. Atmos. Environ 2009, 43, 170–181.

[11] WHO Guidelines for Indoor Air Quality: Dampness and Mould; World Health Organization Regional Office for Europe: Copenhagen, Denmark, 2009.

[12] Reducing Risks, Promoting Healthy Life; World Health Organization: Geneva, Switzerland, 2002.

[13] C.C. Lee, Fuzzy Logic In Control Systems: Fuzzy Logic Controller, IEEE Transactions, 1990, Vol20, No. 2.

[14] L. A. Zadeh, Fuzzy Logic Computing With Words, Pieee Transactions On Fuzzy Systems, 1996, Vol.4, No.2.

[15] K. M. Passino, S. Yurkovich, Fuzzy Control, 1998, Addison Wesley.

[16] M. Du, T. Fan, W. Su, H. Li, Design Of A New Practical Expert Fuzzy Controller In Central Air Conditioning Control System, Ieee Pacific-Asia Workshop On Computational Intelligence And Industrial Application, 2008.

[17] M. S. I. Md., S. Z. Sarker, K. A. A. Rafi, M. Othman, Development Of A Fuzzy Logic Controller Algorithm For Air Conditioning System, ICSE Proceedings, 2006.

[18] Gerla G., Fuzzy Logic: Mathematical Tools for Approximate Reasoning Kluwer Academic Publishers Dordrecht, 2001.

[19] Abuswer, M.; Amyotte, P.; Khan, F.; Morrison, L, An optimal level of dust explosion risk management : Framework and application. Journal of Loss Prevention in the Process Industries 2013, 26, 1530–1541.

[20] M. Zile, Analysis Of Job Accidents Transferred To The Judiciary And Creation Project Of Causality Modeling, University Of Cukurova, 2013, Publication No: E / 2013/600, Isbn: 978-605-01-0469-1, Adana.

[21] M. Zile, Forensic Cases Based On Maintenance And Repair Projects For Causing Approaches In Work Accidents And Measures To Be Taken, University Of Cukurova, 2015, No: E/2015/635, Isbn: 978-605-01-0710-4, 626, Adana.

Detection of DDOS Attacks in Network Traffic Using Deep Learning

A. SUNGUR UNAL[1], M. HACIBEYOGLU[1]

[1]Necmettin Erbakan University, Konya/Turkey, aysegulsngr@gmail.com
[1]Necmettin Erbakan University, Konya/Turkey, hacibeyoglu@konya.edu.tr

***Abstract* - In the literature, machine learning algorithms are frequently used in detecting anomalies in network traffic and in building intrusion detection systems. Deep learning is a subfield of machine learning that trains a computer-based system to perform humanitarian tasks, such as disease diagnosis, speech recognition, image recognition, fraud detection, and making predictions. In the experimental study, NSL-KDD dataset was used for evaluating the performance of the proposed deep learning based DDoS detection model. NLS-KDD dataset contains normal network traffic and 23 different DDoS attacks that consists of 41 features. In the experimental study two different experiments are carried out. Firstly, the proposed deep neural network detected the Dos attacks with 0.988 classification accuracy. In the second experiment, the number of features of NSL-KDD is reduced to 24 by examining the previous feature reduction research on NSL-KDD dataset. The proposed deep neural network classified the all cyber-attacks with 0.984 classification accuracy. The 10-fold cross validation is used for all experiments. As a result, the proposed deep learning based DDoS detection achieved good performance.**

***Keywords* - Deep learning, intrusion detection systems, distributed denial of service, classification.**

I. INTRODUCTION

The number of cyber-attacks is increased with the developing technology. Personal and institutional computer systems can face a variety of threats and danger by malicious people, such as information theft, spoofing and denial of service. This can create material and moral damages for both individuals and institutions. For this reason, it is necessary to take security measures.

Security mechanisms have been established to prevent security weaknesses against a variety of threats that may violate information security. Intrusion detection systems are a security mechanism which is created for this purpose.

Catak and Mustacoglu, proposed a model for the detection of malicious network attacks with machine learning and deep learning technologies by using a data set from Cyber Security Laboratory in Australia Cyber Security Center [1].

Kaya, examined performance of artificial neural networks (ANN), K-nearest neighbor algorithm (KNN), support vector machines (SVM) and decision tree by using the datasets KDD CUP99 and NSL-KDD. In determining DOS attacks, KNN, decision trees and ANN have achieved good performances [2].

Yuan et al. have proposed DeepDefense technique to detect DDOS attacks by using the ISCX2012 dataset and found that the error rate is lower than the traditional machine learning methods [3].

Vijayasarathy tried to detect DOS and DDOS attacks using the Naive Bayes method for TCP and UDP with a system designed for real-time and feasibility in his work [4].

Uslu has created two different decision trees using the standard ID3 and the newly proposed method using the KDD CUP99 data set for DOS detection in his work. He compared the success rates of the attacks detected by these two methods and found that the decision tree structure constructed with the new methods is 3% more successful than the ID3 algorithm [5].

Guven uses the SVM, ANN, decision trees, bayesian networks and KNN machine learning algorithms using KDD CUP99 and NSL-KDD datasets to determine the sensitivity, selectivity, precision and F-score of intrusion detection systems [6].

Saied et al. attempted to detect known and unknown DDOS attacks in the real-time environment. In order to detect DDOS attacks, artificial neural networks connected to certain characteristic features that distinguish real traffic from attack traffic are used. [7]. According to this study:

- The detection of DDOS attacks by deep learning methods that are becoming popular nowadays,
- Deep learning methods enable faster and higher success with large data sizes.

II. INTRUSION DETECTION SYSTEM

An intrusion detection system is a warning system that detects an attack if the system's confidentiality, integrity or accessibility is damaged, by monitoring all events that may occur in a computer system or computer network.

Intrusion detection systems are similar to security alarms which are used in our daily lives. Attacks against computer networks can be understood by detecting abnormalities in network traffic. The network traffic is monitored and compared against the database where the attack signatures are already located, and the attack is alerted.

There are 4 basic attack types on the network.

1. Denial of Service Attack (DoS): It is to make the system unserviceable by sending more requests than the system can respond to. The most well-known DOS example attacks can be given as SYN floods, Smurf, UDPstorm, Pingflood, Neptune, Mailbomb attacks [8].

2. User Root Attack (U2R): An attacker with access to a regular user account on the system uses some security vulnerabilities (by sniffing passwords, dictionary attacks or social engineering) to provide root access to the system [9].

3. Remote Local Attack (R2L): An attacker who has the ability to send packets to a machine over a network, but does

not have an account on this machine, gains some security advantage and gains local access as the user of that machine [8].

4. Probing Attack: An attempt to gather information about a computer network to prevent security checks [8].

III. DDOS

A common method for performing Distributed Denial of Service (DDOS) attacks is when an attacker sends a packet flow to a victim; this flow consumes some important resources and makes it unusable for legitimate customers of the victim. Another common approach is that an attacker sends several malformed packets that confuse an application or protocol in the victim machine and force it to freeze or restart [10].

IV. DATA SET

The NSL-KDD data set was used to train and test our system. KDD CUP 99 is the most commonly used data set for anomaly detection. However, the various disadvantages and various statistical analyzes in the KDD CUP 99 dataset affected the accuracy of many IDS determinations modeled by researchers. NSL-KDD, a new data set consisting of the selected records of the full KDD CUP 99 data set was created.

The advantages of NSL-KDD over the original KDD dataset are:

• There are no unnecessary records in the learning set. For this reason, classifiers will not be directed to records more often.

• The number of records selected from each difficulty level group is inversely proportional to the percentage of records in the original KDD data set. As a result, the classification ratios of different machine learning methods vary within a wider range, making it more efficient to correctly evaluate different learning techniques.

• The number of records in the train and test sets is reasonable and this makes it convenient to run the entire set without having to randomly select a small section. As a result, the evaluation results of different research studies will be consistent and comparable [11].

In each record there are 41 properties that reveal different properties of the stream, and each has an attack type or a label that is normally assigned. These types of attacks are grouped as DoS, Probe, R2L and U2R. Attributes are listed on the tables [12].

1. Basic Attribute: This category encompasses all attributes extracted from a TCP / IP connection. Many of these features cause a latent delay in detection.

Table 1. Each Network Connectivity Vectors Basic Features

No.	Feature	Type
1	Duration	Numeric
2	Protocol_type	Nominal
3	Service	Nominal
4	Flag	Nominal
5	Src_bytes	Numeric
6	Dst_bytes	Numeric
7	Land	Binary
8	Wrong_fragment	Numeric
9	Urgent	Numeric

2. Context Properties: Contrary to the majority of DoS and Probing attacks, R2L and U2R attacks do not have any suitable attacks against the frequently encountered sequential die. For this reason, DoS and Probing attacks contain many links with some hosts or hosts in a very short time; but R2L and U2R attacks are embedded in the data partitions of the packages and normally only contain a single link. To detect such attacks, some features are needed to look for suspicious behavior in the data partition (for example, unsuccessful logon attempts). These properties are called content properties [9].

Table 2. Each Network Connectivity Vectors Related Features

No.	Feature	Type
10	Hot	Numeric
11	Num_failed_logins	Numeric
12	Logged_in	Binary
13	Num_compromised	Numeric
14	Root_shell	Binary
15	Su_attempted	Binary
16	Num_root	Numeric
17	Num_file_creations	Numeric
18	Num_shells	Numeric
19	Num_access_files	Numeric
20	Num_putbound_cmds	Numeric
21	Is_hot_login	Binary
22	Is_guest_login	Binary

3. Traffic Properties: This category contains properties calculated according to a window range and is divided into two groups:

a) "Same Host" Features: The same target host with the current connection has features that only review the last 2 seconds of connectivity, and such as protocol behavior or service calculates statistics about.

b) "Same Service" Features: It examines only the last 2 seconds of connections with the same service as the current connection.

The two "traffic" modes mentioned above are called time based. However, there are several slow scan attacks that scan the host machine (or ports) for a time interval greater than 2 seconds, for example, every minute. As a result, these attacks do not create attack patterns with a 2-second time window. To solve this problem, the "same host" and "same service" features are recalculated, but rely on the link window of 100 links instead of a 2 second time interval. These features are called link-based traffic features [11].

Table 3. Traffic Characteristics for Each Networking Vectors Time

No.	Feature	Type
23	Count	Numeric
24	Srv_count	Numeric
25	Serror_rate	Numeric
26	Srv_serror_rate	Numeric
27	Rerror_rate	Numeric
28	Srv_rerror_rate	Numeric
29	Same_srv_rate	Numeric
30	Diff_srv_rate	Numeric
31	Srv_diff_host_ rate	Numeric

Table 4. Host Based Traffic Features in Network Connection Vectors

No.	Feature	Type
32	Dst_host_count	Numeric
33	Dst_host_srv_count	Numeric
34	Dst_host_same_srv_rate	Numeric
35	Dst_host_diff_srv_rate	Numeric
36	Dst_host_same_src_port_rate	Numeric
37	Dst_host_srv_diff_host_rate	Numeric
38	Dst_host_serror_rate	Numeric
39	Dst_host_srv_serror_rate	Numeric
40	Dst_host_rerror_rate	Numeric
41	Dst_host_srv_rerror_rate	Numeric

V. DEEP LEARNING

Deep learning is one of the machine learning techniques and machine learning is done in one layer at the same time in many layers. A group of machine learning algorithms are used at the same time to produce results in a single operation. Deep learning is the use of multi-level "deep" neural networks with advanced technology to create feature detection systems with large numbers of untagged training data. Figure 1 shows the input layer, multiple hidden layers and output layers.

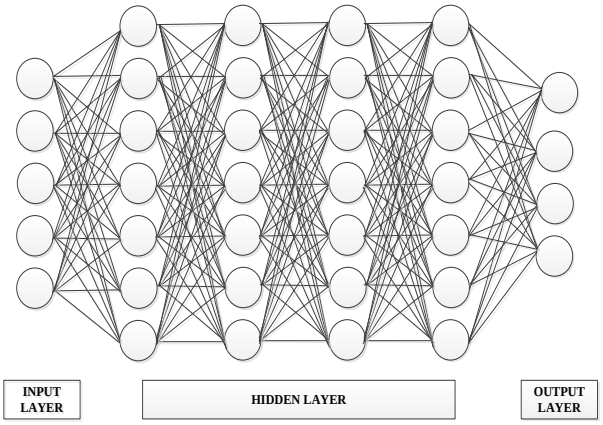


Figure 1. Deep Learning Neural Network

Deep learning methods learn the characteristics that distinguish them from a large number of data inputs. This feature should be adequately trained so that the learning can be done successfully. Feature learning consists of layers. The sublayers have less distinctive features while the uppermost layers have more distinctive features. The lower level features are the basis for building the upper level features. This type of learning is different from learning a machine. This is because the machine is trained with the characteristics determined by the people at the learning stage. In other words, while machine learning algorithms are human dependent, deep learning is human independent. This approach is an important influence on the success of deep learning.

Deep learning has many effects on our life. These;

1. Increasing the amount of data: Over the years the widespread use of the internet has caused it to be produced and stored digitally in very large sizes. The use of these great data has been achieved through deep learning systems [13].

2. Increase GPUs and processing power: Powerful and efficient parallel calculations can be done by GPU (Graphics Processing Unit) calculation. GPUs are being used to train deep learning algorithms using much larger training sets in a much shorter time and with much less data center infrastructure [13].

3. Increased depth: Increased processing power enables deep models to be used in practice. Deep learning models are also models with multi-layered structure. We can connect with the visual system of the human brain in order to understand the depth. The signals coming to the brain through the nerves in the eyes are processed by hierarchical evaluation in a layered structure. In the first layer the signal follows, the more local and basic features of the image, such as the edges and the corners, are recognized. In the next layer these edges and corners are brought together to form the mouth and nose, then faces on the next layer. On the next layer, the characteristics of the whole of the view like the settlement of the person and objects can be recognized. Many deep learning systems work on this principle [13].

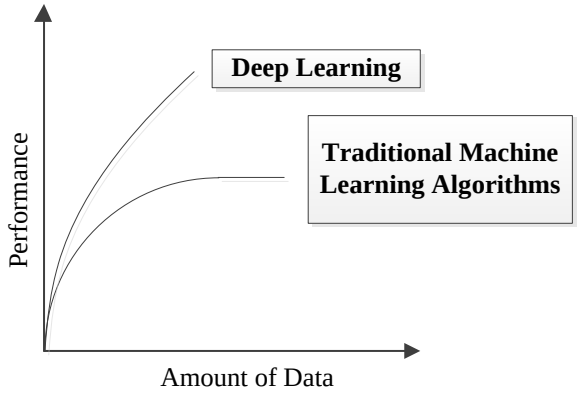


Figure 2. Performance Graph of Deep Learning and Machine Learning Algorithms

Increasing number of layers brings a deeper decision mechanism. Deep learning can work with more layers than traditional artificial neural networks. In the classical ANN

structure, each cell is connected to all the cells in the previous and subsequent layers. There are mathematical operations that need to be computed for each connection. As the number of layers and the number of cells increase, these processes require a high amount of CPU (Central Processing Unit) power. The CPU power in personal computers is insufficient to create a deep network structure. With the development of the GPU, the development of deep learning has also been accelerated. The difference between the GPU and the CPU is how a process is processed. CPUs contain several cores. Each core has a high processing capacity and processes are processed in a serial manner distributed to these cores. GPUs have hundreds of cores. The processing capacity of each kernel is less than CPU. However, it has high parallel processing power and can perform many simultaneous operations. This processing capacity of the GPU is an important point for implementing deep learning. In this way, the performance of training with a very large amount of data has increased. This is a point that distinguishes deep learning from traditional machine learning algorithms. Because in traditional machine learning algorithms, a high amount of data causes the success to increase some and then the success to remain constant. This success is shown in the graphic in Figure 2 [14].

VI. EXPERIMENTAL RESULTS

In order to evaluate the performance of the deep learning techniques for DDoS detection problem the proposed models are implemented using Python Keras library [15]. All experiments are carried out using a computer with an Intel Core i7 3840QM@2.80 Ghz processor with 16 GBs of memory operated on Microsoft Windows 8 system. In all experiments, the well-known evaluation measures classification accuracy (CA), recall and F-measure are obtained using a 10-fold cross validation for the proposed deep learning model. These evaluation measures are calculated in terms of true positive (TP), true negative (TN), false negative (FN) and false positive (FP) that are given in Eq. (1).

$$CA = (TP + TN)/(TP + TN + FP + FN)$$
$$Precison = (TP)/(TP + FP) \quad (1)$$
$$Recall = (TP)/(TP + FN)$$

The NSL-KDD dataset contains four main attack types and 22 different attacks classes which are shown in Table 5.

Table 5. Attack types of NSL-KDD

4 Main Attack Types	22 different attacks
Denial of Service (DoS)	back, land, neptune, pod, smurf, teardrop
Remote to User (R2L)	ftp_write, guess_passwd, imap, multihop, phf,spy, warezclient, warezmaster
User to Root (U2R)	buffer_overflow, perl, loadmodule, rootkit
Probing	ipsweep, nmap, portsweep, satan

Two different experiments were carried out in the experimental study. In the first experiments, DoS attacks are labeled as "True" and the other attacks are labeled as "False" and the deep learning model shown in Figure 3 is used.

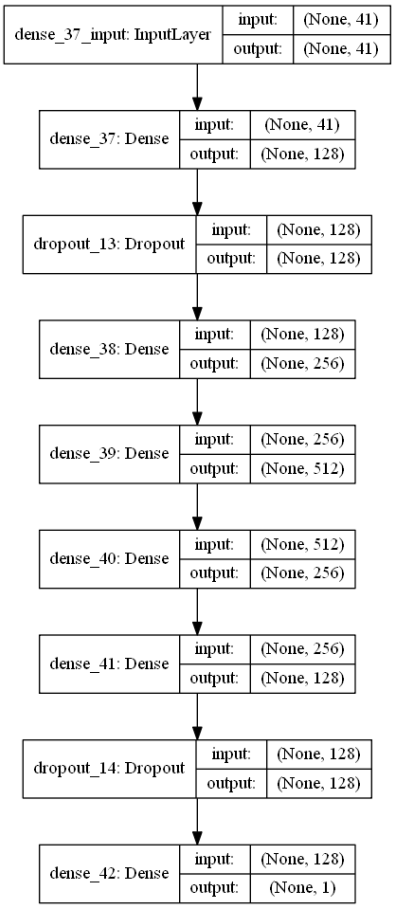


Figure 3. Proposed model for first experiment

For the first experiment, the number of epoch is set to 20 in the training process, binary_crossentropy is used as loss function and adam is used as optimizer. The evaluation measure results of the first experiment are shown in Table 6.

Table 6. Evaluation measures for DoS classification

CA	Precision	Recall
0.988	0.982	0.984

In the second experiments, deep learning is used to classify all cyber-attacks in NSL-KDD dataset. Prior the classification, most relevant features of DoS attacks are selected as a data pre-processing according to the previous literature [16]. The selected features are shown in Table 7.

Table 7. Attack types of NSL-KDD

Class Label	Relevant Features
Land	7
Smurf	2,3,5,23,24,27,28,36,40,41
Neptune	4,25,26,29,30,33,34,35,38,39

Teardrop	8
Back	10,13

For the second experiment, the number of epoch is set to 20 in the training process, sparse_categorical_crossentropy is used as loss function and adam is used as optimizer. The number of input is increased to 25 and the number of output is increased to 23. The evaluation measure results of the second experiment are shown in Table 8.

Table 8. Evaluation measures for DoS classification

CA	**Precision**	**Recall**
0.948	0.945	0.948

VII. CONCLUSION

In this paper, deep learning approach is proposed to detect DDoS cyber-attacks and NSL-KDD dataset is used for the experiments. A deep learning model is proposed for two different experiments. In the first experiment, our model achieved 0.988 CA for DoS attacks. In the second experiment, proposed model achieved 0.948 CA for all types of cyber-attacks. For future work, deep learning approach will be used with other machine learning approaches in a hybrid way and the parameters of deep learning will be tuned for improving the deep neural.

REFERENCES

[1] Çatak, F. Ö., Mustaçoğlu, A. F., (2017). Derin Öğrenme Teknolojileri Kullanarak Dağıtık Hizmet Dışı Bırakma Saldırılarının Tespit Edilmesi,2017. The 5th High Performance Computing Conference.

[2] Kaya, Ç. (2016). Saldırı Tespitinde Makine Öğrenmesi Tekniklerinin Performans Analizi.

[3] Yuan, X., Li, C., & Li, X. (2017, May). DeepDefense: Identifying DDoS Attack via Deep Learning. In Smart Computing (SMARTCOMP), 2017 IEEE International Conference on (pp. 1-8). IEEE.

[4] Vijayasarathy, R. (2012). A system approach to network modeling for DDoS detection using a Naïve Bayesian classifier.

[5] Uslu, N. Celal. (2009). Veri Madenciliği ile Bilgisayar Ağlarında Yeni Bir Saldırı Tespit Algoritması

[6] Güven, E. N.(2007). Zeki Saldırı Tespit Sistemi Tasarımı ve Gerçekleştirilmesi.

[7] A. Saied, R. E. Overill ve T. Radzik, "Detection of known and unknown DDoS attacks using Artificia Neural Networks", Neurocomputing, 172, pp.385-393, 2016.

[8] Mukkamala, S., Janoski, G., & Sung, A. (2002). Intrusion detection using neural networks and support vector machines. In Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on (Vol. 2, pp. 1702-1707). IEEE.

[9] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009, July). A detailed analysis of the KDD CUP 99 data set. In Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on (pp. 1-6). IEEE.

[10] Mirkovic, J., & Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. ACM SIGCOMM Computer Communication Review, 34(2), 39-53.

[11] Datti, R., & Verma, B. (2010). B.: Feature Reduction for Intrusion Detection Using Linear Discriminant Analysis. In International Journal on Engineering Science and Technology.

[12] Dhanabal, L., & Shantharajah, S. P. (2015). A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. International Journal of Advanced Research in Computer and Communication Engineering, 4(6), 446-452.

[13] Genç Ö., 2016, Keras ile Derin Öğrenmeye Giriş [online], https://medium.com/turkce/keras-ile-derin-%C3%B6%C4%9Frenmeye-giri%C5%9F-40e13c249ea8, [Available Date: 11 Mart 2018]

[14] Buber, E., 2017, Derin Öğrenme Nedir? [online], Cyber Security with Machine Learning, https://cybrml.com/2017/06/06/derin-ogrenme-uygulamalari/ [Available Date: 06 Şubat 2018]

[15] Keras, https://keras.io/

[16] Noureldien, N. A., & Yousif, I. M. (2016). Accuracy of machine learning algorithms in detecting DoS attacks types. Science and Technology, 6(4), 89-92

Data-Driven Estimation of Direction of Gravity from a Single Image

B. Z. TURKKOL ABUZARIFA[1] and Y. GENC[2]

[1] Bayburt University, Bayburt/Turkey, bzturkkol@bayburt.edu.tr
[2]Gebze Technical University, Kocaeli/Turkey, yakup.genc@gtu.edu.tr

***Abstract* – Direction of gravity is a natural way of orienting oneself in an unknown environment. Human beings do this with equilibrioception. It would be beneficial to estimate the direction of gravity from a single image for many tasks such as autonomous driving and augmented reality where the knowledge of location of the agent is very important.**

Extracting this knowledge from an image usually requires a reference to be identified. For example, a traffic light can give away the direction of gravity as it has to be placed in an environment in a specific way with respect to the gravity. Reference-based approaches require a lot of hand modeling for solving the problem. When there are a lot of images with ground truth data is available, one can model these references implicitly using machine learning techniques.

We propose to use a set of images along with readings from inertial magnetic unit (IMU) taken from a smart camera observing an indoor environment. This data includes a lot of images as well as ground truth labels for gravity direction extracted from the IMU readings. The data is fed to a convolutional deep neural network to estimate the gravity directions formulated as regression as well as classification problem. We show that this modeling works quite well with a few hundred images when we formulate the estimation as a classification problem. The details of the networks trained and the results obtained are presented in depth. Further research with more images but with less accurate ground truth data is underway.

***Keywords* - Camera Calibration, Motion Estimation, Deep Learning, Convolutional Neural Networks, IMU.**

I. INTRODUCTION

Estimation of the location of a camera from a single image is a well-studied problem in the field of computer vision [1]. It is also known as external camera calibration [2] and used in many applications such as autonomous navigation, augmented reality and robotics.

Most approaches to this problem requires the matching of image features to that of scene features given on some sort of a reference object in the scene. These references could be actual objects (such as a marker or a car) or a set of uniquely identifiable features (such as straight lines [3] or surface patches [4]). This requirement restricts the use in many applications where a predefined set of features might be difficult to extract. Similarly, a set of known objects may be hard to find in the scene.

It would be desirable to estimate the camera location based on the image without prior modeling of the scene. The data-driven methods can achieve this goal. Given a set of images with known poses, data driven models can build an implicit model of the scene (see Figure 1). When a new image is seen, the learned model model can estimate the pose of the camera by looking for implicitly learned features.

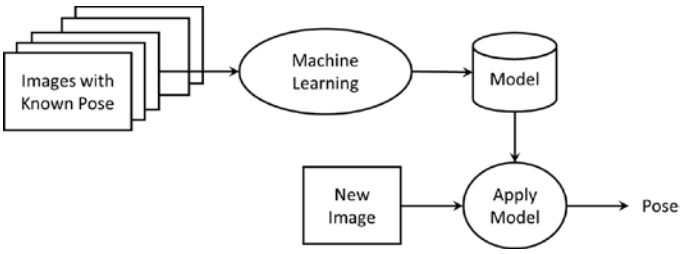


Figure 1 Direct estimation of pose from data using machine learning models.

We propose a data-driven modeling finding a component of the pose of a camera. In particular, we build a deep-learning model for estimating the direction of gravity given a single image. Posed as a classification problem, we show that the method works very well in indoor environments. Following the trend in application of deep learning in computer vision and other fields, our model comprises of a convolutional neural network trained with a new data set collected for this paper.

II. BACKGROUND

Estimating of the direction of gravity can be an important component of the pose of a camera. Such an estimation requires existence of a certain distribution of scene features. For example, lines in man-made environments tends to be distributed vertically and horizontally [5, 6]. Furthermore, other objects in the scene, for example, tiles and chairs, are placed in certain orientations in accordance with the direction of gravity.

Hand modeling these relations for a man-made scene might be difficult. Instead, such a model can be learned using machine learning techniques. When there is a lot of data deep learning methods can capture the relevant features for such a modeling task. For example, AlexNet [7] can capture features and their relationship to each other for object detection tasks of thousands of objects with millions of images. ResNet architecture provides even better performance by solving the problem of diminishing gradients with skip connections allowing very deep networks [8].

Gravity direction estimation can be formulated as a regression problem. Direct estimation of motion parameters using deep learning has been proposed in the field of homography estimation [9] or image path orientation estimation [12]. When applied to gravity estimation problem, we have found that regression approach does not work very well when there are

not many images available for training. Instead, converting the problem into classification by discretizing the continuous values space, one can achieve better performance.

III. Proposed Method

We propose a data driven method for estimating the direction of gravity given an image (see Figure 2). A set of images with known gravity directions are collected for training a model using machine learning.

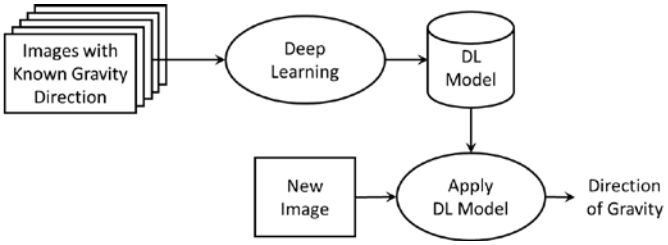


Figure 2 Data-driven learning-based estimation of direction of gravity from a single image.

The direction of gravity of the camera where the image are taken can be represented with two angles. These are pitch (θ) and yaw (α) angles (see Figure 3 for the configuration for a smartphone). We assume that the images are taken from a smartphone that is held upright and screen pointing towards the user and camera towards to opposite directions. In this case, the pitch is the angle between the direction of gravity and the phone's in plane orientation as seen from the user. The yaw is the angle between the direction of gravity and the phone's orientation towards the user.

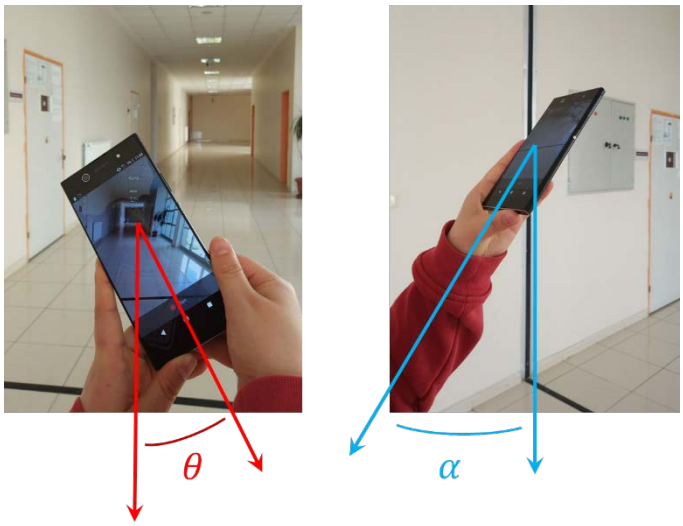


Figure 3 The geometry of camera with respect to the environment where the pitch and yaw angles are indicated with θ and α respectively.

The straightforward approach to modeling the direction of gravity is to estimate the two angles given the input image I:

$$[\theta\ \alpha\ ] = f(I),$$

where f is a regression function to be modeled. A convolutional deep neural network for regression can be applied [9].

An alternative to regression is to formulate the problem as a classification problem. In this case the two angle measurements are discretized and each value pair is designated as a class. For example, if $\theta \in [-50^o, +50^o]$ is split into 5° angles and $\alpha \in [-30^\circ, +30^o]$ is split into 10° angles, we will have 20×6 distinct value pairs as our classes (e.g, $[-50^\circ\ -30^o\ ]$ labeled as Class-0 and $[50^\circ\ 30^o\ ]$ labeled as Class-120).

Both of these models are learned using a convolutional neural network (CNN). In particular, we follow ResNet [8] and AlexNet [7] architectures as they are shown to perform very well in both classification and regression problems.

CNNs are well-known classifiers for image data classification and regression. ResNet [8] is one of the most popular CNN architectures. Because of residual blocks number of parameters to be learnt by the network is reduced. We use a modified ResNet architecture with 16 residual blocks, each block having 2 convolutional layers (see Figure 4).

IV. Experiments and Results

The proposed method has been tested on real data acquired from a two-floor building. The image acquisition is done while walking in the hallways of the two floors (see Figure 4 for the map of the first floor – the second floor is slightly different than the first one with less number of doors and a wider hallway). The camera is handled pointing forward towards the walking direction (see Figure 3). The corresponding angles between the camera and the gravity is called pitch (θ) and yaw (α) respectively. For the sake simplicity, we have restricted the yaw to be $-30^o, 0^o$ or 30° for our experiments.

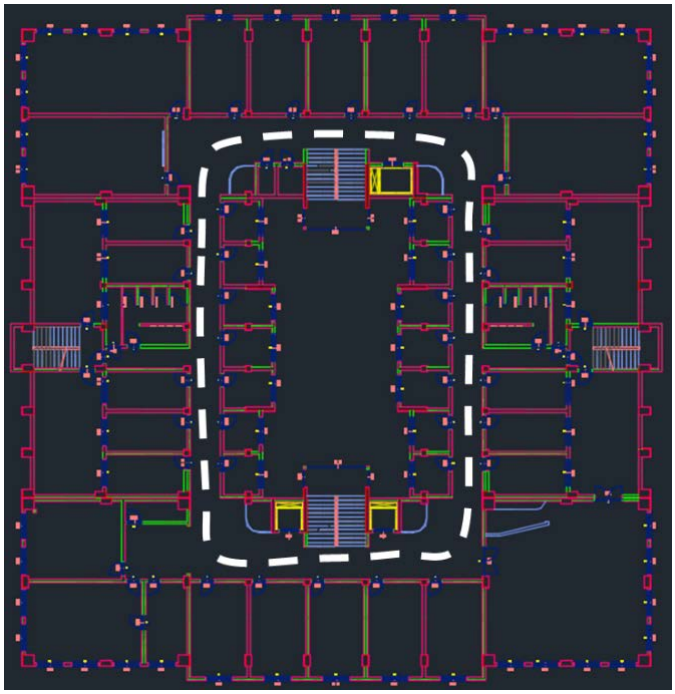


Figure 4 The floorplan of the first floor of the building where the data is collected. The white dashed lines indicate the path followed for image acquisition. The second floor is similar in architecture with smaller number of doors and a wider hallway.

The images are collected using a smartphone (Sony Xperia). To establish ground truth, the IMU (inertial-magnetic unit) sensor readings from the camera are recorded while saving the images. Sample images and corresponding angles collected from the camera data can be seen in Figure 5.

$\theta = 0^{\circ}, \alpha = 0^{o}$ $\theta = 50^{\circ}, \alpha = -30^{o}$ $\theta = -42^{\circ}, \alpha = 30^{o}$

Figure 5 The training and test images are taken in a hallway of a building. Images with varying pitch (θ) and yaw (α) angles are shown.

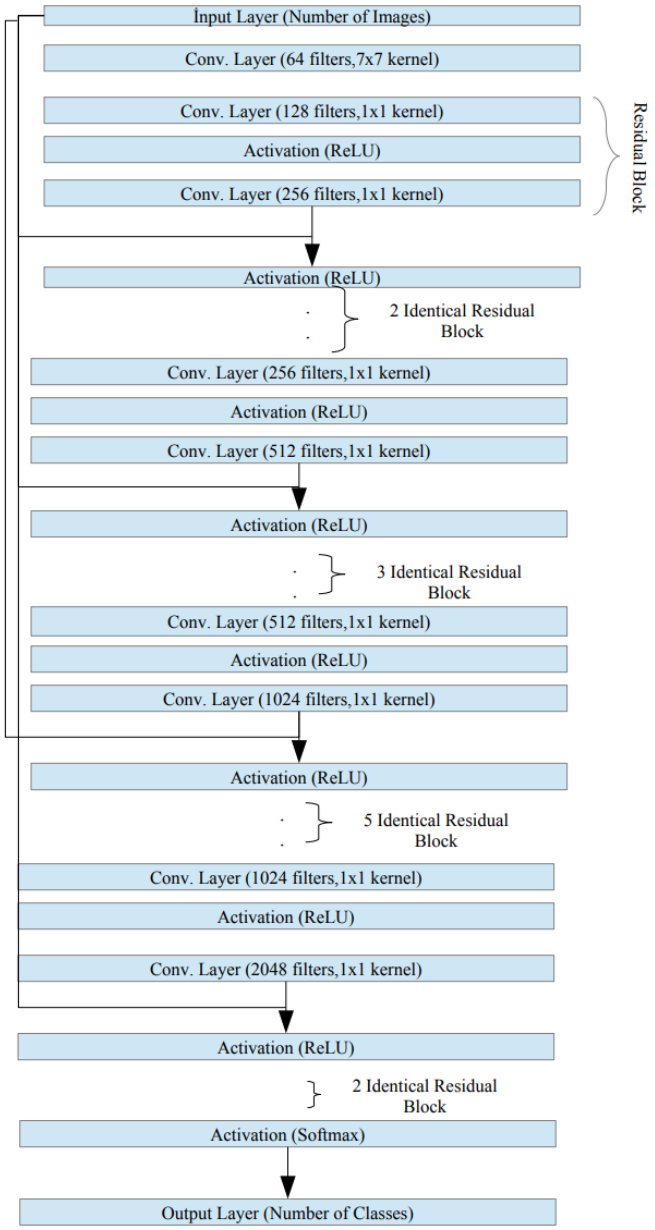


Figure 6 Deep Learning Architecture used in the classification problem to estimate the direction of gravity.

The collected training dataset includes 19,212 images of resolution 128x128 pixels. These images are collected on the first floor. The original RGB image of resolution 1920x1080 pixels are reduced to grayscale images of 128x128 pixels. We do not use RGB values as our building does not have a lot of color variations. The test images are taken on the other floor. We have 413 test images of the same resolution covering the entire floor. The data is restricted to have three distinct yaw values (i.e., be $-30^{\circ}, 0^{o}$ and 30°) and pitch values between -50° and 50°.

The direction of gravity estimation is converted to a classification problem by splitting the pitch and yaw angles into distinct values. We have split the interval first into 128 pieces corresponding to 2° of resolution on pitch and three distinct yaw angles as indicated earlier. The model estimation for this resolution is labeled DELTA2. In other case, we have split the interval into 61 pieces corresponding to 5° resolution of pitch and 3 distinct values for yaw. The model estimation for this resolution is labeled as DELTA5.

Our classifier is based on the ResNet architecture [8]. We have 16 residual blocks with 2 convolutional layers. These are after an input and a convolutional layer (see Figure 6). We used Categorical Cross Entropy as loss function, SGD [10] as optimizer, ReLU [11] in all convolutional layers and finally SoftMax in the last layer as activation functions. An alternative model is used following the success of AlexNet architecture [7].

The ResNet model has about 16M parameters. During training we run the optimization process for 20 epochs with batch sizes 64 for this model. For the AlexNet model which has about 82M parameters we run the process also for 20 epochs batch sizes 128. Our implementation is based on Keras 2.0.8 using TensorFlow backend with Python 2.7.6 and NumPY 1.11.0. The models are trained on two Nvidia Tesla K20Xm GPUs spending around 300 and 100 seconds to train per epoch for Resnet and AlexNet respectively.

Table 1 Training and test result for DELTA2 on ResNet and AlexNext models.

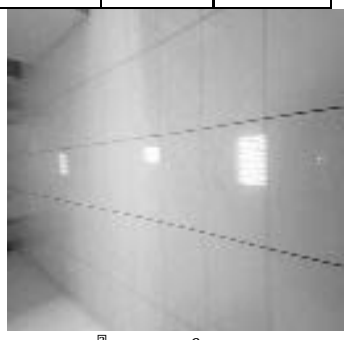
	Rank1	Rank2	Rank3	Rank4	Rank5
ResNet (Validation /Test)	%88	%99	%100	%100	%100
	%93	%97	%98	%99.3	%99.5
AlexNet (Validation /Test)	%64.7	%81.9	%81.9	%81.9	%83
	%82.3	%92.7	%92.8	%92.8	%92.9

Table 2 Training and test result for DELTA5 on ResNet and AlexNext models.

	Rank1	Rank2	Rank3	Rank4	Rank5
ResNet (Validation	%74	%98	%99	%100	%100

/Test)	%96	%98	%99.7	%99.8	%99.9
AlexNet (Validation	%64.5	%74.9	%75.6	%77	%79
/Test)	%82	%88	%88.7	%88.9	%89

$\theta = -30^{\circ}, \alpha = -30^{o}$



$\theta = -90^{\circ}, \alpha = 30^{o}$

Figure 7 Sample results of the classifier on the test data. The left image is predicted correctly in Rank1 while the right image is predicted only in Rank5. The right image is clearly ambiguous to even a human eye in terms of the direction of gravity. The best estimate for right image is -90 with a probability of 0.45.

Results for ResNet and AlexNext variants are shown in Table 1 and Table 2 Table 1. These results suggest that the classification works well for this classification problem. We have also compared this method against and algorithm where the gravity direction estimation is posed as a regression problem. A network like AlexNet is trained and an average error of 16.4° was obtained. This is clearly much bigger than even Rank5 results shown above for DELTA2.

V. Conclusion

A data-driven method for single image gravity direction estimation is proposed. Learning from the data in an indoor setting, our method provides very good accuracy with gracefully degraded performance Posing the prediction of the gravity estimation as a classification problem instead of a regression problem allowed us to converge to a good solution during the training. A convolutional neural network emulating ResNet [8] gave close to 100% performance within 8° error in estimating the direction of gravity in estimating the direction of gravity.

We are currently improving our model to predict the full heading direction of a moving camera given three consecutive frames. We are also using less accurate ground truth data for gravity direction but many more measurements.

References

[1] Richard Hartley and Andrew Zisserman. Multiple view geometry in computer vision. Cambridge university press, 2003.

[2] Roger Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses." IEEE Journal on Robotics and Automation 3.4 (1987): 323-344.

[3] Philip David et al. "Simultaneous pose and correspondence determination using line features." Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on. Vol. 2. IEEE, 2003.

[4] David G. Lowe. "Object recognition from local scale-invariant features." The proceedings of the seventh IEEE international Conference on Computer Vision, Vol. 2, 1999.

[5] Andrés Almansa, Agnès Desolneux and Sébastien Vamech. "Vanishing point detection without any a priori information." IEEE Transactions on Pattern Analysis and Machine Intelligence. 25.4 (2003): 502-507.

[6] A. Tai et al. "Vanishing point detection." British Machine Vision Conference BMVC 1992. Springer, London, 1992. 109-118.

[7] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in Neural Information Processing Pystems. 2012.

[8] Kaiming He et al. "Deep residual learning for image recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[9] Daniel DeTone, Tomasz Malisiewicz and Andrew Rabinovich. "Deep image homography estimation." arXiv preprint arXiv:1606.03798 (2016).

[10] A. Plakhov and P. Cruz. A stochastic approximation algorithm with step size adaptation. *Journal of Mathematics and Sciences*, *120*(1), 2004, 964–973.

[11] Vinod Nair and Geoffrey E. Hinton. "Rectified linear units improve restricted boltzmann machines." *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*. 2010.

[12] Philipp Fischer, Alexey Dosovitskiy and Thomas Brox. "Image orientation estimation with convolutional networks." German Conference on Pattern Recognition. Springer, Cham, 2015.

Live Target Detection with Deep Learning Neural Network and Unmanned Aerial Vehicle on Android Mobile Device

Ali Canberk ANAR, Erkan BOSTANCI, Mehmet Serdar GUZEL
SAAT Laboratory
Department of Computer Engineering
Ankara University
Ankara, Turkey
canberkanar@gmail.com, ebostanci@ankara.edu.tr, mguzel@ankara.edu.tr

***Abstract*—This paper describes the stages faced during the development of an Android program which obtains and decodes live images from DJI Phantom 3 Professional Drone and implements certain features of the TensorFlow Android Camera Demo application. Test runs were made and outputs of the application were noted. A lake was classified as seashore, breakwater and pier with the accuracies of 24.44%, 21.16% and 12.96% respectfully. The joystick of the UAV controller and laptop keyboard was classified with the accuracies of 19.10% and 13.96% respectfully. The laptop monitor was classified as screen, monitor and television with the accuracies of 18.77%, 14.76% and 14.00% respectfully. The computer used during the development of this study was classified as notebook and laptop with the accuracies of 20.04% and 11.68% respectfully. A tractor parked at a parking lot was classified with the accuracy of 12.88%. A group of cars in the same parking lot were classified as sports car, racer and convertible with the accuracies of 31.75%, 18.64% and 13.45% respectfully at an inference time of 851ms.**

Keywords—drone, deep learning, computer vision tensorflow

I. INTRODUCTION

Unmanned aerial vehicle (UAV) is a type of aircraft which is remotely controlled via a control unit. UAVs are currently being used in many different areas such as border patrol, noticing terrorists, destruction of targets, intelligence, fire extinguishing, traffic control, filming, cargo services, agricultural activities, prevention of smuggling, and personal use.

The main aim of this study was to create a simple system which autonomously detects certain objects and informs the user about the situation. The idea was mainly inspired from the successful operations, a military grade UAV has accomplished on national defense.

In order to get a thought on the methods that will be developed to achieve our goals, literature research regarding to pre-done studies about the topic was done. The projects about Discovery, Surveillance and Intelligence, Unmanned Air Systems Projects were found related to the subject.[1]

The image processing system of the researched project, which is written in C++ programming language, is activated automatically without any human assistance. When the UAV geographically enters the search area, the image processing system automatically starts. The image processing algorithm uses the OpenCV library, a powerful and well-known tool for image processing. The images sent to the UAV ground station are processed.

For target detection, the existence of the target object within the acquired image is checked firstly. For this purpose, noise reduction algorithms are applied to the image, converting it to HSV color space format. After the conversion, MSER Blob Detection Algorithm is applied, revealing target candidates considering their aspect ratios.

For the autonomous flight function, Auto-Pilot System Pixhawk was used. Pixhawk is an open source software which allows its' user to make changes to the source via its interface.

To put the idea into practice, the classification and detection of various live objects viewed by using an unmanned aerial vehicle was decided to be performed. In the hardware part of the study, a DJI Phantom 3 Professional UAV was used to obtain images and a Samsung Galaxy S2 tablet PC was used to control the UAV and process the images received.

Various researches have been carried out to resolve the software needs of the study. To work out the goals, it was decided to integrate the classification and detector features of the Tensorflow Android Camera Demo application with the study. For the integration process; Android Studio development environment, Android NDK toolkit, Gradle, CMake and Ndk-Build compilers were used. Android Camera Demo is an open source application shared by the Tensorflow community via Github[14].

Tensorflow is a very powerful and open source deep learning library developed by the Google Brain Team [16]. The library, written in Python and C ++ languages, supports a large variety of operating systems.[13] With Tensorflow, a computer becomes capable of object classification, object detection, voice recognition, cluster analysis and much more.

The classification feature of the Tensorflow Android Camera Demo application sorts the identified objects in the frame according to ascending prediction rate, listing the top 3 together with their precision percentages at the upper part of the screen. The Detector feature detects and marks the identified objects within the image.

Tensorflow libraries written in C++ programming language was compiled during the development of the application together with the FFMPEG library which also depends on C++

language. The FFMPEG library was used during the decoding of the frame obtained from the UAV.

DJI Mobile SDK Library was used for the communication of the developed application with the UAV. The library supports devices with Android or iOS operating systems. The developed application was registered to the DJI Developer System together with the developer information and intended use in order to gain access permit to the SDK.

The rest of the paper is structured as follows; the developed system is demonstrated at Section II and the experimental results are shown at Section III.

II. DEVELOPED SYSTEM

A. DJI SDK Library Activation

In order for the developed application to obtain image frames from the UAV, DJI Mobile SDK Library was used. The developed application was activated using the DJI API Key, to gain access to the library. To acquire the API key, a developer account was registered to the DJI Developer system, where the name of the developed application, package name and purpose of use was specified to register the application yielding the 24-digit API activation key.

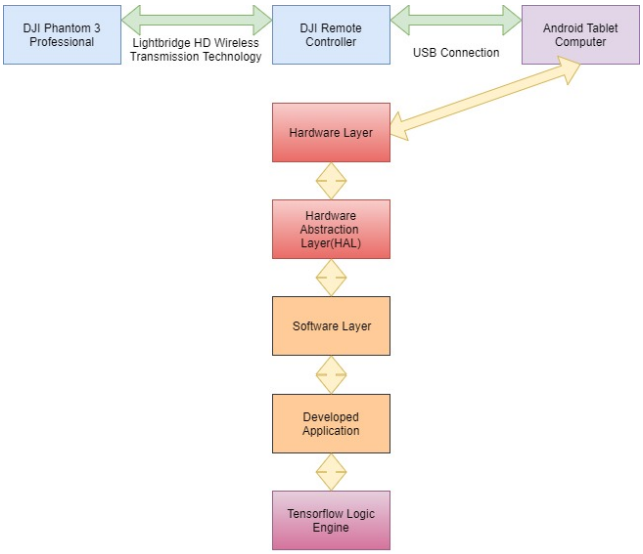


Fig. 1 Developed System's Hierarchical Block Diagram

The resulting key is then included in the software development process as a meta-data through the agency of the AndroidManifest.xml file. Library activation is performed by `ConnectionActivity` class of the 'app' module which is the main component of the study.

B. Tensorflow Library

Tensorflow library calculations are done using data flow diagrams. In this process, diagram nodes represent mathematical operations and diagram vertices represent data strings (tensors) that represent inter-node flow. The usage areas of the system are quite flexible. Image classification, object detection, sound classification, pattern recognition and pattern generation according to a specific rule are just a few of these areas of use.

Tensorflow library allows machine learning through artificial neural networks. The logic engine is trained and conclusions are made with the experience gained from the training. The input data received during application is directed to the Tensorflow layers. Neurons within these layers may be activated according to the input data. If the input image reaches the top layer by activating the neurons within lower layers, a judgment about the content of the frame can be made.

An image database pre-trained by the Tensorflow community was used in the study. This database contains images categorized according to their contents. The database, which contains at least 500 images in each category, has become a standard of image processing performance. The success of an image processing application is measured by its performance on the imagenet[15] database. There are 14.197.122 images in the database at the moment.

C. Obtaining Permissions Required from the Android Operating System

The developed application uses features of the Android device such as the internet access, location services, camera, storage read and write, screen timeout prevention and USB port access. Due to the Android operating system security procedures, the application had to be authorized to access these features. To do so, features needed were declared at the Android Manifest file. After that, the device user was prompted to grant required permissions via a popup.

D. Application Design

The visual interface of the developed application is created through the activity classes it contains. Each activity class retrieves the design of the user interface it uses from an XML file. This file defines the interface elements and their locations. The interface elements are manipulated via Java code found within the relevant activity class.

```
DJIVideoStreamDecoder.getInstance().setYuvDataListener(CameraActiv
ity.this);
```

The developed application's home page, as shown Figure 2, is implemented by `ConnectionActivity` class. This class fulfills the DJI SDK library activation, connection to the UAV control unit and the retrieval of the UAV information. The result of the library activation process is notified to the end user via a Toast message. The UAV connection information and the version of the connected UAV are displayed on the activity through textview elements. In addition to these functions, another feature of the home screen is that it provides access to other activities via buttons. Desired activity is triggered after the user presses one of the buttons.

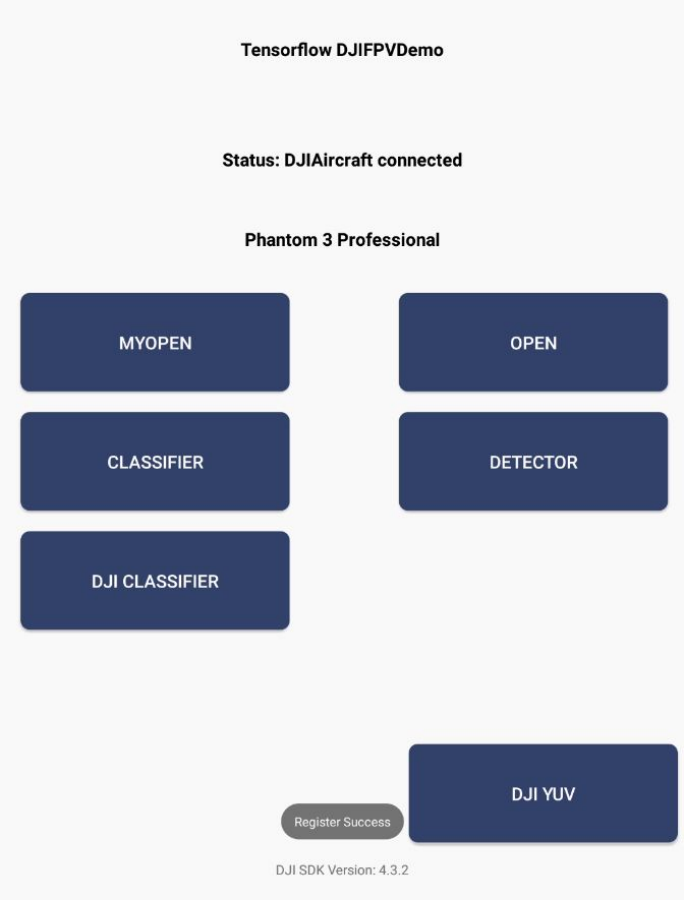


Fig. 2 Connection Activity

In the classification activity, image frames taken from the UAV is processed to classify the objects within the frame. The largest element of this activity is the textview element as can be seen in Figure 3. Through this element, images taken from the UAV is displayed to the user. The result of the image processing operation is presented to the user through the blue `RecognitionScoreView` view element located at the top of the activity. The first three predictions with the highest accuracy are listed in the `RecognitionScoreView` element.

The image, which is displayed via the textview object, is only used for enhancing the user experience. Image processing operation is done in the background, by decoding received image frames. The classification process takes at least 300 milliseconds. In the meantime, the user is presented with a dynamic screen via the `textview` element.

The details are viewed by adding a layer with invisible background over the existing activity. It is activated and deactivated by pressing the 'volume down' key of the Android device. At the lower right corner of the screen, decoded image which was sent to the Tensorflow Image Classification Function is displayed. At the lower left corner, resolution of the image frame taken from the UAV, resolution of the image sent

Fig. 3 Classifier Activity with Details Layer Activated

to the Tensorflow function, resolution of the device's screen, camera's rotation degree and the time elapsed during the processing of the last image frame are displayed. In the center of the screen, the top 10 Tensorflow functions, which occupy most of the processor time, are shown in table format with information of all other functions used.

```
// The callback for receiving the raw H264 video data for camera
                                    live view
mReceivedVideoDataCallBack = new VideoFeeder.VideoDataCallback()
                                    {

                                    @Override
    public void onReceive(byte[] videoBuffer, int size) {/*
     !!!!!!!!!!!!!!! FRAME RECEIVED !!!!!!!!!!!!!!!!!!!!!!!!*/
     //Log.i(TAG, "VideoDataCallBack videoBuffer Received of H.264
                                    TYPE");


                        if (mCodecManager != null) {
            DJIVideoStreamDecoder.getInstance().parse(videoBuffer,
 size);//sends video to decoder for YUV(MUST BE BEFORE SURFACE
                                    DECODER!!!)
                mCodecManager.sendDataToDecoder(videoBuffer,
 size);//flushes videobuffer !!MUST BE AFTER YUV DECODER PARSER!!
                                    }
                                    }
```

Fig. 4 Registering the CameraActivity class as YUV Data Listener

Classification activity inherits `CameraActivity` class. In this class, YUV image data which is broadcasted by DJI image stream decoder is captured through `IYuvDataListener`, depicted in Figure 4. After relevant format conversions, the image is sent to Tensorflow logic engine.

```
DJIVideoStreamDecoder.getInstance().setYuvDataListener(CameraActi
vity.this);
```

Fig. 5 Sending Image acquired from the UAV Camera to Decoders

Detector Activity also inherits `CameraActivity` class like the Classification activity. The difference is that instead of the Tensorflow Classify function, the Tensorflow Yolo Detector function is called to identify the objects previously taught in the database by training and to mark them in the `textview` element.

E. Image Acquisition

Within the `CameraActivity` class, an object of the `LegacyCameraConnectionFragment` class is created. The camera functions are carried out through this object. The raw H264 formatted image is captured and sent to both `DJIVideoStreamDecoder` and `DJICodecManager` [12] decoder. `DJIVideoStreamDecoder` broadcasts image in YUV data format and `DJICodecManager` feeds the `textview` element used to present the image to the user.

F. Image Processing

The `CameraActivity` class captures the YUV data broadcasted by the `DJIVideoStreamDecoder` via the `onYuvDataReceived` function of the `IYuvDataListener` event listener. First of all, the function checks whether a previously started image processing process is continuing. If the processing of the previous image frame is not completed, the function is terminated via the return command and the current YUV data is dropped as shown in Figure 6.

```
@Override
public void onYuvDataReceived(final byte[] yuvFrame, int width,
int height) {
/*  Log.d(TAG,
          "onYuvDataReceived: frame index: "
                  +
DJIVideoStreamDecoder.getInstance().frameIndex);*/

  if (isProcessingFrame) {
    LOGGER.w("Dropping frame!");
    return;
  }


  final byte[] bytes = YuvToYuv420Sp(yuvFrame, width, height);
```

Fig. 6 State Control of Image Processing and Triggering of Format Transformation

As YUV data are received, it is checked whether an image processing event is occurring or not. If not, the image obtained in YUV data format via the listener is converted to ARGB8888 image format. After that, image is sent to the Tensorflow logic engine. The function finally indicates that there is data ready to be processed to the Tensorflow logic engine, by triggering the `processImage` function. Doing so, the processing of the image frame begins.

III. EXPERIMENTAL RESULTS

After the application was developed, a test run was made at Ankara University Computer Engineering Department's parking lot. Objects such as tractor, lake, pier and a variety of vehicle types (racer, sports car, convertible and etc.) were classified by the classifier activity. Vehicles were also detected and marked by the detector activity of the developed application. For testing purposes, a few cat pictures, toy taxi, keyboard laptop computer, UAV controller's joystick were displayed to the UAV's camera manually and all of the so mentioned were spotted by the developed application.

TABLE I. CLASSIFICATION ACCURACY TEST

Object	**Accuracy**
Joystick	19.04%
Laptop	11.68%
Laptop Monitor	14.76%
Tractor	12.88%
Sports Car	28.74%
Racer	27.16%
Convertible	13.45%
Cars in General	23.17%
Cat	22.79%
Seashore	24.44%

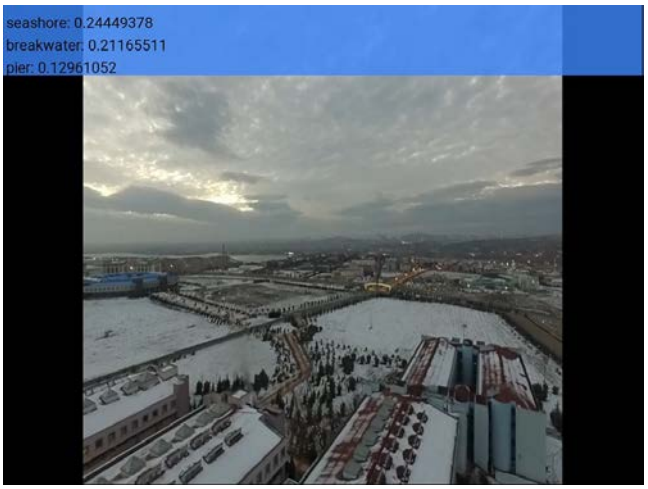


Fig. 7 Classification Activity Sea Edge Detection (UAV Hovering Over Ankara University Computer Engineering Department)

Fig. 8 Classification Activity Vehicle Detection (UAV Hovering Over Ankara University Computer Engineering Department Parking Lot)

Fig. 9 Classification Activity Vehicle Detection with Details Layer Activated (UAV Hovering Over Ankara University Computer Engineering Department Parking Lot)

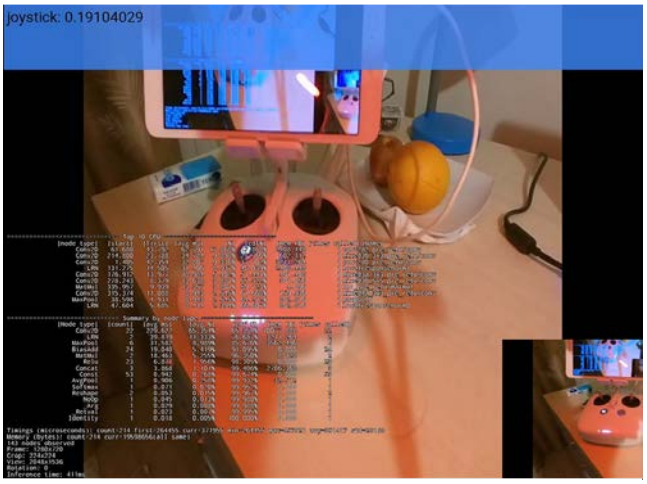


Fig. 10 Classification Activity Controller Unit's Joystick Detection

IV. CONCLUSION

The intention of this study was to develop an Android application, which analyzes and interprets objects within the image frames, acquired by DJI Phantom 3 Professional UAV.

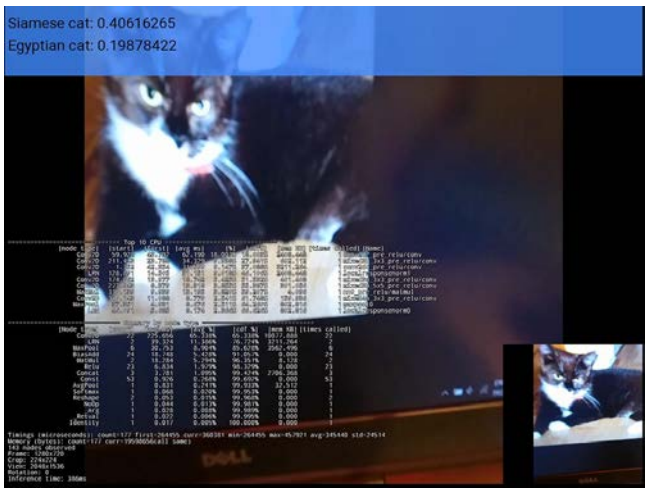


Fig. 11 Classification Activity Cat Detection

Fig. 12 Classification Activity Notebook Computer Detection

Various researches have been carried out on the subject and it was decided to use Artificial Neural Networks trained by Machine Learning for the analysis and interpretation of the images. To do so, the Tensorflow sample Android application which includes all the necessary functions and libraries was used during application development.

Integration of Tensorflow sample Android application with DJI Phantom 3 Professional UAV was made in the scope of this study. The pre-trained Inception-V3 model was used for classification and Mobilenet model was used for detection of objects. As a result, an Android application that acquires image frames from the UAV, classifying objects up to 1000 categories and detecting up to 171 different types of objects was discovered.

The developed application is planned to be improved during future studies. First of all, it is aimed to reconfigure the process of transmitting the acquired image to the Tensorflow logic engine, reducing the number of image format conversions. After that, the calibration problem occurring during marking of the detection objects will be solved and the DJI application widgets will be added to the application. Finally after these objectives are satisfied, certain objects will

be marked with an attached laser pointer after being detected by the system.

REFERENCES

[1] Öznalbant Z. Filik T. Gerek Ö.N. "KGİ(Keşif, Gözetleme ve İstihbarat) İHS(İnsansız Hava Sistemleri) Projeleri Kapsamında Anadolu Üniversitesinde Yapılan Çalışmalar". IX. TURKISH NATIONAL AIRCRAFT AEROSPACE ENGINEERING CONGRESS PAPERS, 168-171

[2] Abadi, Barham, Chen, Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16).2016. TensorFlow: A System for Large-Scale Machine Learning.; 265-278

[3] Goldsborough P., A Tour of TensorFlow, 2016.

[4] Abeywickrama S., Jayasinghe L., Fu H., Yuen C., ICASSP 2018. RF-based Direction Finding of Small Unmanned Aerial Vehicles Using Deep Neural Networks

[5] Baylor D., Breck E., Cheng H., Fiedel N., Foo C. Y., Haque Z., Haykal S., Ispir M., Jain V., Koc L., Koo C. Y., Lew L., Mewald C., Modi A. N., Polyzotis N., Ramesh S., Roy S., Whang S. E., Wicke M., Wilkiewicz J., Zhang X., Zinkevich M., GOOGLE INC., KDD 2017 Applied Data Science, TFX: A TensorFlow-Based Production-Scale Machine Learning Platform

[6] Kovalev V., Kalinovsky A., Kovalev S., Minsk: Publishing Center of BSU PATTERN RECOGNITION AND INFORMATION PROCESSING (PRIP'2016). Deep Learning with Theano, Torch, Caffe, Tensorflow, and Deeplearning4J: Which One is the Best in Speed and Accuracy?

[7] Howard A. G., Zhu M., Chen B., Kalenichenko D., Wang W., Weyand T., Andreetto M., Adam H., GOOGLE INC. 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

[8] Ivanova D., Kadurin V., Proceedings of the International Conference on Information Technologies (InfoTech-2017). Mobile Application For Pattern Recognition Based On Machine Learning

[9] Sheppard, Rahnemoonfar, School of Engineering and Computing Sciences, Texas A&M University-Corpus Christi (Access Year: 2017). Real-time Scene Understanding for UAV Imagery based on Deep Convolutional Neural Networks

[10] Yoo J., Hong Y., Yoon S. (Access Year: 2017). Autonomous UAV Navigation with Domain Adaptation

[11] "CS 20SI: Tensorflow for Deep Learning Research", access time 30 December 2017, https://web.stanford.edu/class/cs20si/

[12] "DJICodecManager", access time 30 December 2017, https://developer.dji.com/iframe/mobile-sdk-doc/android/reference/dji/sdk/Codec/DJICodecManager.html

[13] "Top Five Use Cases Of Tensorflow", access time 30 December 2017, https://www.exastax.com/deep-learning/top-five-use-cases-of-tensorflow/

[14] "TensorFlow", access time 30 December 2017, https://github.com/tensorflow/tensorflow

[15] "ImageNet ", access time 30 December 2017, http://image-net.org/index

[16] "Google Brain Team", ", access time 21 December 2017, https://research.google.com/teams/brain/

Deep Learning Based Vehicle Detection on Cross-Roads

M. GENÇER[1] and N. AYDIN ATASOY[1]

[1] Karabuk University, Karabuk/Turkey, muratgencer1@hotmail.com
[1]Karabuk University, Karabuk/Turkey, nesrinaydin@karabuk.com

Abstract **- Although the number of lanes of roads in residential cities is constant, the number of vehicle in traffic is increasing every year. This causes traffic congestion at certain times, such as before and after work hours. Thus, instead of traditional methods, intelligent systems have become a necessity to control traffic lights. For this problem, there are traffic signaling applications developed with using image processing and artificial intelligence techniques in literature. In this study, an application was developed to provide more detailed data for traffic signaling applications. Used Faster R-CNN model was trained and tested in Karabük and trained model detected 76 of 79 vehicles in 23 test frames and achieved 96% success.**

Keywords **– Deep learning, convolutional neural networks, cntk, vehicle detection.**

I. INTRODUCTION

INTELLIGENT traffic systems (ITS) have been developed for effective time management, reducing traffic congestion, and control traffic on roads and informing people, such as drivers. ITS's are commonly used for controlling traffic of cross roads [1]. As seen in Figure 1, common areas formed by the intersection or combination of two or more roads are called "cross roads" [2]. In traditional systems, management of traffic lights depends on constant time-loop and traffic congestion increases in some directions at different times. It causes time and energy loss. If traffic lights are controlled by intelligent systems, time and energy loss can be minimized. For traffic congestion management, there are developed applications by using image processing and artificial intelligence techniques in literature [3].

Deep learning has shown its first major impact in the area of object recognition with the convolutional neural network model in 2012 ImageNet competition [4]. Actually, artificial neural network architectures with multiple hidden layers are called deep neural networks [5]. Nowadays, deep learning algorithms and applications are developing with increasing data and computational performance. Thus, deep learning algorithms are used in many areas, from voice recognition, object detection, natural language processing, robotics, virtual reality to autonomous vehicles, and the obtained results are successful.

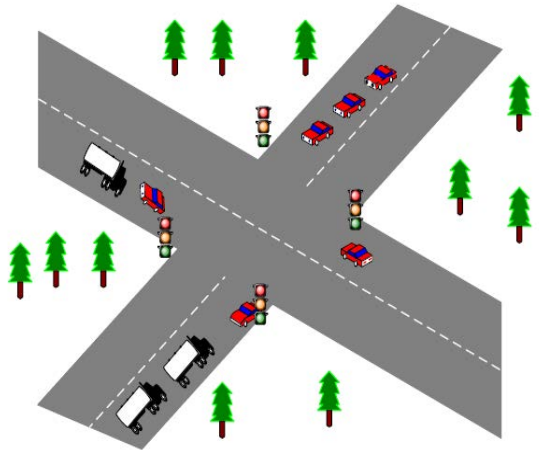


Figure 1: Representational example of a cross road.

Deep learning is the most popular approach to develop artificial intelligence for machines which can sense and understand the real-world [6]. Deep learning structure explores the complexity of large data sets using a backpropagation algorithm to show how parameters between layers must change [7].

In this study, an application was developed to provide more detailed data to ITS. Thus, which type of vehicles cause traffic congestion at different hours, can be detected. The obtained results are useful information for intelligent traffic light control systems. The developed application was tested at cross road of 1002. and 1003. streets 100. Yil in Karabük. As a result, the trained model detected 76 of 79 vehicles in 23 test frames and achieved 96% success.

II. METHOD

There are 3 basic deep neural network models, Fully Connected Neural Network (FCN), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). To develop application with using the models, open-source libraries are existed. Caffe – UC Berkeley, CNTK – Microsoft, TensorFlow – Google, Torch – MXNet are the some of most common libraries with developer groups respectively [8]. CNN is powerful model for image classification and object detection [9]. CNN is effective for computation and it includes convolution and pooling layers. Feature mapping extraction is performed at convolution layer and size reduction is performed

at pooling layer [10].

In this study, one of the CNN architecture of AlexNet and region-based CNN (Faster R-CNN) structure are preferred. Faster R-CNN, achieves approximately real-time rates, if the time spent in regional proposals is neglected [11]. Additionally, it provided the best results for car and bus detection in object detection category in Pattern Analysis, Statistical Modelling and Computational Learning Visual Object Classes Challenge 2012 (PASCAL VOC2012) [12].

III. Application

In all experiments in this study AMD 1100T CPU, 8GB 1333 MHz memory, Nvidia 1050ti with 4GB memory graphic card are used as hardware.

A. *Preparing Development Environment and Data*

CNTK (Cognitive Toolkit), is an artificial intelligence development kit which developed by Microsoft and made available to developers as open source on April 2015 [13]. It can run on Linux and Windows. We preferred Windows version and followed documentation [14]. Setup files are extracted in new directory named "local" in main directory and setup was done with "install.bat" file which located in "local/cntk/Scripts/install/Windows". Python and other libraries are installed with completed setup.

PyCharm ide is setup to develop software easily [15] and to use Python version which comes with CNTK, related paths must add to PyCharm ide.

Images must be labelled to use Faster R-CNN model. Labelling process was performed with Visual Object Tagging Toolkit (VOTT) [16] which is developed by Microsoft. VOTT separates the images in 3 categories as training data, test data and unused data. It reserves 20% sample of the tagged frames randomly as a test dataset. If any vehicle is not labelled in a frame, it is unused data. After process is completed, it generates an output file that includes label names and locations of objects. The generated file was moved "DataSets" directory in CNTK setup directory. The performed processes are given in Figure 2.

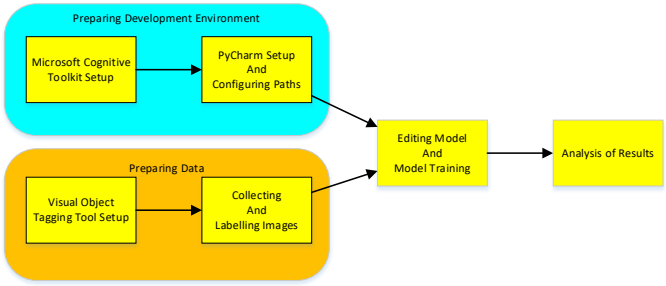


Figure 2: General flowchart of the application.

B. *Creating Model*

Example model of Faster R-CNN which comes with CNTK, was edited for our dataset. Configuration file was created and located in "configs" directory. For example, computation is performed according to working time, size of dataset and labelled dataset. Model file, detected vehicles in test data and detection ratio are given as output files.

IV. Developing Graphical User Interface

Graphical user interface (GUI) given in Figure 3, was developed with using PyQt library [17] for model creation, model selection, and inference processes.

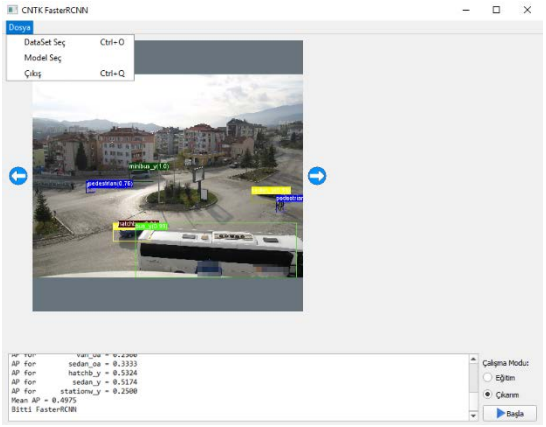


Figure 3: Developed graphical user interface.

Report file which is created after inference process is shown in Figure 4. The first 4 columns show coordinates of detected vehicles in current frame and the last column shows type of the vehicle.

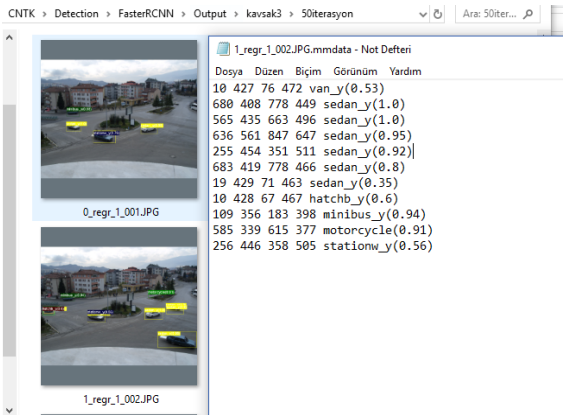


Figure 4: Content of results file.

A report file is generated for each frame of the video image.

In following studies, GUI will be improved, and parameters (i.e. number of iteration, path of report files etc.) will be set on the GUI.

V. Results and Discussions

In this study, 336 frames were used. 313 of them are used for training, 23 of them were used for testing. There were total 1571 labelled example used. Training process was run for 50 iterations and the loss function results shown in Figure 5. Total loss consists of region proposal network regression loss, detector regression loss [18].

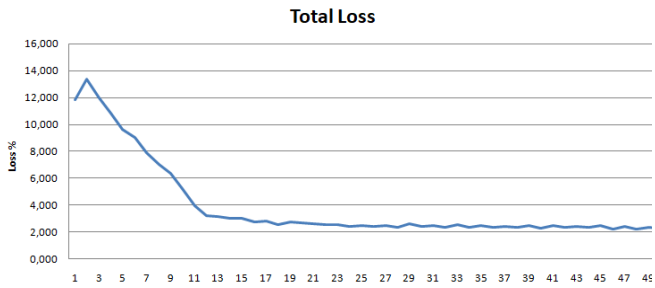


Figure 5: Total loss results.

We tried to run training for 20 iterations, but the obtained results were not as well as 50 iterations. Figure 6 shows comparison of result of 20 and 50 iterations. 50 iterations need more time but achieve better labelling. The trained model detected 76 of 79 vehicles in 23 test frames and achieved 96% success. The success rate of vehicle type detection was 49% average, as shown in Figure 7. The success rate can be increase with more labelled examples.

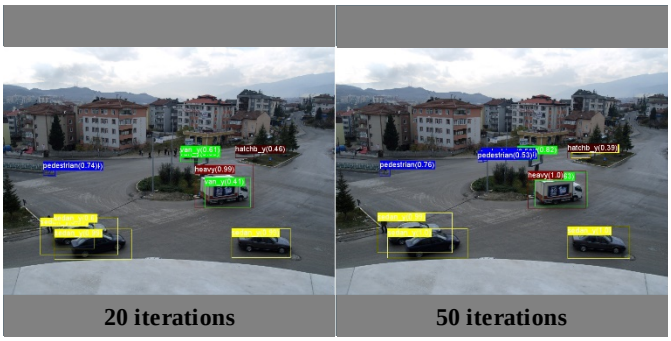


Figure 6: Comparison of 20 and 50 iteration success.

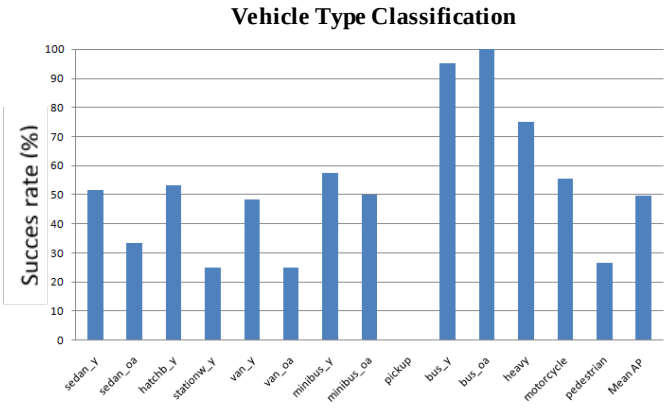


Figure 7: Success rate of vehicle type detection.

If lightning is enough good, model can work at night images as seen in Figure 8. It is expected that the results will be better when model night images are trained. Studying on night images is continued.

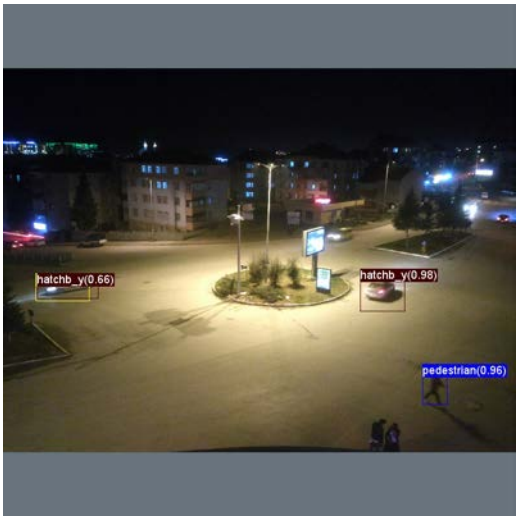


Figure 8: Vehicle detection at night images.

In this study, vehicles were detected on each frame. In future works, vehicle coordinates in different report files will be associated. Thus, flow traffic and traffic congestion can be detected. After that, management of traffic lights can be organized according to the model results.

REFERENCES

[1] Hamza Şamil Dağüstü, "Trafik Yönetiminde Kavşak Trafiğinin Kontrolü İçin Bir Sinyal Zamanlama Modeli", *Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi*, Sf. 7, İstanbul, 2010.

[2]Internet: http://www.trafikkurallari.net/kavsak_nedir.php, avaliable at: 25.03.2018

[3] O.A. Erdem, "Kavşak trafik sinyalizasyon sistemi için bulanık mantık tabanlı gerçek zamanlı denetleyici tasarımı ve uygulaması." *e-Journal of New World Sciences Academy, Natural and Applied Sciences*, vol 2 (4), pp. 241-255, 2007.

[4] Özkan İnik, Erkan Ülker, Derin Öğrenme ve Görüntü Analizinde Kullanılan Derin Öğrenme Modelleri, *Gaziosmanpaşa Bilimsel Araştırma Dergisi (GBAD)*, Vol:6,Num:3,Pages: 85-104,2017.

[5]Doğa Barış Özdemir, Ahmet Cumhur Kınacı, Görüntülerdeki Araba Nesnelerinin Belirlenmesi İçin Derin Öğrenme ile Bir Model Eğitilmesi, 20. *Akademik Bilişim Konferansı*, Karabük, Türkiye.

[6] Ferhat Kurt, Derin Öğrenme Nedir?, *Akademik Bilişim 2016* – Adnan Menderes Üniversitesi.

[7] Yann LeCun, Yoshua Bengio & Geoffrey Hinton,Deep learning, *NATURE*, 28 May 2015,Vol:521, Pages: 436-444.

[8] Shaohuai Shi, Qiang Wang, Pengfei Xu, Xiaowen Chu,Benchmarking State-of-the-Art Deep Learning Software Tools, *Cloud Computing and Big Data (CCBD), 2016 7th International Conference on*,16-18 Nov. 2016,IEEE, Macau, China.

[9] Andrej Karpathy Li Fei-Fei, Deep Visual-Semantic Alignments for Generating Image Descriptions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 39, Issue: 4, Page(s):664 - 676 April 1, 2017.

[10] Internet: http://ufldl.stanford.edu/tutorial/supervised/Pooling/, avaliable at: 25.03.2018

[11] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 39, Issue: 6, Page(s): 1137 - 1149, June 1, 2017.

[12] Internet: http://host.robots.ox.ac.uk:8080/leaderboard/displaylb.php?challengeid=11&compid=4, avaliable at: 25.03.2018

[13] Internet: https://github.com/Microsoft/CNTK, avaliable at: 25.03.2018

[14] https://docs.microsoft.com/en-us/cognitive-toolkit/setup-windows-binary-script, avaliable at: 25.03.2018

[15] Internet: https://www.jetbrains.com/pycharm/, avaliable at: 25.03.2018

[16] Internet: https://github.com/Microsoft/VoTT, avaliable at: 25.03.2018

[17] Internet: https://wiki.python.org/moin/PyQt, avaliable at: 25.03.2018

[18] Internet: https://docs.microsoft.com/en-us/cognitive-toolkit/object-detection-using-faster-r-cnn, avaliable at: 25.03.2018

An Empirical Comparison of Data Mining Tools and Migrating Birds Optimization Algorithm on Medical Diagnosis

Muyesaier AIHAITI[1], Ali Fuat ALKAYA[2] and Ramazan ALGIN[3]

[1]Marmara University Göztepe. Istanbul. Turkey, muyassar1.9@gmail.com
[2]Marmara University Göztepe. Istanbul. Turkey, falkaya@marmara.edu.tr
[3]Florida International University Miami. Florida. USA, ralgin@fiu.edu

Abstract **- Today, breast cancer is the most common non-skin cancer affecting women. Because mammograms do not give clear findings until a certain age for the diagnosis of breast cancer, ultrasounds are applied as an imaging technique. Then, according to the findings in ultrasounds, imaging techniques like mammograms or other detection techniques are applied. Patient will be required mammograms if ultrasound findings look suspicious or the patient is above a certain age. Patients will have biopsy according to the mammogram findings and other risk factor values. Each step in this process will lead the patient anxiety, and expenditures. But the results of not performing these processes will lead one case of cancer and leads the patient more serious problems and expenditures. Therefore, every step of the process is vital importance to predict the most accurate information available. Computer-aided diagnostics (CADx) models are helpful in working on a huge number of variables, and related the risk factors and risk estimation together. Many CADx models provide support for experts by evaluating the mammographic findings and are used by radiologists to increase the detection rate of missed cancer patients. In this study, application of well-known data mining techniques is done on a public breast cancer data set. As a novelty, Migrating Birds Optimization (MBO) which is a fairly new metaheuristic is also applied on the same data set and the results are discussed. It is shown empirically that the MBO meta-heuristic presents preferable performance to the well-known data mining techniques.**

Keywords **– BC, CADX, MBO, DATA MINING, META HEURISTICS**

I. INTRODUCTION

CURRENTLY, breast cancer is the leading non-skin cancer type causing death among women population all over the world. Breast cancer (BC) can be described as the cells growing and acting abnormally in the texture of the breast that hard to control.

According to statistics about 231,840 new cases of invasive breast cancer are diagnosed among women in the US during 2015 and about 2,350 new cases men. BC is the most frequently diagnosed cancer among women (when skin cancer is excluded) [1].

Early detection of tumor cells in the breast is vital for the treatment and life quality of the patients. There are many clinical tests and medical imaging techniques for diagnosis of BC. Among them, mammography, X-ray photography of the breast taken using a low-dose X-ray machine, is the most common imaging technique used for diagnosing BC. However, mammography has some limitations. Mammograms do not give clear findings until a certain age. In such cases, ultrasounds are applied as a supportive imaging technique. Then, according to the findings in ultrasounds, imaging techniques like mammography may be applied if ultrasound findings look suspicious. Patients will have biopsy according to the mammogram findings and other risk factor values. Each step in this diagnosis process will lead the patient both emotional and financial costs. However, early diagnosis of BC is essential for the treatment. Therefore, every step of the diagnosis process has vital importance in gathering information for the prediction of whether the patient has BC or not.

Although clinical prediction is the most important step in the diagnosis of BC, evaluating the related information and the findings from the imaging techniques is a very though and challenging task for radiologists who must balance the relative contributions of numerous risk factors, especially in some cases. Radiologists often predict the presence of a BC by using probabilities calculated with heuristic methods on the basis of their training and experience. Albeit the valuable assistance and necessity of these heuristic methods, they can lead to systematic errors [2] in some cases due to their biased nature. Moreover, estimation of breast cancer must put in to account combination of more than 100 risk factors to diagnose a single test which is very difficult task if only rely on experience. Nevertheless, accurate prediction of clinical outcomes is important to successful decision making and can lead to better patient care. Accordingly, assistance of computer-based decision support models becomes very valuable in solving this complex problem. In the literature, current computer models used in the area of breast-cancer diagnosis can be categorized into three groups: i) prognostic, ii) computer-aided detection (CAD) and iii) computer-aided diagnostics (CADx) models. Prognostic models use retrospective risk factors and clinical information to predict breast cancer risk [3] during a time interval for future treatment or decisions. CAD models [4] are used to locate and identify abnormalities in radiologic images, leaving the interpretation of them to the radiologists. CADx models [5] are used by radiologists to classify the case as benign or malignant

by producing a probability that shows the possibility of having BC using the information about the findings detected and characterized on a mammogram or some other imaging technique either by radiologists or CAD systems.

Computer models can provide assistance and convenient in processing a large number of variables, and bridging the gap between risk estimation and risk factors. For an example, computer models can be used by radiologist to increase the accuracy of mammography inspections in detection and diagnosis [6].

The aim of this study is developing a CADx model to get accurate diagnostics for BC. Though most of the CADx models in the literature are developed by using Artificial Neural Network (ANN), logistic regression (LR) and decision tree (DT), meta-heuristics Migrating Birds Optimization (MBO) [7] were rarely applied to this domain. In this study ANN, LR and decision tree models are built as CADx models and tested. Furthermore, MBO is also used to build classification CADx models as a novelty. The dataset used in this study is the Wisconsin Breast Cancer dataset (WBCD), derived from the University of California Irvine (UCI) machine learning data repository. The performance of the meta-heuristic model is compared with the traditional data mining models built in this study and in the literature. It is shown that the meta-heuristic model show better performance.

The remaining part of the paper is organized as follows: Section 2 gives a brief literature survey on the diagnosis of breast cancer. Section 3 gives the details of the dataset used and the modeling phase. Section 4 shows the results and includes a discussion about the findings of the study. Section 5 concludes the study.

II. PREVIOUS STUDIES ON THE DIAGNOSIS OF BC

Data mining techniques are wide range used to the early diagnosis of breast cancer besides clinical tests. When the literature is surveyed, it can be seen that data mining techniques have been used for the diagnosis of BC shows good performance. One of the most common data mining technique for the diagnosis of BC is Artificial Neural Networks (ANN) due to their nonlinear nature. ANN is a modeling techniques that inspired from the human brains structure. ANN can be used in modeling any complex pattern, so they are well suited for diagnostics of BC problem [8]. A neuron is the basic component of a neural network. It gets inputs, operate inputs (sums them and applies a (usually nonlinear) transfer function) and then generates the outcomes which can be input of other neurons or a model prediction. In a neural network neurons are connected to each other systematically. Another technique is Logistic Regression is used for predicting binomial or multinomial results by deciding the contribution of multiple factors to the outcomes. Logistic regression is a member of generalized linear models. This technique is used to predict the risk of breast cancer and also helps managing the huge amount of information to make better decisions for cancer detection [9]. Decision Tree is also a popular technique used for BC diagnostics. In the literature there are many decision tree algorithms like: C&RT, ID3 and C5.0 methods use impurity measures to choose the splitting attribute and the split value/s. As an impurity measurement, C&RT [14] uses Gini index, ID3 [13] uses information gain and the successor, C5.0 uses gain ratio. Unlikely, CHAID uses chi-square and F statistics to choose the splitting variable [14].

In the literature, there are also some meta-heuristics exploitations to train ANNs which are used for diagnosing BC. In one of those, MBO is both successful and has robust characteristics in ANN training by using its unique benefit mechanism [10].

III. MODELING APPROACHES USED

Using real data for BC diagnostic problem is very important. In this paper we used the same data set to assess the performance of different classic and meta-heuristic classification models. The data set used here is the Wisconsin Breast Cancer data set [11]. Originally, this data set had 699 records but some of them include missing values. We used this data set after eliminating the records with missing values and id number attribute. After this elimination, there are 683 records each having ten attributes in this data set. Table 1 gives the attribute names and domains.

Table 1: ATTRIBUTES OF DATA SET

Attribute Name	Domain
Clump thickness	1-10
Uniformity of cell size	1-10
Uniformity of cell shape	1-10
Marginal adhesion	1-10
Single epithelial cell size	1-10
Bare nuclei	1-10
Bland chromatin	1-10
Normal nucleoli	1-10
Mitoses	1-10
Class attribute	Benign, malignant

Every sample has nine different attributes; they used to measure the changing with nine different attributes. The nine attributes are numbered from 1 (normal state) to 10 (the most abnormal state). The class attribute is defined as 2 for benign and 4 for malignant diagnose. Therefore, the given information can be stored in a two dimensional matrix where each row stores a patient and each column stores an attribute data.

We use data mining classification methods like Artificial Neural Network (ANN) with the attributes given in Table 2. ANN has six different kinds of methods namely; Dynamic, RBFN, Multiple, Quick, Prune and Exhaustive Prune methods claimed in Clementine 12.0. Radial basis function network (RBFN) is a unique neural network that contains three layers namely; input layer, hidden layer and output layer. The input and output layers have similarity to those multilayer perceptron.

A single neural network is trained, when the quick method is chosen to make model. While training the network the back-propagation method is used. In a case where dynamic method is chosen, during the training process by increasing number of neurons till aimed accuracy is achieved to increase the

performance, the topology of the network is changed. There are two steps of training dynamic method: i) finding the topology and ii) training the final network. When the multiple methods are chosen, pseudo-parallel mode is used in training. Initialization of each specific network is done and all neural network are trained. The prune method is the opposite of the dynamic method. The prune method begins with large network and prunes useless neurons from the input and hidden layers.

Table 2: ATTRIBUTES OF ANN

Model		Quick	Prune	Multiple	RFBN
Model Parameters	Alpha	0.9	0.9	0.9	0.9
	Initial Eta	0.3	0.3	0.3	20
	High Eta	0.1	0.1	0.1	0.4
	Low Eta	0.01	0.01	0.01	1.0
	Eta Decay	30	30	30	
	Layer 1	20	20	20	
	Persistent	200	200	200	30
	Sample%	50%	50%	50%	50%

The decision tree model used is the C5.0 decision tree algorithm for which the impurity measure is Entropy, pruning severity is 75% and tree depth is 7.

Logistic regression is a well-established statistical method for predicting binomial or multinomial results by determining the contribution of multiple factors to the outcomes. In our LR model the usage of partitioned data, calculating variable importance and calculating raw propensity scores are set to false, multinomial procedure, constants to be included in equation as true are selected.

Those models mentioned above were designed on IBM SPSS clementine 12.0 environment to get CADX classification models.

In addition to the aforementioned methods, we also exploited the MBO algorithm to predict BC diagnosis. The MBO algorithm is a neighborhood search method. The algorithm starts with a number of initial solutions corresponding hypothetically to birds in a "V" formation. In this algorithm, each solution is tried to be improved by its neighbor solutions by starting with the first solution (corresponding to the leader bird), and progressing on the lines towards the tails. In our implementation, we define a solution vector which stores the weight of each variable on the forecast. The elements of the x vector are limited in the range [-1,1] for calculation convenience. Again in our implementation, a neighbor is obtained by adding a random value within [-*radius*, *radius*] to each weight where *radius* is a new parameter to MBO tailored for BC diagnosis. If the new weight value exceeds the bounds, the new value is set to the nearest bound. In the MBO algorithm, if an improvement is brought by the best neighbor bird (solution) then the current solution is replaced by that one. This process also brings a benefit mechanism which is shortly sharing the best unused neighbors with the birds that follow. Once all of solutions are improved (or tried to be improved) by neighbor solutions, this operation is repeated a number of times (tours) after which the first bird (leader bird) becomes the last, and one of the second birds moves the first place and another loop starts. In our implementation the algorithm is stopped after 100000 solutions (including neighbor solutions) are created.

In the MBO, there are basically four parameters (*nob*, *non*, *nof*, *olf*) but inherent to the tackled problem *radius* and *threshold* parameters are also included to the design. In an extensive set of computational experiments, we tried the values given in Table 3 where the best ones are shown in bold for each variable.

Table 3: PARAMETERS OF MBO

Parameters	Values
nob	{**5**, 11, 21, 51, 101}
non	{**3**, 5, 7}
nof	{5, **10**, 20}
olf	{**1**, 2, 3}
threshold	{0.5, 1, 1.5, 2, 2.5, …, **12,5**, … , 89,5}
radius	{0.01, 0.02, 0.03, 0.04, **0.05**}

Those methods implemented in the Java language. In its design, MBO is implemented to minimize the misclassification cost which is given in Equation (1).

$$\text{Cost} = 1-(TP+TN)/\ (TP+FP+TN+FN) \qquad (1)$$

where

- TP-True positive, these are cases in which we predicted yes (they have the disease), and they do have the disease.
- FP- False positive, we predicted yes, but they don't have the disease
- TN- True negative, we predicted no, and they don't have the disease.
- FN- False negative, we predicted no, but they do have the disease.

IV. RESULTS AND DISCUSSIONS

In this study, SPSS classic modeling techniques' performance is compared with the recently proposed MBO algorithm. Data Mining and heuristic modeling techniques are used in different encoding data types like temperature encoding and binary encoding data types. We can get those data from the original decimal dataset. The performance of each model is getting from confusion matrix which is popular in the area of machine learning and particularly the problem of statistical classification. Confusion matrix [12] contains information about actual and predicted classifications generated by a classification system. Performance of such systems is commonly measured using the data in the matrix. The following table (Table 4) shows the confusion matrix for two class classifiers. Accuracy is the proportion of the total number of predictions that were correct. It is determined using the following equation.

$$\text{Accuracy} = (TP+TN)\ /\ (TP+FP+TN+FN) \qquad (2)$$

Table 4: THE CONFUSION MATRIX

		predicted	
		positive	negative
actual	positive	TP	FP
	negative	FN	TN

Looking at the results on Table 5, it is clearly seen that the MBO is highly successful in modeling and diagnosis. On the other hand, most ANN versions produced good accuracy results for the data sets, but it could not outperform the MBO. Considering the accuracy, the MBO algorithm shows the best and the LR algorithm shows the worst performances. A support system related to medical decision-making was proposed in this study, using the ANN, LR, DT models and MBO optimization algorithms collectively for the diagnosis of BC.

Table 5: PERFORMANCE OF EXPLOITED TECHNIQUES

Model name	Temperature Data Set	Binary Data Set
LR	89.82%	89.82%
DT	95.21%	93.41%
ANN	94.61%	97.01%
ANN1	97.01%	95.81%
ANN2	93.41%	95.81%
ANN3	95.81%	95.81%
ANN4	97.01%	96.41
ANN5	95.21%	95.21
ANN6	96.41%	95.81%
ANN7	96.41%	96.41%
ANN8	97.01%	96.41%
ANN9	97.01%	97.01
ANN10	97.01%	96.41%
ANN11	95.21%	96.41%
MBO	97.30%	98.2%

V. CONCLUSION

In this paper, the performance of a recently proposed metaheuristic on a medical diagnosis problem is presented by comparing with the performance of classical data mining tools. The compared data mining tools are ANN (with several versions), DT and LR. Comparison study is conducted through an extensive set of computational experiments using a benchmark public data. Even though MBO algorithm is a recently introduced metaheuristic algorithm, its performance outperforms the other aforementioned tools. In figures, MBO is at least 1.1% and up to 8.3% better than the other tools. This shows that MBO can also be used in support systems related to medical decision making problems. Therefore, a further study may include application of MBO to other diagnosis problem cases.

REFERENCES

[1] 2American Cancer Society (2015). Cancer Facts & Figures 2015. Atlanta: American Cancer Society;

[2] Akay, M. F. (2009). Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications, 36(2), 3240-3247.

[3] Bornstein, B. H., & Emler, A. C. (2001). Rationality in medical decision making: a review of the literature on doctors' decision-making biases. Journal of evaluation in clinical practice, 7(2), 97-107.

[4] Noble, M., Bruening, W., Uhl, S., & Schoelles, K. (2009). Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. Archives of gynecology and obstetrics, 279(6), 881-890.

[5] Baker, J. A., Kornguth, P. J., & Floyd Jr, C. E. (1996). Breast imaging reporting and data system standardized mammography lexicon: Observer variability in lesion description. AJR. American journal of roentgenology, 166(4), 773-778.

[6] Burnside, E. S., Rubin, D. L., Fine, J. P., Shachter, R. D., Sisney, G. A., & Leung, W. K. (2006). Bayesian network to predict breast cancer risk of mammographic microcalcifications and reduce number of benign biopsy results: initial experience. Radiology, 240(3), 666-673.

[7] Duman, E., Uysal, M., & Alkaya, A. F. (2012). Migrating Birds Optimization: A new metaheuristic approach and its performance on quadratic assignment problem. Information Sciences, 217, 65-77.

[8] Ayer, T., Chhatwal, J., Alagoz, O., Kahn Jr, C. E., Woods, R. W., & Burnside, E. S. (2010). Comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radiographics, 30(1), 13-22.

[9] Chhatwal, J., Alagoz, O., Lindstrom, M. J., Kahn Jr, C. E., Shaffer, K. A., & Burnside, E. S. (2009). A logistic regression model based on the national mammography database format to aid breast cancer diagnosis. American Journal of Roentgenology, 192(4), 1117-1127.

[10] Makas, H., & Yumusak, N. (2013, November). A comprehensive study on thyroid diagnosis by neural networks and swarm intelligence. In Electronics, Computer and Computation (ICECCO), 2013 International Conference on (pp. 180-183). IEEE.

[11] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[12] Kohavi, R. (1998). Glossary of terms. Machine Learning, 30, 271-274..

[13] Sahin, Y., Bulkan, S., & Duman, E. (2013). A cost-sensitive decision tree approach for fraud detection. Expert Systems with Applications, 40(15), 5916-5923.

[14] Wheeler, R., & Aitken, S. (2000). Multiple algorithms for fraud detection. Knowledge-Based Systems, 13(2-3), 93-99.

A New IPv6 Addressing Strategy to Mitigate Reconnaissance Attacks

S. A. ABDULLAH[1]

[1]Sultan Qaboos University, Muscat, Oman, shubair@squ.edu.om

Abstract- It has been widely assumed by the research community that the network reconnaissance attacks in IPv6 networks are unfeasible because they would take tremendous effort to perform address scanning of 2^{64} hosts in an IPv6 subnet. However, recent research has revealed feasibility of these attacks by investigating a number of native IPv6 networks. The research concluded that an intelligent attacker could easily reduce the target search space by predicting the network host addressing schemes when performing the scanning. This indeed enhances security concerns and undermines the chances of IPv6 being deployed. This paper overviews the IPv6 addressing strategies currently used and proposes a new replacement strategy to mitigate reconnaissance attacks. The new strategy is evaluated against some reconnaissance attack approaches. The experimental results confirm the effectiveness and validation of the new addressing strategy in terms of the mitigation of reconnaissance attacks.

Keywords - IPv6, IPv4-IPv6 Transition, IPv6 Address Strategies, Security, Reconnaissance Attacks

I. INTRODUCTION

To solve the address exhaustion of IPv4, the Internet Protocol version 6 (IPv6) is introduced in 1998 [1]. The large address space is one of the main developments in IPv6 because it extremely enhances the address space from 32-bit to 128-bit. The IPv6 addresses are written using 32 hexadecimal numbers. These numbers are grouped in 8 blocks of 2 bytes each separated by a colon ":". To simplify the written forms of addresses, some shortening techniques are employed. The technique being used is to omit repetitive 0's, i.e. omitting a string of zeros and leading zeros. In addition, if there is a hexadecimal number of 4 zeros, it will be converted to single zero. The shortening techniques are explained in Figure 1.

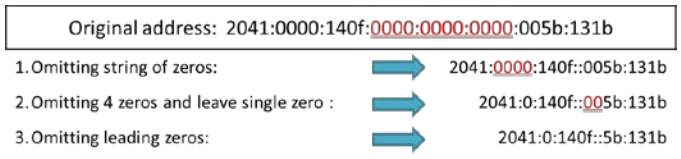


Figure 1: IPv6 Address Shortening Techniques

The high-order left bits of an IP address specify the network whereas the rest bits specify a network particular node. All addresses in a subnet have the same first N bits, and those N bits are called "prefix". The prefix is denoted as "/N". For example: in the address 2041:0:140F::5B:131B/64, the first 64 bits "2041:0:140F:0" represent the address of the network and the rest bits "0:0:5B:131B" represent the node. There are four main IPv6 address types, Unicast, Global Unicast, Unique Local, and Link-local. The IPv6 global unicast addresses are globally routable and similar to IPv4 public addresses. The unique local addresses work like IPv4 private addresses.

Despite that the Internet is almost at the stage of IPv4 exhaustion, the deployment rate of IPv6 is still far from expectation, it barely reached 22.20% in February 2018 [2]. The vulnerabilities of IPv6 are still big concern, and as the deployment of IPv6 proceeds, security issues concurrently raise [3, 4]. Network reconnaissance of IPv6 global unicast addresses is an attempt to discover the services available in the network systems and identify the vulnerabilities. Attackers usually perform address-scanning as initial stage to more severe cyber-attacks, i.e. DoS, or as a prerequisite stage in ransomware attacks as seen in WannaCry ransomware [5]. A traditional way of address scanning is carried out by probing an entire address range of a target subnet. Since the typical IPv6 subnet is a/64, scanning an entire address range, which is equal to 2^{64} would take a tremendous effort, and as a result, network reconnaissance attacks are infeasible in IPv6. However, recent research revealed feasibility of these attacks by analyzing the addressing methods currently adopted in some IPv6 networks [4]. Some IPv6 addressing strategies produce predictable IPv6 addresses and thus attackers could easily perform address scanning by reducing the target search space. Therefore, there is a need for addressing scheme that creates highly structured and unpredictable IPv6 addresses. This paper introduces a new IPv6 addressing strategy to mitigate reconnaissance attacks called Segment Extended Unique Identifier (SEUI-64 bits). The generated addresses are unpredictable, arbitrary, and assigned based on the subnets.

The rest of the paper is organized as follows. Section II reviews some related works; the research methodology is explained in section III. Section IV presents the experimental results. The discussion of the results is presented in Section V. Finally, the paper is concluded in Section VI.

II. LITERATURE REVIEW

In IPv6 networks, each interface has to be assigned a global routable IPv6 address, which is globally unique in the Internet and usually noted as /64. An IPv6 address is divided into two parts: prefix and interface ID. The prefix part is obtained from an ISP or Regional Internet Registry (RIR) and usually divided into two small parts: the global prefix (48 bits) and the subnet ID (16 bits). For attackers, it is easy to get a prefix part of an IPv6 address. For example, they can run the following command in Windows OS command prompt to get the IPv6 address of a website server www.server.com:

C:\> nslookup -query=AAAA www.server.com

The second part, the interface ID part identifies uniquely an interface of a particular node. Figure 2 shows an example of global unicast IPv6 address.

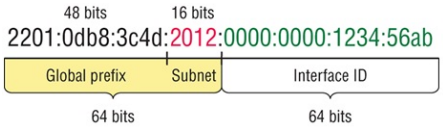


Figure 2: Global unicast IPv6 address

Regardless of assignment method, an IPv6 address is configured for network interfaces by learning the prefix part and appending a locally generated interface ID for it. The IPv6 addresses on a network are assigned in several ways, manual assignment (static IPv6), DHCPv6 (stateless and stateful modes), and stateless address autoconfiguration (SLAAC) [6]. Typically, servers and routers are manually configured while the SLAAC and DHCPv6 assignments are used to configure IPv6 addresses for the workstations. In all cases, network administrators must be careful not to generate predictable addresses that are vulnerable to reconnaissance attacks.

As stated in RFC published by F. Gont and T. Chown [7], despite the full freedom to select the interface ID in the manual assignment, the patterns currently used, such as using IPv4 addresses and service port, lead to generate predictable addresses. They also stated that the use of DHCPv6 could lead to predictable IPv6 addresses because DHCPv6 servers assign addresses by generating sequentially interface IDs from a specific predefined range. The use of SLAAC is the best choice in terms of generating unpredictable IPv6 addresses. The basic idea of SLAAC is that every host joining a network will learn the prefix part from the router advertisement messages while the interface ID is generated using the neighbor discovery protocol. The method of generating interface ID used in the SLAAL is called EUI-64 bits, short for Extended Unique Identifier. EUI-64 bits is totally based on the MAC address of the interface [8], and that could be considered as vulnerable. Given a MAC address detected in one subnet, sequential or nearby MAC addresses might be seen in in the same network [7]. Moreover, the node generates the same interface ID part when it joins a new network making it vulnerable to privacy related attacks and easy for an attacker to attack it [9, 10].

IPv6 addressing schemes used during the process of IPv4-IPv6 transition did not significantly investigated by the research community. This might refer to the fact that the scalability of the SLAAC mechanism, which is the most adopted scheme [11]. However, the IPv6 addressing assignment strategies have been covered extensively in the wireless sensor networks (WSN), IoT, and Mobile IPv6. The breaches in privacy caused by employing the EUI-64 bits addressing scheme have been addressed particularly. Narten et. al. pointed out in their RFC [12] that a non-changing interface IPv6 address may allow an attacker to correlate unrelated information with a particular node. Haddad mentioned that using SLAAC can reveal the mobile nodes location to an attacker [13]. Koodli's study concluded that the roaming status of a node could be revealed to an attacker [14].

In summary, the IPv6 addressing mechanisms used do not provide a deployable IPv6 addresses that are high unpredictable and immune to reconnaissance attacks, despite that some methods, such as the SLAAC, being scalable and reliable. The main issue is the way being followed to generate the interface ID, the second part (64 bits) of an IPv6 address. Therefore, there is a need to an efficient addressing scheme that can be integrated in the SLAAC process in order to reduce the probability of reconnaissance attacks as well as privacy breach attacks to the lowest possible level.

III. METHODOLOGY

The SEUI-64 bits addressing strategy is introduced in this paper as a substitution of EUI-64 bits. Its primary objective is to generate unpredictable IPv6 addresses automatically for network interfaces associated with servers or workstations. The unpredictability provides the IPv6 addresses with immunity to reconnaissance attacks by maximizing the search space as well as to other malicious attacks such privacy breach. In addition, the IPv6 addresses generated by SEUI-64 bits define the gateway that the node is connected to, which gives information about the physical location.

A. Autoconfiguration Protocol Overview

This section provides an overview of the typical steps of SLAAC that take place when an interface autoconfigures itself with IPv6 address, and highlights the time of running SEUI-64 addressing scheme. SLAAC is enabled by the Neighbor Discovery protocol and starts immediately after a node is connected to the network. It involves two phases: link-local IPv6 address configuration and global unicast IPv6 address. The first phase generates link-local IPv6 address to provide the node with connectivity to all other nodes on the same link. It involves three steps:

1) Link-local IPv6 address generation: Appending the interface ID, which is generated using EUI-64, to the link-local prefix (FE80::/10).
2) Duplicate address detection: Verifying the uniqueness of the address by sending a neighbor solicitation message.
3) Link-local IPv6 address assignment: If the address is unique, it will be assigned to the interface.

The second phase generates global unicast IPv6 address to provide full IPv6 global connectivity. It involves four steps:

1) Router Advertisement: The node sends router solicitation message to prompt on-link routers to send router advertisement message that contains the prefix.
2) Global unicast IPv6 address generation: Appending the interface ID, which is generated using EUI-64, to the prefix received from the router.
3) Duplicate address detection: Verifying the uniqueness of the address by sending a neighbor solicitation message.
4) Global unicast IPv6 address assignment: If the address is unique, it will be assigned to the interface.

The proposed SEUI-64 bits will be a substitute to EUI-64 bits. It is triggered by the node to create an interface ID that will be used to generate the link-local and the global unicast IPv6 addresses. Triggering SEUI-64 one time at the beginning of SLAAC process will assure high execution speed and

unpredictability in both link-local and global unicast IPv6 addresses.

B. Generation of Interface ID

The interface ID generated is a link-dependent identifier for an interface. It is a string of 64 bits derived from the network segment interface's link-layer address (MAC address). To increase the interface ID unpredictability, SEUI-64 algorithm selects the Organizational Unique Identifier part of the network segment interface's MAC address and combines it with other two parts. Table 1 explains all these parts.

Table 1: Interface ID parts

Part	Name	Length	Source
1	OUI	3 bytes	Organizational Unique Identifier part of the network segment interface's MAC address
2	STag	2 bytes	Segment Tag: to be set according to the network administrator's desire
3	ABytes	3 bytes	Arbitrary

The Organizational Unique Identifier (OUI) is the first three bytes of a MAC address for a network interface. It uniquely refers to the specific vender or manufacturer for the interface's device. To assure the uniqueness, the OUI is assigned globally by the IEEE to venders. For example "00:00:0A" is owned by Omron and "CC:46:D6" is owned by Cisco Systems Inc. A unique interface identifier, which is 3 bytes long, is appended to OUI bytes to form a 6 bytes MAC address. Figure 3 shows an example of interface MAC address associated with device manufactured by Cisco.

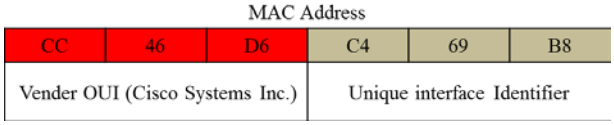


Figure 3: Interface MAC address

SEUI-64 bits strategy takes OUI bytes from the MAC address of the gateway interface, and it will be executed once to create IPv6 addresses for all interfaces connected to the gateway. The OUI three bytes will be taken without any process and set to the first three bytes of the interface ID of the firstly connected node. For the other nodes connected to the gateway, the OUI three bytes will be shuffled at the level of bits before setting them to the first three bytes of the interface ID of other nodes.

The second part is the STag, which is a tag number of two bytes length for a subnet. The value of this part is selected according to the network administrator's desire, i.e. tag number for each router interface. The firstly connected node's interface ID will take the two bytes of STag without any process, while for the rest nodes, the STag two bytes will be randomly selected from the range 0 to STag value. For example, if the segment tag is set to 00:AA, then STag value of the firstly connected node's interface ID will be the same, and the STag values for other nodes' interface ID will be a random number between 0 and 00:AA.

The last part is called ABytes short for arbitrary bytes. ABytes is a string of 24 bits (3 bytes) selected randomly at each time a new interface ID is generated. Figure 4 shows three examples of IPv6 address generated using SEUI-64 bits.

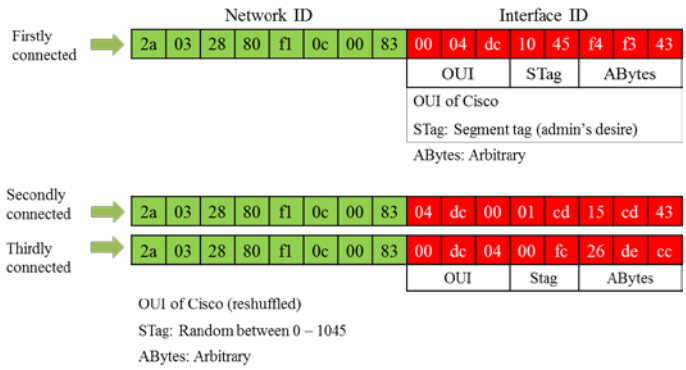


Figure 4: Example of SEUI-64 bits IPv6 Addresses

IV. Experimental Results

The SEUI-64 bits interface ID generation strategy is intended to mitigate reconnaissance attacks in IPv6 network. Since it does not rely on observation and analysis of attack behaviors, it gives the intrusion detection systems chances to detect zero-day address-scanning attacks. By invalidating the attacker's assumptions about the nodes' addresses, SEUI-64 bits strategy increases the probability of connection failures, thus making the attack more detectable [4]. This section investigates the effectiveness of SEUI-64 against sequential scanning attacks as they seem to be more applicable in IPv6 networks than other types of scanning [15]. Specifically, this section tries to improve the detectability of brute-force remote address scanning attacks with the presence of IPv6 addresses produced by SEUI-64 strategy. To quantify the effectiveness, the percentages of failed scans and successful scans have been calculated for a variety of IPv6 environments with different numbers of nodes. A failed scan means probing a nonexistence IPv6 address and a successful scan means probing an existence IPv6 address. Assume S denotes the number of scans generated by attacker, S_{FA} denotes the number of failed scans, and S_{SU} denotes the number of successful scans.

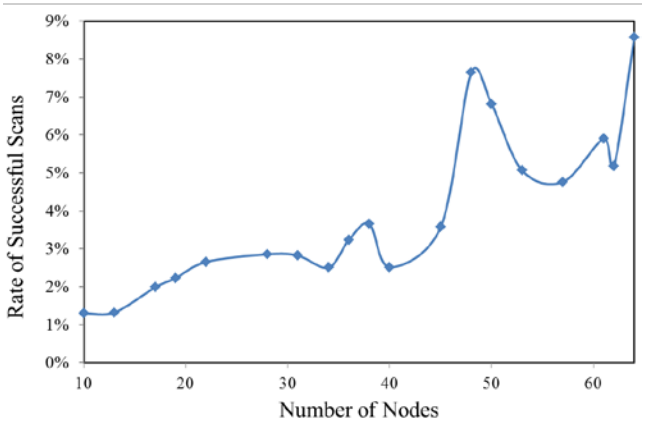


Figure 5: Rate of successful scans in various IPv6 environments (normal sequential scanning)

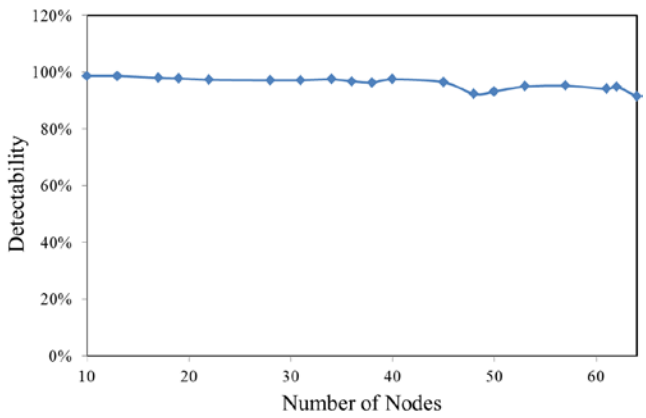


Figure 6: Rate of detectability in various IPv6 environments (normal sequential scanning)

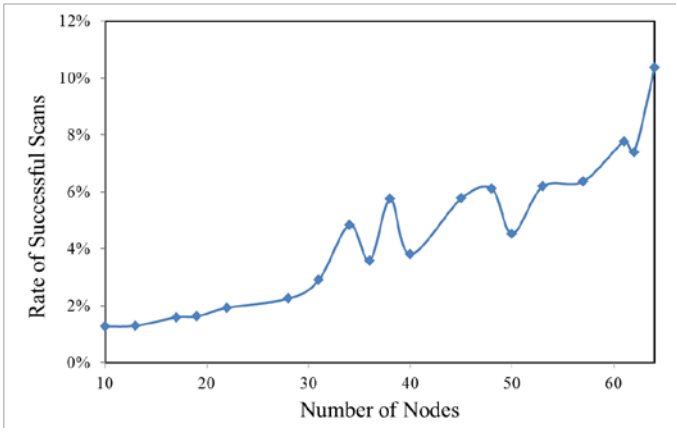


Figure 7: Rate of successful scans in various IPv6 environments (local-preference sequential scanning)

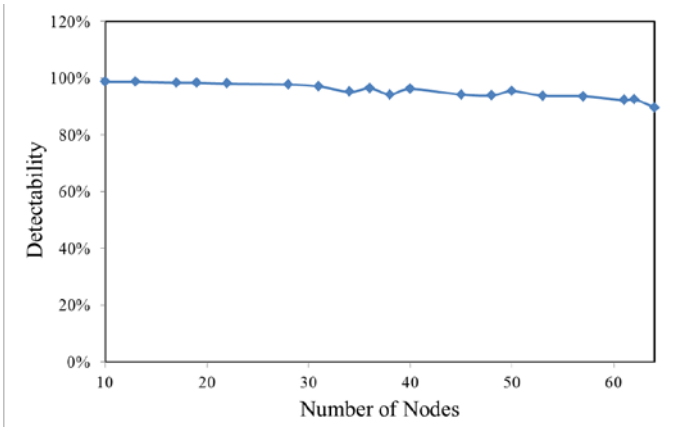


Figure 8: Rate of detectability in various IPv6 environments (local-preference sequential scanning)

The percentage of successful scans is defined as S_{SU} / S, and the percentage of failed scans is defined as S_{FA} / S. Since the number of failed scans, or connections, is a strong indicator of suspicious address-scanning attacks, the percentage of failed scans could measure how much the attack is clear for intrusion detection systems, thus this percentage is established as the detectability of an address-scanning attack. By fixing the prefix part to 2041:0:140F:0/64, a simulation program is developed to act as address-scanner by creating a sequential IPv6 address search space from 2041:0:140F::/64 - 2041:0:140F::1:2C/64, while the SEUI-64 bits strategy is used to addressing the nodes. To avoid any bias in the results, which may appear by creating a large number of addresses, the address-scanning simulator was executed for only 2-3 seconds during all of the experiments. The number of IPv6 addresses generated by the simulator was between 307-315 addresses. The number of IPv6 network environments was 19, and the numbers of nodes ranged from 10 to 64 nodes. Figure 5 and Figure 6 show the rate of successful scans and the rate of detectability respectively in various IPv6 environments.

Another experiment has been conducted with consideration of local-preference sequential scanning that utilizes sequential nature of IP addresses of nodes in same network segment. Local-preference means when an attacker chooses the starting point of IPv6 address search space, he chooses an address close to the IPv6 address obtained initially in order to increase the potential number of scanning nodes in the network segment. In contrast with the first experiment, no changes were made except the starting point of the search space. Figures 7 and 8 show the rate of successful scans and the rate of detectability.

V. DISCUSSION

The SEUI-64 bits strategy presented in this paper mitigates address-scanning attacks and helps in detecting them in IPv6 networks. In general, the experiment results have revealed that avoiding the use of predictable addresses is very crucial to mitigate IPv6 address scanning attacks. As showed in Figures 6 and 8, in the situation of normal and local-preference sequential scanning attacks, the predictability rate, which is the percentage of failed scans, was between 92.97% - 98.71%. Obviously, the unpredictability of IPv6 addresses enforces the attackers to create these high rates of failed scans, which will help the monitoring systems to arise suspicious alarms. In order to prove the high unpredictability level feature of IPv6 addresses, the equation (1) is used to calculate the total number of possible

values of the OUI part, the first three bytes of the interface ID created by SEUI-64 bits strategy:

$$OUI_s = \frac{fact(n)}{(fact(d_0).fact(d_1))} \dots\dots\dots\dots (1)$$

Where OUI_s is the size of the target search space the results from the OUI three bytes, n is the number of bits in OUI part, which is 24 bits, d_0 and d_1 denote the number of zero's and one's in OUI part value respectively. For example, if the Cisco vendor identifier CC:46:D6 is used, shuffling its equivalent binary number 110011000100011011010110 will result in 2,704,156 possible values for OUI part, which reflects very high unpredictability. The same high rates can be found in the second and third parts (STag and ABytes) of the interface ID. The rates might hit 2^{16} for STag and 2^{24} for ABytes if the maximum values (FF:FF and FF:FF:FF) are selected initially for these parts.

For the local address-scanning attacks, there is variety of network vectors that can be utilized by attackers to perform reconnaissance attacks. Even if the unpredictable addresses are used, they could still rely on, e.g. solicited-node multicast address as a resource for network reconnaissance. Therefore, employing intrusion prevention systems at the perimeter along with the unpredictable address will enhance the mitigation process.

VI. CONCLUSION

Network reconnaissance of IPv6 global unicast addresses is an attempt to discover the services available and identify the vulnerabilities of the network systems. Attackers usually perform address-scanning attacks as initial stage to more severe cyber-attacks, i.e. DoS. Moreover, these attacks have become a prerequisite stage in many ransomware attacks. Despite large size of the target search space of a /64 host subnet, some of current employed practice such as using IPv6 addresses based on MAC identifiers (SLAAC) and sequential addresses when using DHCPv6 allow attackers to reduce the large size of the target space when performing an address-scanning attack. In this paper, a new IPv6 addressing strategy to mitigate reconnaissance attacks called Segment Extended Unique Identifier (SEUI-64 bits) has been introduced. The SEUI-64 bits strategy is intended to substitute the EUI-64 bits mechanism that is currently used within the SLAAC process. It generates an interface ID by combining three parts, (1) the OUI: three bytes derived from the MAC address of the interface of the gateway, (2) the STag: two bytes selected by network administrators to tag the segment, and (3) the ABytes: three arbitrary bytes. The parts are subject to special processes before appending to a /64 global prefix in order to generate IPV6 address. The addresses generated have been evaluated against some reconnaissance attack approaches, and the results confirmed the effectiveness and validation of SEUI-64 bits strategy in terms of producing highly unpredictable IPv6 addresses which might mitigate the reconnaissance attacks.

REFERENCES

[1] S. E. Deering, "Internet protocol, version 6 (IPv6) specification," 1998.

[2] (13/2/2018). *IPv6 Adoption - Google Internet Statistics*. Available: https://www.google.com/intl/en/ipv6/statistics.html

[3] A. Shubair, "Survey of security issues in IPv4 to IPv6 tunnel transition mechanisms," *International Journal of Security and Networks*, vol. 12, pp. 83-102, 2017.

[4] J. H. Jafarian, E. Al-Shaer, and Q. Duan, "An effective address mutation approach for disrupting reconnaissance attacks," *IEEE Transactions on Information Forensics and Security*, vol. 10, pp. 2562-2577, 2015.

[5] J. M. Ehrenfeld, "Wannacry, cybersecurity and health information technology: A time to act," *Journal of medical systems*, vol. 41, p. 104, 2017.

[6] S. Hagen, *IPv6 Essentials, 3rd ed.* California, USA, Print: O'Reilly Media, Inc, 2014.

[7] F. Gont and T. Chown, "Network Reconnaissance in IPv6 Networks RFC No 7707," 2016.

[8] R. Asati, H. Singh, W. Beebee, C. Pignataro, E. Dart, and W. George, "Enhanced Duplicate Address Detection RFC No 7527," 2015.

[9] H. Rafiee and C. Meinel, "SSAS: A simple secure addressing scheme for IPv6 autoconfiguration," in *Privacy, Security and Trust (PST), 2013 Eleventh Annual International Conference on*, 2013, pp. 275-282.

[10] S. Groat, M. Dunlop, R. Marchany, and J. Tront, "The privacy implications of stateless IPv6 addressing," in *Proceedings of the Sixth Annual Workshop on Cyber Security and Information Intelligence Research*, 2010, p. 52.

[11] T. Savolainen, J. Soininen, and B. Silverajan, "Ipv6 addressing strategies for iot," *IEEE Sensors Journal*, vol. 13, pp. 3511-3519, 2013.

[12] T. Narten, R. Draves, and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6 RFC No 4941," 2007.

[13] W. Haddad, E. Nordmark, F. Dupont, M. Bagnulo, and B. Patil, "Privacy for mobile and multi-homed nodes: MoMiPriv problem statement," *Internet Draft*, 2005.

[14] R. Koodli, "IP Address Location Privacy and Mobile IPv6: Problem Statement RFC No 4882," 2007.

[15] C. C. Zou, D. Towsley, and W. Gong, "On the performance of Internet worm scanning strategies," *Performance Evaluation*, vol. 63, pp. 700-723, 2006.

An Application of Temperature and Relative Humidity Data Obtaining by RF Communication

H. BAKIR[1], Ü. AĞBULUT[2], M.S. BAŞARSLAN[2*]

[1] Duzce University, Duzce/Turkey, hsynbakr@gmail.com
[1] Duzce University, Duzce/Turkey, umitagbulut@duzce.edu.tr
[2] Doğuş University, Istanbul/Turkey, mbasarslan@hotmail.com

Abstract **- Measurement and control of environmental variables such as temperature and relative humidity have a significant application in science, industry, agriculture, healthcare and controlling and automation technological processes. These two environmental parameters are critical to continuously measure and keep in desired ranges for real working conditions. This paper aims to obtain the values of temperature and relative humidity via wireless communication which is one of the most common communication methods. The transmission of the data was performed using RF receiver and RF transmitter modules. The SHT11 sensor was used to measure both temperature and relative humidity of the environment for the same point. This application is realized and successfully tested in this study. The transfer of data to the computer is successfully provided. Obtained results can be used and applied to increase a system's life in automation and control technologies.**

Keywords **– RF Communication, RS 232 serial Communication, Telemetry, Wireless Data Transfer.**

I. INTRODUCTION

Continuous measurement and control of natural variables are of great importance in many sectors such as food stores, health sector (hospitals, drug stores, pharmacies), greenhouses, etc. The ones of the most important variables affecting many issues in nature are temperature and relative humidity. Hence these values should be under control. Depending on the areas, the values of temperature and humidity should increase or decrease. If these values are not convenient for areas, negative results will probably occur.

In this study, temperature and relative humidity data are continuously measured and these measured data are transmitted via RF (radio frequency) to a computer.

The SHT11 sensor is used to measure the temperature and humidity values of the environment. The advantage of SHT11 sensor is ability to measure both temperature and humidity values at the same point without using ADC due to it's digital output. This sensor provides data transmission by 2 pins [1-4]. PIC16F877 was used to evaluate the data at the transmitter and send it to the receiver. The PIC C-Compiler program is used to program PIC. The obtained data from the sensor comes to PIC and PIC evaluates this program data and sends it to the RF transmitter module for transmission via the output ports.

The ATX-34 module is used for the RF transmitter circuit. The data is transmitted from the antenna of the RF transmitter to the antenna of the RF receiver, and then to the ARX-34D RF receiver module. The data coming back to the PIC from the RF receiver circuit is evaluated and communicated to the computer. The USB to TTL device used for serial data transmission is used to transfer the information to the computer. The data measured by the interface program created after transmission of data to the computer is displayed on the screen.

From past to present, mankind are busy at their work and the time is significant. In some application, there is to give rise to the necessity of using technology in order to gain time especially going and coming to remote distances [5-8]. The aim of this paper is to take the data from a greenhouse where is in 150 meter distance via RF.

II. MATERIAL AND METHOD

In greenhouses, controlling of the temperature and humidity values has an important role to increase the products' life-cycle [8, 9]. There are many methods to obtain data such as; RF, GRPS, Bluetooth, Wi-Fi, Zigbee etc. Using true communication method is depend on the working-conditions. The compare of the some wireless communication methods is given in Table 1 [10].

Table 1: Comparing of some wireless communication methods.

	RF	GPRS	Wi-Fi	Bluetooth
Focusing Area	**Monitoring and control**	Data and audio transfer in remote distances	Web, Email, Video	Alternative to cable
Covering Area	**200 m.**	1000+ m.	1-100 m.	1-10+ m.
Advantages	**Cost and commonly using**	Commonly using and quality	Speed and flexibility	Cost, applicability

As seen in Table 1, RF and GPRS communication methods can be used for 150 m. distance. In this study, RF communication is preferred owing to it's cost. The communication system mainly consists of 2 separate circuits

and they are showed in Figure 1.

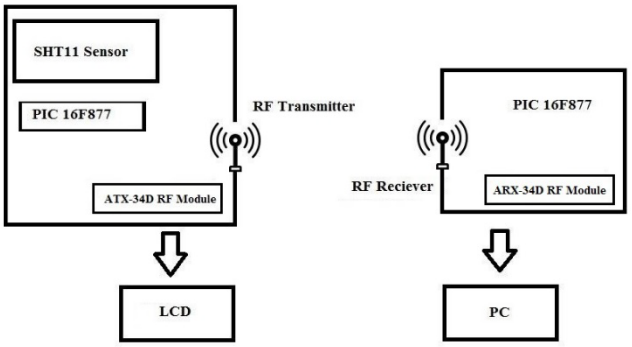


Figure 1: General view of RF communication circuit.

In transmitter circuit, temperature and relative humidity values are measured and then sending to a microcontroller to be evaluated. From microcontroller to RF antenna (ATX-34D transmitter module) is transmitted. The receiver circuit consists of a microcontroller that allows the transmitted data to be received by the RF receiver (ARX-34D) and made meaningful for transmission to the computer, and a USB to TTL module that allows the data to be transferred to the computer. The schematic view using in this study is given in Figure 2.

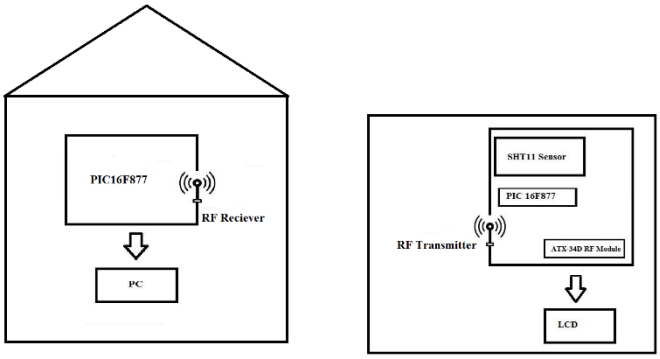


Figure 2. Schematic view of system.

A. *Transmitter Circuit*

Transmitter circuit consists of various diodes, resistors and capacitors with 9 V DC supply voltage input socket, LCD, LM-7805 voltage regulator, PIC16F877 microcontroller, 4 MHz crystal oscillator, red LED, SHT11 temperature and humidity sensor, ATX-34D RF transmitter module and transmitter antenna. With the proper connection and programming of these elements, the first phase of data transmission is realized. The operations performed in the transmitter circuit are shown in Figure 3.

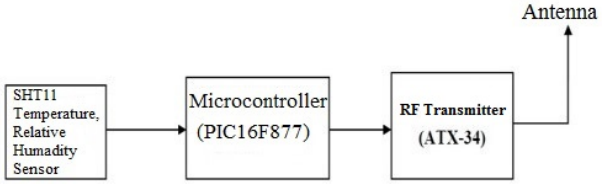


Figure 3: Data flow-chart of transmitter circuit.

1) *SHT11 Temperature and Relative Humidity Sensor*

Thermistors commonly referred to as PTC and NTC are used as temperature sensors for temperature control. The thermistors supply a nonlinear analogue output depending on the temperature. For this reason, their use in sensitive temperature measurements is causing problems. The SHT11 sensor is preferred because of its important properties for wireless data transmission. SHT11 is a single chip temperature and humidity multi sensor module comprising a calibrated digital output. This sensor has included a capacitive polymer sensing element for temperature sensor and relative humidity.

Technical properties of SHT11 sensor are given in Table 2 and Table 3 [11].

Table 2. SHT11 Technical properties of SHT11 sensor.

Relative Humidity			
Parameter	Condition	Value	Units
Resolution	12 bit	0.04	%RH
	8 bit	0.7	%RH
Accuracy tolerance	typ	± 1.8	%RH
	Max		%RH
Repeatability		± 0.1	%RH
Hysteresis		± 1	%RH
Nonlinearity		<0.1	%RH
Response time	63%	8	s
Operating Range	Extended	0 to 100	%RH
Long Term Drift	typ	<0.25	%RH

Table 3. SHT11 Technical properties of SHT11 sensor.

Temperature			
Parameter	Condition	Value	Units
Resolution	14 bit	0.01	°C
	12 bit	0.04	°C
Accuracy tolerance	typ	± 0.2	°C
Repeatability		± 0.1	°C
Operating Range	Extended	-40 to 125	°C
Response time	63%	5 to 30	s
Long Term Drift	typ	<0.02	°C

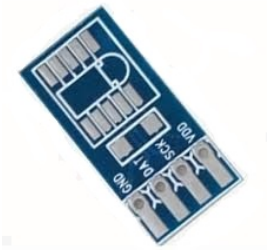


Figure 4. SHT11 Temperature and humidity sensor.

As seen in Figure 4, the pins of SHT11 sensor are:

- GND pin
- DATA pin
- SCK pin
- VDD pin.

This sensor is often used in smart home appliances, humidifiers, desiccants, test and measurement systems, greenhouse systems, agricultural irrigation, cold storage, drug stores, pharmacies, hospitals, and food storage areas.

2) A.4. ATX-34 RF Transmitter Module

The ATX-34 RF transmitter module, produced by Udea Electronics, operates at 433 MHz in the UHF band. The ATX-34 RF transmitter module is preferred because of its high frequency stability, low power consumption, economical efficiency and physical size for short distance data transmission applications. The ATX-34 RF transmitter module was shown in Figure 5.

Figure 5: ATX-34 RF transmitter module.

The use of antennas is very important because it is not possible to send data over long distances without an antenna. The antenna, which can be attached to the ATX-34D RF transmitter module, is a simple cable length 17.3 cm long. In communication systems, the minimum length of the antenna to be used for the transmission of a signal must be at least four times the wavelength of the signal.

B. Receiver Circuit

The receiver consists of 9 V supply voltage, LM7805 voltage regulator, receiver antenna, ARX-34D RF receiver module, PIC16F628A microcontroller, red LED, USB to TTL converter and various capacitors and resistors. The data sent by the transmitter circuit is first transmitted to the RF receiver module, then to the PIC and from there to the computer via the USB to TTL converter module. The block diagram of the receiver circuit was shown in Figure 6.

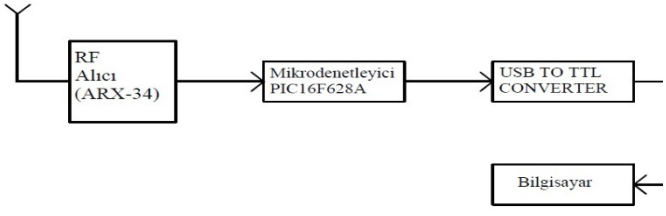


Figure 6. Data flow-chart of receiver circuit.

1) ARX-34D RF Receiver Module

The ARX-34 RF receiver module, manufactured by Udea Electronics, operates at UHF band at 433 MHz. The simplest antenna that can be connected to this module is a 17.3 cm long conductive cable as in the RF receiver. The antenna size is the same as stated in the RF transmitter circuit. The ARX-34 RF receiver module was shown in Figure 7.

Figure 7: ARX-34 RF receiver module.

2) Serial Communication

Serial communication is the type of communication in which information is transferred over a single channel. Serial communication is divided into synchronous and asynchronous. In synchronous communication, the transmitted and received data must be in harmony. In synchronous communication, there is a line that carries the Clock signal for the receiver and transmitter to operate at the same bit time, meaning that there is no need for Start and Stop bits for communication. In asynchronous communication, there are only data lines. Instead of the clock line, there are Start and Stop bits used at the beginning and end of the data packet. Many protocols are used for serial communication. The most suitable of these protocols is RS-232.

3) USB-to-TTL Converter

The USB-to-TTL converter board has 6 pins. These are VCC, GND, TXD, RXD, SET and CS pins. The USB-to-TTL converter schematic diagram was shown in Figure 8.

Figure 8. USB-to-TTL converter schematic diagram.

III. DESIGN

Design of receiver and transmitter circuits was done in Proteus program. Proteus and ARES drawings of the transmitter circuit were shown in Figure 9 and Figure 10. Proteus and ARES drawings of the receiver circuit were shown in Figure 11 and Figure 12.

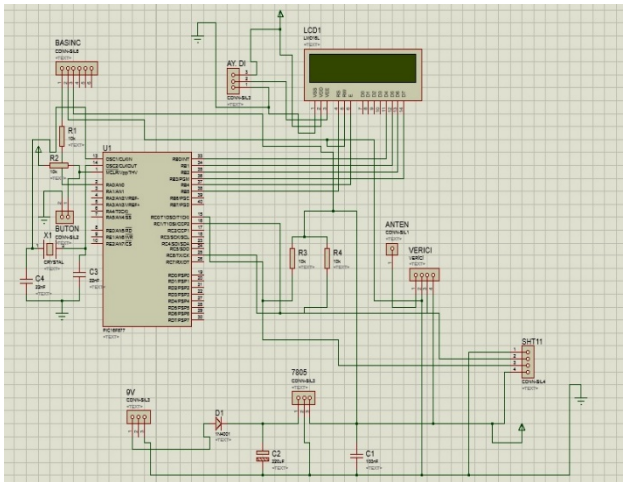


Figure 9: Transmitter circuit Proteus drawing.

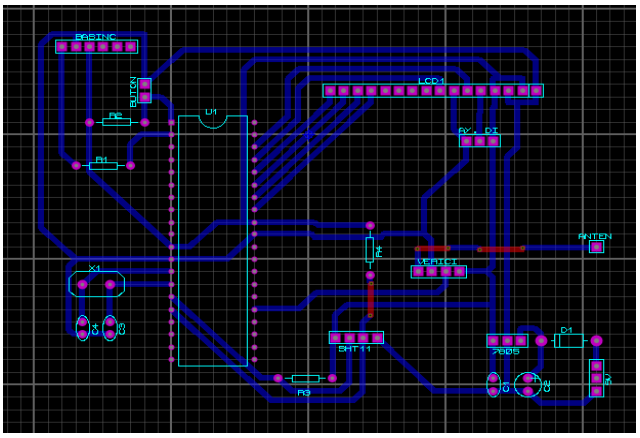


Figure 10: Transmitter circuit ARES drawing.

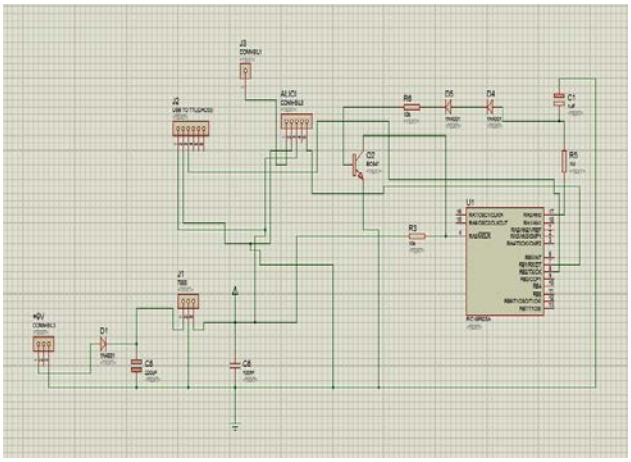


Figure 11: Receiver circuit Proteus drawing.

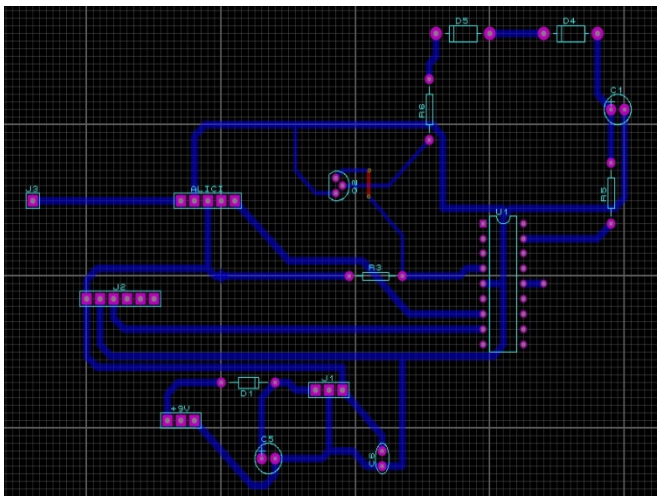


Figure 12: Receiver circuit ARES drawing.

The PIC C Compiler program is used to create software for the PIC16F877 and PIC16F628A microcontrollers in the transmitter and receiver circuits. With the prepared software, the temperature and humidity values were transferred from the greenhouse to the computer in the control center using RF communication.

IV. TESTING THE RF COMMUNICATIONS SYSTEM

Continuous measurement of temperature and humidity values are important in many areas [6-8]. In this study, temperature and humidity values of the greenhouse were measured and transmitted to a computer in the control center via RF communication. Receiver and transmitter circuit boards were shown in Figure 13 and Figure 14.

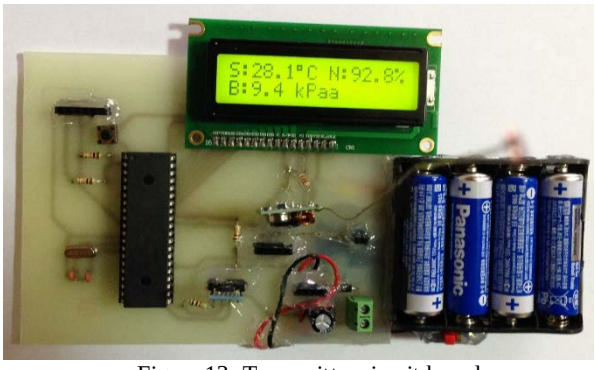


Figure 13: Transmitter circuit board.

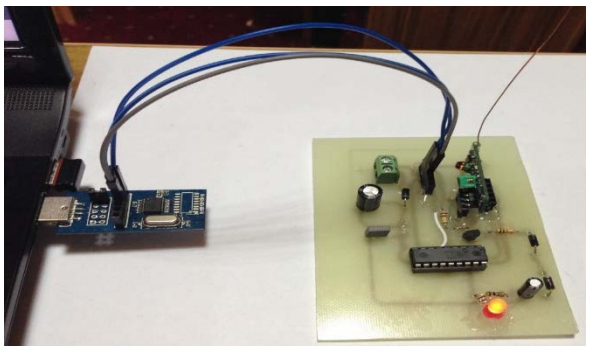


Figure 14: Receiver circuit board.

As shown in Figure 15, the data transferred to the computer is observed with the interface program created with Visual Basic. In the interface program, the information of the data, the port settings and the recording interval of the measurements are displayed.

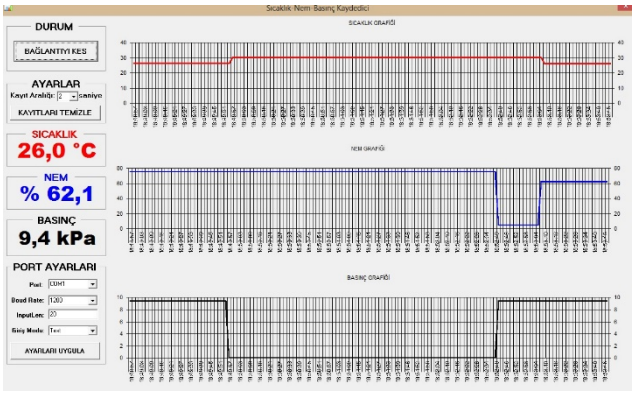


Figure 15: Screen shot of the interface program.

The monitoring of the temperature and humidity data with the Termite program was shown in Figure 16.

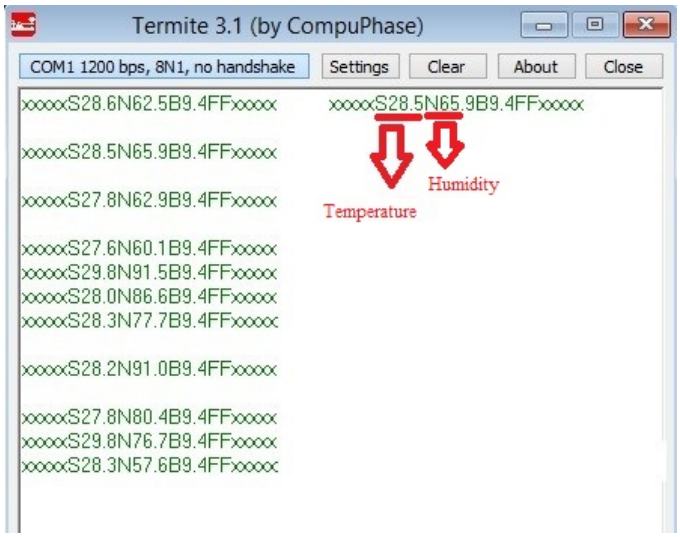


Figure 16: Observation of temperature and humidity data with termite program.

V. CONCLUSION

Intelligent sensors have many advantages such as; combining two measurement elements, cost, integrating ADC etc. In this study, it is successfully performed obtaining temperature and relative humidity data from a greenhouse. In this regard, the time of transferring data between control cabin and greenhouse is reduced less than a second. This study can apply other sectors and systems to control them and save the time. For more complicated applications such as rugged area, windy environment and more remote distances can be used other communication methods. Besides these, specific ranges can be identified and only when these ranges are exceeded, the system can automatically send the informational message or mail to users. Hence, it is no continously need to follow up the screen for data.

REFERENCES

[1] C. Meng, M. Li, and A. C. Li, "Digital temperature and humidity sensor SHT11 based on I~ 2C bus and its application in the single-chip microcomputer system," *International Electronic Elements*, vol. 2, no. 016, 2004.

[2] Q. Wang, P. Chen, Y. Ma, H. Lin, L. Cui, and H. Liu, "Development of high-precision temperature and humidity testing cabinet and its measure and control system," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 27, no. 2, pp. 203-207, 2011.

[3] M. Shuying, M. Yuquan, C. Lidong, and L. Shiguang, "Design of a new measurement and control system of CO 2 for greenhouse based on fuzzy control," *In Computer and Communication Technologies in Agriculture Engineering (CCTAE) 2010 International Conference On* IEEE, pp. 128-131.

[4] J. A. Zhong, "Temperature and humidity monitoring system for young silkworm base on semiconductor refrigerating heat pump," *Transactions of the Chinese Society of Agricultural Engineering*, vol. 28, no. 11, pp. 183-188, 2012.

[5] B. C. Wu, "Design of Load Monitoring and Control System for Demand Response," 2017.

[6] C. J. Lambert, D. T. Lockman, and I. A. Nielsen, "Wireless Communication for Greenhouse Environment," 2017.

[7] Q. Bai, C. Jin, "The Remote Monitoring System of Vegetable Greenhouse," *In Computational Intelligence and Design (ISCID) 2017 10th International Symposium on IEEE*, pp. 64-67

[8] M. A. Dayioğlu, "Seralar İçin Bluetooth Tabanlı Kablosuz Ölçüm Sisteminin Tasarımı: Prototip Geliştirme ve Uygulama," *Tarım Makinaları Bilimi Dergisi,* vol. 9, no. 2, 2013.

[9] K. Jeyashree, G. Chandrabalan "Monitor and Control of Environment for Greenhouse Using Sensor Networks," *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE),* vol. 5, no. 3, 2016.

[10] SHT11 relative Humadity and Temperature User Guide, (2017 December).[Online] Access:https://www.mouser.com/ds/2/682/Sensirion_Humidity_Sensors_SHT25_Datasheet-1274212.pdf

Optimization of Sensor Deployment for *k*-coverage in Wireless Sensor Networks

R.ÖZDAĞ[1] and M. CANAYAZ[2]

[1] Van Yüzüncü Yıl University, Van/Turkey, rozdag@yyu.edu.tr
[2] Van Yüzüncü Yıl University, Van/Turkey, mcanayaz@yyu.edu.tr

Abstract **- Wireless Sensor Networks (WSN) are used for the monitoring of objects in various fields of application as well as the monitoring of military and civilian environments. The energy consumption of sensors and the optimization of network lifetime in WSNs are among the important problems that are constantly investigated and for which solutions are developed by linear programming method. Furthermore, different algorithmic solutions have been developed to perform the dynamic deployment of the nodes efficiently for the solution of this problem. The proposed solutions require that the targets in the network are covered by a minimum number of sensors. *k*-coverage, that determines the degrees of coverage of the targets in the area of interest, is an important criterion in determining the number of sensors covering each target after the deployment of the sensors. Because the coverage of the targets by a minimum number of sensors and the minimization of the intersection area of the sensor increase the lifetime of the network by optimizing the energy consumptions of the sensors.**

In this study, the dynamic deployment approach based on the Whale Optimization Algorithm was proposed to provide the optimum solution to the *k*-coverage problem of WSNs by ensuring that the targets in the area are covered by a minimum number of sensors. This approach, that performs the effective dynamic deployment of sensors by covering the maximum number of target points and ensuring the minimum degree of k-coverage, was compared with the MADA-EM in the literature. Simulation results have shown that this approach is optimum and can be recommended in the solution of the *k*-coverage problem by ensuring that the targets are covered by a minimum number of nodes.

Keywords **- Wireless Sensor Networks, *k*-coverage Problem, Sensor Deployment, Sensor Intersection, Whale Optimization Algorithm**

I. INTRODUCTION

WSNs are distributed systems consisting of sensor nodes with a small size, low cost and limited battery working together in the fulfillment of certain and specific tasks such as environment monitoring and object tracking [1]. Sensors should be efficiently deployed in the area to perform critical tasks such as surveillance and monitoring in military and civilian applications.

The coverage and energy efficiency for WSNs are the most important problems encountered in the distribution of sensors. Coverage is a criterion of the ability of sensors to monitor the area where they are distributed better, and energy efficiency is a criterion of the ability of these nodes to perform perceiving longer in the area of interest. The optimum dynamic deployment of sensors in WSNs will ensure that smaller numbers of nodes are used in the area of interest by preventing the excessive distribution of nodes. In this case, the energy consumption of the nodes will be reduced and the lifetime of the network will increase [2]. Therefore, the optimization of the number of nodes distributed provided that the coverage requirement of the whole area of interest or certain target points is met is an important criterion in ensuring the continuity of the network. The continuity of the network enables the longer surveillance and monitoring of the objects in the area by sensors.

Coverage problem in WSNs is classified as area coverage, target coverage and barrier coverage. Area coverage [3] aims to cover the entire area, target coverage [4] aims to cover specific target points (target set) in the area, and barrier coverage [5] aims to minimize the possibility of penetration that cannot be detected along the barrier in the area. According to the coverage problems, the distribution of nodes in the area of interest is performed using the random or deterministic approach. It is not possible to manually perform the distribution of sensors within the area in war or disaster environments. For this reason, the initial distributions of the sensors can be performed randomly from the aircraft through airway. In this case, sensor distribution should be performed with a deterministic approach in order to ensure the Quality of Service (QoS) of the network [6]. When the initial distributions of the sensors are performed randomly, the target points in the area are ensured to be covered in a balanced manner by performing the efficient dynamic deployment of the nodes deterministically.

In the literature, geometric approaches such as Art Gallery Problem and Circle Covering Problem have been developed and formulated for coverage problem. Art Gallery Problem [7-8] determines the number of observers required to cover the art gallery so that each target point in the art gallery can be monitored by at least one observer. It also assumes that an observer can monitor all the target points within a field of view. Another approach used in computational geometry is the Circle Covering Problem, which allows the arrangement of identical circles on the plane to completely cover a plane [9]. However, the approaches developed for the solution of these problems are

not directly compatible for the application of WSNs. Because they have a limited sensing radii and battery power due to the structure of the sensors. However, in Art Gallery Problem, the visibility of observers is infinite as long as no barrier is encountered [10].

Wang et al. [11] studied the coverage problem by considering the k-coverage requirement for a quality observation. In this study, the coverage level of all intersection points was determined by the coverage level of the area. They proposed a heuristic to create the subset of sensors that can provide *k*-coverage. Zhou et al. [12] proposed a greedy heuristic for k-coverage problem, and the main idea of their study was to choose a candidate path. These two studies in the literature [11-12] aim to create coverage set instead of distributing sensors into subset in a way to ensure *k*-coverage for each subset created.

Huang et al. [13] formulated the coverage problem as a decision problem to determine the coverage status of each target point by at least *k* sensors.

Many studies based on heuristic algorithms and developed to optimize the *k*-coverage problem have been conducted. Wang et al. [14] proposed a new approach consisting of the combination of Particle Swarm Optimization (PSO) and Simulated Annealing (SA) algorithms to ensure an energy-efficient coverage in WSNs. Mini et al. [15] used the Artificial Bee Colony (ABC) algorithm to find the optimum positions of the nodes distributed in a 3-dimensional plane. Wang et al. [16] optimized *k*-coverage problem with Biogeography Based Optimization (BBO) algorithm. Maleki et al. [17] optimized the *k*-coverage problem by combining the PSO with the Differential Evolution (DE) algorithm (PSO+DE) and concluded that the PSO+DE algorithm increased the lifetime of the network.

In this study, it was aimed to perform the efficient dynamic deployments of mobile nodes distributed randomly in the area of interest based on the area coverage problem. For this purpose, it was targeted that all the Grid (target) points in the area would be covered by a minimum number of nodes. The *k*-coverage problem of WSNs was optimized using the Maximum Area Detection Algorithm based on Whale Optimization Algorithm (MADA-WOA) [18] studied in the literature. Furthermore, it was compared with Maximum Area Detection Algorithm based on Electromagnetism-Like Algorithm (MADA-EM) [19] studied in the literature to show the effectiveness of the nodes deployed by MADA-WOA in optimum energy consumption by covering the minimum number of target points.

In this study, *k*-coverage problem of the network is described in Section 2. The developed algorithmic structure of MADA-WOA and MADA-EM for the optimization of the *k*-coverage problem is explained in Section 3. The experimental studies performed and the findings obtained are presented in Section 4, and final results are explained in Section 5.

II. PROBLEM DEFINITION

In this study, the optimum deployments of the sensors distributed in a 2-dimensional square area will be ensured based on the Circle Covering Problem to cover the whole area, and *k*-coverage degrees of all target points in the area will be optimized.

A. Circle Covering Problem

It is used in the applied simulation of sensor nodes in accordance with the Geometric calculation problem and the nature of the coverage problems of WSNs. For the optimum coverage of the square [20] or square [21] area (Figure 1) defined by the nodes distributed in WSNs, the placements of these nodes can be calculated based on the Circle Covering Problem.

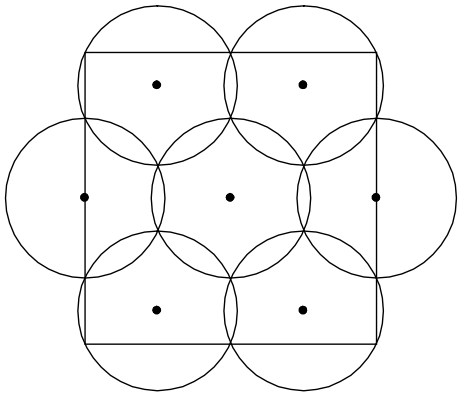


Figure 1: According to the Circle Covering Problem, node placement for optimum coverage of a square area.

B. Degree of Coverage of the Targets

The degree of coverage is defined as the number of sensors covering each target point within the coverage area of the nodes. When it is assumed that an area *A* is covered by the S_1, S_2, and S_3 sensors, the degrees of coverage of the target points covered by these nodes within the *r* sensing radius are presented in Figure 2 [10].

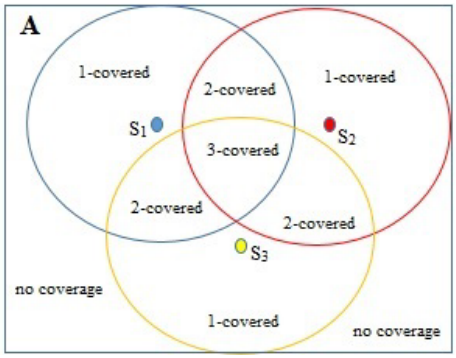


Figure 2: Degrees of Coverage.

C. Intersection and Overlap Conditions of Sensors

Sensor intersection refers to the intersection of the areas perceived by the nodes at a radius distance with each other. When it is assumed that more than one node is distributed within the area *A*, whether the sensors have formed intersection

is determined according to the Euclidean distance between the nodes. If the sensing radius of S_1 node in the area of interest is r_1 and the sensing radius of S_2 node is r_2. Where, the S_1 node is in the (x_1,y_1) coordinate of the area of interest and the S_2 node is in the (x_2,y_2) coordinate, the Euclidean distance between these nodes is calculated by the Equation (1).

$$d(S_1,S_2)=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2} \quad (1)$$

Condition 1 (sensor intersection): Where, it is $r_1=r_2$, and if it is $r_1+r_2<d(S_1,S_2)$, each target point within the intersection area is covered by two nodes since the coverage areas of these nodes are intersecting (so, it is 2-covered) and k=2 (Figure 2).

Condition 2 (sensor overlap): Where, it is $r_1\neq r_2$, and if it is $(r_1-r_2>d(S_1,S_2))\wedge(r_1>r_2)$, the nodes are overlapped since the S_1 node completely covers the S_2 node. In this case, it is k=2 for all target points covered by the S_2 node.

III. METHOD

In this study, the Maximum Area Detection Algorithm approach developed for the solution of the area coverage problem based on both the Whale Optimization Algorithm and the Electromagnetism-Like Algorithm in the literature will be briefly explained in this section.

A. Whale Optimization Algorithm

The Whale Optimization Algorithm developed as a meta-heuristic optimization algorithm [22] was designed by being inspired by the bubble-net strategy used by humpback whales while moving towards their prey or hunting. It was developed based on a swarm-based approach such as Firefly Algorithm, Ant Colony Algorithm and Cricket Algorithm [23]. Since humpback whales have their own special hunting strategy, they gather their prey together by creating bubble clouds under water and keep their prey in this cloud until they get to the surface of the water. 3 strategies have been used in the modeling of the hunting method with the Whale Optimization Algorithm. These are Surrounding the prey, Moving towards the prey and Seeking for the prey [22].

B. Electromagnetism-Like Algorithm

It is a meta-heuristic and population-based global optimization algorithm developed based on the pushing and pulling mechanism between charged particles in the Electromagnetic field [24]. The load values of the particles are firstly determined by calculating the fitness function for each particle studied on set of sample points in the solution space. The load value of each particle determines the magnitude of the pushing and pulling forces applied to the other particles. The resultant force is calculated by taking the sum of the forces applied to each particle by the others, and the direction of movement of the particle is determined in the direction of this force. Thus, the positions of the particles in the solution space are updated so that they are ensured to converge towards the optimum solution.

C. Maximum Area Detection Algorithm

For the solution of coverage problem in WSNs, MADA-WOA and MADA-EM in the literature aim to optimize the coverage rate of the sensors in the area of interest. The minimization of the number of grid points covered by more than one node at the intersection that occurs when nodes are deployed for this purpose constitutes the objectives of these studies [18-19]. Therefore, with this algorithmic approach developed to ensure that all Grid points in the area are covered by a minimum number of nodes, the k-coverage level of all the target points is optimized and the solution of the area k-coverage problem is solved.

In designing this algorithm, the coverage statuses of the Grid points in the area are calculated according to the locations where the nodes that are dynamically deployed are placed, and the fitness function (f_x) values of each node are determined. At this stage, the Euclidean distance between the nodes located in the area and each P Grid point is first calculated by Equation (2). If the P Grid point is within the sensing radius distance of the s^i node, the CS value of the P Grid point is equalized to the binary value of 1, otherwise it is determined that this Grid point is not covered by equalizing to 0 (Equation 3 [25]). Then, f_x values of the sensors are calculated according to the coverage status of all Grid points in the area by more than one node.

$$d(s^i,P)=\sqrt{(x^i-x^p)^2+(y^i-y^p)^2} \quad (2)$$

$$CS_P^i=\begin{cases}1, if\cdot d(s^i,P)\leq r\\ 0, otherwise\end{cases},\forall i\in s \quad (3)$$

In the calculation of f_x, updating is performed for each Grid point covered by taking into account the following cases [19].

Case 1: If the Grid point is covered by only one node, only f_x updating of that node is performed.

Case 2: If the Grid point is covered by more than one node and is within the intersection area of the nodes, depending on the fact that the nodes that cover this Grid point are optimum sensors;

Case 2.1: If any of the nodes that cover this Grid point is the optimum sensor, f_x updating of none of these nodes is not performed.

Case 2.2: If all of the nodes that cover this Grid point are not the optimum sensor, f_x updating of all of these nodes is performed.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, simulation studies were carried out in the MATLAB® environment. The dynamic deployments of mobile nodes distributed randomly in the area were performed by MADA-WOA and MADA-EM, and both the calculated k-coverage level of each Grid point in the entire area and the number of Grid points covered in the intersection area for each k detected were compared. As it was required by the area

coverage problem, common parameters were used in the simulation of both algorithms to ensure that all Grid points in the area can be covered under equal conditions and a reliable comparison could be performed (Table 1).

Table 1. Common parameters of the compared algorithms

Parameters	Defined Value or Range
Area size (A)	100 $\times$ 100 meter (10000 m^2)
Number of targets in the area	10201
Distance between targets	1 meter
Number of mobile sensors (m)	10 – 100
Node sensing radius (r)	7 meter
f_x update interval	25 units
Monte Carlo simulation number	5

Provided that the number of nodes distributed randomly is between 10 and 100, the mobile nodes with a detection radius of 7 meters were deployed by both algorithms and placed in the area. There are a total of 10201 Grid points in a 2-dimensional area consisting of 100 $\times$ 100 meters. The f_x value of each node was determined by calculating the number of Grid points that the nodes covered with respect to their position after their dynamic deployments. Furthermore, the number of nodes covering the Grid points was also calculated and the k-coverage degree of each Grid point was determined. The maximum k-coverage degrees determined by the number of mobile nodes deployed by the algorithms compared in this study are presented in Figure 3.

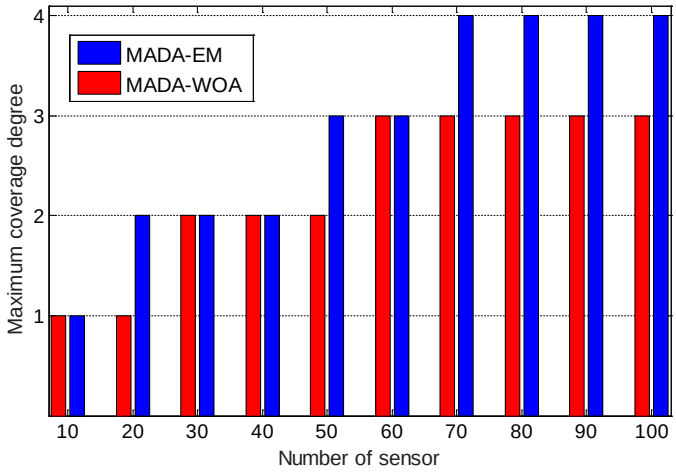


Figure 3: Maximum k-coverage degrees calculated by the number of sensors

While the maximum k-coverage degree was 4 when mobile nodes were distributed by MADA-EM between 70 and 100, the maximum k-coverage degree was 3 when nodes were distributed by MADA-WOA between 60 and 100. Therefore, the maximum k-coverage degree reached by MADA-WOA due to the intersection of fewer nodes compared to MADA-EM in the node deployments performed was minimized. In conclusion, the energy consumptions of the nodes are optimized by preventing the Grid points in the area from being covered by excessive nodes (Figure 3).

The numbers of Grid points covered in the intersecting area for k=2, k=3 and k=4 determined in the algorithms compared were calculated and analyzed separately between Figures 4 and 6.

For k=2 (Figure 4), while the number of Grid points covered was 0 since there was no sensor intersection when MADA-WOA deployed 10 and 20 nodes, the number of Grid points covered was 0 when only 10 nodes were deployed by MADA-EM. In the algorithms compared, the number of Grid points covered in the intersecting area when node deployment was performed between 30 and 100 was calculated to be less by 7% and 91% when 100 nodes and 30 nodes were deployed, respectively, by MADA-WOA.

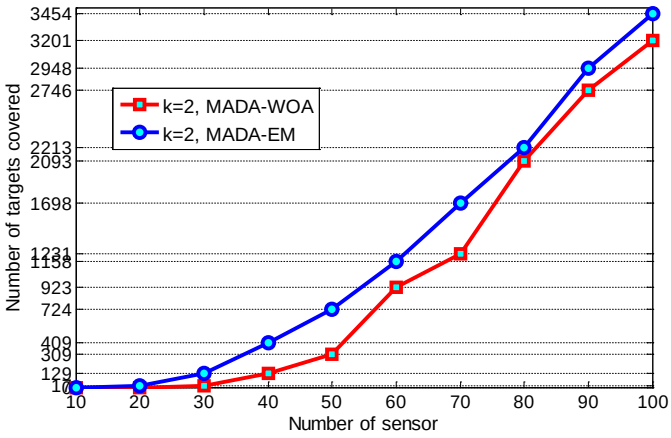


Figure 4: Change graphic of the number of targets covered by the algorithms compared for k=2 by the number of sensors.

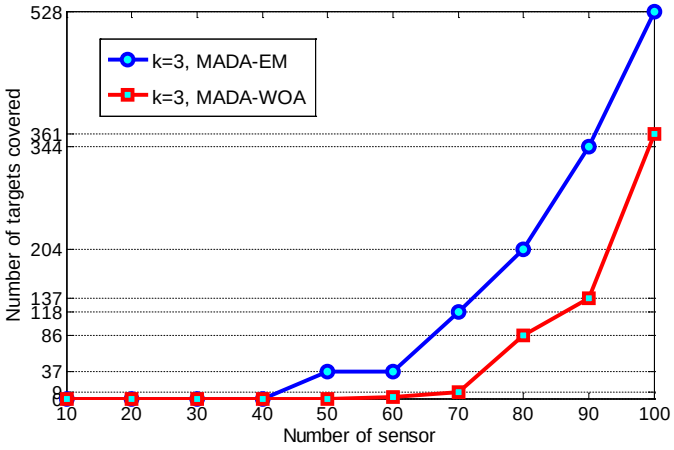


Figure 5: Change graphic of the number of targets covered by the algorithms compared for k=3 by the number of sensors.

For k=3 (Figure 5), while the number of Grid points covered was 0 when node deployment was performed by MADA-WOA between 10 and 50, the number of Grid points covered was 0 when node deployment was performed by MADA-EM between 10 and 40. In the algorithms compared, the number of Grid points covered in the intersecting area when node deployment was performed between 60 and 100 was calculated to be less by 31% and 94% when 100 nodes and 60 nodes were deployed, respectively, by MADA-WOA.

For k=4 (Figure 6), while the number of Grid points covered was always 0 when node deployment was performed by

MADA-WOA between 10 and 100, the number of Grid points covered was 0 when node deployment was performed by MADA-EM only between 10 and 60. Therefore, no intersecting node was detected in the deployment performed by MADA-WOA.

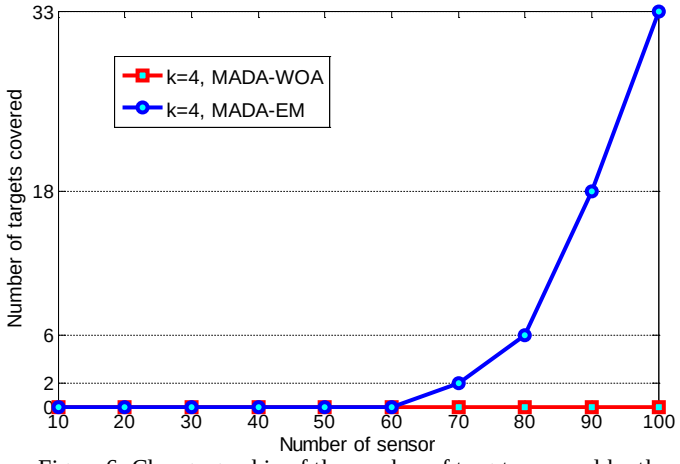


Figure 6: Change graphic of the number of targets covered by the algorithms compared for k=4 by the number of sensors.

In the algorithms compared for k=2, k=3 and k=4 (Figure 7), the Grid points covered by more than one node by MADA-WOA were calculated to be always at a lower rate compared to the total number of Grid points in the area (10201) when the sum of the number of Grid points covered in the intersecting area formed in the area of interest was taken.

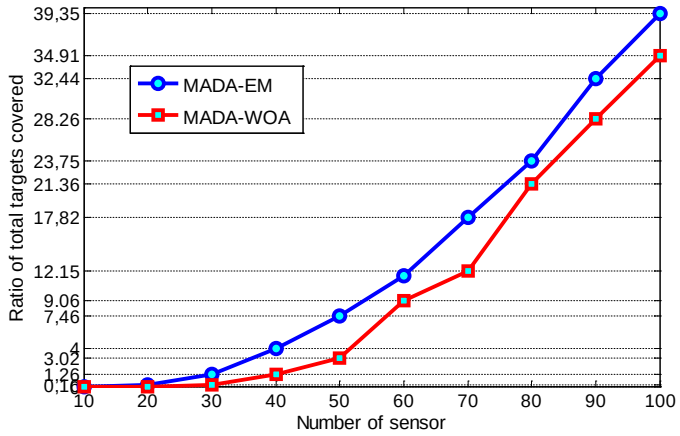


Figure 7: Change graphic of the ratio of total number of targets covered in intersecting areas for k=2, k=3 and k=4 to the whole area.

V. CONCLUSION

The fact that all Grid points in the area were ensured to be covered by a minimum number of nodes by applying MADA-EM and MADA-WOA approaches in the literature to the k-coverage problem, and the minimization of the number of Grid points in the resulting intersecting area constituted the focal point of this study. The Maximum Area Detection Algorithm approach developed for the solution of the coverage problem was also applied to the k-coverage problem, and the simulations made with MADA-EM and MADA-WOA were compared. According to simulation results, optimum results were achieved in calculating the maximum k-coverage degree with MADA-WOA. While all Grid points in the area were covered by maximum 4 nodes in the deployment performed with MADA-EM, it was determined that they were covered by maximum 3 nodes in the deployment performed with MADA-WOA. In the analyses performed for k=2, k=3 and k=4, it was determined that the number of Grid points within the intersecting area formed by the nodes after the deployment performed with MADA-WOA was always lower compared to MADA-EM. Furthermore, when total Grid points in the intersecting areas were analyzed, it was determined that Grid points were always covered by a lower rate with MADA-WOA compared to total area. Therefore, it was concluded that MADA-WOA was a more effective approach in reaching the minimum k-coverage degree and minimizing the number of Grid points covered in the intersecting areas.

In the next study, it is considered to study k-coverage with MADA-WOA for the target coverage problem.

ACKNOWLEDGMENT

This project is supported by Van Yüzüncü Yıl University Scientific Research Projects Coordination Unit (YYU BAP) under project number FBA-2017-5831.

REFERENCES

[1] O. Banimelhem, M. Mowafi, and W. Aljoby, "Genetic Algorithm BasedNode Deployment in Hybrid Wireless Sensor Networks," *Communications and Network*, vol. 5, pp. 273–279, 2013.

[2] R. Özdağ, "Kablosuz algılayıcı ağlarda hedef kapsama problemi için algılayıcı dağıtımı ile ağın yaşam süresinin optimizasyonu," *J Fac Eng Archit Gaz*, vol. 32, pp. 1155–1167, 2017.

[3] Z. Abrams, A. Goel, and S. Plotkin, "Set K-Cover Algorithms for Energy Efficient Monitoring in Wireless Sensor Networks," in *Conf. Third International Symposium on Information*, pp. 424–432, 2004.

[4] M. Cardei, M. Thai, Y. Liand, and W. Wu, "Energy-Efficient Target Coverage in Wireless Sensor Networks," in *Conf. IEEE INFOCOM 2005*, pp. 1976–1984.

[5] H. Tan, Y. Wang, X. Hao, Q. S. Hua, and F. C. M. Lau, "Arbitrary Obstacles Constrained Full Coverage in Wireless Sensor Networks," *in Conf. WASA 2010*, pp. 1-10.

[6] R. Özdağ, "Optimization of Target Q-Coverage Problem for QoS Requirement in Wireless Sensor Networks," *Journal of Computers*, vol. 13, pp. 480-489, 2018.

[7] J. O'rourke, *Art gallery theorems and algorithms*. Oxford University Press, 1987.

[8] M. D. Berg, O. Cheong, M. V. Kreveld, and M. Overmars, *Computational geometry: algorithms and applications*. Springer Press, 2008.

[9] R. Williams, *The Geometrical Foundation of Natural Structure: A Source Book of Design*, New York: Dover Publications, 1979.

[10] N. Yeasmin, "k-coverage Problems and Solutions in Wireless Sensor Networks: A Survey," *International Journal of Computer Applications*, vol. 100, pp. 1–6, 2014.

[11] X. Wang, G. Xing, Y. Zhang, C. Lu, R. Pless, and C. Gill, "Integrated coverage and connectivity configuration in wireless sensor networks," *in Conf. 2003 1st International Conference on Embedded Networked Sensor Systems*, pp. 28–39.

[12] Z. Zhou, S. Das, and H. Gupta, "Connected k-coverage problem in sensor networks," *in Conf. 2004 IEEE Ic Comp Com Net*, pp. 373–378.

[13] C. F. Huang and Y. C. Tseng, "The coverage problem in a wireless sensor network," *Mobile Netw. Appl.*, vol. 10, pp. 519–528, 2005.

[14] X. Wang, J. J. Ma, S. Wang, and D. W. Bi, "Distributed Particle Swarm Optimization and Simulated Annealing for Energy-efficient Coverage in Wireless Sensor Networks," *Sensors*, vol. 7, pp. 628–648, 2007.

[15] S. Mini, S. K. Udgata, and S. L. Sabat, "Artificial Bee Colony Based Sensor Deployment Algorithm for Target Coverage Problem in 3-D Terrain, *in Conf. 2011 Lect Notes Comput Sc*, pp. 313–324.

[16] G. Wang, L. Guo, H. Duan, L. Liu, and H. Wang, "Dynamic Deployment of Wireless Sensor Networks by Biogeography Based Optimization Algorithm," *J. Sens. Actuator Netw.*, vol. 1, pp. 86–96, 2012.

[17] I. Maleki, S. R. Khaze, M. M. Tabrizi, and A. Bagherinia, "A New Approach for Area Coverage Problem in Wireless Sensor Networks with Hybrid Particle Swarm Optimization and Differential Evolution Algorithms," *International Journal of Mobile Network Communications & Telematics*, vol. 3, pp. 61–75, 2013.

[18] R. Özdağ, and M. Canayaz, "A New Dynamic Deployment Approach Based On Whale Optimization Algorithm in the Optimization of Coverage Rates of Wireless Sensor Networks," *European Journal of Technique*, vol. 7, pp. 119–130, 2017.

[19] R. Özdağ, "Kablosuz Algılayıcı Ağlarda Alan Kapsama için Dinamik Düğüm Dağıtımı ile Yeni bir Meta-sezgisel Yaklaşım," *in Conf. 2016 4th International Symposium on Innovative Technologies in Engineering and Science*, pp. 1513–1522.

[20] A. Heppes and J. B. M. Melissen, "Covering a Rectangle with Equal Circles," *Periodica Mathematica Hungarica*, vol. 34, pp. 65–81, 1996.

[21] K. J. Nurmela and P. R. J. Ostergard, "Covering a square with up to 30 equal circles," Research Report A62, Helsinki University of Technology, Laboratory for Theoretical Computer Science, Espoo, Finland, June 2000.

[22] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Adv. Eng. Software*, vol. 95, pp. 51–67, 2016.

[23] M. Canayaz, A. Karci, "Cricket behaviour-based evolutionary computation technique in solving engineering optimization problems, *Applied Intelligence*, vol. 44, pp. 362–376, 2016.

[24] S. I. Birbil, S. C. Fang, "An Electromagnetism-like Mechanism for Global Optimization," *J Global Optim.*, vol. 25, pp. 263–282, 2003.

[25] R. Özdağ, A. Karcı, "Sensor Node Deployment Based on Electromagnetism-Like Algorithm in Mobile Wireless Sensor Networks," *Int. J. Distrib. Sens. N.*, vol. 2015, 15 pages, 2015.

Cost-effective logging using SDN architecture

B. BABAYIGIT[1] and S. KARAKAYA[2]

[1] Erciyes University, Computer Engineering Department, Kayseri/Turkey, bilalb@erciyes.edu.tr
[2] Erciyes University, Graduate School of Natural and Applied Sciences, Kayseri/Turkey, sinankarakaya1992@gmail.com

***Abstract* – Existing networking systems are hardware-based and rely on inflexible architectures. In recent years, Software defined networks (SDNs) has emerged as a new paradigm for next-generation networks. SDNs are proposed to separate control plane and data forwarding plane in the traditional networks to reduce the increasing complexity of the network equipment. This separation enables a programmable and flexible hardware infrastructure via OpenFlow protocol and provides great opportunities in terms of reducing operating cost, simplifying network management tasks, gathering network statistics, and accelerating innovation.**

Logging is a structure that records the events of a system and the users in the system. With logging operations, the actions making by each device can be recorded. Especially in places where more than one person connects to the internet such as hotels, cafes, restaurants, student dorms and companies etc., it is necessary to monitor and take logs of the events. Already, by the Law No 5651, to combat certain crimes committed on the internet, it has become obligatory to take logs to institutions or people who provide collective internet services.

In this paper, a logging and monitoring system has been designed and implemented. For logging and monitoring proceses, sFlow, OpenFlow, Floodlight, Open vSwitch, node.js, and MySQL technologies have been used. sFlow technology is used to monitor the networks. sFlow standard gives complete visibility into the use of networks enabling performance optimization, usage, and defense against threats. OpenFlow is a protocol used in SDN environments to enable the SDN controller to interact with the data forwarding plane of the network devices. Floodlight software is used by SDN controller for network operations. Open vSwitch software which enables to handle the traffic loads is installed at Raspberry Pi 3 hardware. The obtained results show that the proposed system can take the flowtable records, transfer the records to node.js via sFlow, and save the records to MySQL via node.js. Finally, the network traffic is successfully monitored and logged.

***Keywords* - Software defined network, sFlow, Floodlight, OpenFlow, logging.**

I. INTRODUCTION

Traditional network infrastructure consists of devices such as switches, routers, and intermediate devices. These devices are based on application-specific integrated circuits to perform dedicated functions and preprogrammed with different rules. Unfortunately, these networking devices cannot be modified and programmed with multiple rules to establish optimal network services, especially in real-time [1].

Recently, Software Defined Networks (SDNs) have been proposed as a promising network architecture to tackle such challenges of traditional networking architecture [1, 2]. The main aim of SDN architecture is to add simplicity, programmability and flexibility features to traditional networking architecture. For this, the control plane is detached from the data plane, and the ability of programming the control plane is introduced according to the application-specific requirements in real-time. As a consequence, these SDN features allow users to develop applications and to configure the network with a predefined policy.

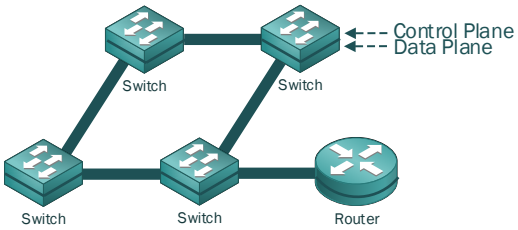
Today institutions and individuals such as universities, schools, corporations, cafes, restaurants, etc. provide internet services to students, customers, and guests. For this reason, the importance of traffic monitoring and logging has increased. Moreover, by the Turkish Law No. 5651, it has become obligatory to control and monitor the internet traffic and to take logs of the internet activities. In addition to keeping track of mobile device use and user internet activities, monitoring and logging can also protect the institutions and business corporations against cybercrime that can be committed [3].

In this paper, a cost-effective logging and monitoring system using SDN architecture is proposed. The proposed programmable SDN network is implemented using Floodlight SDN controller, Open vSwitch, and is tested sFlow monitoring tool.

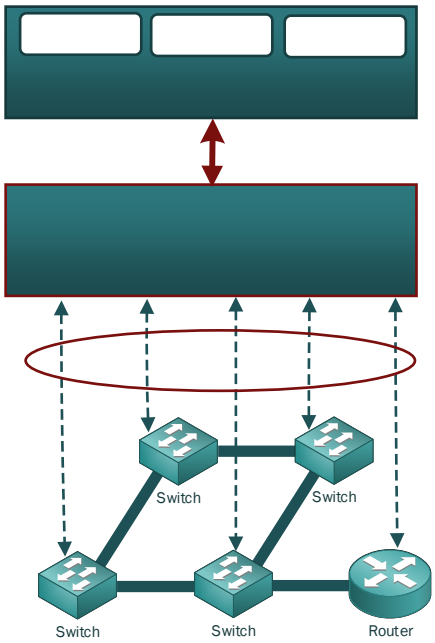
The remainder of the paper is organized as follows. Section 2 briefly describes traditional and SDN architectures. Section 3 explains the proposed logging and monitoring system. Section 4 presents the monitored and logged traffic results. Finally, Section 5 concludes the paper.

II. SOFTWARE DEFINED NETWORKS

Software-defined networks (SDNs) are regarded as one of the future networking paradigms which offers flexible, programmable, and virtualizable network architecture via a centralized logical network controller [1, 2]. With the logically centralized controller, it is easier to apply user-defined policies and optimize network configurations. The main idea of SDN is to separate control plane software from data plane hardware. Separating the control decisions from the network hardware (switches and routers) and enabling the hardware programmable via software-defined OpenFlow interface make the network management agile and flexible. SDNs improve network management, reduce operating cost, and adapt to gradually expanding network size situations. Traditional and SDN networking architectures is illustrated in Figure 1.



(a)



(b)

Figure 1: Networking architectures. (a) Traditional, (b) SDN.

As depicted from the bottom to the top of Figure 1, the SDN architecture has three main planes. These planes are data, SDN controller, and network applications.

Data plane consists of network (forwarding) devices such as switches and router. These network devices forward the traffic (flows) based on the policy of SDN controller plane. An SDN controller plane, acting as the brain of the SDN system, comprises of one or more software-based SDN controllers and manage the flow control of the network elements. The centralized SDN controller collects global information on network resources. This is also essential for network optimization. Finally, the application plane includes a variety of business and network applications and is responsible for policy enforcement.

In SDN architecture, there also two application programming interfaces (APIs): Northbound API and Southbound API. The Northbound API is used for management or development, and interfaces between the SDN controller plane and the network applications plane. The Southbound API connects the SDN controller and the network devices.

OpenFlow [4] is the Southbound communication protocol. This protocol standardizes information exchange between the control and data forwarding planes. An OpenFlow-enabled switch has three main parts: Flow Table, Secure Channel, and OpenFlow Protocol. An OpenFlow switch maintains one or more flow tables containing a list of flow entries. SDN controller manages and configures the network devices through the secure channel using OpenFlow Protocol standard. When an OpenFlow switch receives a packet, incoming packets are compared with flow entries. If there is a match, the packet is processed according to the entry actions. If the is no match, the switch is sent an OpenFlow message to the controller which conveys the headers of the packet.

As can be seen from Figure 2, the secure channel is established over a connection between controller and the switch (host). The controller collects the real-time flow information from the corresponding switch, connects to the hosts (switches) via secure channel, and then the controller builds a graph-based model of the network; finally, adds, updates, and deletes the stream entries to flow tables using the OpenFlow protocol. The OpenFlow switch compares the incoming data packet with the stream table headers. If the packet header fields (left side of Figure 2) and the flow diagram match the entry, the packet is forwarded to the required location. If not found, the package controller is sent. The controller broadcasts the new rule entry to the entire network.

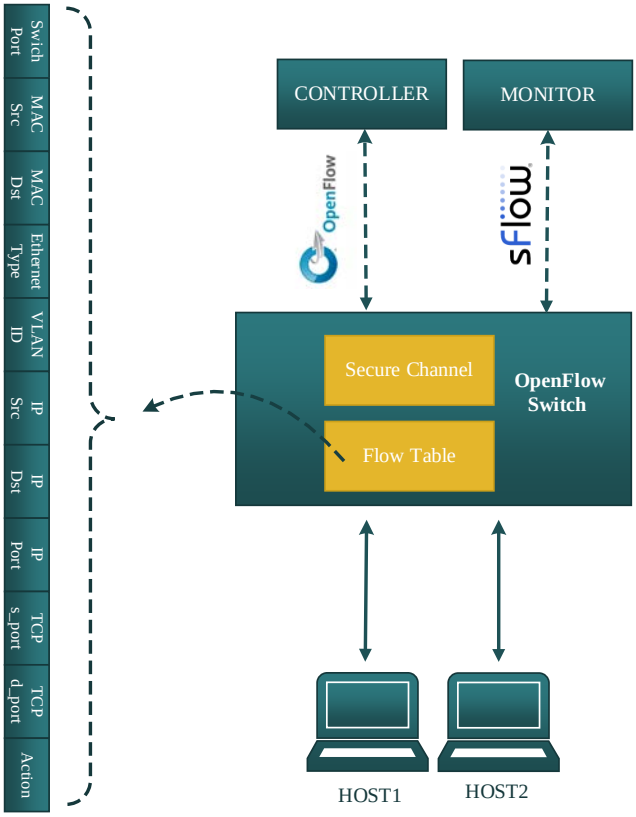


Figure 2: Diagram for sFlow and OpenFlow functions.

III. LOGGING PROCESS

A. *Logging and Monitoring*

As organizations and business corporations provide free

internet access to their customers and guests, taking logs has become increasingly important than ever. So, logging and monitoring the internet activities through log management tools are necessary for security. Log management has three processes: log generation, log storage, and log monitoring. In log generation, the activities are logged. In log storage, the logs are retained and disposed off. While in log monitoring, the logged activities and traffic are reviewed.

Clearly, without specifying exact regulations, unauthorized access to data, obtaining and recording data, collection and storage of personal information by logging can do happen. Accordingly, regulations and laws are mandatory for a good log retention policy. The Law No. 5651 (The Internet Law of Turkey) [3] aims at combating crimes committed by misuse of the Internet environment and the measures against the harmful content. Law No. 5651 also defines the liabilities of content providers, hosting providers, access providers, and public use providers. More specifically, in Article 7(2) of Law No. 5651, the public use providers who provide individuals with the possibility of using the Internet in a certain location and for a certain period time, are obliged to investigate and monitor the information accessed and must take measures to prevent access to the content that constitutes a crime.

B. Proposed Logging and Monitoring Process with SDN

In this section, a cost-effective SDN architecture equipped with logging and monitoring features is explained. The diagram of the architecture for monitoring the network traffic and the steps for logging process are depicted in Figures 2 and 3, respectively. The implementation details and the explanations of each step for the proposed systemare given below.

For the logging and monitoring system, Floodlight [5] is used as the SDN controller and Raspberry Pi 3 [6] is used as the networking device (switch). First, Open vSwitch [7] software is installed on Raspberry Pi 3 to convert Raspberry Pi 3 into OpenFlow-enabled software switch. Then, a bridge is created for sharing the internet and USB to Ethernet adapter is added to this bridge. Also, IP of the SDN controller is assigned to Open vSwitch.

In the monitoring and the logging processes, OpenFlow [4] and sFlow [8] technologies are used for real-time traffic data collection. sFlow is an open source monitoring tool for OpenFlow networks and used for flow-based data gathering. However, OpenFlow is used to define the flows to be monitored by sFlow.

After configuring Raspbery Pi 3, logging and monitoring processes start which consist of sFlow-RT, sFlow, node.js, and MySQL steps, as shown in Figure 3.

As depicted in Figure 3, because of the flexibility provided by SDN and sFlow-RT, the sFlow application (flow-graph module) is installed to monitor the traffic with Raspberry Pi 3. Flow-graph module provides real-time visibility in the traffic flows. Then traffic data is sent to node.js using jQuery AJAX POST request. Received request from node.js insert into MySQL. Thus, the traffic data has been monitored and logged.

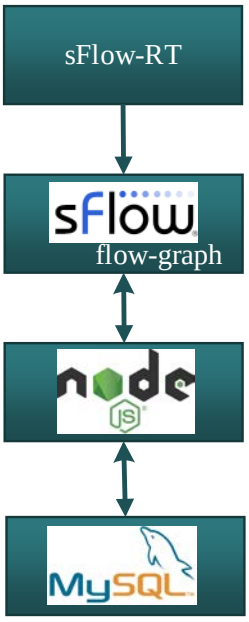


Figure 3: Steps for logging and monitoring.

IV. RESULTS

This section presents the obtained results of the proposed system explained at Section 3. The fields in the flow table header can be selected by the user as in Figure 4. ipsource, ipdestination, stack, and bytes are selected. These selected fields are monitored using sFlow-RT as shown in Figure 5.

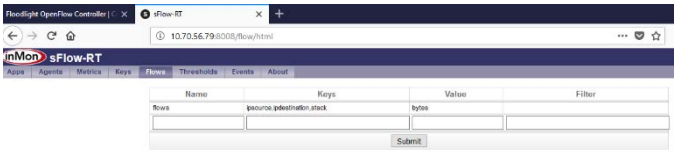


Figure 4: Flow field selection.

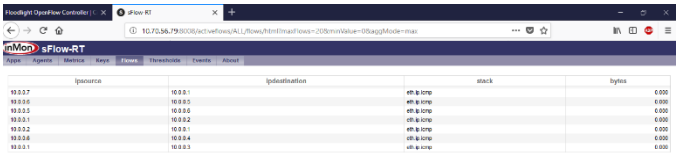


Figure 5: Monitoring of the selected field of the traffic.

But, as can be seen from Figure 5, the records monitored are all text-based. To make a graph-based flow, sFlow's flow-graph module is used. Graph-based output of flow-graph is shown in Figure 6.

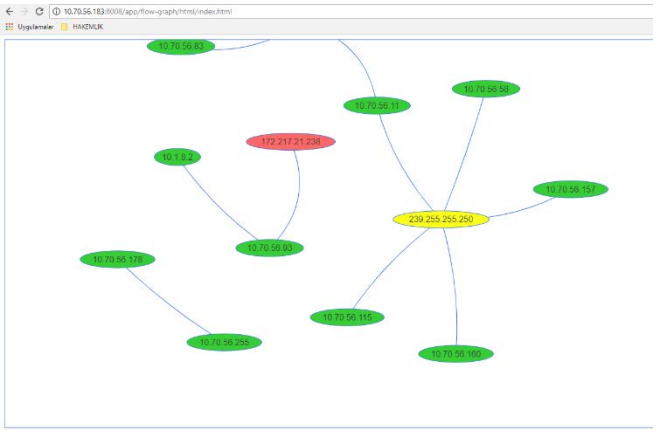


Figure 6: Graph-based output of Flow-graph.

Although, the traffic flows are monitored, the system cannot record the fields. Thus, we added codes to flow-graph module of sFlow. By this way, the traffic data is sent to node.js and output console. Graph-based outputs of flow-graph after adding AJAX and console codes for logging is shown in Figure 7.

Figure 7: Graph-based outputs of Flow-graph after adding AJAX and CONSOLE codes for logging.

At last, traffic outputs are received from node.js application and saved to MySQL via node.js. MySQL output of the logs is shown in Figure 8.

link	from	destination	createdAt
10.70.56.93-172.217.21.238	10.70.56.93	172.217.21.238	2018-03-19 21:15:03
10.70.56.93-212.101.122.34	10.70.56.93	212.101.122.34	2018-03-19 21:19:38
10.70.56.93-212.101.122.34	10.70.56.93	212.101.122.34	2018-03-19 21:23:55
10.70.56.93-212.101.122.34	10.70.56.93	212.101.122.34	2018-03-19 21:24:30
10.70.56.93-212.101.122.34	10.70.56.93	212.101.122.34	2018-03-19 21:25:36
10.70.56.115-224.0.0.22	10.70.56.115	224.0.0.22	2018-03-19 21:25:36
10.70.56.93-85.111.27.87	10.70.56.93	85.111.27.87	2018-03-19 21:26:00
10.70.56.93-193.255.88.9	10.70.56.93	193.255.88.9	2018-03-19 21:26:16
10.70.56.58-224.0.0.252	10.70.56.58	224.0.0.252	2018-03-19 21:26:33
10.70.56.255-10.70.56.58	10.70.56.255	10.70.56.58	2018-03-19 21:26:41
10.70.56.160-10.70.56.255	10.70.56.160	10.70.56.255	2018-03-19 21:26:49
10.70.56.178-10.70.56.255	10.70.56.178	10.70.56.255	2018-03-19 21:27:28
10.70.56.115-224.0.0.251	10.70.56.115	224.0.0.251	2018-03-19 21:31:33
10.70.56.83-224.0.0.252	10.70.56.83	224.0.0.252	2018-03-19 21:34:37
10.70.56.83-224.0.0.252	10.70.56.83	224.0.0.252	2018-03-19 21:42:31
10.70.56.93-193.255.88.9	10.70.56.93	193.255.88.9	2018-03-19 21:46:40
10.70.56.178-10.70.56.93	10.70.56.178	10.70.56.93	2018-03-19 21:59:55
10.70.56.178-10.70.56.93	10.70.56.178	10.70.56.93	2018-03-19 21:59:57
10.70.56.11-10.70.56.183	10.70.56.11	10.70.56.183	2018-03-19 21:59:57
10.1.8.2-10.70.56.93	10.1.8.2	10.70.56.93	2018-03-19 22:00:09
10.70.56.200-239.255.255.250	10.70.56.200	239.255.255.250	2018-03-19 22:00:32
10.70.56.115-239.255.255.250	10.70.56.115	239.255.255.250	2018-03-19 22:00:38
10.70.56.93-193.255.88.9	10.70.56.93	193.255.88.9	2018-03-19 22:00:42
192.168.0.10-192.168.0.10	192.168.0.10	192.168.0.10	2018-03-19 22:00:46
10.70.56.178-239.255.255.250	10.70.56.178	239.255.255.250	2018-03-19 22:00:50
10.70.56.115-10.70.56.255	10.70.56.115	10.70.56.255	2018-03-19 22:01:00
10.70.56.255-10.70.56.57	10.70.56.255	10.70.56.57	2018-03-19 22:01:06
10.70.56.11-10.70.56.255	10.70.56.11	10.70.56.255	2018-03-19 22:01:25
10.70.56.178-10.70.56.255	10.70.56.178	10.70.56.255	2018-03-19 22:01:37
10.70.56.58-239.255.255.250	10.70.56.58	239.255.255.250	2018-03-19 22:01:55
10.70.56.11-239.255.255.250	10.70.56.11	239.255.255.250	2018-03-19 22:02:53
10.70.56.157-239.255.255.250	10.70.56.157	239.255.255.250	2018-03-19 22:03:09
10.70.56.178-10.70.56.93	10.70.56.178	10.70.56.93	2018-03-19 22:10:36
10.70.56.11-10.70.56.183	10.70.56.11	10.70.56.183	2018-03-19 22:10:36
192.168.0.10-192.168.0.10	192.168.0.10	192.168.0.10	2018-03-19 22:10:36
10.70.56.58-239.255.255.250	10.70.56.58	239.255.255.250	2018-03-19 22:12:28
10.70.56.178-10.70.56.93	10.70.56.178	10.70.56.93	2018-03-19 22:12:29
10.70.56.11-10.70.56.183	10.70.56.11	10.70.56.183	2018-03-19 22:12:29
10.70.56.115-10.70.56.255	10.70.56.115	10.70.56.255	2018-03-19 22:12:29
10.70.56.200-239.255.255.250	10.70.56.200	239.255.255.250	2018-03-19 22:12:29
10.70.56.178-239.255.255.250	10.70.56.178	239.255.255.250	2018-03-19 22:12:29
10.70.56.255-10.70.56.57	10.70.56.255	10.70.56.57	2018-03-19 22:12:29
192.168.0.10-192.168.0.10	192.168.0.10	192.168.0.10	2018-03-19 22:12:29
10.70.56.57-239.255.255.250	10.70.56.57	239.255.255.250	2018-03-19 22:12:29
10.70.56.11-10.70.56.255	10.70.56.11	10.70.56.255	2018-03-19 22:12:29
10.70.56.11-224.0.0.252	10.70.56.11	224.0.0.252	2018-03-19 22:12:29
10.70.56.160-239.255.255.250	10.70.56.160	239.255.255.250	2018-03-19 22:12:29
10.70.56.157-239.255.255.250	10.70.56.157	239.255.255.250	2018-03-19 22:12:29
10.70.56.83-239.255.255.250	10.70.56.83	239.255.255.250	2018-03-19 22:12:29

Figure 8: MySQL output of the logs.

V. CONCLUSION

SDN is a new networking architecture which is simple, flexible and manageable than the traditional network, and allows to develop applications to control the network. Today, logging and monitoring of the connected device (computers, mobile phones, etc.) use and user internet activities are very important for institutions or people who provide collective internet services. In this paper, a cost-effective SDN architecture equipped with logging and monitoring features has been designed and implemented. The proposed system is capable of creating a log and monitoring in real time.

As future work, we want to create a more secure network infrastructure by redirecting (feedback) the logs to the SDN controller.

REFERENCES

[1] B.A.A. Nunes et al. "A survey of software-defined networking: Past, present, and future of programmable networks," IEEE Commun. Surv. Tutor., vol. 16, no. 3, pp. 1617–1634, 2014.

[2] S. Bera, S. Misra, and A. V. Vasilakos, "Software-Defined Networking for Internet of Things: A Survey," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1994–2008, 2017.

[3] Official Gazette, Law No. 5651, The Internet Law of Turkey. Available: http://www.resmigazete.gov.tr/eskiler/2007/05/20070523-1.htm.

[4] N. McKeown et al., "OpenFlow: Enabling innovation in campus networks," ACM SIGCOMM Comput. Commun. Rev., vol. 38, no. 2, pp. 69–74, 2008.

[5] Project Floodlight. Accessed on Mar. 27, 2018. [Online]. Available: http://www.projectfloodlight.org/floodlight/.

[6]Raspbian for Raspberry Pi. Accessed on Mar. 27, 2018. [Online]. Available: https://www.raspberrypi.org/downloads/.

[7] Open vSwitch. Accessed on Mar. 27, 2018. [Online]. Available: http://www.openvswitch.org.

[8] sFlow, Accessed on Mar. 27, 2018. [Online]. Available: http://www.sflow.org/.

Optimizing Cellular Networks for Adaptive Video Streaming

Ali Fuat ALKAYA[1], Talha DEMIR[2], Burak GORKEMLI[3], and Sinan TATLICIOGLU[4]

[1]Marmara University Göztepe. Istanbul, Turkey, falkaya@marmara.edu.tr
[2]TUBİTAK, Gebze, Kocaeli, Turkey, demir.talha@icloud.com
[3]Argela, Istanbul, Turkey, burak.gorkemli@argela.com.tr
[4]Argela, Istanbul, Turkey, sinan.tatlicioglu@argela.com.tr

***Abstract* - In this paper we present a cross-layer optimization method for adaptive video streaming in cellular networks. The proposed method makes use of a new scheduling algorithm, tailored to adaptive video streaming applications, which provides better allocation of radio resources in order to increase the video quality experienced by the users. After introducing our novel scheduling algorithm, we compare its performance with the state-of-the-art, using DASH as the adaptive video streaming technique and LTE as the cellular network technology. The results show that introducing DASH-awareness into the network increases the streamed video quality and the network utilization. We believe that the proposed novel cross-layer optimization method can also be applied to future cellular network technologies, such as 5G, where application-aware resource allocation will be crucial in supporting the required quality-of-service (QoS).**

***Keywords* – LTE, DASH, MPEG-DASH, video streaming on cellular networks.**

I. INTRODUCTION AND BACKGROUND

ADAPTIVE video streaming is becoming a standard for many video streaming applications such as YouTube, Netflix, and Hulu [1]. A typical example is the dynamic adaptive video streaming over HTTP (DASH). On the other hand, the number of users connected to the internet with their mobile devices over cellular networks is increasing very fast [2].

In a typical DASH environment, the streaming client decides the quality of the video to be streamed by requesting the video stream that matches its available bandwidth. However, in cellular networks, it is observed that the client being the decision-maker results in inefficient usage of network resources and unfair allocation of available bandwidth [3]. Therefore, for a cell-based optimization over multiple concurrent DASH flows and fair allocation of resources, a scheduler should take the control and allocate resources. In our study, we assume the MAC scheduler in the eNodeB takes the charge since it has the required information for obtaining a cell-based optimum allocation of radio resources for its connected users.

In this study, we propose an intelligent, DASH-aware scheduling algorithm for adaptive streaming of video content for cellular users. Simulation results show that the proposed algorithm is able to improve the video quality experience by providing smoother video streams with more quality and increasing bandwidth utilization.

A. LTE Network Architecture

In an LTE network architecture, the base stations are called the eNodeBs. Due to the scarcity of radio resources, it is very important for the LTE system to allocate the available resources efficiently, in order to support as many users as possible, while maintaining a required QoS per user. Resource sharing is performed by the schedulers running on eNodeBs, whereas QoS is maintained by Admission Control (AC) and Congestion Control (CC) schemes.

The scheduler is a key element in wireless networks, where it is mostly the entity that determines the overall behavior of the system. It is the responsibility of the scheduler to share the available radio resources between different users to achieve an efficient resource utilization. The scheduler determines, for each 1 ms subframe, which users are allowed to transmit using which frequency resources at what data rate. The radio resource that is available for a user in the downlink 3GPP LTE system is defined in both frequency and time domains and is called a resource block (RB). LTE defines a number of channel bandwidths. The channel bandwidths that have been chosen for LTE are 1.4, 3, 5 10 15 and 20 MHz whereas their corresponding number of resource blocks are 6, 15, 25, 50, 75 and 100.

In this system, each user reports its instantaneous downlink channel conditions (e.g. channel quality indicator, CQI) to the serving eNodeB at each transmit time interval (TTI) where higher CQI values usually indicate better channel quality.

A generalized model of the packet scheduling algorithm in the downlink 3GPP LTE system is given in Fig. 1. From the figure, it can be seen that each user is assigned a buffer at the serving eNodeB. Packets arriving into the buffer are time stamped and queued for transmission based on a first-in-first-out basis. At each TTI, the packet scheduler determines which users are to be scheduled based on a packet scheduling algorithm. In this system, there is a possibility that a user may be allocated zero, one or more than one RBs at each TTI as shown in Fig. 1.

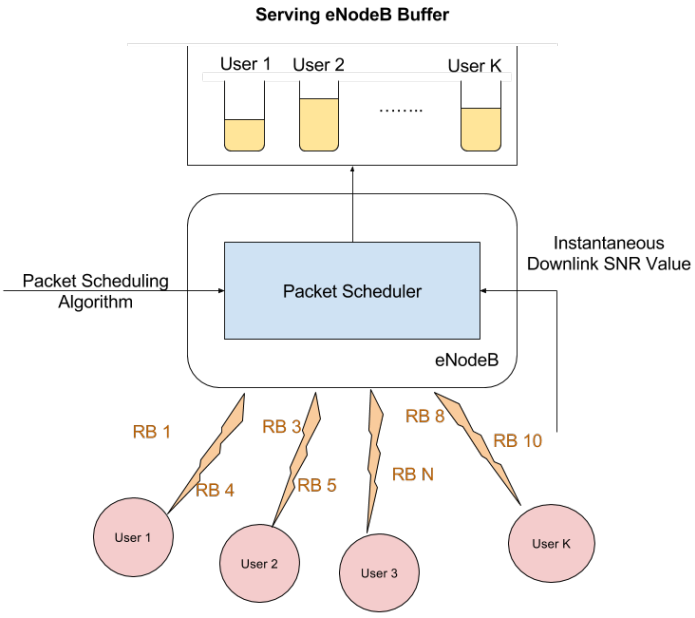


Figure 1. Working Mechanism of Packet Scheduler in the Serving eNodeB

B. DASH

Dynamic Adaptive Streaming over HTTP (DASH), also known as MPEG-DASH, is an adaptive bitrate streaming technique that enables high quality streaming of media content over the Internet delivered from conventional HTTP web servers. DASH works by breaking the video content into a sequence of small HTTP-based file segments, each segment containing a short interval of playback time of content that is potentially many hours in duration, such as a movie or the live broadcast of a sports event. The content is made available at a variety of different bit rates, that is, alternative segments are encoded at different bit rates covering aligned short intervals of play back time. The available segments of the video file and other information including the duration of each segment are packed in a special XML based manifestation file, having ".mpd" extension. Once the request for a video stream is received by the server, it starts the streaming process by sending the manifestation file first.

While the content is being played back by a DASH client, the client monitors the current network conditions and selects the segment to request next from the available segments stated in the manifestation file. The client selects the segment with the highest bit rate possible that can be downloaded without causing a pause in the video playback. Thus, a DASH client can seamlessly adapt to changing network conditions, and provide high quality play back with fewer re-buffering events [4]. Table 1 describes the amount of bandwidth and number of resource blocks that each bit rate requires [5].

In the literature, very few studies consider DASH over LTE. Wei et al. offer that a wireless DASH (WiDASH) proxy concept which locates at the edge of Internet and wireless networks [6]. WiDASH proxy splits a long TCP connection between DASH server and wireless user into one wired TCP and multiple parallel wireless TCPs. Sánchez et al. studied the benefits of adopting the DASH based on the Scalable Video Coding (SVC) as key technology for spreading video streaming over the Internet and making it a ubiquitous application [7]. Galanopoulos et al. present an optimization framework that formalizes the inherent trade-off between the user's perceived quality of wireless video and the energy consumption cost of the network [8].

TABLE 1. AMOUNT OF BANDWIDTH AND NUMBER OF RESOURCE BLOCKS EACH BITRATE LEVEL REQUIRES

Type	Video Bitrate, Standard Frame Rate (24, 25, 30)	# of average required resource blocks in a TTI
2160p (4k)	35-45 Mbps	87-112
1440p (2k)	16 Mbps	40
1080p	8 Mbps	20
720p	5 Mbps	12
480p	2.5 Mbps	6
360p	1 Mbps	3

Our study is more similar to [8] in the context of LTE resource allocation. However, in that study the main focus is minimizing the energy consumed by the base stations. Also, the bit rate levels and assignment of RBs to users are not elaborated. On the other hand, we allocate resources according to bit rate levels stated in manifestation files, with the objective of distributing good quality to UEs in view of the fact that their minimum requirements and fairness policy.

The rest of the manuscript is as follows: Section 2 gives the details of the proposed cross-layer optimization algorithm. Section 3 gives the results of the simulation studies and section 4 presents the conclusive remarks together with some possible future works.

II. PROPOSED SCHEDULING ALGORITHM

A. Model

In our scenario, there are a number of users simultaneously watching DASH video content over LTE network. We assume that the eNodeB is capable of parsing the manifestation file of each DASH streamer and based on the current network conditions and the content of the manifestation file, it assigns resource blocks to users at each time slot.

Eventually, for each time slot the problem turns out to have the objective of choosing a rate vector that maximizes a weighted sum of all users' rates.

In that system suppose that there are U number of users and N number of resource blocks. Let x_{un} be the binary variable having value of 1 if resource blocks is assigned to a user u, 0 otherwise. Then we can write the optimization model as follows:

$$\text{Maximize } \sum_{u=1}^{U}\sum_{n=1}^{N} w_u x_{un} \quad (1)$$

$$\sum_{u=1}^{U} x_{un} \le 1 \qquad n = 1, \cdots, N \quad (2)$$

$$\sum_{n=1}^{N} x_{un} \le h_u \qquad u = 1, \cdots, U \quad (3)$$

$$\sum_{n=1}^{N} x_{un} \ge l_u \qquad u = 1, \cdots, U \quad (4)$$

$$pt_u - ct_u \le 1 \qquad u = 1, \cdots, U \quad (5)$$

Where w_u is some kind of weight (such as the SNR or CQI values) assigned to user u, h_u is the number of resource blocks

that user u requires at maximum, and l_u is the number of resource blocks that user u needs to watch at the lowest bitrate written in its mpd file. Equation (2) states that each resource bloks can be allocated to at most one user, (3) and (4) state that each user must be allocated the resource blocks required at most and at least, respectively. Equation (5) states that the bitrate type should not drop more than one level where pt_u is the bitrate type at the previous subframe and we assume that it is given, whereas ct_u is the current bitrate type assigned to user u which and defined in the following discrete function:

$$ct_u = \begin{cases} 1 & if\ 0 \le \sum_{n=1}^{N} x_{un} \le 3 \\ 2 & if\ 4 \le \sum_{n=1}^{N} x_{un} \le 6 \\ 3 & if\ 7 \le \sum_{n=1}^{N} x_{un} \le 12 \\ 4 & if\ 13 \le \sum_{n=1}^{N} x_{un} \le 20 \end{cases} \quad (6)$$

Apparently, equation 6 is derived by making use of Table 1.

B. Algorithm

The algorithm that we used in our test scenarios assigns RBs to users as follows:

1. Given the RB demands of each user (calculated based on the demanded bitrates) and total number of RBs to be allocated for DASH users, give the amount of RBs to each user to watch their stream content at the lowest bitrate defined in the manifestation file.
2. For assigning the remaining unassigned RBs, the scheduler selects the lowest bit rate among next level bit rates of UEs by reading from their MPD files and it determines this lowest bitrate as next upgrade bitrate level. In this way, scheduler updates the target throughput table and selects a new next upgrade bitrate level every millisecond.
3. In the next TTI, scheduler first allocates RBGs to appropriate UEs according to the target throughput table and the CQI values reported by UEs. Then, again it compares for the lowest bitrate among the next bit rate levels of UEs to update target throughput table and this process continues till all UEs reach max bit rates specified in their manifestation files or all RBG are allocated for UEs.

This procedure uses the content of the mpd files of the DASH streaming users and can be considered as a DASH-aware algorithm. The rationale behind this procedure is allocating enough resources to each user so that all of the users can be streamed at least at the lowest bit rates as stated in their manifestation files. On the other hand, if any UE cannot exceed its target throughput in 100 TTI, target bit rate of UE is downgraded the previous bit rate level and target throughput table and next upgrade bitrate level of UE are updated to proper values.

III. SIMULATIONS, RESULTS AND DISCUSSION

Computational tests are conducted on the ns-3 simulation tool running on a computer with Intel i7 processor, 32 GB RAM and 1TB HDD. The tests are realized with 3 and 5 users who are streaming DASH application. We assume that the server streams data towards all UEs as much as possible as if the client requests proper DASH contents consecutively. Besides, we also assume that the data bit rate of UEs are controlled by the MAC scheduler.

Our MAC scheduler is implemented over a PSS (Priority Set Scheduler). PSS controls the fairness among the UEs by a specified target bit rate. Technically, resource blocks are assigned as groups of four resource blocks in LTE technology [9]. PSS assigns available RBGs (Resource Block Groups) to DASH UEs with a metric that is directly proportional to (i) target bit rate which represents target throughput level (ii) achievable bit rate from the RBG (which is calculated according to the CQI) thus providing fairness and is inversely proportional to the square of the last average bit rate (represents previous status of the UE). On the other hand, in addition to the aforementioned metric factors, the proposed MAC scheduler controls the UEs' target bit rates dynamically and gradually according to the last average throughput of UEs.

In our scenario, different stable conditions were maintained during every 10 seconds. In the scenario, the number of RBGs allocated to UEs are 13, 10, 8, 6, 4, 2, 4, 6, 8 and 10 for 10 consecutive 10 second intervals (from 1 to 100). We assume that there are up to five users streaming DASH content from five different servers hosted in YouTube, Twitch, Netflix, Dailymotion and iTunes, respectively. The bit rates that servers are capable of streaming are taken from their official systems. We compared our proposed scheduling algorithm with Proportional Fair Scheduler (PFS). In PFS, each UE's instantaneous channel quality is high relative to its own average channel condition over time and PFS divides RBGs to each UE equally. As a result, all of the UE throughput values become approximately the same. Throughput request levels are not important for PFS.

In our first test scenario, there are three UEs streaming DASH video. First we apply PFS scheduling policy (Fig. 2a) and then we apply our proposed scheduling policy (Fig. 2b). We observe that when PFS scheduling policy is applied, every UE gets the same throughput as expected but not all UEs need that much throughput. Then we apply our DASH-aware scheduling policy and we observe that UE1 gets a larger throughput than the other UEs and enjoys a better video quality while the other DASH streaming UEs keeping their video quality unaffected even though they have less throughput with respect to what they had when PFS scheduling policy was applied. When we run the same test scenarios with four and five DASH streaming users we get the similar results as shown in Fig. 3.

IV. CONCLUSION

In this work we propose a DASH-aware scheduling policy, which aims to distribute limited radio resources intelligently among clients so that some clients, if not all, can stream video at higher bitrates, compared to the case when a DASH agnostic scheduling policy is used. The proposed algorithm has the knowledge of all DASH content to be streamed by the clients, and it utilizes it to distribute the radio resources accordingly, so that the clients can receive higher quality video, as shown in the results. Other algorithms that do not have this knowledge, like PFS, may end up assigning resources to clients that cannot increase the received video quality. It may be argued that providing the DASH knowledge to the scheduler is a challenging issue, but we believe the realization of 5G technology will boost the network intelligence, enabling us to manage the network with increased application-awareness.

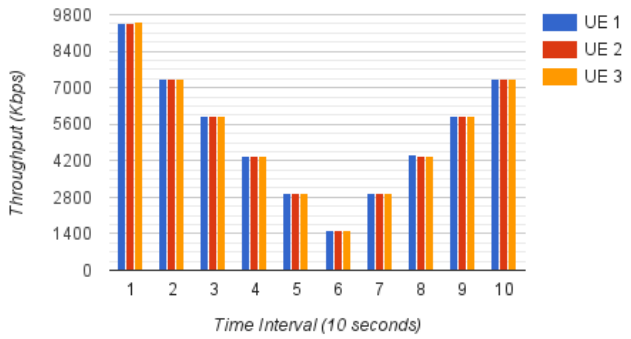


Figure 2a. Test results when PFS scheduling policy is applied (for three UEs).

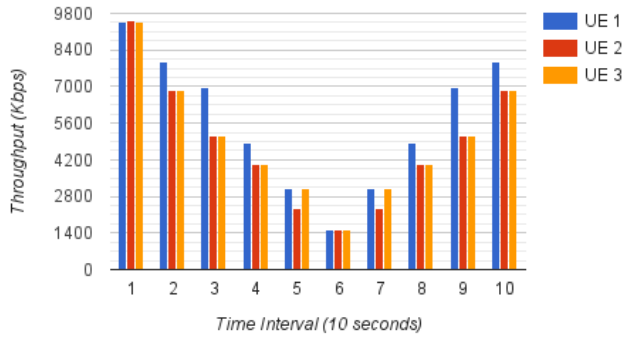


Figure 2b. Test results when our new DASH-aware scheduling policy is applied (for three UEs).

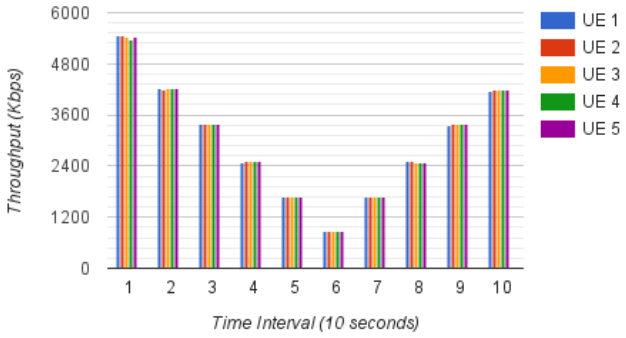


Figure 3a. Throughput distribution of the five DASH streaming UEs when PFS scheduling policy is applied.

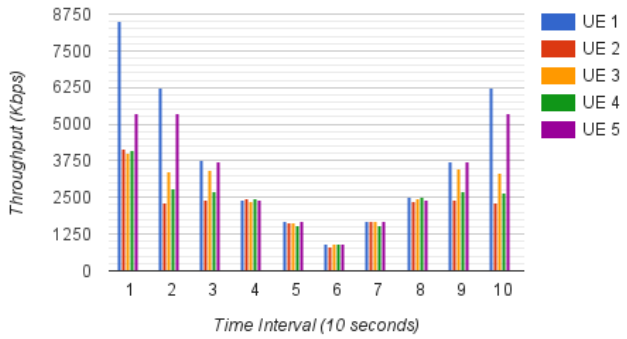


Figure 3b. Throughput distribution of the five DASH streaming UEs when our new DASH-aware scheduling policy is applied.

REFERENCES

[1] DASH – It's Happening All Around You. [Online]. Available: https://www.qualcomm.com/news/onq/2014/04/02/dash-its-happening-all-around-you.

[2] Cisco white paper, Cisco visual networking index: Global mobile data traffic forecast update, 2015-2020. [Online]. Available:http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[3] Akhshabi, Saamer, et al. "What happens when HTTP adaptive streaming players compete for bandwidth?." Proceedings of the 22nd international workshop on Network and Operating System Support for Digital Audio and Video. ACM, 2012.

[4] Information technology — Dynamic adaptive streaming over HTTP (DASH), INTERNATIONAL STANDARD ISO/IEC 23009-1 Second edition 2014-05-15

[5] Recommended upload encoding settings, YouTube Help. [Online]. Available:https://support.google.com/youtube/answer/1722171?hl=en

[6] Wei Pu, Zixuan Zou, and Chang Wen Chen, "Video adaptation proxy for wireless dynamic adaptive streaming over HTTP" Proceedings of 2012 IEEE 19th International Packet Video Workshop, Munich, Germany, May 10-11, 2012.

[7] Yago Sánchez, Cornelius Hellge, Thomas Schierl, Werner Van Leekwijck, Yannick Le Louédec, "Scalable video coding based DASH for efficient usage of network resources", InThird W3C Web and TV Workshop 2011 Sep (pp. 19-20).

[8] A. Galanopoulos, G. Iosifidis, A. Argyriou and L. Tassiulas, "Green video delivery in LTE-based heterogeneous cellular networks," World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2015 IEEE 16th International Symposium on a, Boston, MA, 2015, pp. 1-9.

[9] LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures, 3GPP TS 36.213 version 13.0.0 Release 13, 2016-05.

Detection of EEG-Based Motor Imagery Tasks with 1D-Local Binary Pattern (LBP) Features

F. KUTLU ONAY[1] and C. KOSE[2]

[1] Amasya University, Amasya/Turkey, funda.kutlu@amasya.edu.tr
[2] Karadeniz Technical University, Trabzon/Turkey, ckose@ktu.edu.tr

Abstract **- EEG signals are commonly used data sources in BCI applications. For this reason, recent studies to analyze the EEG signals in the most accurate way are increasing rapidly. When features are extracted from EEG signals, the use of methods sensitive to local variations is of great importance for correct classification of the signals. In this study, 1D-local binary pattern (LBP) method which is sensitive to local changes was applied to motor imager/movement EEG signals and the obtained features were classified with the k-NN and SVM classifiers. Accordingly, in the case of using the k-NN method, the lowest 99.98%, and highest 100% classification accuracy was obtained.**

Keywords **– motor imagery, one-dimension local binary pattern, EEG, k-NN, SVM.**

I. INTRODUCTION

PEOPLE have to communicate with the environment in order to continue their lives. However, some people may not have enough physical conditions to perform this interaction. The normal ways to perceive and express oneself may be lost or damaged by some diseases or accidents. For this reason, alternative systems called BCI are being produced, developed and become popular research topics in recent times. Researchers have focused primarily on developing signal acquisition techniques to understand neurophysiological activity in the human brain and optimizing signal processing techniques for faster and more accurate BCI systems.

BCI systems are classified as invasive and non-invasive according to the used signal acquisition techniques. Among these methods, EEG is preferred because of its low cost and risk and portability advantages. In addition, the temporal resolution of EEG signals is very high, which has been effective in its use in online BCI applications [1].

The neurological mechanisms used in BCI applications may be a reaction that occurs as a result of an event in the brain, or it may be an activity independently produced by an object. In BCI applications, these activities are transformed into control signals to interact with the outside world. In this study, motor images and EEG signals recorded during motor movement were used.

Since the EEG signals are very long, it is time consuming and costly to process whole of the signal. For this reason, effective feature extraction methods are needed to find features that can best represent the properties of the signal. In this context, it may be easier to classify the signals if methods that are sensitive to local changes of the signal are used and the 1D-LBP method is one of these methods. In the existing literature, there are successful studies for 1D-LBP method of EEG signals which were recorded various purposes (such as epileptic seizure detection, Alzheimer's disease, Parkinson's disease, sleep apnea syndrome, biometric identification, and mental activity detection etc.).

In this study, the 1D-LBP method was used to extract features from EEG signals recorded for 4 different motor imagery/movement tasks. The k-NN and SVM methods were used as the classifier and the average classification accuracy was 100% and 97.20%, respectively. In addition, all of the recording channels were not used, and channel selection was made using the developed clustering method.

II. METHODS

A. Details of the Dataset

Physiobank Motor Imagery/Movement (MMI) EEG dataset [2-3] were recorded from 64 channels which were placed according to the 10-10 system and digitized with 160 Hz sampling frequency. This dataset was created for 6 different tasks with the participation of 109 volunteers. Two of these are baseline tasks recorded with eyes open and closed. The 3rd and 4th tasks were obtained respectively during the opening and closing of the right or left fist and the imagination of this movement. The 5th and 6th tasks were recorded during the opening and closing of two fists or two feet and the imagination of this movement, respectively. Each subject repeats this process with 14 runs.

B. 1D-Local Binary Patterns

In the feature extraction process, the feature vector is obtained by extracting the related data from raw data. Thus, the data is represented in a smaller size. In this study, the 1D-LBP method was used to represent the EEG signals. In general, the LBP method is a commonly used transformation method in image processing applications due to its sensitivity to local variations, structural stability, and high classification success [4]. The development of 1D-LBP is completely based on the working principle of a 2D-LBP method and has recently been used for the analysis of non-stationary and non-linear biomedical signals. Because local changes in the EEG signal may contain important information about the entire of the signal and should be examined [5].

In the production of the 1D-LBP binary sequences, the

current sample in the signal (S) is compared to its P/2 neighbors on the left and right. The decimal values of the binary sequences generated and this value is the new value representing the corresponding instance. The mathematical equations of the 1D-LBP methods are as follows: Equations (1-3) [6]:

$$i \neq c \quad \text{and} \quad t = S_i - S_c, \tag{1}$$

$$LBP_P = \sum_{i=1,}^{P} F(t)2^{i-1}, \tag{2}$$

$$F(t) = \begin{Bmatrix} 1, t \geq 0 \\ 0, t < 0 \end{Bmatrix} \tag{3}$$

Where P is the number of neighbors and in this study, it was set to 8. In there, the current sample will be compared with its 4 neighbors on the right and left. In the comparison process, the current sample value is subtracted from the neighbor sample values. If this difference is greater than and equal to 0, the new value of that neighbor is set to 1 and otherwise, it is set to 0 [7]. For each instance in the signal, there are now 8-digit binary sequences and their decimal equivalents. These decimal values are in the range of 0-255 for 8 neighborhoods. A sample taken from the dataset used in the study and the calculation of the 1D-LBP result is simulated in Figure 1. The histogram information obtained from the frequency distributions of the values in the new dataset is used as features since it is distinctive for the data.

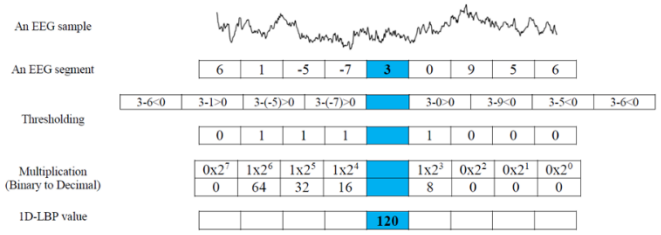


Figure 1: Demonstration of the 1D-LBP method on a sample.

C. Classification

Classification is the assignment of correct class labels to the features obtained from the data. In the study, k-nearest neighbors (k-NN) and support vector machines (SVM) methods, which are widely used and successful methods, were used to measure the classification accuracy.

According to the k-NN method, the label value is determined by looking at the k-neighbors of the current instance. The label value of the majority of k-neighbors of the instance is found, in this case, the same value is assigned to this. In order to classify correctly, the most appropriate value of k should be determined and the most conventional method for this is cross-validation. The distance is an important parameter when the distance between neighbors is measured. Several methods of distance measurement have been used in the literature [8]. Among these, the highest classification performance is the Euclidean distance method. Euclidean distance between the X and Y points considered as N-dimensional (d) Equation (4):

$$d = \sqrt{(X_1 - Y_1)^2 + \ldots + (X_N - Y_N)^2} \tag{4}$$

SVM is one of the other successful classification methods. It can be adapted to all linear and non-linear classification problems and multi-class problems. The real-time applications are not suitable for linear classification. In this case, the data are defined in input space are mapped to the higher dimensional space by the kernel functions and can be linearly separated in the new space. It is important to find appropriate kernel function and its optimal parameters. Radial basis function (RBF) is a kernel function which is the most used and successful kernel in the existing literature [9]. RBF kernel is expressed as Equation (5).

$$K(x, y) = e^{-\gamma |x - x_i|^2} \tag{5}$$

Here, γ refers to kernel size.

III. Results and Discussion

A. The channel reduction

The EEG dataset used in the study was recorded from 64 channels. However, to disable channels that are not related to motor movements, not all 64 channels were used. Unlike the conventional channel selection methods [10,11] available in the literature, channel reduction was performed by combining the channels that control the same brain activity and taking the averages. Table 1 lists the functions and their control regions [12]. Within this information, the channels were combined and the number of channels were reduced from 64 channels to 21. The combined channels in the same region were indicated by the same color in Figure 2.

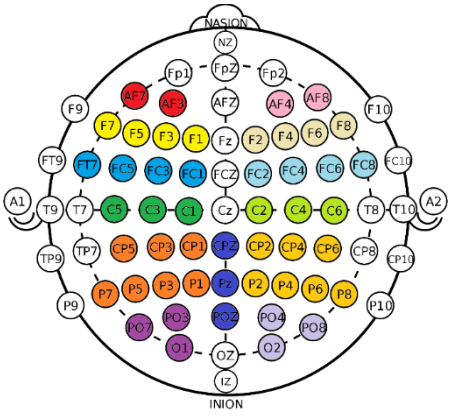


Figure 2: The combined channels on the 10-10 system.

Table 1. Brain functions and their control areas.

Brain lobe	Related function(s)
Frontal Lobe	with reasoning, planning, parts of speech, movement, emotions, and problem solving
Parietal Lobe	movement, orientation, recognition, perception of stimuli
Temporal Lobe	perception and recognition of auditory stimuli, memory, and speech

Occipital Lobe	vision
Cerebellum	coordination

All of these 21 channels will not participate in the classification. Because using all of the channels does not always improve the performance of the classification. For this reason, the channels which are related to the problem were selected by inspiration from the literature [10,11] from the obtained 21 channels. Accordingly, it was decided to use 13 channels in the study and the channels used in obtaining the 13 electrodes were shown in Figure 3.

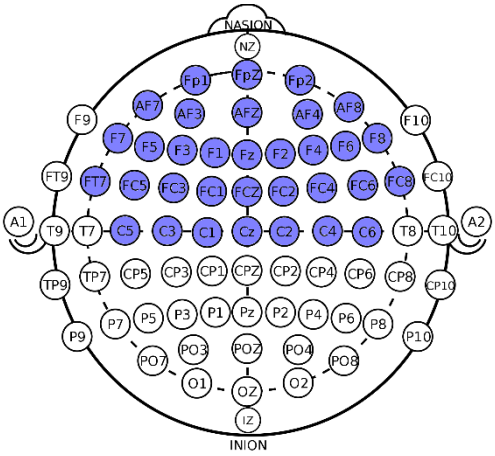


Figure 3: The electrodes used to form 13 electrodes.

B. 1D-LBP Features

In the used dataset, records were taken for 6 different tasks, but in this study, 4 tasks were taken into consideration. Accordingly, in the following stages, the motor imagery/movement tasks will be mentioned with new notations: RFist (right fist), LFist (left fist), BFist (both fist), and BFeet (both feet). 1D-LBP was applied to each task data and histogram features were obtained. The distribution graph for four classes of histograms were given in Figure 4 (a-d).

C. Classification Results

In this section, the selected optimal parameters for used classifiers and the classification results will be described.

For the K-NN, 10-fold cross validation (10-FCV) method was used with the optimum k parameter determined and this value was searched for between 1-25. In addition, Euclidian distance metric was used as the most successful method in the literature. Accordingly, it was found that all classification accuracies were obtained with k = 1 and the highest one was 100% for RFist – LFist, BFist – LFist and LFist – ALL matchings. For other matchings, the classification accuracies were found around 99%. For this reason, it can be concluded that the k-NN classifier for this study is a successful classification method.

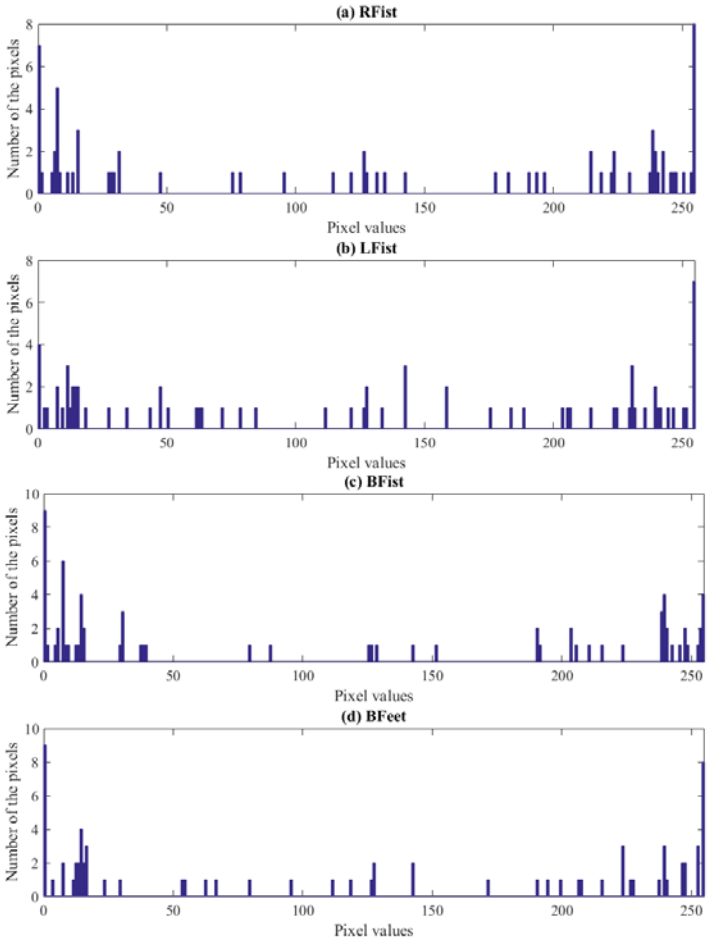


Figure 4: Histogram distributions of the (a) right fist (b) left fist (c) both fists (d) both feet.

Another classifier was RBF-based SVM and the best σ value affecting the γ parameter for the RBF kernel was searched between 0.1-2.5 by the 10-FCV method. Accordingly, the best classification performance was achieved for BFeet-RFist classification with σ = 2.5, and average classification accuracy (ACA), specificity and, sensitivity was obtained 97.2 (±0.72) %, 100% and %94.40, respectively for classification of BFeet and RFeet.

In summary, 4 tasks were selected in this study. For each task histogram information was obtained and used as the feature vector and applied to the classifier. The classification results were obtained by mapping for each task. Accordingly, the highest performance was found to be 100% with k-NN.

D. Comparison of the study with others

The results of the studies using the 1D-LBP method on motor imagery signals were given in Table 2 to see where this study is in the literature. As a result, it is seen that the achievement of this study is the highest with 100%.

Table 2: The classification results of the studies which used 1D-LBP method.

Author(s)	Dataset	Classifier(s)	ACA (%)
This study	Physiobank MMI	k-NN	100
		MLANN	100
Xu et al. [13]	ECoG-based MMI	Ordinary least squares	95
Turk and Ozerdem [14]	5 mental tasks	k-NN	87.30

As another comparison, the study results using the same dataset in different ways were given in Table 3. Accordingly, this study is one of the two studies that yield the highest classification result.

Table 3: The classification results of the studies worked with the same dataset.

Author(s)	Method(s)	Classifier(s)	ACA (%)
This study	1D-LBP	k-NN	100
		MLANN	100
Starkey et al. [15]	FWSOM	SVM	100
Kim et al. [16]	STFT	LDA	94.20
Lee et al. [17]	EMD	RF	80.05
Gopan et al. [18]	Kurtosis, IQR and MAD	k-NN	77.92

IV. CONCLUSION

The classification of EEG-based MMI signals is a popular research topic in recent years and many successful methods have been proposed in these studies. For the classification of these signals, there are many studies that obtain high classification accuracy usually by extracting features with sub-bands in the frequency domain. Contrary to these studies, in this study, the 1D-LBP method is applied which reveals the selective properties of the signal in time-space. This method, which is known to give successful results in biomedical signal processing studies, has been used to classify 4 different motor movements.

Another issue highlighted in the study was that not all channels need to be involved in classification. For this, a new EEG-based channel selection algorithm was proposed. Rather than applying conventional channel selection algorithms, a new channel editing operation was performed. New channels were obtained by combining and averaging channels that control the same functions. This was done by looking at the literature that controlled the function.

Finally, all these results suggest a successful feature extraction and channel reconstruction method to classify motor imagery EEG signals recorded for 4 different tasks with 100% classification accuracy.

REFERENCES

[1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.

[2] G. Schalk, D. J. Mcfarland, T. Hinterberger, N. Birbaumer, J. R. Wolpaw, and D. J. Mcfarland, "BCI2000: A General-Purpose Brain-Computer Interface (BCI) System," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, 2004.

[3] A. L. Goldberger *et al.*, "PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals.," *Circulation*, vol. 101, no. 23, pp. E215-20, Jun. 2000.

[4] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns."

[5] N. Chatlani and J. J. Soraghan, "Local binary patterns for 1-D signal processing," 2010.

[6] Ö. F. Ertuğrul, Y. Kaya, and R. Tekin, "A novel approach for SEMG signal classification with adaptive local binary patterns," *Med. Biol. Eng. Comput.*, vol. 54, no. 7, pp. 1137–1146, Jul. 2016.

[7] B. Amir, *Control Engineering and Applied Informatics*, vol. 16, no. 4. 1999.

[8] J. Walters-Williams and Y. Li, "Comparative Study of Distance Functions for Nearest Neighbors," in *Advanced Techniques in Computing Sciences and Software Engineering*, Dordrecht: Springer Netherlands, 2010, pp. 79–84.

[9] R. Sharma and R. B. Pachori, "Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions," *Expert Syst. Appl.*, vol. 42, no. 3, pp. 1106–1117, Feb. 2015.

[10] J. Yang *et al.*, "Channel selection and classification of electroencephalogram signals: An artificial neural network and genetic algorithm-based approach," *Artif. Intell. Med.*, vol. 55, no. 2, pp. 117–126, Jun. 2012.

[11] L. He, Y. Hu, Y. Li, and D. Li, "Channel selection by Rayleigh coefficient maximization based genetic algorithm for classifying single-trial motor imagery EEG," *Neurocomputing*, vol. 121, pp. 423–433, Dec. 2013.

[12] "Brain Structures and Their Functions." [Online]. Available: http://serendip.brynmawr.edu/bb/kinser/Structure1.html. [Accessed: 07-Mar-2018].

[13] F. Xu, W. Zhou, Y. Zhen, Q. Yuan, and Q. Wu, "Using Fractal and Local Binary Pattern Features for Classification of ECOG Motor Imagery Tasks Obtained from the Right Brain Hemisphere," *Int. J. Neural Syst.*, vol. 26, no. 6, p. 1650022, Sep. 2016.

[14] O. Turk and M. S. Ozerdem, "Mental activity detection from EEG records using local binary pattern method," in *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017, pp. 1–4.

[15] M. Arvaneh, C. Guan, K. K. Ang, and C. Quek, "Robust EEG channel selection across sessions in brain-computer interface involving stroke patients," in *The 2012 International Joint Conference on Neural Networks (IJCNN)*, 2012, pp. 1–6.

[16] H. S. Kim, M. H. Chang, H. J. Lee, and K. S. Park, "A comparison of classification performance among the various combinations of motor imagery tasks for brain-computer interface," in *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, 2013, pp. 435–438.

[17] K.-B. Lee, K. K. Kim, J. Song, J. Ryu, Y. Kim, and C. Park, "Estimation of Brain Connectivity during Motor Imagery Tasks using Noise-Assisted Multivariate Empirical Mode Decomposition," *J. Electr. Eng. Technol.*, vol. 11, no. 6, pp. 1812–1824, Nov. 2016.

[18] Gopika Gopan K, N. Sinha, and Dinesh Babu J, "Statistical feature analysis for EEG baseline classification: Eyes Open vs Eyes Closed," in *2016 IEEE Region 10 Conference (TENCON)*, 2016, pp. 2466–2469.

Development of A Multi-Objective Optimization Via Simulation Approach for Inventory Control Systems and Supplier Selection

A. T. DOSDOGRU[1], A. BORU[1], and M. GOCKEN[1]

[1]Adana Science and Technology University, Adana/Turkey, adosdogru@adanabtu.edu.tr
[1]Adana Science and Technology University, Adana/Turkey, aboru@adanabtu.edu.tr
[1]Adana Science and Technology University, Adana/Turkey, mgocken@adanabtu.edu.tr

Abstract **- Inventory control models are one of the most mature fields in the area of supply chain management but it is not obvious how to extend inventory control models under different types of problems where each problem has specific features and includes many different limitations and constraints. In addition, the dynamism of the real-world interactions occurring in inventory control systems should be modeled with much details, realities, and complexities. At this point, Simulation Optimization (SO) models along with modern computing power provide a significant opportunity to respond effectively in a dynamic and stochastic environment. To remain competitive and viable in today's business climate, SO can be explicitly taken into account as a complementary tool. In this study, SO is used to develop a multi-objective search for the inventory control system with supplier selection where objectives of the SO model are minimizing total supply chain cost and maximizing average service level. The optimal values of the initial inventory, reorder point, and order-up-to level for supply chain members are determined by proposed SO model. The proposed model has an ability to analyze dynamics of the supply chain members and to transform it into a clear structure for supply chain decisions. This is very important in real-world settings because proposed models lead to a large number of variants to manage a wide range of decisions in the supply chain.**

Keywords **- Inventory control systems, Simulation Optimization, Multi-objective optimization problem.**

I. INTRODUCTION

Competition among companies has become fiercer and fiercer in recent years. Therefore, the new economic environment is described by short product life cycles, increased customer demand for customized goods, quick response to customers' wants and needs in terms of quality and quantity. Considering the basic characteristic of the current market economy, it can be said that a continuous feedback is needed between business strategy and performance measurement systems. Related with performance measurement systems, many researchers feel that continuous improvement in an organization depends on "measuring, measuring, and measuring again" [1]. Each individual organization can have specific performance measures reflecting its fundamental purpose and its environment. Hence, advantages, disadvantages, and limitations of each performance measures should be analyzed to increase the supply chain performance.

Beamon [2] emphasized that researchers have mostly used two different performance measures for supply chain models: (1) cost; and (2) a combination of cost and customer responsiveness. Without doubt, cost is one of the most effective performance measures of the supply chain. Companies always strive to find ways to minimize cost. To track and manage cost, financial performance measures are generally utilized. This help companies to see where the money is spent [3]. Performance measures based on cost can be found in [4].

Customer responsiveness is related with being prompt and right in responding to customers. Being right is a value that is clear - customers obtain the thing that ensures their needs and wants. On the other hand, the value also depends considerably on the pace with which the response is fulfilled [5]. Performance measures based on customer responsiveness can be found in [4].

The cost-responsiveness performance of supply chain including measures based on cost and customer responsiveness can be represented with efficient frontier. Although a basic strategic choice for supply chain members is the level of responsiveness they seek to ensure, companies must make a trade-off between cost and responsiveness [6]. Herein, inventory management problems can be considered as one of the multi-objective optimization problem where optimal inventory can be determined under conflicting objectives. Multi-objective problem includes simultaneous optimization problems with at least two objective functions which are conflicting with each other. The solutions of this kind of problem are typically provided in the Pareto-optimality sense. Thus, they are known as Pareto optimal and also denoted as non-inferior or efficient or non-dominated. The solutions cannot be improved in any objective function without causing degradation in at least one other objective. Multiple Pareto-optimal solutions may be available for the multi-objective problem but two Pareto-optimal solutions cannot be comparable for a given problem without some preference orderings that depend on the decision maker. It is desirable to determine all the Pareto-optimal solutions since different decision makers with different preferences can prefer different Pareto-optimal solutions [7]. Assume that the problem at stake has three objectives whose objective space $\{Obj_1, Obj_2, Obj_3\}$. A solution is efficient of order 2 if it is Pareto efficient in all the three subspaces of the objectives taken 2 at the time, namely $\{Obj_1, Obj_2\}$, $\{Obj_1, Obj_3\}$, and $\{Obj_2, Obj_3\}$[8].

In recent years, most of the supply chain optimization problems include multi-objective models. Among these models, nondominated sorting genetic algorithm II (NSGA-II) is the most commonly used model for simulation-based optimization [9]. NSGA-II outperforms over other multi-objective genetic algorithm variants by using a fast nondominated sorting procedure, an elitist-preserving approach and a parameter-free sharing operator [10]. By maintaining a diverse population of members throughout their search, NSGA-II is able to provide Pareto optimal solutions that satisfy multiple conflicting objectives [11]. In this study, NSGA-II provides a multi-objective search for the inventory control problem with supplier selection where objectives are total supply chain cost minimization and average service level maximization.

Of the diverse operations involved among the supply chain members, purchasing is one of the most significant activities because it gives us the opportunity to reduce total supply chain cost. Certainly, supplier selection is arguably the single most important task within the purchasing function [12]. Hence, researchers have continued to investigate models including different aspects of the supplier selection and inventory control system. Basnet and Leung [13] solved the multi-period inventory lot sizing problem using an enumerative search algorithm and heuristic model where one or more supplier can be selected in each of periods for the purchase of products. Haq and Kannan [14] considered not only multi-echelon inventory control model but also supplier selection in built to order supply chain system using fuzzy analytical hierarchy process and Genetic Algorithm (GA). In the study, unlimited supplier capacity and deterministic demand are considered. Moghadam et al. [15] used a fuzzy neural network to forecast the periodical demand rates. Also, GA is utilized to improve the inventory control models based on which the proper suppliers are determined. Mendoza and Ventura [16] used a mixed integer nonlinear programming model to solve stationary inventory control policy and supplier selection under serial supply chain system where constant demands must be met without shortage. Choudhary and Shankar [17] determined the most proper suppliers, carriers, and optimal timings and procurement lot-sizes simultaneously using an integer linear programming model. Choudhary and Shankar [18] addressed the development of multi-objective integer linear programming model that allows the decision maker to find which supplier to procure from, how much to procure from the selected suppliers and in which periods, and which carrier is to be chosen for delivering procured items over the planning horizon. Shadkam and Bijari [19] presented generalized data envelopment analysis, discrete event simulation, and cuckoo optimization algorithm for the supplier selection and order quantity determination. Pazhani et al. [20] addressed the supplier selection and inventory control problem simultaneously using mixed integer nonlinear programming model that determined the optimal inventory levels for each supply chain stage. At the initial stage, allocation of orders among the suppliers is specified. Then, shipments from the suppliers to the manufacturer and between the stages using truckloads are defined in the study.

In the light of previous studies, it can be said that multi-objective SO are widely used to solve complex supply chain problems, but not many SO based solution methodologies exist to investigate inventory control problem and supplier selection simultaneously. We used SO model to identify the values of reorder point, order-up-to level and initial inventory in distribution center (DC)s and Suppliers, and to determine the proper set of Suppliers for DCs. To the best of our knowledge, there is no clear methodology is available related with multi-objective (R, s, S) policy and supplier selection using SO model that applies the NSGA-II to determine a set of optimal solutions.

The remainder of the study is organized as follows. The (R, s, S) policy and inventory control system are stated in Section 2. Section 3 describes the proposed SO methodology that includes the details of both the optimization phase and the simulation phase. In Section 4, an analysis of multi-objective (R, s, S) policy is given. Finally, Section 5 provides concluding remarks for the results obtained in this research.

II. (R, S, S) POLICY AND INVENTORY CONTROL SYSTEM

In this paper, each DC and each Supplier uses the (R, s, S) policy. Review period (R) is 5 days for all supply chain members. However, an order is not always placed in each review period. If the inventory level decreases to the reorder point or lower, we order enough to raise it to order-up-to level. Supply chain members have specific initial inventory, reorder point, and order-up-to level values separately.

In this supply chain, the customer population includes a large group of individuals acting independently and therefore a Poisson demand is appropriate (rate parameter is 50). DCs face stochastic customer orders and Suppliers receive only the replenishment orders from DCs. Incoming order is fully satisfied when it is lower than the current inventory level of supply chain members. If incoming order is not completely satisfied, both DCs and Suppliers use outsourcing to recover order. Note that outsourcing models and lost sales models are conceptually equivalent. However, outsourcing is more suitable than lost sales models for real-world applications [21].

To model the supply chain members, values of the cost components and lead time are obtained from the observations of past customer behavior and fitted to probability distributions using the knowledge and experience of researchers. It is determined that cost components are uniformly distributed. For a given case, order processing time and transportation time are also uniformly distributed. However, processing time has a triangular distribution. DCs require a triangular processing time with endpoints (1, 3) and mode at 2 minutes to reflect the processing of the product into the stores and on the shelves while Suppliers require a triangular processing time with endpoints (3, 7) and mode at 5 minutes for serving DCs. The order processing time of DCs and Suppliers that is uniformly distributed on the interval [2, 5] hours. Order processing time should be thought as the time spent processing order before it is filled (i.e., some routine

paperworks and arrangements). Also, transportation times (from Suppliers to DCs) are uniformly distributed on the interval [1.25, 3] days.

Note that i denotes the DC in the system, $i = 1, 2, 3, \dots, I$ and j denotes the number of Suppliers in the system, $j = 1, 2, 3, \dots, J$. Also, n specifies a period (month). Average holding cost of supply chain members, which is uniformly distributed on a range between \$2 and \$5, is proportional to the remaining inventory quantity over period n ($h_i X_{in}^+$ or $h_j X_{jn}^+$). Outsourcing cost of supply chain members ranges from \$80 and \$100 and is charged proportional to the outsourced customer order quantity at over period n ($k_i X_{in}^-$ or $k_j X_{jn}^-$). Processing cost ($p_i P_i$ or $p_j P_j$) is charged proportional to the processing time to use any supply chain member for processing activity. Processing cost of each DCs ranges from \$5 and \$10 while processing cost of each Supplier is ranges from \$50 and \$75. It is because processing times of Suppliers are high and more costly than DCs. Order cost per use (c_i or c_j) that ranges from \$50 and \$100 is the cost charged, or accrued, to the cost of any order that is placed at any supply chain member irrespective of the time spent in there. Order processing cost (O_i or O_j) includes order processing cost rate and cost per use. Order processing cost rate, which is proportional to the order processing time, ranges from \$2 and \$5. Cost per use is the one-time cost that is accrued each time any supply chain member is used, regardless of the usage duration. Cost per use of each DC ranges from \$10 and \$20 while cost per use of each Supplier ranges from \$100 and \$150. Note that, cost per use of each Supplier is high so supplier selection becomes more important to decrease the total cost. Our proposed problem can simply be represented in Equation (1-2). Total supply chain cost is represented as TSCC. Average service level is denoted as ASL.

$$minimize\ TSCC = \sum_{n=1}^{Periods\ Considered} (TSCC_n) = \sum_{n=1}^{Periods\ Considered} \{\sum_{i=1}^{I} h_i X_{in}^+ + I\{X_{in} \le s\}\,(k_i X_{in}^- + p_i P_i + c_i + O_i)\} + \{\sum_{j=1}^{J} h_j X_{jn}^+ + I\{X_{jn} \le s\}\,(k_j X_{jn}^- + p_j P_j + c_j + O_j)\}) \quad (1)$$

$$maximize\ ASL = \frac{\sum_{a=0}^{Per\ Arrival} min\,(1, \frac{Current\ Inventory\ Level}{Incoming\ Order\ Quantity})}{Total\ Number\ of\ Incoming\ Orders} \quad (2)$$

Equation (1) defines the objective of minimizing the $TSCC$ comprising of average holding cost, outsourcing cost, order cost per use, processing cost and order processing cost. Equation (2) represents the maximization of ASL performance of each supply chain member.

Note that DCs and Suppliers have several specific features. First, single product flows through the supply chain members. Second, inventory control parameters remain the same across the entire finite time horizon. Third, stochastic lead time and poisson demand process are used. Fourth, each customer order is supplied only by a single predetermined DC and each DC replenishment order is supplied only by a single Supplier. Selected Supplier, which is determined after the optimization phase among the candidate Suppliers, remains the same over the simulation run. Suppliers' inventories are replenished from unlimited sources. Fifth, if incoming order is not completely satisfied, both DCs and Suppliers use outsourcing to recover order.

III. Simulation Optimization

SO model includes two fundamental phases: (1) A simulation phase is utilized to evaluate the performance of the candidate solutions, (2) An optimization phase (NSGA-II) is used to determine a set of optimal solutions. In NSGA-II, chromosomes include two parts and the first part of the chromosome represents supplier selection for DCs. Length of the supplier selection part is equal to the number of DCs. The second part of the chromosome represents a determination of the values of initial inventory, reorder point, and order-up-to level of each DC and each supplier, respectively.

NSGA-II forms a combined parent and offspring population $R_t = P_t \cup Q_t$ where R_t is of size $2N$. Herein, non-dominated sorting is utilized to sort the population R_t and to preserve the elitist-individuals. Candidate solutions are divided into small groups using non-dominated sorting where no one solution is superior to the other solution. Then, small groups are sorted and best of these small groups is selected. In the combined population (include all previous and current population members), solutions belonging to the best nondominated set F_1 are of best solutions. All members of the set F_1 can be selected for the new population set P_{t+1} if the size of set F_1 is smaller than N. Subsequent nondominated front is used to select the remaining members of the population P_{t+1}. The process proceeds until no more sets can be accommodated. Thus, the set F_l is the last nondominated set. The population size can be smaller than the count of solutions in all sets from F_1 to F_l. To select exactly N population members, crowding distance sorting is used to sort the solutions of the last set F_l. Basically, crowding distance sorting is utilized to maintain a diverse population of candidate solutions. Large average crowding distance will provide better diversity in the population. A series of subsets from the small groups is extracted by crowding distance sorting. Then, subsets are sorted and transferred to next generation solutions. After, selection, crossover, and mutation are used to form a new population Q_{t+1} of size N. More details can be found in [10, 11, 22].

In crossover operator, two individuals are taken and their chromosomes are randomly cut to generate two "head" parts and two "tail" parts. Then, the tail parts are swapped over to create two new full-length chromosomes. Also, this is denoted as single point crossover. In literature, crossover rate is typically applied between 0.6 and 1.0 and it is considered to be 0.8 in this study. After applying crossover, mutation is randomly performed to each child individually with a small probability and we set mutation rate as 0.05. In this way, a small amount of random search is provided.

At the end of the each iteration, the best chromosome is extracted and used for the evolution of the next iteration. This procedure is continued until the desired number of iterations is reached. In each iteration, simulation output is returned to the NSGA-II as the most recent fitness function to be evaluated, and NSGA-II once more tries to determine better decision variables to improve model performance. Thus, optimization and simulation are sequential in nature. The configuration parameters within the simulation are optimized using NSGA-II that also identifies the new configurations within each iteration. The performance values to the NSGA-II are provided by the results of the simulation experiment [23].

Simulation is a remarkable recourse to analyze the performance of inventory control systems. However, one of the major challenges of applying simulation into SO is that it is not clear how to extend simulation models to a more general problem setup. On the other hand, your skills and knowledge related can reasonably reduce the computational burden or improve the solution [24]. In this study, simulation modeling framework is formed by using Simio (Version: 7.121.12363) is based on intelligent objects and is created to simplify model structure by promoting a modeling paradigm shift from the process orientation to an object orientation [25].

IV. ANALYSIS OF MULTI-OBJECTIVE (R, S, S) POLICY

In this study, we focused on SO based (R, s, S) policy and supplier selection simultaneously in a two-echelon supply chain under outsourcing system and stochastic environment. In today's fiercely competitive world, effective inventory control system is imperative to meet high customer expectations while maximizing total supply chain profitability. Also, companies have to work with their supply chain members to remain competitive. At this point, supplier chain member can satisfy strategic objectives in supply chain using simulation based NSGA-II. In Table 1, determined values of the initial inventory, reorder point, and order-up-to level are given for supply chain members. In proposed SO model, each objective function along the Pareto-front is only improved by degrading the other objective function. In this case, none of the Pareto-optimal solutions is exactly better than other solutions and therefore one of them can be considered as an acceptable solution. In this study, there are two non-dominated solutions that can be chosen for a decision. Supplier1 is selected to satisfy the replenishment orders of DC2 and DC3 in solution 1 and solution 2. Supplier2 is chosen for DC1 in solution 1 while Supplier 3 is selected for DC1 in solution 2. The average service level of two non-dominated solutions is given in Table 2.

Table 1: Optimal values of inventory control parameters for two nondominated solutions.

		Solution 1		**Solution 2**
Initial Inventory	DC1	7803	DC1	7803
Reorder Point		565		565
Order Up To Level		2015		2015
Initial Inventory	DC2	7803	DC2	7803
Reorder Point		825		825
Order Up To Level		1179		1889
Initial Inventory	DC3	7803	DC3	7803
Reorder Point		787		787
Order Up To Level		1889		1179
Initial Inventory	Supplier1	2371	Supplier1	2371
Reorder Point		787		787
Order Up To Level		2812		2812
Initial Inventory	Supplier2	6217	Supplier3	6217
Reorder Point		590		590
Order Up To Level		3166		3095

Table 2: The average service levels for two non-dominated solutions.

Solution 1		**Solution 2**	
DC1	0,999952	DC1	1
DC2	1	DC2	1
DC3	1	DC3	1
Supplier1	0,992779	Supplier1	0,996648
Supplier2	0,975249	Supplier3	0,974824

In this paper, total supply chain cost includes five different cost components including average holding cost (1), order cost per use (2), outsourcing cost (3), order processing cost (4) and processing cost (5). The analysis of the total value of cost components over twelve periods is given in Table 3 in which S1 denotes solution 1 and S2 represents solution 2. Note that total supply chain cost of solution 1 is lower than solution 2.

According to the results, the largest share for DCs is the total average holding cost. Even the minimum one is accounted for 57.24%. Thus, holding cost is the most strategic issue despite various cost types involved in DCs. Similarly, total average holding cost is the most critical cost component for Supplier2 and Supplier3. On the other hand, total processing cost whose share is at least 57.23% is the most strategic cost components for Supplier 1. Note that share of the outsourcing cost is generally low when compared with the other cost components.

Table 3: Cost analysis of DCs and Suppliers.

		1	2	3	4	5
DC1	S1	94504	26887	88	36434	2890
	S2	94992	26934	0	36407	2904
DC2	S1	91202	27063	0	36507	3086
	S2	98860	27026	0	36508	2933
DC3	S1	95721	27015	0	36439	2945
	S2	89114	26994	0	36492	3088
Supplier1	S1	71456	2359	11346	3168	118212
	S2	71267	2348	4686	3144	117312
Supplier2	S1	162699	627	12105	834	32611
Supplier3	S2	159677	586	14747	806	31884

Table 4: Quantity based analysis for each supply chain member.

		TWMOQ	TPMOQ
DC1	Solution 1	18023	1
	Solution 2	18020	-
DC2	Solution 1	18024	-
	Solution 2	18032	-
DC3	Solution 1	18029	-
	Solution 2	18031	-
Supplier1	Solution 1	21570	124
	Solution 2	21621	54
Supplier2	Solution 1	10072	134
Supplier3	Solution 2	9666	146

Total number of wholly met order quantity over twelve periods (TWMOQ) and total number of partially met order quantity over twelve periods (TPMOQ) are given in Table 4. The number of partially met order for supply chain members is at most 2 orders. There is no order that is totally outsourced. The results of this study show that managers can increase company's competitiveness and responsiveness by using SO. Also, integration of inventory control system and supplier selection is important for managing and controlling inventories under dynamic environments.

V. CONCLUSION

The model with (R, s, S) policy is highly relevant in practice, as inventory level is periodically checked in order to control the behavior of supply chain. However, it is challenging in the case of modeling due to the presence of complex interactions between supply chain members. Herein, companies not only need to operate at a lower cost but also operate at a better service level to be distinguished from competitors and stand out in the marketplace. Therefore, most of the real-world problem can be considered as multi-objective that are generally conflicting with each other. At this point, our proposed SO can allow managers to notice the possibilities and trade-offs between conflicting objectives before choosing the most appropriate solution to implement on inventory control system and supplier selection. In addition, we proposed a realistic and flexible (R, s, S) policy to accurately capture of dynamic behavior of supply chain. The paper shall also be of great value to supply chain practitioners to extend their research avenues into this exiting area.

Further research studies can introduce more than two objective functions to the inventory management problem such as lead time minimization, inventory level minimization, stockout probability minimization, order met probability maximization, totally met order quantity maximization, and so on. The proposed model can also be extended to include different supply chain networks (e.g. multi-echelon supply chain including suppliers, manufacturers, DCs, retailers, and customers), and different cost structures (e.g., discount pricing structures). A discussion of the model with fuzzy data (e.g. ordered fuzzy numbers) and investigation of demand patterns (e.g. smooth, erratic, intermittent, and lumpy demand) can also make the SO model more pragmatic in approach and scope. To improve the performance of NSGA-II algorithm, many researchers have put forward some improved NSGA-II models (e.g. dynamic crowding distance and gradient-based hybrid operator) that may also be used as a future work.

REFERENCES

[1] L. Lapide, "What about measuring supply chain performance," *Achieving Supply Chain Excellence through Technology*, 2(2), pp. 287-297, 2000.

[2] B. M. Beamon, "Measuring supply chain performance," *International Journal of Operations & Production Management*, 19(3), pp. 275-292, 1999.

[3] C. Elrod, S. Murray, and S. Bande, "A review of performance metrics for supply chain management," *Engineering Management Journal*, 25(3), pp. 39-50, 2013.

[4] B. M. Beamon, "Supply chain design and analysis: models and methods," *International Journal of Production Economics*, 55(3), pp. 281-294, 1998.

[5] A. Chavosh, A. B. Halimi, S. Soheilirad, A. Ghajarzadeh, and A. Nourizadeh, "Customer responsiveness and export performance of selected electronic equipment export companies in Malaysia," *in International Conference on Social Science and Humanity*, pp. 124-127, 2011.

[6] S. Chopra, and P. Meindl, *Supply chain management, strategy, planning, and operation*. 3rd ed. Upper Saddle River, N.J.: Pearson Prentice Hall, 2007.

[7] P. N. Poulos, G. G. Rigatos, S. G. Tzafestas, and A. K. Koukos, "A Pareto-optimal genetic algorithm for warehouse multi-objective optimization," *Engineering Applications of Artificial Intelligence*, 14(6), pp. 737-749, 2001.

[8] S. T. Khu, F. di Pierro, D. Savic, S. Djordjevic, and G. A. Walters, "Incorporating spatial and temporal information for urban drainage model calibration: An approach using preference ordering genetic algorithm," *Advances in Water Resources*, 29, pp. 1168-1181, 2006.

[9] M. G. Avci, and H. Selim, "A multi-objective, simulation-based optimization framework for supply chains with premium freights," *Expert Systems with Applications*, 67, pp. 95-106, 2017.

[10] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, 6(2), pp. 182-197, 2002.

[11] K. S. Jeong, M. Kim, H. Jo, J. A. Gim, D. K. Kim, and G. J. Joo, "Search of optimal locations for species-or group-specific primer design in DNA sequences: Non-dominated sorting genetic algorithm II (NSGA-II)," *Ecological Informatics*, 29, pp. 214-220, 2015.

[12] K. S. Bhutta, and F. Huq, "Supplier selection problem: a comparison of the total cost of ownership and analytic hierarchy process approaches," *Supply Chain Management: An International Journal*, 7(3), pp. 126-135, 2002.

[13] C. Basnet, and J. M. Y. Leung, "Inventory lot-sizing with supplier selection," *Computers & Operations Research*, 32(1), pp. 1-14, 2005.

[14] A. N. Haq, and G. Kannan, "Design of an integrated supplier selection and multi-echelon distribution inventory model in a built-to-order supply chain environment," *International Journal of Production Research*, 44(10), pp. 1963-1985, 2006.

[15] M. R. S. Moghadam, A. Afsar, and B. Sohrabi, "Inventory lot-sizing with supplier selection using hybrid intelligent algorithm," *Applied Soft Computing*, 8, pp. 1523-1529, 2008.
[16] A., Mendoza, and J. A. Ventura, "A serial inventory system with supplier selection and order quantity allocation," *European Journal of Operational Research*, 207(3), pp. 1304-1315, 2010.
[17] D. Choudhary, and R. Shankar, "Joint decision of procurement lot-size, supplier selection, and carrier selection," *Journal of Purchasing & Supply Management*, 19(1), pp. 16-26, 2013.
[18] D. Choudhary, and R. Shankar, "A goal programming model for joint decision making of inventory lot-size, supplier selection and carrier selection," *Computers & Industrial Engineering*, 71, pp. 1-9, 2014.
[19] E. Shadkam, and M. Bijari, "Multi-objective simulation optimization for selection and determination of order quantity in supplier selection problem under uncertainty and quality criteria," *International Journal of Advanced Manufacturing Technology*, pp. 1-13, 2015.
[20] S. Pazhani, J. A. Ventura, and A. Mendoza, "A serial inventory system with supplier selection and order quantity allocation considering transportation costs," *Applied Mathematical Modelling*, 40(1), pp. 612-634, 2016.
[21] F. Chu, and C. Chu, "Single-item dynamic lot-sizing models with bounded inventory and outsourcing," *in IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 38(1), pp. 70-77, 2008.
[22] Q. Ma, D. Xu, P. Iv, and Y. Shi, "Application of NSGA-II in parameter optimization of extended state observer," *in Challenges of power engineering and environment, Springer Berlin Heidelberg*, pp. 587-592, 2007.
[23] M. Tripathi, G. Kuriger, and H.-da. Wan, "An ant based simulation optimization for vehicle routing problem with stochastic demands," *in Proceedings of the 2009 Winter Simulation Conference*, pp. 2476-2487, 2009.
[24] W. D. Kelton, J. S. Smith, D. T. Sturrock, M. Göçken, and A. T. Dosdoğru, *Simio & simulation: modeling, analysis, applications: Turkish Translation*. 3rd edn. Simio LLC, 2015.
[25] C. D. Pegden, "Simio: a new simulation system based on intelligent objects," *in Proceedings of the 2007 Winter Simulation Conference*, pp. 2293-2300, 2007.

Smart Traffic Signal and Routing System for Emergency Vehicles

S. OZDEM[1], S. OZUM[1]

[1] Hitit University, Corum/Turkey, selimozdem@gmail.com
[1] Hitit University, Corum/Turkey, songulozum@hitit.edu.tr

Abstract **- The increasing population ratio in cities and the increase in vehicle and pedestrian traffic due to this population also brought problems with the transition of emergency vehicles such as ambulance and fire brigade. The aim of this study is to present solutions for problems with traffic congestion in vehicles with transit superiority in an emergency. It is also intended to provide wireless communication between traffic signaling lights in operation and visual warning monitors used in designated locations and emergency vehicles. When the route is determined via the mobile device on the emergency vehicle, the traffic light and the predetermined locations are communicated and the other drivers in the traffic will be alerted in an audible and visual manner. Since these alerts will make in advance in the area where the traffic intensity is, traffic accidents would be resolve and the emergency vehicles would be continue without losing time.**

Keywords **- Smart traffic control system, microcontroller, and emergency vehicle**

I. INTRODUCTION

Recent evolution in wireless communications has opened new fields of research such as network connectivity in environments where the wired solutions are not possible. In additional, the continuous increase in the congestion level on public roads in many countries is becoming a major concern. Thus, wireless networks have been attracted great attention due to the applications that they offer, from road safety to traffic control [1,2].

Nowadays the rate of road traffic is increasing day by day in developing countries. Traffic system will be more complicated to traffic emergency vehicles in the modern world. The emergency can be occur any way, time and on any location. In that case be speed is important.

The critical consequences of traffic problem is the delay of emergency vehicles such as, ambulance, fire brigade, police vehicles [3-6].

This paper presents an approach to solutions through a mobile application for problems with traffic congestion in vehicles with transit superiority in an emergency. This paper will be great help to make free flow of emergency vehicles without getting stuck into the traffic. The application will be developed for the Android platform and controlled via a mobile device, with access to GPS (Global Positioning System) technology, Radio Frequency (RF) modules and Internet.

II. PROPOSED STUDY

In our proposed system, we will try to develop a Smart traffic control system that consists of:

- Cloud Server: It will locate microcontroller by using RF modules that will monitor traffic signal activities according to emergency vehicle arrival. Also, it will check the emergency vehicle in case two vehicles arrive from opposite directions. For this system microcontroller will be used.
- Android App: It will be designed to set the priority of the emergency vehicles. Also, it will be used to set destination address and accordingly route will be decided. The data will be send to cloud server using by RF modules.
- GPS: Mobile GPS will be used for determine the emergency vehicle route.

III. METHODOLOGY

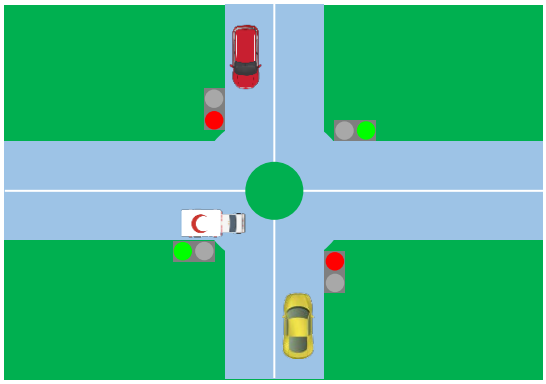


Figure 1: Ambulance passes through traffic junction.

The proposed system aims to overcome the problem of traffic jam at the Traffic Signal system is introduced. Firstly, the density of vehicle will be calculate on the road for flow traffic smoothly without congestion. Then, Priority Based Signaling will be developing which helps to give the priority to the emergency vehicles. This can be seen in Figure 1. This approach will be helpful to overcome the traffic jam problem and reducing the congestion. It also helps to arrive the emergency services like fire brigade, ambulance or police at right time. Smart traffic control system ensures significant benefits ensure like reach faster to the destination, less congestion and saving fuel consumption.

Figure 2: Schematic representation of the system.

Figure 2 shows that emergency vehicle alert that will transmit the vehicle geolocation which will then be received by the road side unit at signal and it will start the green light and this light will be in green state till the vehicle is within the Global Positioning System (GPS) range. This system will be useful for the emergency vehicles such as ambulance, fire brigade, and police. For that strategy, the proposed system will be built in real time. Cloud computing plays important role between emergency vehicles and the traffic signals. This paper is to establish the communication between the traffic signals and the emergency vehicles so that the traffic signal can respond to the arrival of the emergency vehicles. When the traffic signals are changes its states according to the position of the emergency vehicle it can able to open a freeway. The traffic signal will be automatically controlled using a mobile app which uses GPS by capturing the latitude and longitude of the emergency vehicle and will be send signal to the local system, so making uninterrupted traffic to the emergency vehicle. Then, the traffic signals can be controlled by cloud server [2].

IV. CONCLUSION

Human life is very precious and so must follow safety measures very careful in all aspects. The concept of smart cities has been introduced to give solutions to challenging issues in metropolitan cities that emergency vehicle arriving.

In this paper mentioned two situation, firstly, calculating the density of the vehicle on the road for the flow of the traffic smoothly without congestion and secondly, developing Priority Based Signaling which will help to give the priority to the emergency vehicles will be studied. The emergency vehicles will be controlled by cloud server in our proposed system. By using of cloud technology, automatic traffic control signal can be possible without human interrupt. Cloud server can be also control the traffic light according to the emergency location and thus reaching the correct location. Using cloud computing will reduce delay hence will help emergency vehicles reach faster to the destination.

This Smart traffic control approach will yield important benefits like especially reach faster to the destination less congestion, saving fuel consumption. The proposed approach will consider both the priority of the vehicles and the density of the vehicles on the road. It also will control the traffic light sequence efficiently and the accuracy of the GPS.

REFERENCES

[1] A. Zaldívar-Colado et al., "Management of traffic lights for emergency services", *Tehnički vjesnik*, vol. 24, pp. 643-648, 2(2017).

[2] Dr. A. Balamurugan et al., "Automated emergency system in ambulance to control traffic signals using IoT", *IJECS*, vol. 4, pp. 11533-11539, April 2015.

[3] Megha A. Tank et al., "Review on smart traffic control for emergency vehicles", *International Journal of Computer Applications*, vol. 112, pp. 1-3, February 2015.

[4] K.Udhayakumar et al., "A Smart Traffic Management System using the Spatio-Temporal Relationships for an Emergency Vehicle", *IJARCST*, vol. 2, pp. 178-183, Jan-March 2014

[5] N. Sharma et al., "Smart Traffic Signal Management for Emergency Vehicles Using FOG Computing", *IJCSN*, vol. 7, pp. 18-21, February 2018.

[6] V. Venkatesh et al., "Smart Traffic Control System for Emergency Vehicle Clearance", *IJIRCCE*, vol. 3, pp. 7242-7246, August 2015.

Modeling of Voltage Sag/Swell Disturbances in Distributed Generation Systems

A.YILMAZ[1] and G. BAYRAK[2*]

[1] Bursa Technical University, Dept. of Electrical and Electronics Eng., Bursa/Turkey, alper.yilmaz@btu.edu.tr
[2] Bursa Technical University, Dept. of Electrical and Electronics Eng., Bursa/Turkey, gokay.bayrak@btu.edu.tr
[*] Corresponding Author: Tel: +902243003507, E-mail: gokaybayrak@gmail.com

Abstract **- Power quality problems due to short circuit faults, are defined as changes in voltage, current, or frequency in the power system. Among the power quality problems in distributed generation systems are island mode operation, harmonics in current and voltage, voltage sags/swells, voltage interruptions and transients. The most common power quality problem in the distribution system is the voltage sags and swells caused by a short circuit fault. In this study, various Power quality problems due to short circuit faults such as line-to-ground, line-to-line and multistage faults are modeled in MATLAB/Simulink environment. In order to detect the voltage sag and swell problems, firstly the voltage sag that occurs in case of failure is examined and the response of the power system is observed. The study also models multi-stage failures as a result of failure to achieve relay synchronization. Examination of the voltage sags and swells clearly revealed the resulting waveforms, the response of the power system to the fault condition. Another advantage of the realized work is that the developed model can be used to measure the performance of the distributed generation system in failure diagnosis and classification studies.**

Keywords **- Distributed Power Generation, Power Quality, Voltage Sags.**

I. INTRODUCTION

The conventional power distribution systems designed with a unidirectional power flow from high voltage to low voltage are transformed into the power distribution systems which have bi-directional power flows by the integration of PV (photovoltaic)-based distributed generation systems. As a consequence of that, the classical analysis, operation, and design methods are not capable of managing this complex grid structure. Also The PV-based distributed generation systems are also expected not to operate out of the defined threshold values of the voltage and frequency by the related standards, to detect islanding and providing an isolated grid in a loss of main, to be in a stable condition in all abnormal grid faults, to have a fault ride through and not to affect the power quality as far as possible.

Power quality parameters are a set of limits that allow an equipment to operate as intended without significant performance and life expectancy loss [1]. Any PQ problem that causes voltage, current, and frequency deviations can lead to failure of consumer equipment or improper operation of electrical equipment. The most common power quality problem in the distribution system is the voltage sags and swells caused by a short circuit faults.

The first step in preparing a reliable algorithm for detecting mitigation and prediction power quality events especially voltage sag and voltage swell occurring in the current grid is to model a power system in which power quality disturbances can be analyzed [2]. Classification based short circuit fault detection methods and field alternating short circuit fault detection methods utilize the fault signals obtained by modeling the electrical distribution grid. In one of these studies, simulation data obtained in the Matlab/Simulink environment was used for detecting the voltage sag and voltage swell events and a new method was proposed by wavelet transform method [3]. In another study, a wavelet transform and an artificial neural network based algorithm have been developed for the detection and classification of short circuit faults in the transmission lines by using the data obtained from the simulation model [4].

There are a number of methods in the literature for power system faults detection, classification and prediction. These methods are examined in two basic categories as model-based and data-oriented methods [5]. The creation of a system model is crucial where it is difficult to obtain experimental or real-time data. Model-based power system fault detection methods can be applied to threshold-based methods, state estimation based methods, fuzzy logic based methods, effective value transformation methods, classification based methods [6,7] and time-frequency transform based methods [8]. In one of the studies, support vector machine, which is one of the classification based methods with a model based approach, detects short circuit faults in the transmission lines [9].

In three-phase power systems, faults between lines and between lines and ground can be examined in two groups, symmetric and asymmetric faults [10]. The faults occurring as a result of the contact of all three lines with each other or with the earth are symmetrical because the phase difference of 120 degrees between the phases continues to be preserved. The faults that occur as a result of contact of the single line with the ground together with the faults that comes into contact with each other or with the ground come into the asymmetrical fault group.

In this paper, the voltage sags and swells caused by a short circuit fault that occur in the low-voltage grid modeled in MATLAB/Simulink environment. Developed electric power distribution model includes different type of short circuit faults such as line-to-ground faults, line-to-line faults and multistage faults. The proposed power distribution system model enables to obtain the desired data as a result of simulation in fault

analysis in distributed power systems, where it is difficult to obtain experimental or real time data.

II. VOLTAGE SAG AND SWELL DISTURBANCES

According to the IEEE 1159 standard [11], voltage sag is defined as a 10% to 90% reduction in the voltage limited to a time interval of half a period to 60 s in a system operating under nominal conditions. Voltage sag events are caused by line energization, switching on of high power motors and short circuit faults.

According to the IEEE 1159 standard, voltage swell is defined as a 110% to 180% increase in voltage; limited to a time interval of half period to 60 s on a system operating under nominal conditions, Voltage swell is not as common as a power quality phenomenon as a voltage sag, , the maneuvers in the power system, such as the disconnection of large inductive loads, the actuation of large capacitor banks or short circuit faults result in voltage swells. Depending on the fault location and system conditions, a fault can lead to voltage sag, swell or interruption. Waveforms of voltage sag, voltage swell and voltage interruption are shown in Fig 1.

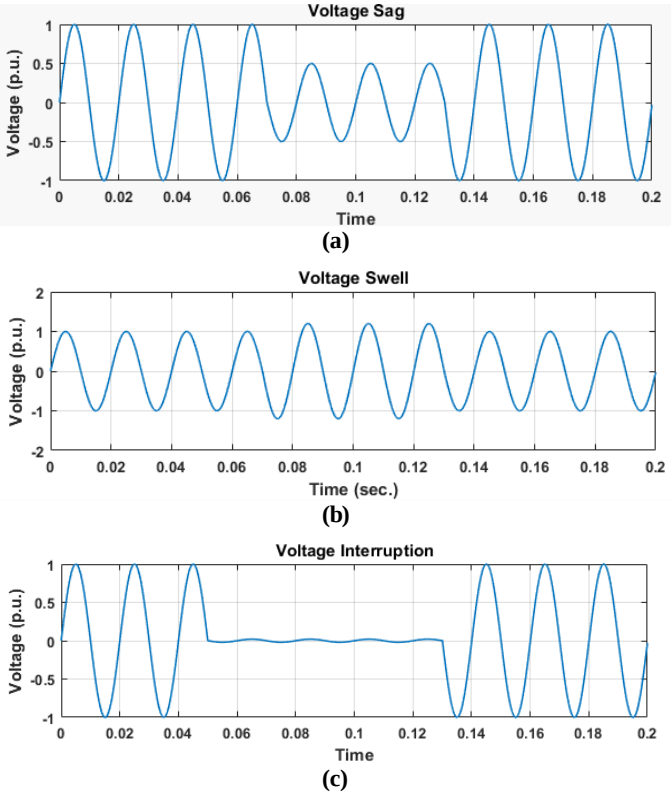


Figure 1: Signal samples of **(a)** Voltage Sag, **(b)** Voltage swell, **(c)** Voltage interruption fault classes used in preliminary operation tests

III. DISTRIBUTED GENERATION SYSTEM MODEL

Since the most common type of failure in power systems is short-circuit faults, the fault is usually used for short-circuit faults. The distributed generation system model was developed using MATLAB / Simulink which is shown in Fig. 2. The simulation model realized includes line-to-line faults, line-to-ground faults, two line-to-ground faults and multistage short circuit faults.

The power distribution system model consist of 10.5 kV main grid, 10 MVA generator, 1MWp grid connected PV plant, 10.5 kV/0.4 kV a delta/star connected step down two windings transformer, inductive and resistive loads (100 kW and kVAr), and squirrel-cage 75 kW induction motor.

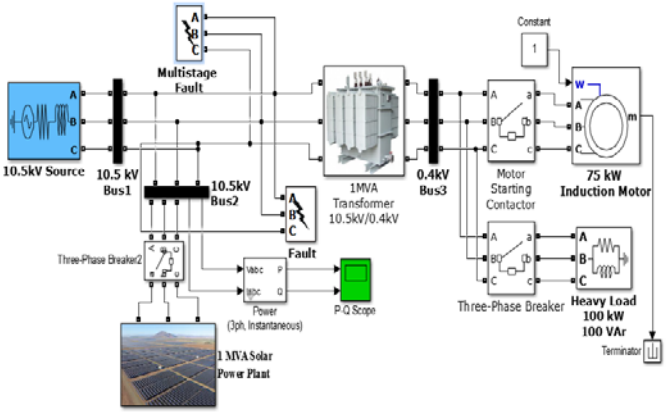


Figure 2: Distributed generation system model

In this paper, effective value conversion was used to observe the effect of power quality disturbances on the voltage.

$$U_{RMS}=\sqrt{\frac{1}{M}\sum_{i=1}^{M}u(i)^2} \quad (1)$$

where M is window size and u(i) is voltage waveform of the i-th sample.

IV. SIMULATION RESULTS

In three-phase systems, faults between phases and between phases and ground can be examined in two groups, symmetric and asymmetric. Fig. 3 shows the effective values of the voltages resulting from the fault occurring as a result of short circuiting of the three phases between 0.15 and 0.25 seconds. The resulting fault is symmetrical because it creates similar effects on all three phases.

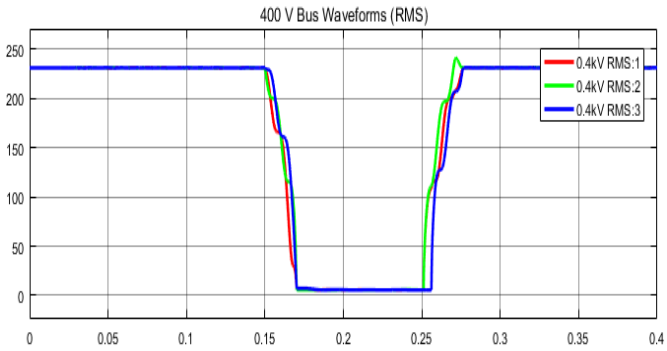


Figure 3: Effective values of voltages that occur as a result of three phase short circuit failure

Fig. 4 shows the effective values of the voltages resulting from the fault occurring as a result of short circuiting of the three line-to-ground between 0.15 and 0.25 seconds. Short-circuit fault resistance 1 ohms is selected in the model. The

resulting fault is symmetrical because it creates similar effects on all three phases.

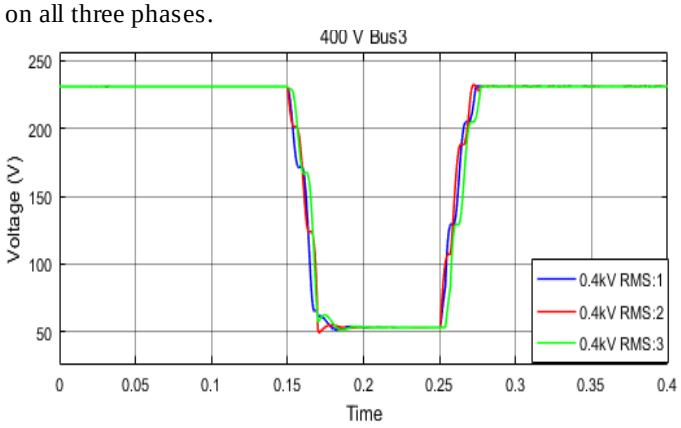


Figure 4: Effective values of voltages that occur as a result of three line-to-ground short circuit fault

In Fig. 5, short-circuit fault resistance 5 ohms is selected in the model which is formed between A-B phases between 0.15 and 0.25 seconds. As a result of the simulation, voltage swell occurs in phase C, while voltage sag is observed in phases A and B. As seen in the figure, two phase faults, single phase and two phase-ground faults show asymmetrical characteristics.

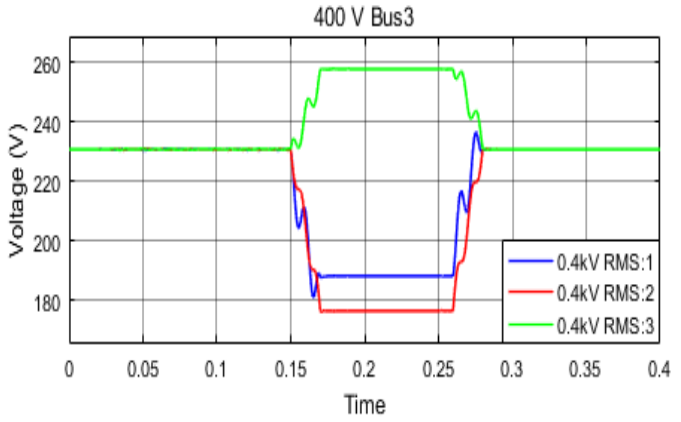


Figure 5: Effective values of phase-to-phase voltage resulting from faults

In Fig. 6, short-circuit fault resistance 1 ohms is selected in the model which is formed between A line and ground between 0.15 and 0.25 seconds. As a result of the simulation, voltage swell occurs in phase B and C, while voltage sag is observed in phases A. As seen in the figure, two phase faults, single phase and single line-to-ground faults show asymmetrical characteristics.

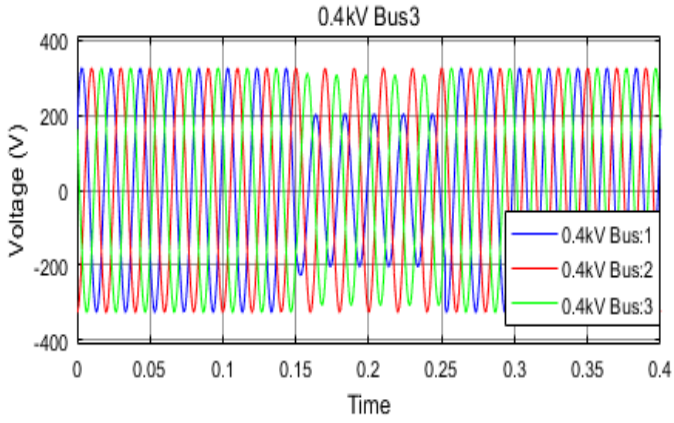


Figure 6: Values of phase-to-phase voltage resulting from single line-to ground faults

As a fault event continues in the power systems, changes in the structure and type of the fault may occur as a result of changes that may occur in the system. This type of fault is referred to as multi-stage faults. The voltage values of the phases in the low-voltage bar during the multi-stage fault are shown in Fig 7. There was a fault between 0.1 and 0.15 seconds in the simulation. Another fault with a short-circuit impedance of 0.1 ohms between 0.15 and 0.25 seconds occurred while the fault was still occurring.

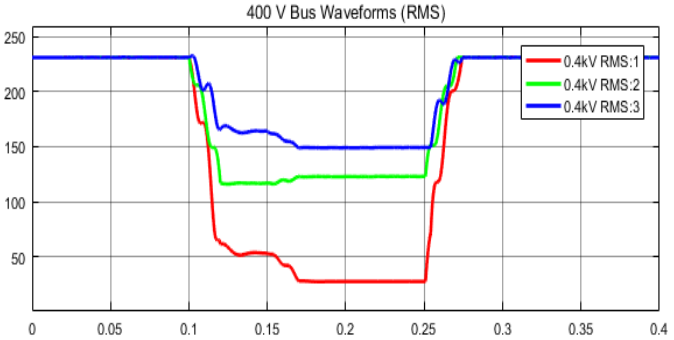


Figure 7: Effective values of voltages generated in a multi-stage 2 phase-ground fault

V. CONCLUSION

In this study, the voltage sags and swells caused of short circuit faults that occur in the low-voltage grid that is fed through both the main grid and the grid-connected PV plant are produced by using developed electrical distribution model in MATLAB/ Simulink environment. Examination of the voltage sag and voltage swell disturbances with simulation clearly revealed the resulting waveforms, the response of the electrical power systemto the shot circuit fault conditions.

The proposed power distribution system model enables to obtain the desired data as a result of simulation in fault analysis in distributed power systems, where it is difficult to obtain experimental or real time data. Advantage of the realized study is also that the developed model can be used to measure the performance of the PV connected distributed generation system in short circuit fault detection and prediction studies.

ACKNOWLEDGMENT

This study is supported by Bursa Technical University Scientific Research Projects Coordination Unit. Project Number: 172L25

REFERENCES

[1] Bollen, M.H.J. and Gu, I.Y.H., 2006, Signal processing of power quality disturbances, John Wiley&Sons, New York.*etection and Estimation.* New York: Springer-Verlag, 1985.

[2] Khokhar, S., Zin, A. M., Mokhtar, A. S., & Ismail, N. A. M. (2014, October). MATLAB/Simulink based modeling and simulation of power quality disturbances. In Energy Conversion (CENCON), 2014 IEEE Conference on (pp. 445-450). IEEE.G. R. Faulhaber, "Design of service systems with priority reservation," in *Conf. Rec. 1995 IEEE Int. Conf. Communications*, pp. 3–8.

[3] Latran, Mohammad Barghi, and Ahmet Teke. "A novel wavelet transform based voltage sag/swell detection algorithm." International Journal of Electrical Power & Energy Systems 71 (2015): 131-139.

[4] Silva, K. M., Benemar A. Souza, and Nubia SD Brito. "Fault detection and classification in transmission lines based on wavelet transform and ANN." IEEE Transactions on Power Delivery 21.4 (2006): 2058-2063.

[5] Hare, J., Shi, X., Gupta, S., & Bazzi, A. (2016). Fault diagnostics in smart micro-grids: A survey. Renewable and Sustainable Energy Reviews, 60, 1114-1124.

[6] Haykin, S. 2008. "Neural networks and learning machines", Pearson, 3rd edition.

[7] Burges, C J.C. 1998. "A tutorial on support vector machines for pattern recognition", Data Mining and Knowledge Discovery, 2, 121-167.

[8] Ray, Prakash K., Soumya R. Mohanty, and Nand Kishor. "Disturbance detection in grid-connected distributed generation system using wavelet and S-transform." Electric Power Systems Research 81.3 (2011): 805-819.

[9] Dash, P.K., Samantaray, S.R., Panda, G. 2007. "Fault classification and section identification of an advanced series-compensated transmission line using support vector machine", IEEE Trans Power Deliv, 22, 67–73.

[10] Saadat, H. (1999). Power system analysis. McGraw-Hill.

[11] IEEE Std 1159, IEEE Recommended Practice for Monitoring Electric Power Quality, 2009.

Efficient Design and Comparative Performance Analysis of PID Controller Applied To Automatic Voltage Regulator Employing Symbiotic Organisms Search Algorithm

E. ÇELİK

Gazi University, Ankara/Turkey, emrecelik@gazi.edu.tr

***Abstract* - This article attempts to solve the problem of efficient design of proportional+integral+derivative (PID) controller applied to automatic voltage regulator (AVR) by employing recently introduced symbiotic organism search (SOS) algorithm for the first time. SOS is a metaheuristic proved recently to be promising by benefitting from the idea of imitating natural phenomena of interactive behavior seen among organisms living together in a similar environment. PID controller design needs proper determination of three control parameters. Such a design problem can be taken as an optimization task and SOS is invoked to find out better controller parameters through the new cost function defined in the paper, which allows to evaluate the control behavior in both time-domain and frequency-domain. For the performance analysis, distinct analysis techniques are deployed such as transient response analysis, root locus analysis and bode analysis. The efficacy of the presented technique is widely illustrated by comparing the obtained results with those reported in some prestigious journals and it is shown that our proposal leads to a more satisfactory control performance from the perspective of both time-domain and frequency-domain specifications.**

***Keywords* - automatic voltage regulator, PID controller, multi-objective optimization, symbiotic organisms search algorithm, performance analysis, cost function**

I. INTRODUCTION

In an electric power network, ensuring a constant voltage level under various circumstances is one of the significant control issues of power systems having a close relation to power quality, grid security and grid reliability. When facing a deviation in grid voltage level, it leads to remarkable changes in the system dynamics and accordingly there may be a deterioration in the performance of the devices connected with this power grid and drop in their life expectancy, because all equipment can operate efficiently only for a particular voltage level termed as nameplate or rated voltage [1, 2]. Moreover, controlling the bus voltage in a local sense has another aspect of regulating reactive power flow, thus rendering it possible to reduce real line losses because of the reactive current components in electric power network. In order to fulfill the aforesaid objectives, automatic voltage regulator (AVR) system is installed in electrical power systems. An AVR is equipment aimed to sustain the output voltage of a synchronous generator (SG) at a desirable voltage level by keeping its excitation voltage under control, where the exciter voltage is regulated to match the voltage drop or rise according to the new conditions [3].

In the hope of implementing and enhancing dynamic response of an AVR system, several control techniques have been studied in the literature based on optimal control, robust control, fuzzy logic, conventional and fractional order proportional+integral+derivative (PID) techniques and adaptive control, which have individual advantages and disadvantages. Among the reported controllers, the classical PID is no doubt the one that is the most preferred owing to its robust performance regardless of variations in system parameters and structural simplicity which requires tuning of only three control parameters, such as proportional gain, integral gain and derivative gain [4]. However, proper determination of PID gains is fairly difficult and there is no universal methodology that assists the operator in designing this controller. When the literature is evaluated, it can be seen that a vast number of artificial intelligence algorithms have been paid much attention by many researchers particularly since 2000, so as to acquire an almost optimal solution in their PID design by exploiting the unique search ability of governed optimization algorithm. In this context, in 2012, artificial bee colony (ABC) algorithm is suggested to enhance the performance of PID-controlled AVR system, where a comparison with particle swarm optimization (PSO) and differential evolution (DE) algorithm are also presented [1]. From the findings, ABC is found to exhibit better performance than the others. Subsequently, many optimizing liaisons (MOL) algorithm, which is the simplified revision of the original PSO, is applied to the same optimization problem of searching for better PID parameters [2]. The results are compared to those in [1] and it is shown that MOL-based PID controller can enhance the system performance with regard to both time-domain and frequency-domain measures. In 2016, biogeography-based optimization (BBO) algorithm is introduced into searching for optimal PID parameters for the concerned control system [5]. Comparative results with ABC-based obtained results in [1] demonstrate that BBO algorithm outperforms the ABC

approach, thereby it yields an improvement in the system dynamic response. In [6], PSO and global neighborhood algorithm (GNA) are adopted to optimize the output response of a PID-controlled AVR system. From the results of transient response analysis, GNA is found to perform better than PSO with regard to settling time and rise time. However, peak overshoot of the response with GNA is greater than that of PSO. An application of chaotic PSO (CPSO) is made in [7] to optimize the AVR system performance. A comparison is also presented with the results obtained by the standard PSO in [6]. It is shown that CPSO-based AVR system performance is improved considering peak overshoot and settling time. However, it can be said that the validation of these two studies [6, 7] is not properly justified because no published work is used for comparison.

Symbiotic organisms search (SOS) algorithm is a relatively straightforward and effective metaheuristic proposed by Cheng and Prayogo in 2014 [8]. In the algorithm, simulation of symbiotic interaction strategies observed amongst organisms in order to keep alive in the ecosystem is realized. A significant advantage of the algorithm is that it requires only two common tuning parameters such as population size and maximum iteration number. Preliminary tests of applying SOS to some mathematical benchmark problems and engineering design problems affirm the excellence of the SOS compared with other remarked optimization algorithms. In addition, superior performance of the SOS for optimizing PI parameters in an off-line sense for a DC servo motor drive is demonstrated based on comparable simulated and experimental results in [9] in comparison with PSO, genetic algorithm and classical Ziegler-Nichols tuning rule. To the author knowledge, it has not been yet addressed in the open literature whether the application of SOS leads to more optimal PID controller gains or not in presence of AVR control application.

In the light of the consequences of the above paragraphs, the author of this article is encouraged to present a unique design methodology for the studied AVR system that improves the trade-off between the dynamic response and the stability margin of the system. For this, the design problem is contemplated as an optimization task and a new composite cost function in the time-domain and frequency-domain is suggested. Then, SOS is invoked to optimize the PID controller gains so that the controlled system may yield the aspired response and stability as depicted by the suggested cost function. Using transient response analysis, root locus analysis and bode analysis, the performance of presented AVR system is widely established in comparison with that based on ABC [1], MOL [2] and BBO [5]. The extensive results reported in this article show that the output voltage profile settles to the unit step reference with the least peak overshoot without compromising on settling time much. This outcome has improved the stability margin of the AVR system compared to other reported approaches.

II. PID CONTROLLER-BASED AVR DESIGN

In spite of many efforts in control engineering field, PID controller or its cousins have been still widely used in various types of control systems [10]. The reason of this wide usage comes from its easily understandable nature, ease of design and robust performance irrespective to model uncertainties with proper tuning of controller parameters [11]. In *s*-domain, the transfer function of a PID controller is expressed by

$$G_{PID}(s) = P + I + D = \frac{U(s)}{E(s)} = K_p + \frac{K_i}{s} + K_d \cdot s \tag{1}$$

where $E(s)$ is the error variable between the desired and real process output which produces the control signal $U(s)$ by computing the sum of proportional term P, integral term I and derivative term D. The three design parameters of this controller, i.e. proportional gain K_p, integral gain K_i and derivative gain K_d, must be tuned jointly by the operator depending upon the plant's dynamics. The resulting response against a unit step input should engage with the given reference with minimal settling time and no sustained oscillation.

As previously mentioned, the design objective of an AVR is to sustain the output voltage of a SG at a certain level. An AVR includes mainly four essential components, such as amplifier, exciter, generator and sensor. In order to investigate the AVR dynamic performance, the following transfer function modelling is assumed, where the major time constants are used and saturation or other nonlinearities are avoided.

A. Amplifier model: The amplifier model is given by a gain K_A and a time constant τ_A,

$$G_{amplifier}(s) = \frac{K_A}{1+\tau_A s} \tag{2}$$

K_A can be varied in the range of 10-40 and τ_A in the range of 0.02-0.1 s.

B. Exciter model: A simplified model of a exciter may be given by a gain K_E and a time constant τ_E

$$G_{exciter}(s) = \frac{K_E}{1+\tau_E s} \tag{3}$$

Standard values of K_E are in the range of 1-10 and τ_E in the range of 0.4-1.0 s.

C. Generator model: The generator is modelled by a gain K_G and a time constant τ_G

$$G_{generator}(s) = \frac{K_G}{1+\tau_G s} \tag{4}$$

K_G and τ_G depend on loading conditions. K_G ranges from 0.7 to 1.0 and T_G from 1.0 s to 2.0 s.

D. Sensor model: The sensor is often modelled simply by a gain K_S and a time constant τ_S

$$G_{sensor}(s) = \frac{K_S}{1+\tau_S s} \tag{5}$$

τ_S normally takes small values ranging over 0.001-0.06 s and K_S is in the neighborhood of 1.0.

In this article, to lead a fair comparison with [1, 2, 5], the same parameter values have been used as $K_A = 10$, $\tau_A = 0.1$, $K_E = 1.0$, $\tau_E = 0.4$, $K_G = 1.0$, $\tau_G = 1.0$, $K_S = 1.0$ and $\tau_S = 0.01$. Thus, the entire AVR transfer function model with that of PID controller is given in Fig. 1.

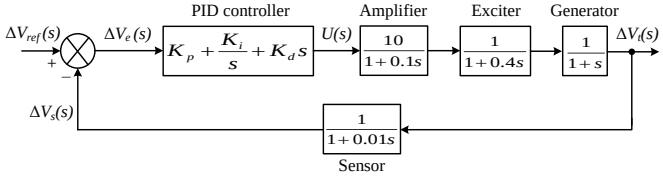


Figure 1: Transfer function model of PID-based AVR control system.

The original terminal voltage step response of the above AVR system without any controller is depicted in Fig. 2, from which the system is observed to be severely oscillating in the beginning and has remarkable error in steady-state. In power systems, such a response is completely unacceptable and cannot be allowed to emerge. As a result, to improve the transient response of the AVR system and eliminate the steady-state error, a controller such as a PID is required to be installed in the concerned system.

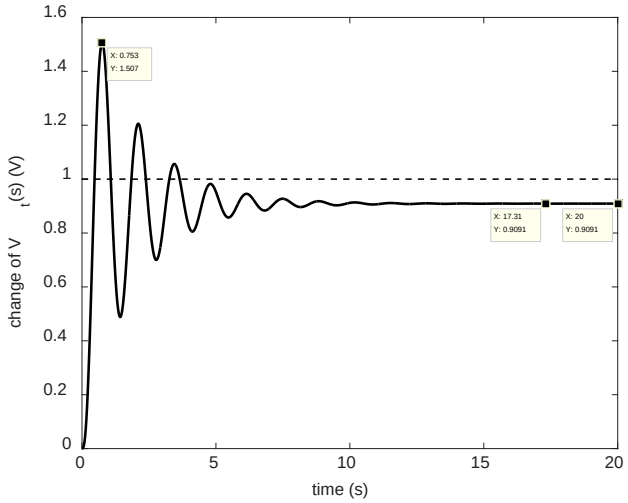


Figure 2: Original terminal voltage step response of an AVR system without any controller.

III. DESIGN OF PID EMPLOYING SOS

In this part, an efficient PID controller design employing the SOS algorithm, which can be shortly referred to as SOS-PID controller, is realized to enhance the voltage response profile of the AVR system while maintaining satisfactory stability margin. The incorporation of SOS in such a problem is primarily owing to the desire of attaining three controller coefficients K_p, K_i and K_d, so that the controlled system can have the desired performance. SOS is a relatively new algorithm proved to be powerful and robust over different kinds of optimization problems [8]. It operates on the basis of three common symbiotic strategies such as mutualism, commensalism, and parasitism developed by organisms. The property of the interaction characterizes the basis of each phase. Mutualism is the symbiotic interaction in which each organism benefits from the other's activity. Commensalism evolves when one organism receives benefits, while the other organism is neutral, and parasitism evolves when an organism gains benefits from a certain interaction at the cost of degrading the other [12]. Through all phases, each organism interacts randomly with other organisms in the ecosystem. After completion of these three symbiotic strategies, SOS tries another generation and is iterated recursively until pre-defined termination criteria are satisfied. The following outline summarizes the afore-said explanations.

Initialization
REPEAT
i = 1;
while *i* is different from *eco_size*
–Mutualism phase
–Commensalism phase
–Parasitism phase
–*i* = *i* + 1;
end while
UNTIL (termination criteria are met)

For further insight into the SOS algorithm procedures, readers are referred to the original study given by [8].

In order to implement SOS algorithm for optimizing the PID controller gains, three design parameters are defined to form an individual organism K by $K = [K_p, K_i, K_d]$ where each member is represented by a real number. Thus, there are three members in an individual to be optimized and each individual may be treated as the PID controller. To assess the performance of distinct PID controllers in the ecosystem, a suitable cost function complying with the requirements and necessities in a control system design must be defined properly. In literature, there exist various performance measures for design of controllers such as integral of absolute error (IAE), integral of squared error (ISE) and integral of time weighted squared error (ITSE) [13]. An important deficiency of IAE and ISE is that they lead to sluggish response due to the accumulated errors regardless of the time. On the other hand, ITSE can solve this problem and increase the dynamic response, but this is not desirable as far as stability margin is concerned as in [3]. Instead, integral of time weighted absolute error (ITAE) has been demonstrated in several studies to exhibit better system performance as compared to its integral-based counterparts [14]. In order to make this study compete with [1, 2, 5], two frequency-domain parameters are combined with the sole ITAE by weighting each term as shown in (6).

$$J = w_1 \int_0^{t_{sim}} t\left|e(t)\right| dt + w_2 \alpha + \frac{w_3}{\beta} \quad (6)$$

where the integral of time multiplied absolute error $e(t)$ is computed numerically up to the sufficiently chosen simulation time t_{sim} and importance of each term in J is set by a weight factor w_j. α stands for the number of the complex poles and β is the sum of the damping ratios corresponding to complex poles. Recommended values of the weighting factors are $w_1 = 0.71$, $w_2 = 0.2$ and $w_3 = 0.09$. As a result, the purpose of the present SOS-based optimization task is to search PID parameters in the range of 0.01-2.0 in order that the minimum value of J could be achieved.

IV. NUMERICAL RESULTS

Here, simulation results obtained after applying the presented technique to the AVR system are provided, and a fair comparison from perspective of transient response analysis, root locus analysis and bode analysis is also made with those based on ABC [1], MOL [2] and BBO [5], which have been published in esteemed journals. In SOS algorithm, only two parameters are set as ecosystem size = 30 and maximum iteration number = 30. Simulations were implemented in the Matlab 8.5.0 (R2015a) software installed on a computer with an Intel core (TM) i5 3.3GHz processor and 8GB memory.

One of the major observations of the present study for analyzing the voltage response curve is given in Fig. 3, where the output voltage profile of the proposed PID-controlled AVR system to the step input perturbation is portrayed in a superimposed manner with those yielded by other approaches *viz*. ABC [1], MOL [2] and BBO [5]. From this figure, it may be seen that our proposal exhibits a better unit step response than the other approaches for the identical AVR system. For instance, peak overshoot of the response is further improved compared with MOL-based response while maintaining nearly the same settling time. This may be taken as a prior indicator that the stability margin of our proposal will be higher than that tuned by MOL algorithm. Also notice that the responses with ABC and BBO cannot be hold at 1.0 pu for the considered simulation time. This is attributed to the use of ITSE as a cost function in those studies, which leads to unrealistic evaluations owing to squaring error.

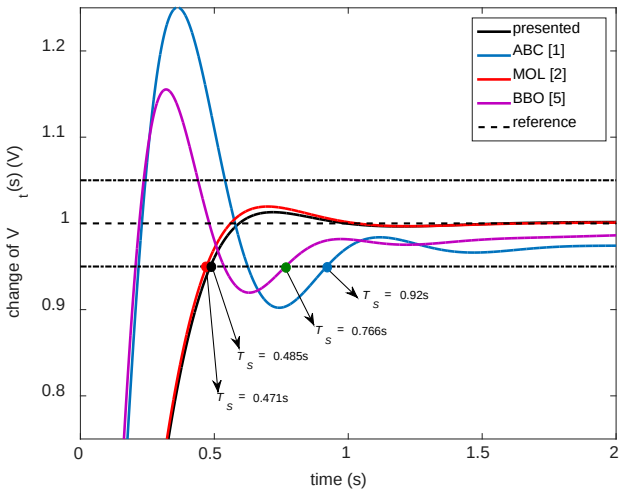


Figure 3: Comparative terminal voltage response profiles.

A comparison of the numerical results of the transient response analysis regarding the time responses in Fig. 3 is reported in Table 1. Transient response analysis covers the time domain performance characteristics such as peak overshoot (M_P), rise time (T_R) and settling time (T_S, 5% bant). In addition, optimized PID controller parameters are also given in their respective sections of Table 1.

Table 1: Comparative controller parameters and corresponding transient response specifications.

	Presented	ABC [1]	MOL [2]	BBO [5]
K_p	0.5693	1.6524	0.5857	1.2464
K_i	0.4097	0.4083	0.4189	0.5893
K_d	0.1750	0.3654	0.1772	0.4596
M_P	**1.013**	1.250	1.020	1.160
T_S	0.485	0.920	**0.471**	0.766
T_R	0.353	0.156	0.343	**0.149**

Bold text indicates result of interest

From Table 1, it may be inferred that SOS-based PID control offers the best value of peak overshoot and comparable settling time with only 2.9% less than MOL-based result. The best result concerning rise time belongs to BBO algorithm.

For the stability concern of the studied AVR system optimized by the proposed technique, root locus analysis is performed and the respective root locus curve is depicted in Fig. 4. As shown, all the closed-loop poles are located at the left side of the s-plane, meaning that the proposed control application is stable.

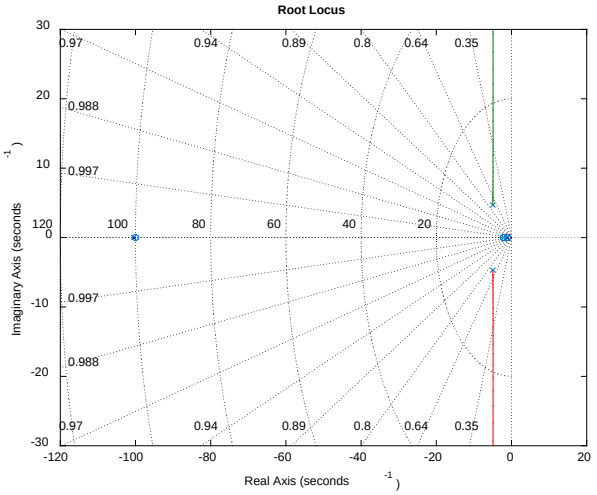


Figure 4: Root locus curve of the studied AVR system.

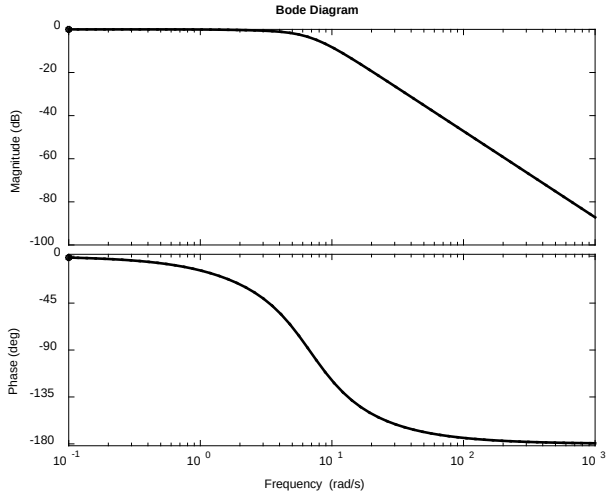


Figure 5: Bode diagram of the proposed AVR system.

The closed-loop poles and their respective damping ratios in Fig. 4 are also computed and gathered in Table 2 in comparison with the other indicated studies. It is manifest that the conjugate poles of the presented AVR system are farther away from the imaginary axis, which makes the system be controlled with the biggest damping ratio as desired.

Table 2: Closed-loop poles and their respective damping ratios of the AVR system optimized by SOS, ABC, MOL and BBO algorithm.

Algorithm	Closed-loop pole	Damping ratio
Presented	−100.48	1
	−1.98	1
	−1.10	1
	−4.97+4.69i	**0.727**
	−4.97−4.69i	**0.727**
ABC [1]	−100.98	1
	−4.74	1
	−0.25	1
	−3.75+8.40i	0.40
	−3.75−8.40i	0.40
MOL [2]	−100	1
	−2.11	1
	−1.06	1
	−4.92+4.72i	0.721
	−4.92−4.72i	0.721
BBO [5]	−100.0	1
	−2.1	1
	−0585	1
	−4.8+10.2i	0.427
	−4.8−10.2i	0.427

So as to investigate the stability of the proposed AVR through another point of view, frequency response or bode analysis of the control system is conducted, and the resulting bode diagram is depicted in Fig. 5. The peak gain, phase margin, delay margin and bandwidth parameter corresponding to this bode plot are tabulated in Table 3.

From Table 3, it is seen that the minimum peak gain, maximum phase margin and maximum delay margin, which are essential factors required for enhanced stability, are provided by both our proposal and that based on MOL algorithm. With regard to bandwidth, its maximum value is offered by using BBO algorithm. As a consequence, as far as peak gain, phase margin and delay margin measures of the bode analysis are concerned, the same performance is achieved by deploying SOS and MOL, and they are the pioneers over the remaining techniques.

Table 3: Peak gains, phase margins, delay margins and bandwidths of different AVR systems using SOS, ABC, MOL and BBO algorithm.

	Presented	ABC [1]	MOL [2]	BBO [5]
Peak gain (dB)	**0.0**	2.87	**0.0**	1.56
Phase margin (deg.)	**180**	69.4	**180**	81.6
Delay margin (s)	**Inf.**	0.111	**Inf.**	0.122
Bandwidth	6.15	12.88	6.34	**14.28**

V. CONCLUSION

Tuning problem of control parameters of a PID controller working in an AVR control application is addressed and tried to be solved in a better fashion by employing SOS algorithm. After a new cost function definition is realized with regard to both time-domain and frequency-domain performance criterion, SOS is invoked subsequently as the powerful and robust optimization algorithm to tune controller gains in a sense that the minimum value of the defined cost function could be achieved. In order to appraise the effectiveness of presented approach, three popular studies are chosen from the literature as a benchmark, then the results are compared with each other under identical conditions from the perspective of diverse analysis techniques such as transient response analysis, root locus analysis and bole analysis. Simulation results show that

the cooperation of introduced cost function and SOS improves the trade-off between the dynamic response and the stability margin of the system. In this context, the presented approach is able to effectively improve the stability degree of the considered AVR systemby further reducing the peak overshoot of the system time response compared with the existing approaches. Moreover, according to the various tests performed by the author, it has been seen that if desired, stability margin of the system can be further improved by increasing w_3 in (6). However, this will degrade the transient-time characteristics such as rise time and settling time. For interested researchers, it would be of interest to use other recent powerful optimization algorithms along with the cost function defined in this paper in the hope of improving the AVR performance further.

REFERENCES

[1] H. Gozde, M.C. Taplamacıoglu, "Comparative performance analysis of artificial bee colony", J Franklin Inst, 348, pp. 1927-1946, 2012.

[2] S. Panda, B.K. Sahu, P.K. Mohanty, "Design and performance analysis of PID controller for an automatic voltage regulator system using simplified particle swarm optimization algorithm for automatic voltage regulator (AVR) system", J Franklin Inst, 349, pp. 2609-2625, 2012.

[3] E. Çelik, "Incorporation of stochastic fractal search algorithm into efficient design of PID controller for an automatic voltage regulator system", Neural Comput Appl, doi: 10.1007/s00521-017-3335-7, 2018.

[4] S. Chatterjee, V. Mukherjee, "PID controller for automatic voltage regulator using teaching–learning based optimization technique", Int J Electr Power Energy Syst, 77, pp. 418-429, 2016.

[5] U. Güvenç, T. Yiğit, A.H. Işık, İ. Akkaya, "Performance analysis of biogeography-based optimization for automatic voltage regulator system", Turk J Electr Eng Comput Sci, 24, pp. 1150-1162, 2016.

[6] H. Gözde, M.C. Taplamacıoglu, M. Arı, "Simulation study for global neighborhood algorithm based optimal automatic voltage regulator (AVR) system", 5th International Istanbul Smart Grids and Cities Congress and Fair, pp. 46-50, İstanbul, 2017.

[7] H. Gözde, M.C. Taplamacıoglu, M. Arı, "Automatic voltage regulator (AVR) design with chaotic particle swarm optimization", International Conference on Electronics, Computers and Artificial Intelligence, pp. 23-26, Bucharest, Romania, 2014.

[8] M.Y. Cheng, D. Prayogo, "Symbiotic organisms search: a new metaheuristic optimization algorithm", Comput Struct, 139, pp. 98-112 2014.

[9] E. Çelik, N. Öztürk, "First application of symbiotic organisms search algorithm to off-line optimization of PI parameters for DSP-based DC motor drives", Neural Comput Appl, doi: 10.1007/s00521-017-3256-5

[10] L.S. Coelho, "Tuning of PID controller for an automatic regulator voltage system using chaotic optimization approach", Chaos Solitons Fractals, 39, pp. 1504-1514, 2009.

[11] D. Guha, P.K. Roy, S. Banerjee, "Study of differential search algorithm based automatic generation control of an interconnected thermal-thermal system with governor dead-band", Appl Soft Comput, 52, pp. 160-175, 2017.

[12] V.F. Yu, A.A.N.P. Redi, C.L. Yang, E. Ruskartina, B. Santosa, "Symbiotic organisms search and two solution representations for solving the capacitated vehicle routing problem", Appl Soft Comput, 52, pp. 657-672, 2017.

[13] M. Zamani, M.K. Ghartemani, N. Sadati, M. Parniani, "Design of a fractional order PID controller for an AVR using particle swarm", Control Eng Pract, 17, pp. 1380-1387, 2009.

[14] S. Padhy, S. Panda, "A hybrid stochastic fractal search and pattern search technique based cascade PI-PD controller for automatic generation control of multi-source power systems in presence of plug in electric vehicles", CAAI Transactions on Intelligence Technology, 2, pp. 12-25, 2017.

Kinect Calibration and Data Optimization For Anthropometric Parameters

M.S.GOKMEN [1], M.AKBABA [2], O.FINDIK [3]

[1] Karabuk University, Karabuk /Turkey, selmangokmen@karabuk.edu.tr
[2] Karabuk University, Karabuk /Turkey, mehmetakbaba@karabuk.edu.tr
[3] Karabuk University, Karabuk /Turkey, oguzfindik@karabuk.edu.tr

Abstract **- Recently, through development of several 3d vision systems, widely used in various applications, medical and biometric fields. Microsoft kinect sensor have been most of used camera among 3d vision systems. Microsoft kinect sensor can obtain depth images of a scene and 3d coordinates of human joints. Thus, anthropometric features can extractable easily. Anthropometric feature and 3d joint coordinate raw datas which captured from kinect sensor is unstable. The strongest reason for this, datas vary by distance between joints of individual and location of kinect sensor. Consequently, usage of this datas without kinect calibration and data optimization does not result in sufficient and healthy. In this study, proposed a novel method to calibrating kinect sensor and optimizing skeleton features. Results indicate that the proposed method is quite effective and worthy of further study in more general scenarios.**

Keywords **- Kinect Sensor, Anthropometry, Joint, Optimization, 3d vision systems**

I. INTRODUCTION

In the last decade, by development of 3d vision systems, have been made available for scientific applications. Most important feature of 3d vision systems (RGBD cameras) is that capturing distance close to real world distances approximately [1]. This feature made Kinect sensor available for medical and biometric applications [2]. In 2010, Microsoft company released a new RGBD camera that called Kinect v1 sensor. After, Microsoft company released Kinect Sensor v2 that contains Time of Flight (ToF) technology in 2014. Cause of Kinect sensor affordable among other sensors, Kinect sensor have been made available that for application development [3]. Another major factor that made Kinect sensor popular, providing for 3d coordinate datas as real-like [4].

Studies made with Kinect sensor, substantially focused on gait recognition, gesture recognition and biometric analysis. Used data sets in this studies that are usually unexceptional and standart [5]. Cause for this, reducing inaccuracy arising out of by Time of Flight technology. Therefore, for improving data quality that obtained from Kinect sensor, calibration and data optimization is essential.

In this study, provided calibration of Kinect sensor and optimization of datas that obtained by Kinect sensor. In the first stage of calibration, measured distance between Kinect sensor and ground. Therefore, aimed to establish height of joint points. The dataset used for Kinect calibration is determined as vertical and horizontal to Kinect sensor on XZ space. Calibration data were collected as 10 times vertically to XZ space and 10 times horizontally to XZ space.

II. RELATED WORKS

Consistency of anthropometric parameters is very important for improvement of studies in gait recognition, gesture recognition and biometric fields. The most common used parameters are temporal parameters, spatial parameters and kinematic parameters in gait and gesture recognition [6]. Spatial and temporal parameters are the intuitive gait features including step length, speed, gait cycle, average stride length, and so on. Kinematic parameters are usually characterized by the joint angles between body segments and their relationships to the events of the gait cycle [6]. Anthropometric parameters are individual features including like bone lengths [7].

Accuracy of spatial and temporal parameters is related to accuracy of kinematic parameters. In situation that joint points coordinate data are incompatible, can not expected being result healthy of length features like stride lengths. In some studies, for preserving of kinematic parameters accuracy, individuals was asked to walking on straight path [8]. Thus, the data collected as least affected from noise of unexpected movements. Movement of contrary to expectations, can effect to consistence of parameters. Therefore, Kinect calibration is necessary for accuracy of parameters that obtained from individuals that in anywhere in Kinect sight [9].

Measurements between joint points are utilized in gait recognition or gesture recognition fields [10]. In related studies, relative distance features (RDF) and vertical distance features (VDF) are particularly used [4][5]. Obtaining coordinate data of joint points as accurate is very important for healthy resulting of study. Thus, importance of Kinect sensor calibration and optimization of data obtained from Kinect sensor that is essential.

Anthropometric parameters are acquired from in consequence of processing of spatial, temporal and kinematic parameters. Anthropometric parameters are fundamental values in performed applications that worked on biometric and identification applications [7]. For accurate of obtained anthropometric parameters from Kinect sensor, consistency of spatial and temporal parameters are crucial. The improperly parameters that obtained from Kinect sensor, results wrong classification and wrong bone lengths calculation. In studies that used anthropometric parameters, does not performed any

application for demonstration of anthropometric parameters accuracy [12]. Reason for this, generation of unstable parameters from non calibrated Kinect sensor. Therefore, unstable and low accuracy parameters must be optimized. Importance of clear and stable joint points that obtained from Kinect sensor, is bone lengths are uniquely by individual. Hence, importance of anthropometric parameters are seen obviously that used in identification applications.

III. METHODOLOGY

Kinect sensor v2 that is used in this study, developed for X-box 360 video console and it connected to pc by usb converter and works with SDK of version 2.0. The Kinect sensor equipped with RGB camera, light emitter and infrared-sensitive depth sensor. Parameters that obtained from Kinect sensor, processes by software library that known NUI API (Natura User Interface). This API made available that 3d data of individuals. The data obtained from Kinect sensor by NUI API that is 25 joint points which occurs coordinates of X,Y and Z that have totally 75 piece of data [11]. Joint points represent as joint points and real world coordinates that belongs to individual. Through usage of SDK, can be achieve data with 30 frame per second. With every captured frame, the data is obtained that contains joint points stored in an array.

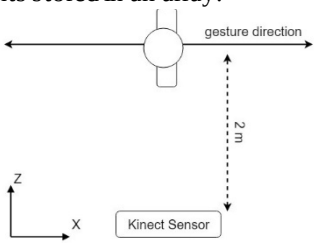


Figure 1 - Horizontal gait direction

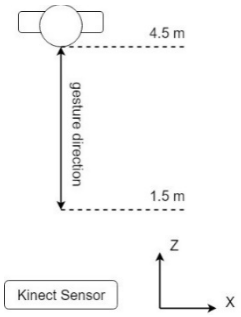


Figure 2: Vertical gait direction

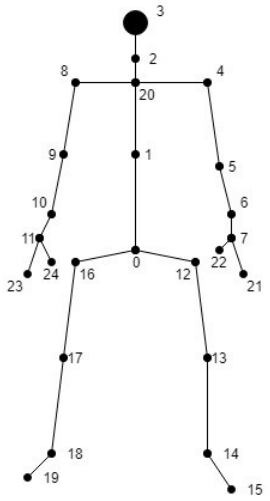


Figure 3 - Joint Points

Joint	Value
Base of the spine	0
Middle of the spine	1
Neck	2
Head	3
Left shoulder	4
Left elbow	5
Left wrist	6
Left hand	7
Right shoulder	8
Right elbow	9
Right wrist	10
Right hand	11
Left hip	12
Left knee	13
Left ankle	14
Left foot	15
Right hip	16
Right knee	17
Right ankle	18
Right foot	19
Spine at the shoulder	20
Tip of the left hand	21
Left thumb	22
Tip of the right hand	23
Right thumb	24

Table 1: Joint Point Index

First, data were collected vertically to Kinect sensor on XZ coordinate space as shown on fig. 2. When collected datas observed, determined joint point coordinates have irregularity on it. The reason for this is angle of inclination between Kinect sensor and X coordinate axis. By variance of inclination angle, Kinect sensor's coordinate axis change according to inclination angle. Height of Kinect sensor from ground represented as h_k on fig. 5.

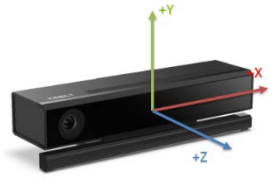


Figure 4 : Kinect coordinate axis

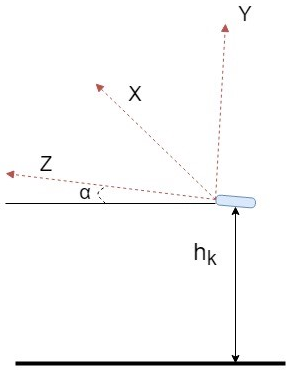


Figure 5: Kinect with inclination angle

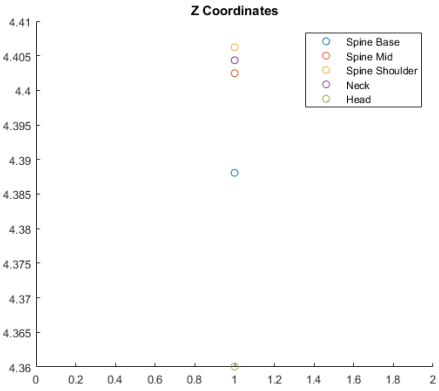


Figure 6: Z coordinates of Joints

In fig. 6, shown Z coordinate values of some specified joint points gathered from individual which standing. Inconsistency that occurs by inclination angle is noticeable in fig. 6.

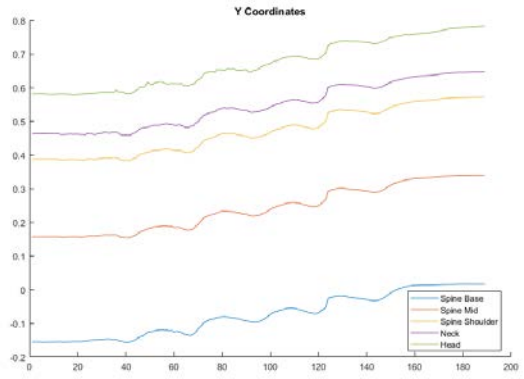


Figure 7: Y coordinate variation

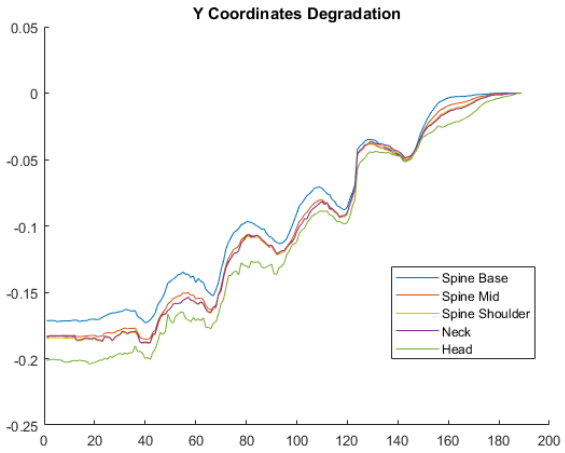
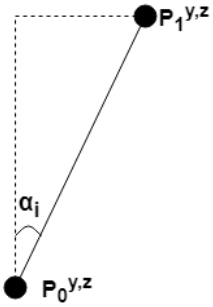


Figure 8: Difference between Y coordinate of last frame and Y coordinate of each frame with raw data

As shown on fig. 7, Y coordinate variations of specified joint points are not consistent caused by inclination angle. To solve this problem, inclination angle must be calculated. In fig. 8, clearly visible variety of Y coordinate values between last frame and each frame.

In our method, while resolving inclination angle, used two spine joint points that are base of the spine and middle of the spine. If inclination angle represents as α, according to every performed gait for calibration, detected least inconsistency on these joint points. Inclination angle was represented as $\propto_m^b$ that for every performed calibration gait. In function that shown at above, b variance is represented as calibration gaits. Inclination angle (α) is calculated for each frame in calibration and represented as $\propto_i$. n variable represents frame number for calibration gaits. m variable shown as 'mean'.



$$\propto_m^b = \frac{\sum_{i=1}^{n} \propto_i}{n} \qquad \textit{Eq.1}$$

Inclination angle that intended to use for calibration ($\propto_g$), represented as geometric mean of each obtained arithmetic mean of inclination angle from calibration gaits.

$$\propto_g = \left(\prod_{i=1}^{n} \propto_m^i\right)^{1/n} \qquad \textit{Eq. 2}$$

By calculation of mean inclination angle ($\propto_g$), could improve inconsistencies on Y and Z coordinate axis. Y and Z coordinate datas that belong to each joint points is represented as in equation at below.

$$J_i^p(i_0{:}i_n) = < P_i^x, P_i^y, P_i^z > \qquad \textit{Eq. 3}$$

In equation at above, i variable represents as index number of each joint point at table 1. In equation 4, new joint point coordinates that calibrated according to $\propto_g$ are represented as in equation at below.

$$J_{i,c}^p(i_0{:}i_n) = < P_{i,c}^x, P_{i,c}^y, P_{i,c}^z > \qquad \textit{Eq. 4}$$

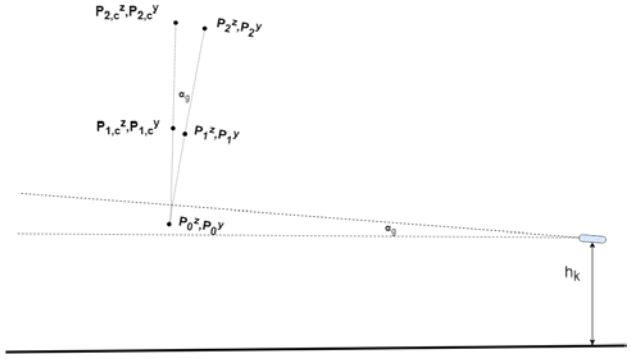


Figure 9: Calibration of Kinect sensor

$$P_{i,c}^z = P_i^y * \sin(\propto_g) + P_i^z \qquad \textit{Eq. 5}$$

$$P_{i,c}^y = P_{i,c}^z * \sin(\propto_g) + P_i^y + h_k \qquad \textit{Eq. 6}$$

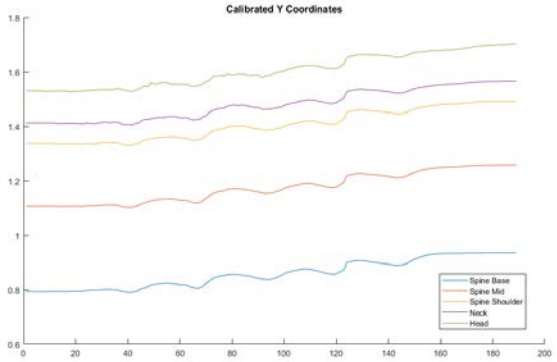


Figure 10: Calibrated Y coordinates

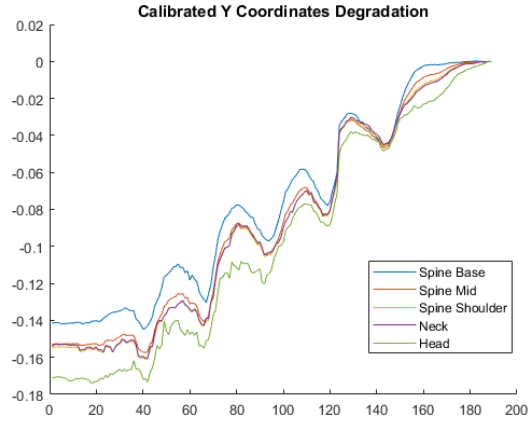


Figure 11: Difference between Y coordinate of last frame and Y coordinate of each frame with calibrated data

In Figure 10, joint points that calibrated has less inconsistency. Whereas, the joint points which located in higher Y coordinates has more inconsistency than located in lower Y coordinates. Cause of this perspective. Depending on perspective, observed that variance of joint point coordinates on individuals which getting closer to Kinect sensor. If assumed that equation $P_{i,c}(i_0{:}i_n) =< C_i^x, C_i^y, C_i^z >$, i represents index number of each joint point.

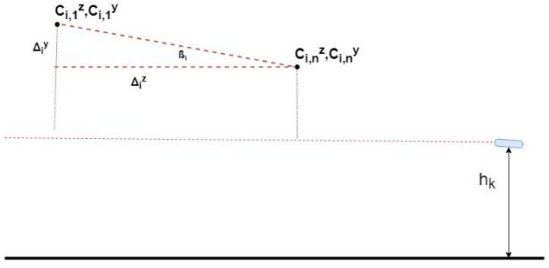


Figure 12: Perspective degree

In figure 12, $C_{i,1}^y, C_{i,1}^z$ and $C_{i,n}^y, C_{i,n}^z$ variables represent that Y and Z coordinate values of specified joint points at first and last

frame. Difference between coordinate values at first and last frame has shown as Δ_i^y and Δ_i^z in figure 12. After implementation of this variable, perspective degree ($ß_i$) is calculated shown as equation 7. After calculation of perspective degree ($ß_i$) for each calibration gait, mean perspective degree is represented as $ß_{i,m}$ in equation 8. In equation 8, b represents each calibration gait and i represents indexes of each joint points. For purpose of calculation of $ß_i$ value, joint points were elected according to Y coordinate values that are going from top to bottom in Y coordinate axis.

$$ß_i = tan^{-1}\left(\frac{C_{i,1}^y - C_{i,n}^y}{C_{i,1}^z - C_{i,n}^z}\right) \qquad \textbf{Eq. 7}$$

$$ß_{i,m} = \frac{\sum_{n=1}^{b} ß_{i,n}}{b} \qquad \textbf{Eq. 8}$$

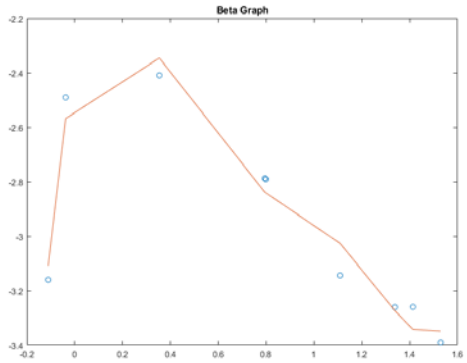


Figure 13: Beta Graph

Variance of perspective degrees according to Y coordinate values of calibrated joint points has shown in figure 13. $P_ß$ polynomial function was obtained by implementation of curve fitting from over $ß_i$ points that shown in figure 13. Through, can be found perspective degree related to height of Y coordinate axis.

$$ß_{i,n}^y = P_ß(C_{i,n}^y)$$
$$C_{i,n}^y = C_{i,n}^y + C_{i,n}^z * \tan(ß_{i,n}^y) \qquad \textbf{Eq. 9}$$

In equation 9, new Y coordinate values of calibrated joint points was calculated with $P_ß$ polynomial equation. n represents frame number in equation and i represents index number of joint points.

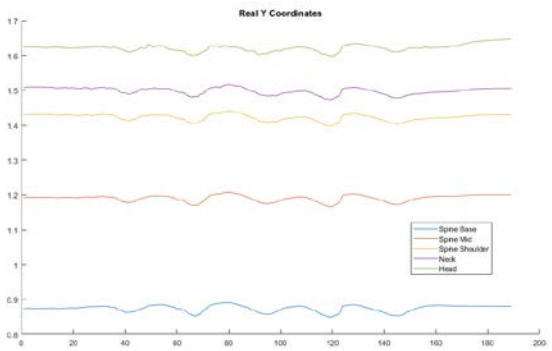


Figure 14: Y coordinate values of optimized data

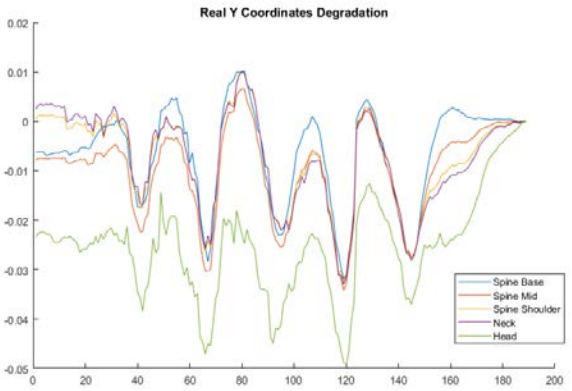


Figure 15: Difference between Y coordinate of last frame and Y coordinate of each frame with optimized data

Experimental results show that, our proposed method optimized data successfully. As shown in figure 14, Y coordinate value of joint points are stable. Reason for big majority of fluctuations is only from gait. Maximum difference variety is clearly seen obviously as 5 cm in figure 15.

IV. CONCLUSIONS

As a result of study, Kinect calibration and joint coordinate data optimization successfully achieved. In study, achieved data optimization of joint point coordinate values over no filter applied on it. This study proposes to contribute to studies will made with Kinect Sensor.

This work supported by Research Fund of the Karabuk University. Project Number: KBÜBAP-17-YL-250

REFERENCES

1. Sarbolandi, H., Lefloch, D., and Kolb, A., "Kinect range sensing: Structured-light versus Time-of-Flight Kinect", ***Computer Vision And Image Understanding***, 1391–20 (2015).
2. Castro, A. P. G., Pacheco, J. D., Lourenço, C., Queirós, S., Moreira, A. H. J., Rodrigues, N. F., and Vilaça, J. L.,

"Evaluation of spinal posture using Microsoft Kinect™: A preliminary case-study with 98 volunteers", ***Porto Biomedical Journal***, 2 (1): 18–22 (2017).

3. Kastaniotis, D., Theodorakopoulos, I., Theoharatos, C., Economou, G., and Fotopoulos, S., "A framework for gait-based recognition using Kinect", ***Pattern Recognition Letters***, 68327–335 (2015).
4. Yang, K., Dou, Y., Lv, S., Zhang, F., and Lv, Q., "Relative distance features for gait recognition with Kinect", ***Journal Of Visual Communication And Image Representation***, 39209–217 (2016).
5. Ahmed, M., Al-Jawad, N., and Sabir, A., "Gait recognition based on kinect sensor", .
6. Yoo, J.-H. and Nixon, M. S., "Automated Markerless Analysis of Human Gait Motion for Recognition and Classification", ***ETRI Journal***, 33 (2): 259–266 (2011).
7. Munsell, B. C., Temlyakov, A., Qu, C., and Wang, S., "LNCS 7585 - Person Identification Using Full-Body Motion and Anthropometric Biometrics from Kinect Videos", (2012).
8. Preis, J., Kessel, M., Werner, M., and Linnhoff-Popien, C., "Gait Recognition with Kinect", .
9. Gianaria, E., Balossino, N., Grangetto, M., and Lucenteforte, M., "Gait characterization using dynamic skeleton acquisition", ***2013 IEEE International Workshop On Multimedia Signal Processing, MMSP 2013***, 440–445 (2013).
10. Zhang, H., Zhong, P., He, J., and Xia, C., "Combining depth-skeleton feature with sparse coding for action recognition", ***Neurocomputing***, 230417–426 (2017).
11. Andersson, V. O. and Araujo, R. M., "Person Identification Using Anthropometric and Gait Data from Kinect Sensor", .
12. Andersson, V. O. and Araujo, R. M., "Full Body Person Identification Using the Kinect Sensor", ***2014 IEEE 26th International Conference On Tools With Artificial Intelligence***, (November): 627–633 (2014).
13. Ball, A., Rye, D., Ramos, F., and Velonaki, M., "Unsupervised clustering of people from "skeleton" data", ***Proceedings Of The Seventh Annual ACM/IEEE International Conference On Human-Robot Interaction - HRI '12***, 225 (2012).
14. Cunado, D., Nixon, M. S., and Carter, J. N., "Automatic extraction and description of human gait models for recognition purposes", .
15. Cunado, D., Nixon, M. S., and Carter, J. N., "Using Gait as a Biometric, via Phase-Weighted Magnitude Spectra", .
16. "IEEE Xplore Full-Text PDF:", **http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=1407887** (2017).
17. Lu, W., Zong, W., Xing, W., and Bao, E., "Gait recognition based on joint distribution of motion angles", ***Journal Of Visual Language And Computing***, 25754–763 (2014).

Some Experimental Studies of Segmentation of Different Medical Images using Color Difference on CIE L*a*b* Color Space

E.IRMAK

Karabuk University, Biomedical Engineering, Karabuk/Turkey, emrahirmak@outlook.com.tr

***Abstract* - Image segmentation is one of the most tedious and challenging research of image processing field and defined as the partitioning of a given image to a finite number of non-overlapping regions such that every homogenous region is connected through a sharp line. When segmenting an image, each pixel is assigned a label so that similar labelled pixels have certain visual characteristics. Image segmentation is a popular image processing task for image interpretation and analysis. Especially, the scope of medical image segmentation includes a wide spectrum of subjects including locating tumors, measuring tissue volumes, planning the treatment, enhancing the medical images, diagnosis, image retrieval etc. For image segmentation and/or other image processes, various color spaces are used: RGB, CMYK, HSI and YIQ. Each color space was proposed for specific purposes and each of them has certain advantages over the others. Color image segmentation provides the user much more information comparing to grayscale image segmentation. This paper performs different medical image segmentations using CIE L*a*b color space. Images are converted to L*a*b color space and color difference idea is used to segment the user selected regions. This paper also covers commonly used color models such as RGB, CMYK, HSI and YIQ.**

***Keywords* - Medical image segmentation, L*a*b color space, Color transformation, Color clustering.**

I. INTRODUCTION

IMAGE segmentation is a stepping-stone to get a much more meaningful and easier image to analyze by changing the visual representation of a given raw image. The objective of image segmentation is to provide a simplified version of original image for further processing. In this restricted sense, image segmentation, which is generally a pre-processing step in image processing and pattern recognition, plays a crucial part in the quality of image analysis. Image segmentation process divides a given image into a finite number of non-overlapping regions such that each region share exact or same characteristics such as pixel intensity, color, boundary, medical/anatomical information, motion, texture, etc. The output image of segmentation process is still the same input image but with a number of segments which collectively form the image. Each segment or region is sorted out to represent the local structure in the image. From this point of view, image segmentation is considered as a clustering problem. The scope of medical image segmentation includes a wide spectrum of subjects including locating tumors, measuring tissue volumes, planning the treatment, enhancing the medical images, diagnosis, image retrieval etc.

Image segmentation techniques can be mainly classified as:

- Histogram thresholding
- Feature clustering
- Region based
- Edge detection
- Fuzzy approaches
- Neural network based
- Genetic algorithm based

image segmentation methods.

The rest of this paper is organized as follows. In Section II, commonly used color models are described. The similarities and the differences between commonly used color spaces are discussed in this section. Section III is devoted for CIE L*a*b color model and experimental results are also demonstrated in this section. Finally, Section IV encompasses the conclusions about the experimental results.

II. COLOR MODELS

The image to be segmented can be grayscale or color image either. Correspondingly the image segmentation can be either grayscale image segmentation or color image segmentation. Although both segmentation processes share mainly similar algorithms and phenomena they are totally different from each other because of human eye's perceive characteristics. There are plenty of color image segmentation spaces in literature; each one has been designated for specific application. Therefore deciding a proper color space is hot topic if image processing field. Color image segmentation is more functional, comprehensive and utilizable than grayscale image segmentation however it is tedious and challenging comparing with grayscale image segmentation.

The world of color spaces is quite broad in scope or content, usually in a straight relation to their specific use. Fabric or paper industry, televisions, press, computers make use of color spaces to fulfill their duties. Although there are so many color spaces, there is neither single all-terrain color

space nor even single way to compare colors valid enough to everyone. The next sections cover the commonly used color models such as RGB, CMYK, HSI, and YIQ.

A. RGB Color Model

A color space can be thought as an intangible mathematical representation model which expresses the colors to be represented as cluster of numbers, typically as three values of color components. This allows reproducing dozens of colors in 2D-3D space. RGB color space is known as the simplest color space and each pixel in an image is denoted by Red, Green and Blue components which are three values of color components as stated previously. Using three primary colors, RGB color space provides users to produce many other colors by adding these three primary colors, being and additive color model. For instance, the combination of R, G, and B colors creates white color having location of (1, 1, 1) whereas the absence of R, G, and B colors creates the black color having location of (0, 0, 0) in the Cartesian coordinate system. The colors that locate on the line joining black to white have equal values of R, G and B components and they represent the grayscale colors [1]. See picture 1. RGB color model is generally used in television and computer screens.

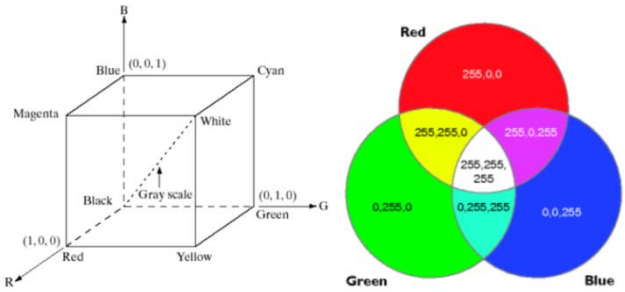


Figure 1: RGB color model.

In RGB color model to compare two colors or to find difference between two colors Euclidean distance is used. Thus,

$$\Delta C = \sqrt{\Delta R^2 + \Delta G^2 + \Delta B^2} \quad (1)$$

B. CMYK Color Model

CMYK, like RGB, models the output of physical devices rather than human visual perception. They have lack of approximating the human vision. In CMYK model, Red, Green and Blue are secondary colors which are primary colors in RGB model whereas Cyan, Magenta, Yellow and Black are primary colors which are secondary colors in RGB model. CMYK model is generally used for printing and is a subtractive color model which means that colors are formed by subtraction rather than addition. Figure 2 shows a comparative representation between RGB and CMYK color models.

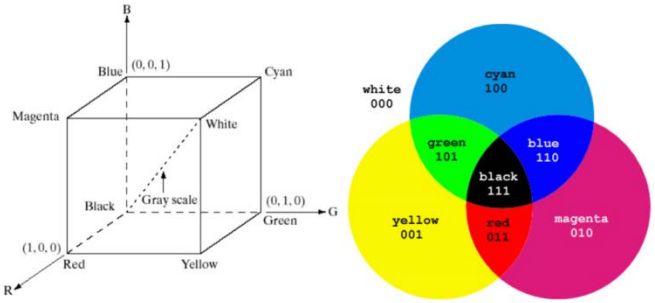


Figure 2: CMYK vs. RGB color model.

C. HSI Color Model

RGB and CMYK color models are ideally well-suited for device implementation rather than human eye perception. There are a number of color models that try to be based on human color perception. A human eye describes a color by its hue, saturation and brightness. Hue is a color attribute that describes pure color, whereas saturation gives a measure of the degree of which a pure color is diluted by white light [2]. Now that brightness is a subjective descriptor, which is not a measurable quantity, HSI color model resolves the intensity component from the color-carrying information which is described by a Hue and Saturation channel. I stands on gray diagonal and models the intensity of color whereas Hue is an angle representing just a single color without any nuance. See Figure 3. That is why Euclidean is not convenient to find difference between two colors rather the following distance would be more convenient [3]:

$$\Delta C = \sqrt{(I_2 - I_1)^2 + S_2^2 + S_1^2 - 2 * S_2 * S_1 * \cos(H_2 - H_1)} \quad (2)$$

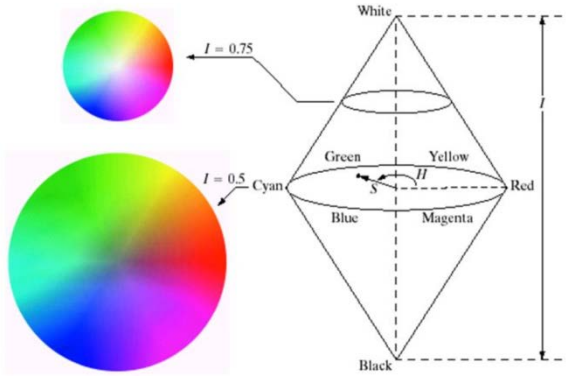


Figure 2: HSI color model with hue, saturation and intensity components.

D. YIQ Color Model

YIQ color model was proposed to separate chrominance which carries color information from luminance which carries grayscale information where Y channel stands for luminance information, I and Q channels stand for color information [4]. YIQ color model is a device dependent model and Y is considered as a combination of red, green and blue intensities while I is considered as hue and Q is considered as saturation. This model is used in color TVs. Figure 4 is a comparative representation between YIQ and RGB color models.

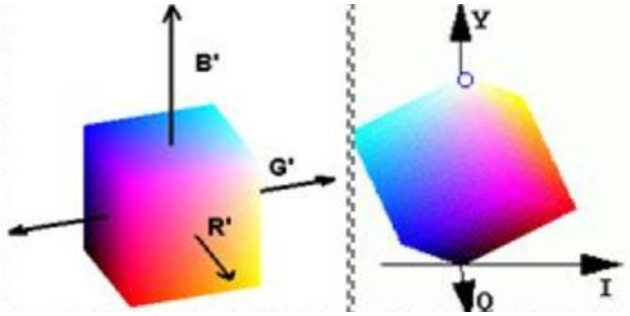


Figure 4: YIQ vs. RGB color model

III. CIE L*A*B* COLOR MODEL AND EXPERIMENTAL RESULTS

CIE L*a*b* color space is commonly considered and known as color-opponent space. In this study the purpose of using CIE L*a*b* color space is that it is more perceptually uniform than RGB or other spaces. The meaning of perceptually uniform is that a simple difference in a color value shall result in a same amount of difference in visual appearance. It contains the whole colors which are said to be visible to the naked human eye. The nonlinear relations for L*, a* and b* are intended to mimic the nonlinear response of the eye [5]. In CIE L*a*b* color space, the vertical axis L* stands for 'Lightness' or 'Luminosity' and its range is 0-100. The first horizontal axis which is represented by a* stands for colors between red-green axis. The idea is that a color cannot be both red and green. The horizontal axis which is represented by b* stands for the color fall along the blue-yellow axis. The idea is again that a color cannot be both blue and yellow. a* and b* layers contain color information whereas L* layer contains luminosity (lightness) information. Considering all the properties and benefits of CIE L*a*b* color space, it can be concluded that the difference between the two points is exactly similar with the human visual system.

In this study color difference is used to for segmentation. Color difference is a well-advised technique to compute the difference (distance) between two colors in color image segmentation science. The color difference is actually a Euclidean Distance represented by ΔE and formulated as follows.

$$\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \qquad (3)$$

(L_1^*, a_1^*, b_1^*) and (L_2^*, a_2^*, b_2^*) are two points having three components: L*, a*, b* in three-dimensional CIE L*a*b* color space.

Figure 5a is a Lower Limbs Angio CT64 image, Figure 5b is segmentation of Abdominal Aorta Branches from Pelvis and Figure 5c is filtering result. Median Filter has been used through the processes. L, a and b channels of original image and Histogram after segmentation can be seen in Figure 6. The data has been used from OSIRIX DICOM image library [6]. Figures 7a and 9a are used from The Cancer Imaging Archieve (TCIA), RIDER Neuro MRI Project [7, 8]. The brain MR image belongs to a patient with recurrent glioblastoma who underwent whole brain 3D FLASH imaging in sagittal plane. Figure 7b and 9b show segmented brain tumor and Figure 7c and 9c show filtering result. Figure 8 and 10 are L, a and b channels of original images and Histogram after segmentations.

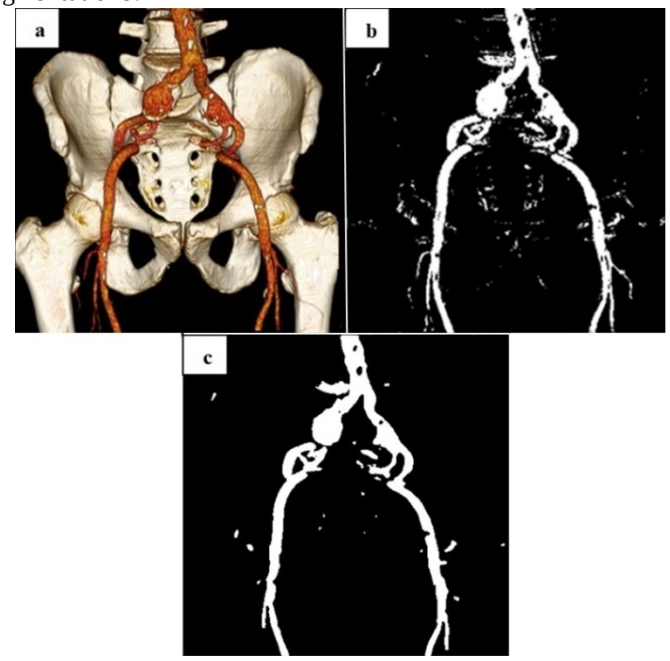


Figure 5: Segmentation of Abdominal Aorta Branches from Pelvis

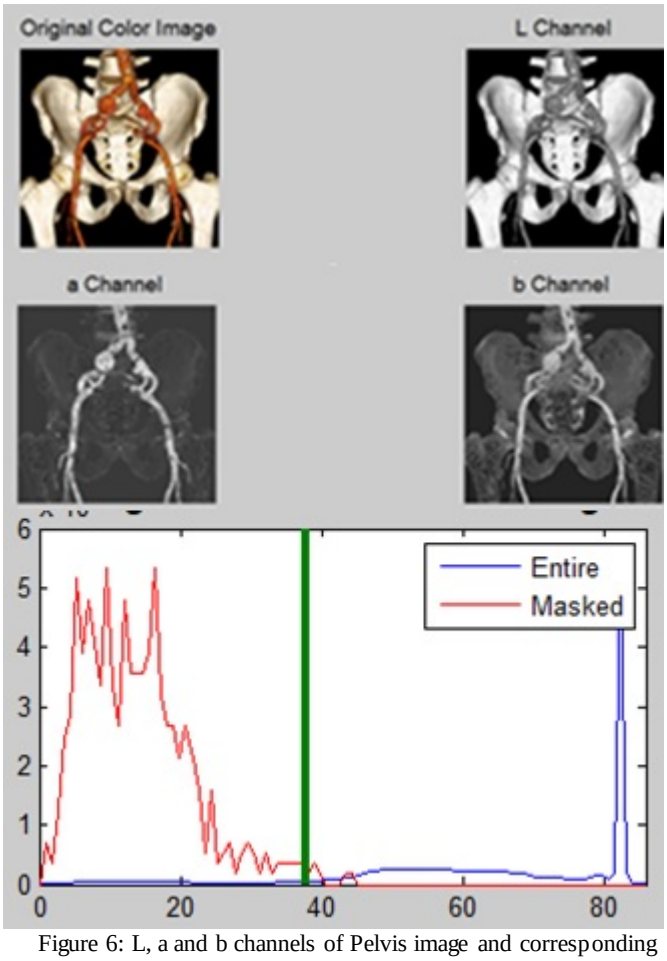


Figure 6: L, a and b channels of Pelvis image and corresponding histogram after segmentation

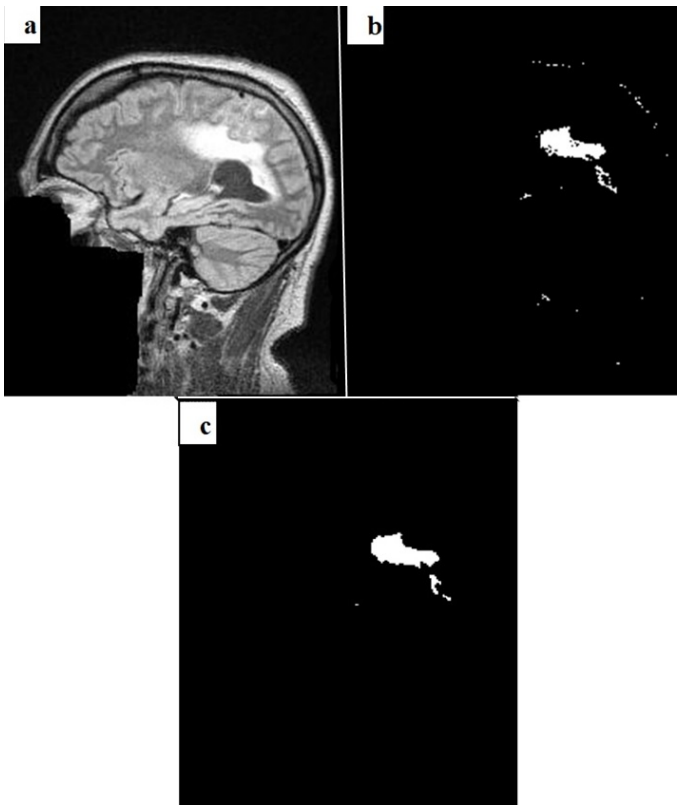


Figure 7: Segmentation of brain tumor

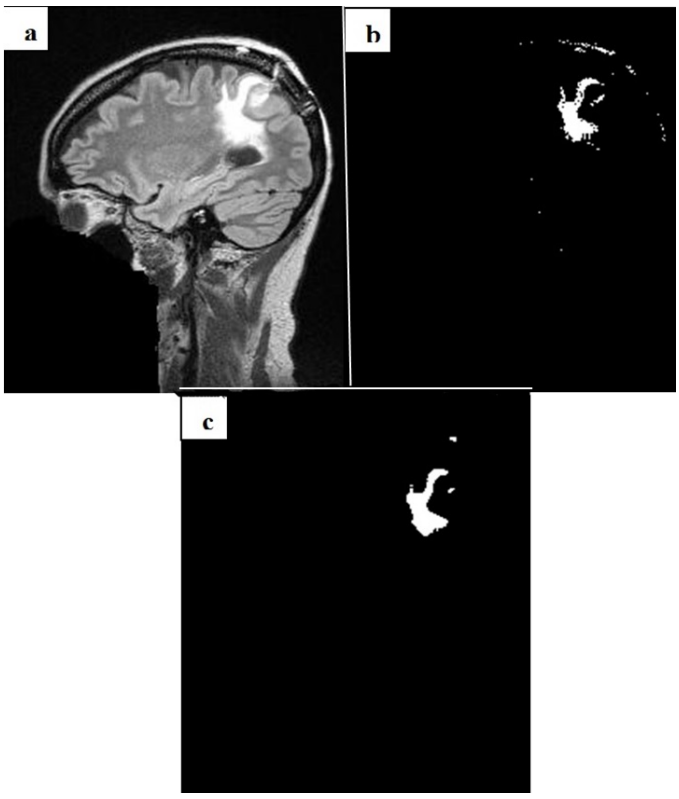


Figure 9: Segmentation of brain tumor

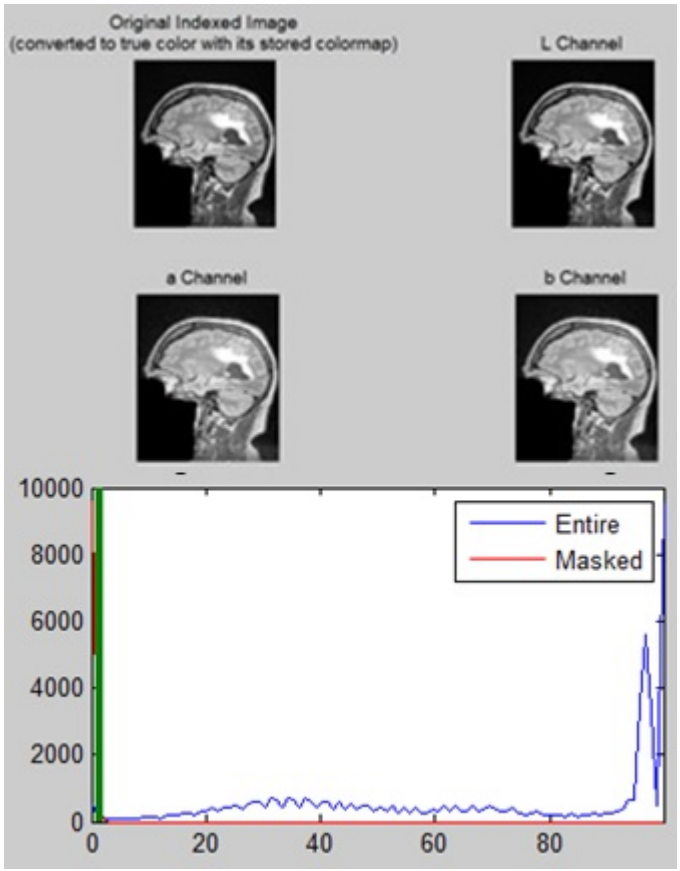


Figure 8: L, a and b channels of Brain image and corresponding histogram after segmentation

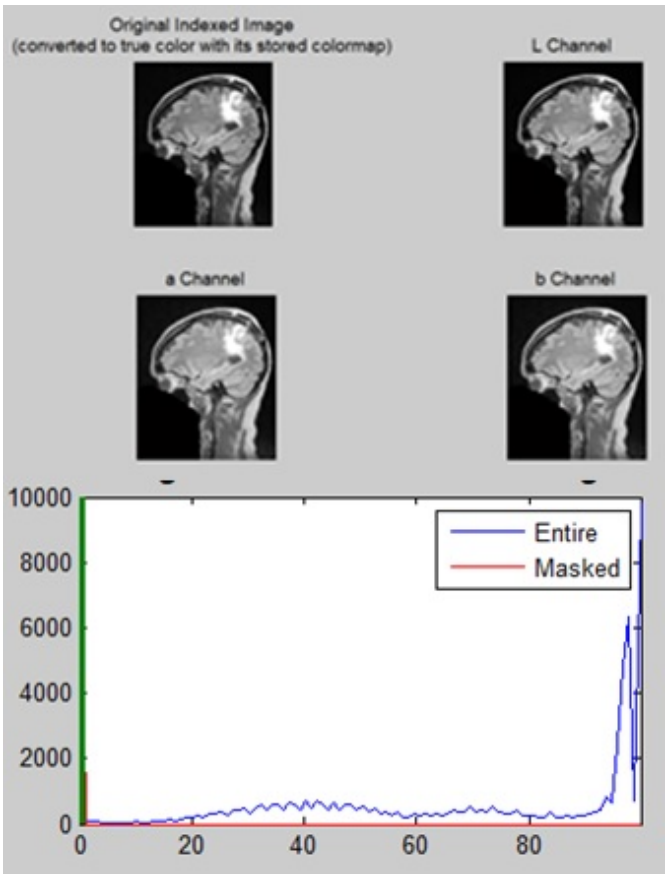


Figure 10: L, a and b channels of Brain image and corresponding histogram after segmentation

Figure 11a is PET-CT study with F18-fluorocholine in patient with prostate cancer and the data has been used from OSIRIX DICOM image library. In this image the liver is segmented from the rest of the image. Filtering result, L, a and b channels and Histogram after segmentation can be seen in the Figure 11 and Figure 12, respectively.

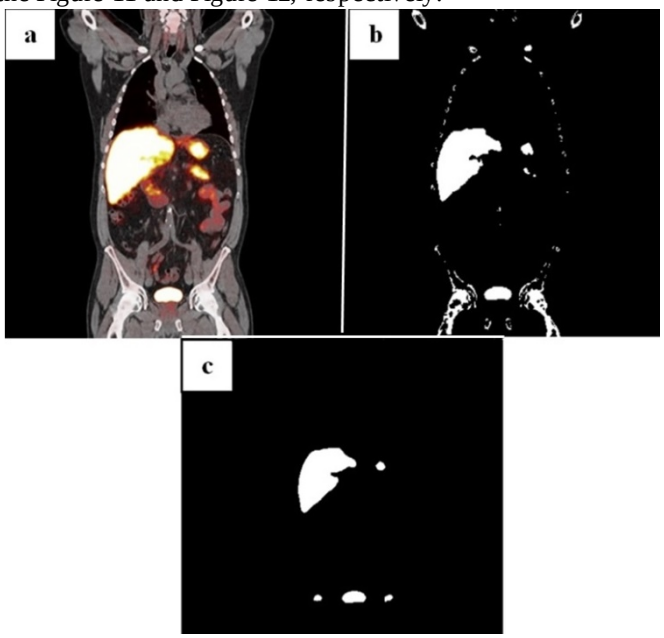


Figure 11: Segmentation of liver from whole body image

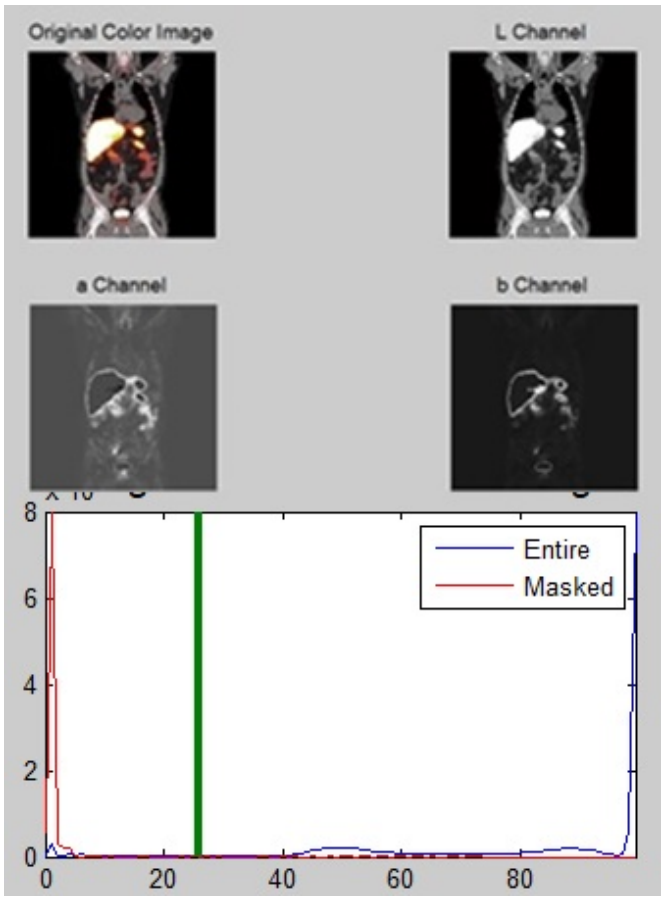


Figure 12: L, a and b channels of whole body image and corresponding histogram after segmentation

IV. CONCLUSION

This paper presented commonly used color models/spaces and four implementations of CIE L*a*b* color space based medical image segmentation. Because CIE L*a*b* color space identifies the whole colors which are said to be visible to the naked human eye, the nonlinear relations of L*, a* and b* are intended to mimic the nonlinear response of the eye. Four different medical images have been used from different modalities. User selected parts have been segmented from the rest of the images individually. This paper is intended to be a comprehensive reference source for the researchers involved in color image segmentation and medical image segmentation fields.

ACKNOWLEDGMENT

This work was supported by Scientific Research Projects Coordination Unit of Karabuk University. Project Number: KBÜ-17-DR-260 [D6].

REFERENCES

[1] P. Ganesan, V. Rajini and R. I. Rajkumar ''Segmentation and edge detection of color images using CIELAB color space and edge detectors''. *Emerging Trends in Robotics and Communication Technologies (INTERACT2010) International Conference IEEE*, pp. 393-397, 3-5 December 2010.

[2] R. F. Gonzalez and R. E. Woods, *Digital Image Processing.* New Jersey, Pearson-Prentice Hall, 3rd Edition, 2008, pp. 401-407.

[3] J. V. Llahi and A. Sanfeliu, "Color constancy and image segmentation techniques for applications to mobile robotics," *A Ph. D Thesis Submitted to the Department of Systems Engineering, Automatics, and Industrial Information Science at the Technical University of Catalonia*, 2015, pp. 117-143.

[4] K. Raval, R. Shukla and A. K. Shah, ''Color image segmentation using FCM clustering technique in RGB, L*a*b*, HSV, YIQ color spaces,'' *European Journal of Advances İn Engineering and Technology*, 2017, 4(3): 194-200.

[5] P. J. Baldevbhai and R. S. Anand, ''Color image segmentation for medical images using L*a*b* color space,'' *Journal of Electronics and Communication Engineering*, 2012,1(2): 24-45.

[6] OSIRIX DICOM Image Library, https://www.osirix-viewer.com/

[7] D. Barboriak, ''Data from Rider Neuro MRI'', *the Cancer Imaging Archieve*, 2015.

[8] K. Clark, B. Vendt, K. Smith, J. Freyman, J. Kirby, P. Koppel, S, Moore, S. Philips, D. Maffitt, M. Pringle, L. Tarbox and F. prior, ''the Cancer Imaging Archieve (TCIA): Maintaining and Operating a Public Information Repository'', *Journal of Digital Imaging*, 26(6): 1045-1057, 2013.

A Mobile Low-Cost Fire Detection System with Infrared Camera

Y. ORTAKCI[1], E. YILDIRIM[2] and O. DERECI[3]

[1] Karabuk University, Karabuk/Turkey, yasinortakci@karabuk.com
[2]Karabuk University, Karabuk/Turkey, emrullahyildirim@windowslive.com
[3]Karabuk University, Karabuk/Turkey, oguzz.dereci@gmail.com

***Abstract* – Forest fire is an important issue that damages thousands of hectares of forest and all creatures inside it. In this study we have developed an early fire detection system running on low-cost, lightweight Raspberry Pi module integrated with an infrared camera. The infrared camera acquires the live video of the fields, then if a fire occurs, our software detects the fire by taking both motion and color characteristics of a flames into account. In our study, there exist two phases: first, motion detection phase is applied to the view, if a motion is detected in the view, then, color detection phase is applied to determine the flames. Obtained results indicates that our fire detection system can detect a forest fire clearly and smoothly without requiring any external equipment and manual intervention.**

***Keywords* – Fire detection, forest fire, infrared camera, object extraction, color detection.**

I. INTRODUCTION

A fire in a city may cause loss of life and property, yet it can be controlled easier than a forest fire which may damages thousands of hectares of forest and all creatures inside it. Early detection of a forest fire is one of the primary issue of the countries. There have been many researches to detect the fires quickly, wisely and high accuracy. The existing detection and warning systems can be grouped into two fundamental categories as sensor-based and vision-based systems.

Sensor-based fire detection systems perform an adequate performance within indoor environment, yet, they cannot succeed the same performance at outdoor environment since they could not detect the smoke, fire and flame instantly in a long distance, while they could detect the temperature and smoke in short distance easily. In addition, a detailed sensor network, whose costs of communication, maintenance, and sustainability are expensive, should be installed in a wide outdoor environment to detect a fire. In contrast, vision-based fire detection systems provide a wider range and quicker detection service than sensor-based systems. The existing vision-based methods capture the image or video of a camera, then transfer them to a remote server for central processing to detect a potential fire case. Thus, they need a high bandwidth internet connection for fast data transfer.

In this study, we have developed a vision-based fire detection system that process the live video to detect forest fires. Our system runs on a Raspberry Pi 3 module, which has an integrated infrared camera, and is capable of detecting the fire instantly without transferring camera views to a remote server. All the process for fire detection is done on this module which is low-cost, lightweight, and runs real-time. Thus, our system can detect a forest fire easier than the existing fire detection system. On the other hand, since Raspberry module is small-scale and lightweight, it can be mounted to a drone that provides mobile fire detection service. Thereby, on the suspicion of a fire occurrence in any region, our module on a drone can check whether a real fire exists in a short time.

II. RELATED WORK

Vision-based fire detection systems are divided into two categories. First one includes the methods running on still images take only color and texture into account to recognise a fire and they have a limitation such as giving wrong alert for the object similar to a fire, yet, it is not a real fire. Second one includes the methods running on the motion videos and they have high accuracy for fire detection since they consider additional fire characteristics such as geometry, vibration, and motion. However, they have a limitation such as computational complicity and expensive hardware requirements.

Some of the vision-based fired detection systems are sensitive to flame, some are sensitive to smoke, and some are sensitive to both. Those system considers color, texture and motion characteristics of a fire. Nguyen et al. developed a video-based fire detection system considering color and motion parameters in a video series [1]. Seebamrungsat et al. presented a fire detection method which uses HSV and YCbCr color model to extract the fire from background in ambient light [2]. Wang and Zhou extracted flame by applying an iterative adaptive threshold technique [3]. Qiu et al. proposed an algorithm which detects a flame clearly using edge detection methods [4]. Santana et al. presented a real-time fire detection algorithm which aims to reduce false alarm rate of fire which frequently occurs at traditional sensor-based fire detection system [5]. To detect the flame, while Lei and Liu used frame differences, median filters and bayes classifier [6], Celik combines the color modelling and background registration [7].

All mentioned studies need to process the views on a remote server to detect a fire case, means that a high-cost system and high bandwidth connection between clients and server is required. However, our proposed system can detect a forest fire case instantly and smoothly on Raspberry Pi 3 module to which an infrared camera is integrated for acquiring real-time views of the field. All the processes to detect a fire in

these views are done in this module which is a low-cost, portable and real-time fire detection tool.

III. METHOD

Our system presents a two-phases algorithm in order to detect a fire in a video view. Figure 1 shows the general workflow of our proposed system.

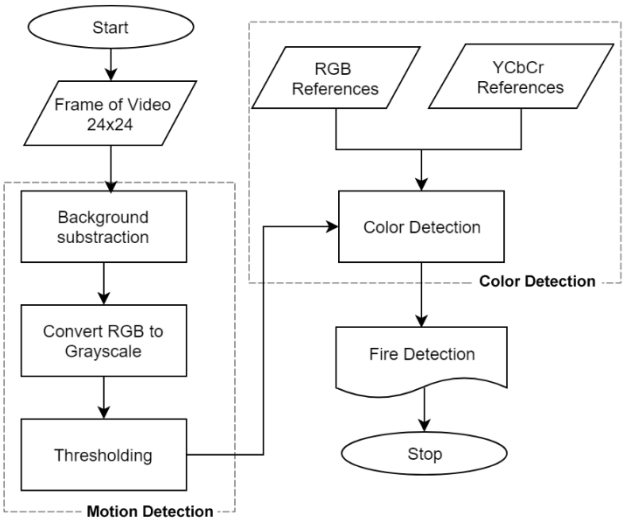


Figure 1: The workflow of our system

The video images are captured 25 times in a second and these captured images are divided to 24x24 pixel-size frames that are represented by matrices. Each frame is processed in two sequential phases in our study. While motion detection is done in the former phase, color detection is done in latter phase. In the first phase, background subtraction, gray-scaling and thresholding, respectively, are applied to each frame to extract the flames from the background and to detect whether there exists a motion. If a motion is detected in the frame, then our method proceeds the second phase which calculates color densities of the frame in RGB and YCbCr color spaces separately. Then, the color densities are compared to a reference color density of a fire. If their color densities are similar to the reference color density and a motion is detected in the corresponding frame, system produces a fire alert.

A. Motion Detection

In our study, motion detection is done in three steps; background subtraction, gray scaling and thresholding, respectively. In background subtraction, we have used $absdiff$ function (Equation 1), which calculates the absolute difference between given two image matrices, provided by OpenCV library. Thus, $absdiff$ function finds out the frames by comparing two images.

$$absdiff\left(|src_1(I) - src_2(I)|\right) \quad (1)$$

where $src_1(I)$ is first image matrix captured at t_1 time, $src_2(I)$ is second image matrix captured at t_2 time. $t_2 = t_1 + 40\ msec$

After background subtraction, we have transformed the color spaces of moving frames to grayscale from RGB. Then we have applied thresholding to the corresponding frame in order to make the moving object clearer by reducing the noise in the frame. Equation 2 has been applied to the grayscale frame to provide thresholding.

$$src_{result}(x,y) = \begin{cases} maxVal, & if\ src_{GrayScale}(x,y) > threshold \\ 0, & otherwise \end{cases} \quad (2)$$

A thresholding example of a frame is given in Figure 2.

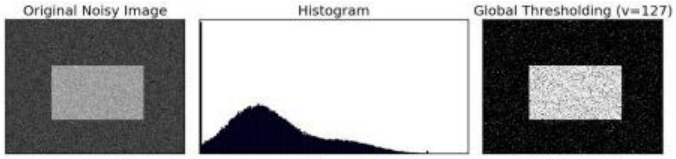


Figure 2: Thresholding of a sample frame

B. Color Detection

Color detection has an important role in the fire detection process. Our system evaluates color density of each moving object by using RGB and YCbCr color spaces to find out whether the pertinent frame includes a potential fire risk. First, RGB value of the frame is calculated, then, frame's color space is transformed to YCbCr as shown in Equation 3 for further analysis [8].

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2568 & 0.5041 & 0.0979 \\ -0.1482 & -0.2910 & 0.4392 \\ 0.4392 & -0.3678 & -0.0714 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (3)$$

After transformation, RGB and YCbCr values of a frame are investigated to recognize a forest fire by checking six conditions, two ones for RGB and four ones for YCbCr. If the frame satisfies all these six conditions mentioned in Table 1, it poses a fire risk.

Table 1: Conditions for fire detection [8]

No	Color Space	Conditions
1	RGB	$R > G > B$
2		$(R > R_{ref})$ **AND** $(G > G_{ref})$ **AND** $(B > B_{ref})$
3	YCbCr	$Y(x,y) \geq Cb(x,y)$
4		$Cr(x,y) \geq Cb(x,y)$
5		$(Y(x,y) \geq Y_{mean})$ **AND** $(Cb(x,y) \leq Cb_{mean})$ **AND** $(Cr(x,y) \geq Cr_{mean})$
6		$(Cb(x,y) \leq Cb_{ref})$ **AND** $(Cr(x,y) \geq Cr_{ref})$

R: Red value of the frame; G: Green value of the frame, B: Blue value of the frame

$Y(x,y)$: Y value of the frame, $Cb(x,y)$: Cb value of the frame, $Cr(x,y)$: Cr value of the frame

Binti Zaidi et al. stated reference values of RGB for a fire detection as $R_{ref} = 180$, $G_{ref} = 70$, $B_{ref} = 100$, $Cb_{ref} = 120$ and $Cr_{ref} = 150$ those they had found by considering the

means of 100 different fire views [8]. The calculation of $Y_{mean}(x,y)$, $Cb_{mean}(x,y)$, and $Cr_{mean}(x,y)$ are shown in Equation 4, 5 and 6, respectively [8].

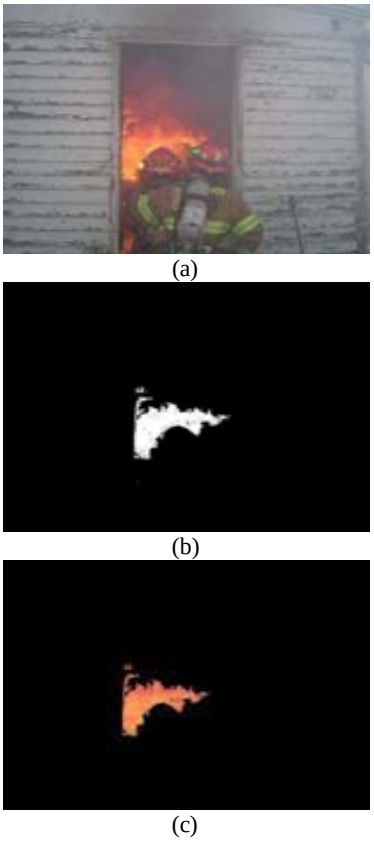
$$Y_{mean}(x,y) = \frac{1}{m*n} \sum_{x=1}^{m} \sum_{y=1}^{n} Y(x,y) \quad (4)$$

$$Cb_{mean}(x,y) = \frac{1}{m*n} \sum_{x=1}^{m} \sum_{y=1}^{n} Cb(x,y) \quad (5)$$

$$Cr_{mean}(x,y) = \frac{1}{m*n} \sum_{x=1}^{m} \sum_{y=1}^{n} Cr(x,y) \quad (6)$$

IV. Experimental Results and Discussion

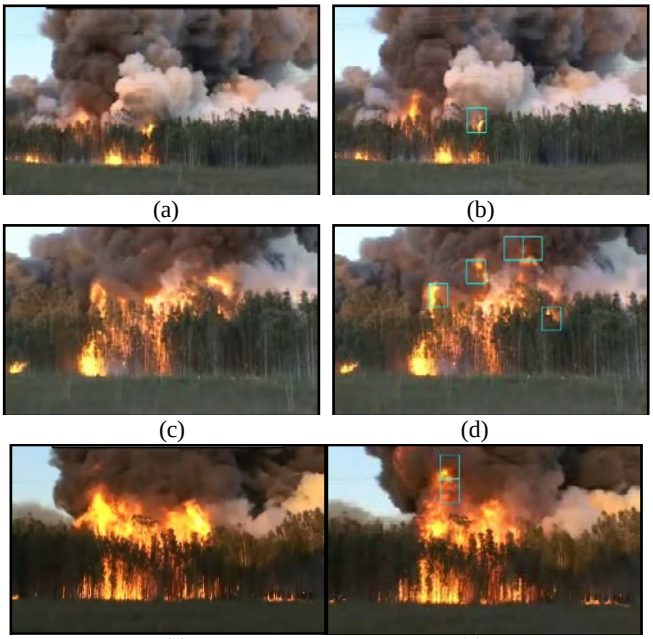
We have run our proposed fire detection software on Raspberry Pi 3 module which includes 1.2 GHz, 4 Core, 64-bit Cortex-A53 multiprocessor. We have tested it on the fire videos provided by Bilkent University [9] and Canadian Wildland Fire Information System [10] which includes some forest fire videos those are taken by a camera on a helicopter. It can successfully detect the flames for all of forest fire videos. For instance, an original fire view is given in Figure 3.a, then, our system can extract the flame from it by applying background subtraction, then, produces the view given in Figure 3.b after gray-scaling and thresholding steps. Finally, our color detection method, whose result is shown in Figure 3.c, is applied, thus, our system succeeds to detect the fire in the image.



(a) (b) (c)

Figure 3. Fire detection steps of a sample fire view

The detected flame parts in a growing forest fire video are taken into turquois rectangle in our system as shown in Figure 4. Figure 4.a, Figure 4.c, and Figure 4.e are original views of the video and the detected fire parts are shown in Figure 4.b, Figure 4.d, and Figure 4.f.



(a) (b) (c) (d) (e) (g)

Figure 4: Sample videos of a growing fire

In Figure 5, there are tail-lamps of a car, which are on and similar to a fire such as a burning car, but the system checks the motion in the corresponding frames and decides that there was no fire even if they have similar color range with a fire.

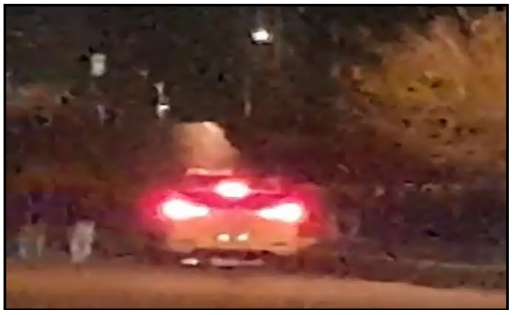


Figure 5: A camera view similar to a fire

The limitation of our proposed system is failure to detect the small-size fire in a video. For instance, in Figure 6, there is a small fire that our system cannot detect. If we could decrease the frame size such as 10x10 in the process, it would be possible to detect such a small fire. Yet, in this circumstance, the microprocessor in our Raspberry Pi module lacks real-time service. However, integrating our module to a drone will eliminate the limitation of detecting small-size fire. Because the drone can converge to the field and get closer views.

Figure 6: A view including a small-scale fire

V. CONCLUSIONS AND FUTURE WORKS

In this study, we have developed a forest fire detection system that runs on a Raspberry Pi 3 module which has an integrated infrared camera. Our proposed system is mobile, lightweight and low-cost and is capable of detecting a fire without requiring to any external equipment and manual intervention. It also does not require to transfer the video views to a remote server and can process the views itself to detect a fire. Thus, it can be easily mounted to drones and these drones fly over the fields where poses high fire risk such as picnic sites, national parks and reserves etc. Besides, our system can be used to tackle false fire reports in such a way that when a fire is reported, a drone with our system can be sent to the fire venue to check whether a real fire exists until firefighters are ready to intervene the fire.

Our system can detect forest fires easily and smoothly by taking both color and motion characteristics of a fire into account. As a result, although our system has a limitation to detect small-scale fire, it is quite sufficient to detect forest fires. Perhaps, it cannot detect a fire at the beginning, but it will grab it if it grows by posing a danger.

As a future work, smoke detection may be added to our system as another fire detection parameter. Thus, our system will be more robust and sensitive for early detection of a fire.

REFERENCES

1. Nguyen-Ti, T., T. Nguyen-Phuc, and T. Do-Hong. *Fire detection based on video processing method*. in *Advanced technologies for communications (atc), 2013 international conference on*. 2013. IEEE.
2. Seebamrungsat, J., S. Praising, and P. Riyamongkol. *Fire detection in the buildings using image processing*. in *Student Project Conference (ICT-ISPC), 2014 Third ICT International*. 2014. IEEE.
3. Wang, W. and H. Zhou. *Fire detection based on flame color and area*. in *Computer Science and Automation Engineering (CSAE), 2012 IEEE International Conference on*. 2012. IEEE.
4. Qiu, T., Y. Yan, and G. Lu, *An autoadaptive edge-detection algorithm for flame and fire image processing*. IEEE Transactions on instrumentation and measurement, 2012. **61**(5): p. 1486-1493.
5. Santana, P., P. Gomes, and J. Barata. *A vision-based system for early fire detection*. in *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*. 2012. IEEE.
6. Lei, W. and J. Liu. *Early fire detection in coalmine based on video processing*. in *Proceedings of the 2012 International Conference on Communication, Electronics and Automation Engineering*. 2013. Springer.
7. Celik, T., *Fast and efficient method for fire detection using image processing*. ETRI journal, 2010. **32**(6): p. 881-890.
8. binti Zaidi, N.I., et al., *Fire Recognicion using RGB and YCbCr Color Space*. ARPN Journal of Engineering and Applied Sciences, 2015. **10**.
9. Cetin, A.E. *Sample Fire and Smoke Video Clips*. [cited 2018 March]; Available from: http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html.
10. *Canadian Wildland Fire Information System*. [cited 2018 March]; Available from: http://cwfis.cfs.nrcan.gc.ca/home.

Detecting Anomalies in Surveillance Videos with Spatio-Temporal Features

K. ÖZ[1] and İ.R. KARAŞ[1]

[1] Karabuk University, Karabuk/Turkey, kadriyeoz@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, irkaras@gmail.com

Abstract **- One of the purposes of video surveillance systems is to detect anomalies which are unexpected situations at a certain location or at a frame. Anomalies can be related to motion or appearance according to its spatial position. In this paper, we propose an anomaly detection system based on spatio-temporal features. Features from Accelerated Segment Test (FAST) is used for detection of corners location. Optical Flow magnitude and orientation of these points is used as spatio-temporal features. A grid is**

to the frames to neutralize the effect of proximity to the camera. Normal patterns are clustered with an unsupervised neural network so called Self-Organizing Maps (SOM). In test videos if extracted features cannot model with normal clusters, associated grid cell will be marked as anomaly

Keywords **- Video Surveillance, Anomaly Detection, Features from Accelerated Segment Test (FAST), Optical Flow, Self-Organizing Maps (SOM)**

I. INTRODUCTION

THE widespread use of cameras and security requirements have increased the interest in surveillance videos which have large volumes of data to analyze. Therefore, traditional human-focused surveillance systems have been transformed into the automatic systems which can detect anomalies unassisted. Anomalies can be described as unexpected patterns, situations, things which are not trivial to model. Thus, in the literature normal patterns are modelled via training data [1] or statistical approaches are used [2] to model abnormal patterns.

First step of modelling normal patterns is extracting features from videos. There are two main approaches in the literature [3] : I) extracting features from pixels of image [4], [5] . II) target based methods [6], [7] . Also there exists some studies which use an hybrid structure of both approaches [8]. After extracting features, there are several methods to model them like Hidden Markov Model (HMM) [9], Markov Random Field (MRF)[10], Latent Dirichlet Allocation (LDA) model [11] , Mixture Of Dynamic Textures (MDT) [12].

In this study, we propose a low-level feature extraction by using image pixel values to avoid occlusions and tracking. Moreover, we combine motion and appearance features by using optical flow histograms according to corner points which detected by FAST method. Normal patterns are clustered with SOM. In the next section we will further discuss our proposed method.

II. PROPOSED METHOD

The overview of our method is illustrated in Figure 1.

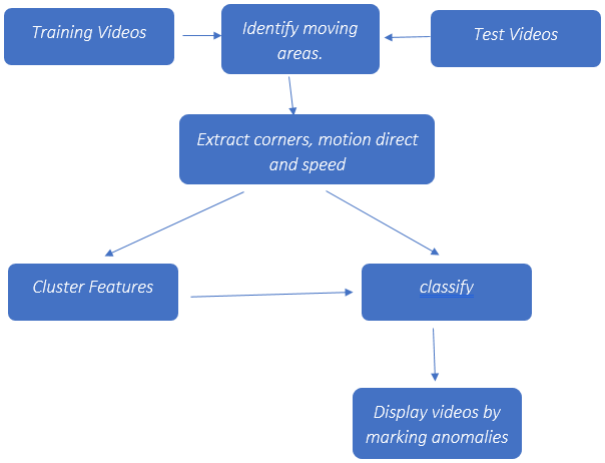


Figure 1. The overview of proposed method.

There are four main phases in our method; identify moving areas, extract features, cluster features and classify. We use non-overlapping grids of video frames as seen in Figure 2.

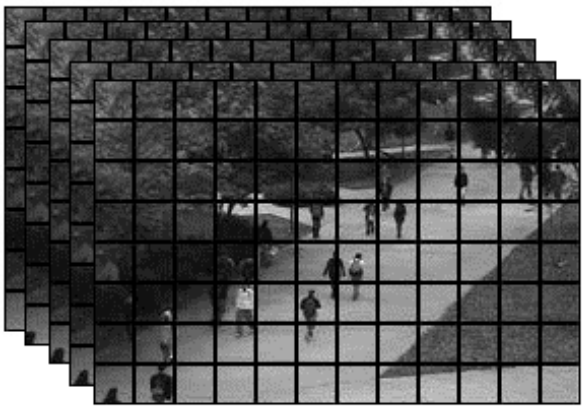


Figure 2. Grid view of frames.

The first step is identifying moving areas, for reducing running time. Motionless areas are eliminating with a basic frame difference method. Pixels that are below the threshold

value, are considered to be 0 and the others are 1, shown in Figure 3. Median filter was used for avoiding noise as shown in Figure 4.

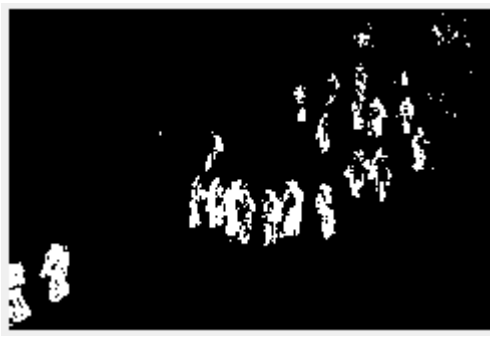


Figure 3: Difference of two frames

Figure 4: Difference of two frames with median filter

The second step is extracting features. To detect corner points Features from Accelerated Segment Test (FAST) is used [13][14]. Then motion direct and speed are calculated with optical flow method [15] at corner points. In a grid cell with detected corner point, optical flow histograms are computed according to orientation of motion vector. Features are created with a 8-bin histogram for each grid cell.

The third step is clustering. Features are clustered using an unsupervised artificial neural network, Self-Organizing Maps (SOM) [16]. SOM is a well-known artificial neural network, which produces low dimensional projection of high dimensional data [17]. The forth step is classifying. Test videos are classified according to clusters. Grid cells are marked as anomalies, if they do not match any of the clusters.

III. Experimental Results

The proposed system is tested in UCSD Peds1 dataset [18]. The features, obtained from the training videos are modeled with SOM to create clusters that model normal conditions for each cell. In the test videos, features are also calculated for each cell to check whether they are included in the relevant clusters. Some samples of detected anomalies as chart and biker in pedestrian walkways and a person in the grass, are shown in Figure 5.

Figure 5: Sample of marked anomalies at video frames

IV. Conclusion

In this paper, an anomaly detection system is proposed, and test results are shown visually. To combine motion and appearance features, optical flow histogram is used at corner points which detected by FAST method. Normal patterns are clustered with SOM. Detected anomalies are shown to user as marked. As future work, feature extracting method will be improved and system will be tested on other datasets.

ACKNOWLEDGMENT

This work was supported by Research Fund of the Karabuk University. Project Number: KBÜBAP-17-DR-457

REFERENCES

[1] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Deep-Cascade: Cascading 3D Deep Neural Networks for Fast Anomaly Detection and Localization in Crowded Scenes," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1992–2004, Apr. 2017.

[2] V. Saligrama and Z. Chen, "Video anomaly detection based on local statistical aggregates," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2112–2119, 2012.

[3] A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A Review of Anomaly Detection in Automated Surveillance," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 42, no. 6, pp. 1257–1272, Nov. 2012.

[4] A. M. R. R. Bandara, L. Ranathunga, and N. A. Abdullah, "A feature clustering approach based on Histogram of Oriented Optical Flow and superpixels," *2015 IEEE 10th Int. Conf. Ind. Inf. Syst. ICIIS 2015 - Conf. Proc.*, no. 12, pp. 480–484, 2016.

[5] R. V. H. M. Colque, C. Caetano, M. T. L. de Andrade, and W. R. Schwartz, "Histograms of Optical Flow Orientation and Magnitude and Entropy to Detect Anomalous Events in Videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 673–682, Mar. 2017.

[6] M. J. V Leach, E. P. Sparks, and N. M. Robertson, "Contextual anomaly detection in crowded surveillance scenes," *Pattern Recognit. Lett.*, vol. 44, pp. 71–79, 2014.

[7] C. Li, Z. Han, Q. Ye, and J. Jiao, "Visual abnormal behavior detection based on trajectory sparse reconstruction analysis," *Neurocomputing*, vol. 119, pp. 94–100, 2013.

[8] A. Surana, A. Nakhmani, and A. Tannenbaum, "Anomaly Detection in Videos : A Dynamical Systems Approach JJRfi J Pinside (X I," pp. 6489–6495, 2013.

[9] J. Snoek, J. Hoey, L. Stewart, R. S. Zemel, and A. Mihailidis, "Automated detection of unusual events on stairs," *Image Vis. Comput.*, vol. 27, no. 1–2, pp. 153–166, 2009.

[10] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates," *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work. 2009*, no. June, pp. 2921–2928, 2009.

[11] S. Kwak and H. Byun, "Detection of dominant flow and abnormal events in surveillance video," *Opt. Eng.*, vol. 50, no. 2, p. 027202, 2011.

[12] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1975–1981, 2010.

[13] E. Rosten, R. Porter, and T. Drummond, "Faster and better: A machine learning approach to corner detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 105–119, 2010.

[14] H. Fradi and J. L. Dugelay, "Spatial and temporal variations of feature tracks for crowd behavior analysis," *J. Multimodal User Interfaces*, vol. 10, no. 4, pp. 307–317, 2016.

[15] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1–3, pp. 185–203, 1981.

[16] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biol. Cybern.*, vol. 43, no. 1, pp. 59–69, 1982.

[17] T. Kohonen, *MATLAB Implementations and Applications of the Self-Organizing Map*. Unigrafia Oy, Helsinki, Finland, 2014.

[18] "UCSD Anomaly Detection Dataset." [Online]. Available: http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm.

Automatic Segmentation and Labelling of 3D Human Activities

R.DURGUT[1], C.OZCAN[2] and O. FINDIK[3]

[1] Karabuk University, Karabuk/Turkey, rafetdurgut@karabuk.edu.tr
[2]Purdue University, West Lafayette, IN/USA, ozcan@purdue.edu
[3]Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

***Abstract* - Recognition and interpretation of human activities are very interesting and hot topics that are frequently studied in the field of computer vision. Especially with the advent and development of the Microsoft Kinect depth sensors, the expansion of the study field has gained momentum in the positive direction. Thanks to RGBD cameras, which also provide depth information in addition to the RGB image, researchers benefit from many advantages in terms of privacy, accuracy and precision. In this study, automatic segmentation of repeated 3D human activity is proposed. A public dataset containing the repeated action sequences are recorded using the RGBD camera. The action sequence in this dataset includes similar and different action information. In order to identify and label each action in sequence, it is necessary to perform the segmentation process. To be able to perform a successful segmentation process, the data must be pre-processed to remove noise. For this purpose, a total variation based noise removal method is used. Human action recognition and detailed error analysis can be performed through the segments derived from the output of this work.**

***Keywords* - Automatic Segmentation, Automatic Labelling, Microsoft Kinect, Skeleton Data, Total Variation, Sliding Windows**

I. INTRODUCTION

HUMAN action recognition is a classification problem that is studied in computer vision and pattern recognition fields in order to automatically determine the performance exhibited by human body [1]. With the identification of the exhibited behaviors, many other useful applications such as video surveillance, human computer interaction, digital assistant, sports video analysis, games and health can be developed [2-6]. In the field of computer vision, human body joint information can be obtained with RGB and RGB-D images [7]. Compared to RGB cameras, RGB-D Cameras provide depth information alongside RGB information, providing greater accuracy in detecting human body joints [8]. Thanks to high-accurate human body joint information, the success of classification process is also increasing.

Studies in computer vision and pattern recognition have gained momentum along with the advent of the Microsoft Kinect which is low cost RGB-D camera [9]. Thanks to the advantages of Microsoft Kinect, action recognition is very popular in the field, but it brings with it some problems. When joint positions are determined, sensor distortions and noise-induced distortions can occur [10]. These distortions can lead to serious errors in the process of motion recognition and interpretation. Various methods are used in the pre-processing step to prevent these distortions for real datasets [11].

There are human action sequences that contain information on different motion sets that are repeated within the problem set studied in the motion recognition field [12]. Particularly, it is possible to carry out the interpretation process by making the temporal segmentation of the action which has the movement belonging to more than one class. With the temporal segmentation, a computational model of action is created by revealing the basic components of action [13]. These basic components uncover the primitives of action in the unsupervised learning step. With the labeling of these primitives, information about the action can be given to people without the need for a learning set.

Human movement and motion primitives are repeated many times in human action applications such as computer games and exercise applications [14]. A number of methods have been proposed to identify temporal segmentation by identifying recurrent parts [12-14]. The Zero Velocity Crossing (ZVC) method uses the angular velocities of the joints to reveal motion segments that form a complex motion set [15]. Due to the high precision it is extremely sensitive to noise. A multi-dimensional segmentation method has also been proposed to decompose the motion into a simple linear dynamic model sequence [18]. In addition, there are different segmentation methods based on principal component analysis (PCA) to distinguish one movement region from the other [17].The segment information obtained from these studies can perform motion recognition using various classification methods (HMM, Cascaded correlation-based classifier) [16, 17]. In order to use these methods, a significant amount of training data is required.

In this study, unsupervised labeling of motion primitives was performed on a real data set containing different actions. The noise in sensor data was removed using a total variation based denoising method. Subsequently, the segments forming the action are identified and tagged according to the joint groups. The data set obtained with Microsoft Kinect is used as the

source data. In addition, the Unity3D game engine is used for visualization.

II. METHODS

This section explains that methods which are used to segmentation and labelling for human action sequences. Total variation based Sparsity Driven Despeckling is used to denoising. The other method, The Sliding Window Method is used to segmentation for multi-dimensional motion sequences.

A. Total Variation Based Noise Filtering

One of the important steps in the success of this study is the need for noise filtering to improve the quality of the data. Noise reduction is used as a pre-processing step to smooth homogeneous regions while preserving important features of the signal. In this study, an adaptation of the Sparsity Driven Despeckling (SDD) algorithm [19], which was based on total variation and previously defined for use on images, is used that operates on one-dimensional signals. Because SDD offers an approach that allows it to be used with one parameter and provides better noise reduction with shorter execution time compared to other methods. The noise reduction on our signal data is defined as the optimization problem as follows:

$$\hat{F} = arg\ \min_{F} J(F) \tag{1}$$

where the cost function $J(F)$ is defined as in below and our goal is to optimize by F.

$$\begin{aligned} &\mathrm{J(F)} \\ &= \sum_{\mathrm{p}} \left(\mathrm{F_p} - \mathrm{G_p}\right)^2 \\ &+ \lambda \Lambda(\left|(\partial \mathrm{F})_{\mathrm{p}}\right|, \mathrm{f}) \end{aligned} \tag{2}$$

where G is the noisy signal, F is the denoised signal, p is the index number of data in the signal, λ is the smoothing level, f is the norm value applied with TV regularization, and $\Lambda(x, f)$ is exponentiation operator for obtaining l_0-norm when $f = 0$ defined as

$$\begin{aligned} &\Lambda(x, f) \\ &= \begin{cases} 0 & x = 0, f = 0 \\ x^f & otherwise \end{cases} \end{aligned} \tag{3}$$

The first term in the cost function refers to the fidelity term which ensures F to be similar to G. The second term is a TV regularization term, which provides penalties on the changes in the signal gradients. In this formula, l_0-norm is obtained when $f = 0$, l_1-norm is obtained when $f = 1$, and fractional norm is obtained when $0 < f < 1$, where they all induce a sparse solution.

B. The Sliding Window Method

The sliding windows approach is one of the most commonly used methods for time series segmentation in the fields of machine learning and pattern recognition [20]. This algorithm is developed for separating the best possible sections starting from the first element to the last element of the time series. By selecting the first element of the partition, it tries to create the longest possible subsequence by adding the incoming partition to the right side in the time series. When the error value at some points reaches a value higher than the user defined error, the partition is terminated and the new partition is created. The element exceeding the error value is determined as the first element of the new segment. By setting the position of the marker, the process repeats all the time series until a partial linear approach is reached. The pseudo code of the algorithm is given in Fig. 1.

Algorithm 1 Sliding Windows Algorithm

```
procedure SEGMENTS=SLIDING_WINDOWS(datas, max_error)
    position = 1
    while Not Finished Segmenting do
        i = 2
        while Calculate_Error(datas[position : position + i]) < max_error do
            i = i + 1;
        end while
        Segments = Add(Segments, datas[position : position + i − 1])
        position = position + 1
    end while
    return Segments
end procedure
```

Figure 1: The pseudo code of sliding windows method.

III. EXPERIMENTAL STUDY

In this study, basic primitive extraction and labeling were performed for real dataset. Dataset includes long motion sequences and different classes of motion. It has been recorded using the Microsoft Kinect depth sensor. The block diagram of the system is shown in Fig. 2.

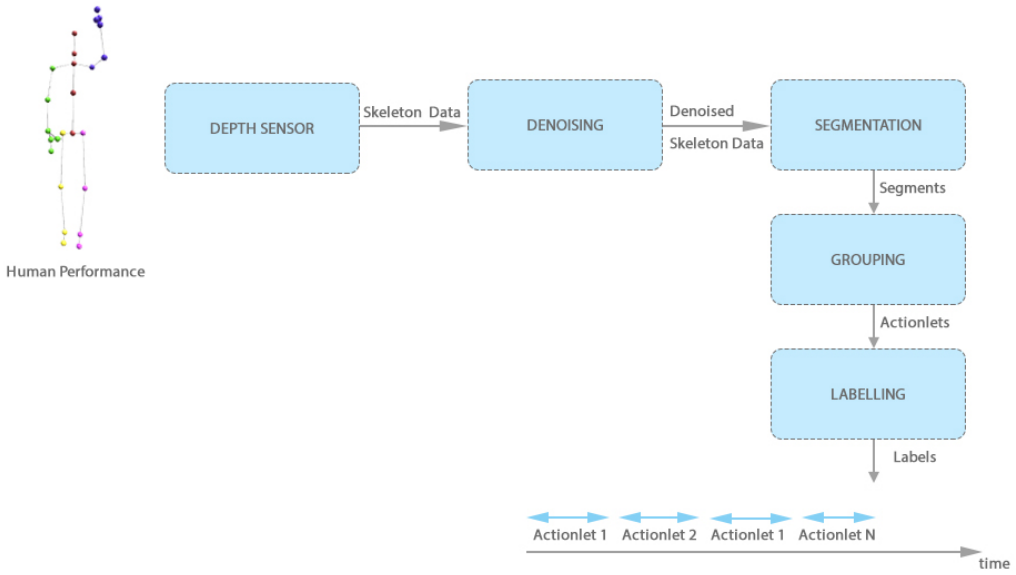


Figure 2. The block diagram of proposed approach.

User performance is recorded in skeleton data format using depth sensor. Denoising is performed in order to minimize the error value of the human skeleton data information caused by the sensor or other sources. It is divided into sub-segments by the sliding windows algorithm on the denoised skeleton information. Parts of each body joint are examined on a group basis to reveal segments of the body part performing the action. As a final step, the operation of the system is completed with the labelling of these segments as unsupervised.

A. Dataset

In this study, we used real dataset which included some errors recorded by Microsoft Kinect. Each action sequence in the dataset contains 5 sub action sequences of 8 aggressive movements (punch kick, etc.). Each movement was performed by 12 different people. Example postures for sample motion sequence are given in Fig. 3.

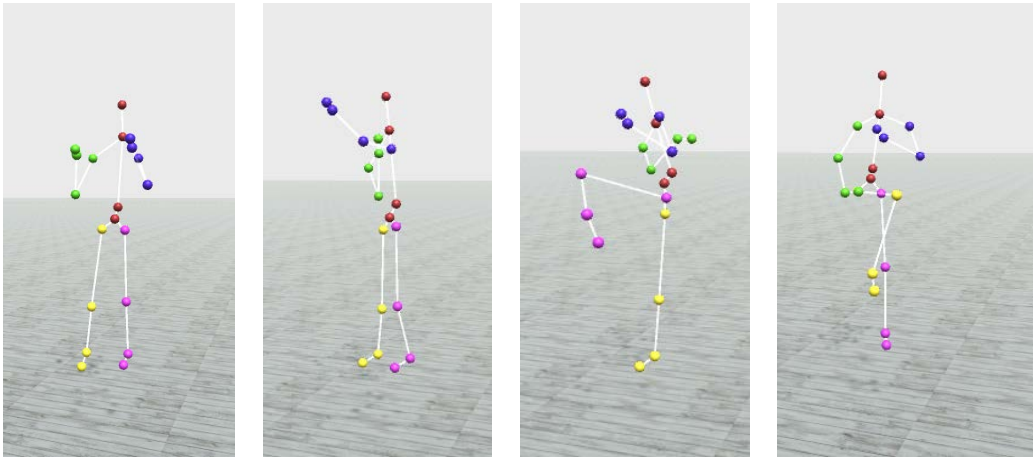


Figure 3. Example of motion sequence for dataset

B. Segmentation

In dataset, joint location information of each joint is stored according to time. The 3D position information of the joint i in the frame t is stored as $P_i^t =< x_i^t, y_i^t, z_i^t >$. The displacement information of the joint is calculated as in eq. 4. This feature information is used for segmentation, including the variation of each joint on the axes depending on the time.

$$D_i^t = P_i^{t-1} - P_i^t \tag{4}$$

When sliding windows method is applied on each axis changes for each joint, the segments in Fig. 4 will be obtained.

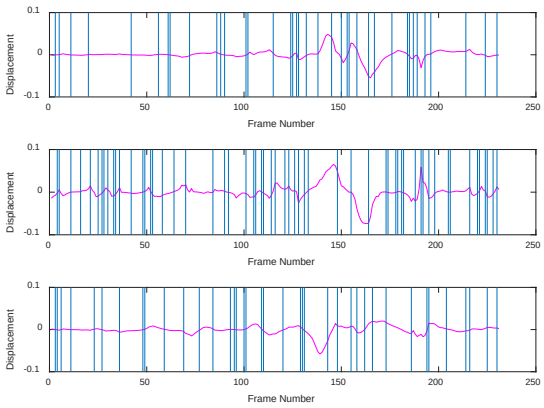


Figure 4. An example segments for joint 8 (Right Arm).

$S_i^j = \{F_k, F_j\}$ contains segment information includes start and end frames for joint *i*. For sensitivity, error value was selected relatively small (0.01). Because the segment change direction specifies the movement direction of the joint, the segments belong the same action direction are merged. If the amount of total change of the segment is greater than the determined threshold value (0.2), it is important section for given sequence. Otherwise, it is labelled as random or insignificant changes. More meaningful segment layout is shown in Fig. 5 when the segments are arranged.

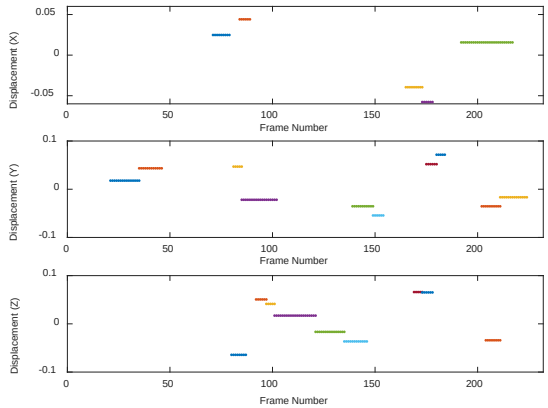


Figure 5. An organized segments for joint 8 (Right Arm).

C. Grouping

There are 20 human joint location recorded by the Microsoft depth sensor. Rather than treating these joint locations individually, grouping can be more efficient. The actions taken by the joints at this point can be more meaningful and there is a possibility to use a simpler language for error indication labeling. The body is divided into 5 basic parts (torso, right arm, left arm, right leg and left leg) when grouping according to joint location. Each part of body has four joints information. When the segment information of the joints in the group is examined observed tendency to act together behavior is observed. When segment information is calculated with the weight of each joint value having same tendency, the view in Fig. 6 is obtained.

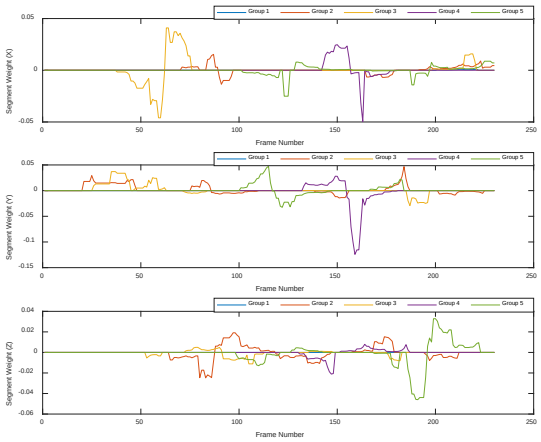


Figure 6. Group segment sequence for example data

D. Labelling

Labeling process can be done using segments which constitutes motion sequence. The visual representation of the changes occurring during the movement of the Right Arm is given in Fig. 7 (label 7, 8, 9 denote the X, Y, Z plane, respectively.). As can be seen in the Fig. 7, the person firstly lifts the right arm, then brings it to the left and forward, then brings it back to the right and back. Therefore, the basic components of the movement performed by the person are revealed on a group basis. Fig. 8 shows the change in movement over all groups. The first two segments clearly indicate that the person raised their hands to up.

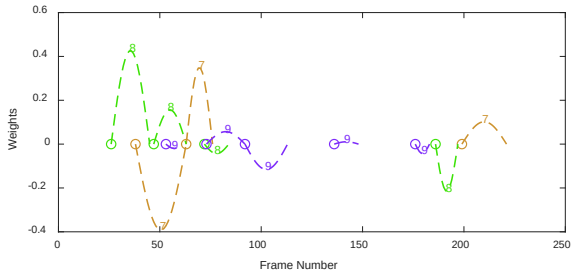


Figure 7. An example of actionlet sequence for right arm

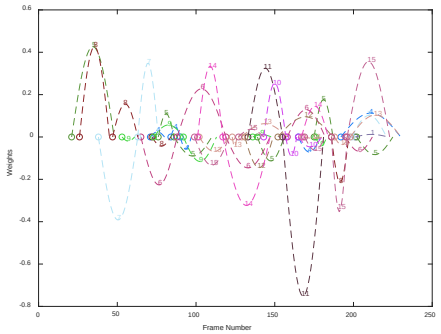


Figure 8. An example of actionlet sequence for full body

Visual presentation with labelling when implemented to Unity3D game engine is shown in Fig. 9. The labels are shown on the screen when the motion is being performed. If there is no movement or if it is meaningless, The Message "No Movement"

is shown. The joints of the same body region are represented by the same color.

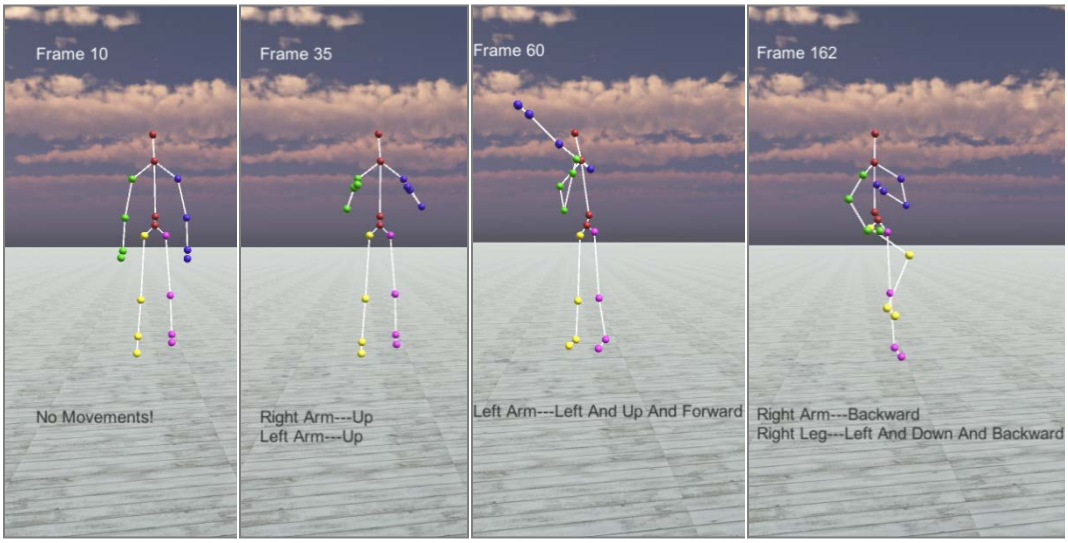


Figure 9 . Visualization of sequences with labels

IV. CONCLUSION

In this study, segmentation and labeling were performed on the real dataset which is recorded by the Microsoft Kinect depth sensor. Noise within motion information is reduced using the adaptation of SDD method. Then, the windows sliding approach is implemented for segmentation on motion sequences. Motion primitives have been obtained using segment information. Then, Motion primitives have been grouped according to relevant body parts. Labelling is utilized to this groups of motion primitives. These features represent action sequence.

REFERENCES

[1] Aggarwal, J. K., & Ryoo, M. S. (2007). Human Activity Analysis : A Review.

[2] Castro, J., Delgado, M.,Medina, J., & Ruiz-Lozano, M. (2011). Intelligent surveillance system with integration of heterogeneous information for intrusion detection. Expert Systems with Applications, 38(9), 11182–11192.

[3] Kim, H., Lee, S., Kim, Y., Lee, S., Lee, D., Ju, J., & Myung, H. (2016). Weighted joint-based human behavior recognition algorithm using only depth information for low-cost intelligent video-surveillance system. Expert Systems with Applications, 45, 131–141. http://doi.org/10.1016/j.eswa.2015.09.035

[4] Wu, M., Chen, T., Chen, K., & Fu, L. (2016). Daily Activity Recognition Using the Informative Features from Skeletal and Depth Data, 1628–1633.

[5] Sharaf, A., Torki, M., Hussein, M. E., & El-Saban, M. (2015). Real-time multi-scale action detection from 3D skeleton data. Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015, 998–1005. http://doi.org/10.1109/WACV.2015.138

[6] Pazhoumand-Dar, H., Lam, C. P., & Masek, M. (2015). Joint movement similarities for robust 3D action recognition using skeletal data. Journal of Visual Communication and Image Representation, 30, 10–21. http://doi.org/10.1016/j.jvcir.2015.03.002

[7] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., … Moore, R. (2013). Real-time Human Pose Recognition in Parts from Single Depth Images. Commun. ACM, 56(1), 116–124. http://doi.org/10.1145/2398356.2398381

[8] Cippitelli, E., Gasparrini, S., Gambi, E., & Spinsante, S. (2016). A Human Activity Recognition System Using Skeleton Data from RGBD Sensors. Computational Intelligence and Neuroscience, 2016. http://doi.org/10.1155/2016/4351435

[9] Zhang, Z. (2012). Microsoft kinect sensor and its effect. IEEE Multimedia, 19(2), 4–10. http://doi.org/10.1109/MMUL.2012.24

[10] Shan, J., & Akella, S. (2014). 3D Human Action Segmentation and Recognition using Pose Kinetic Energy. In Advanced Robotics and its Social Impacts (ARSO).

[11] Gu, Y., Do, H., Ou, Y., & Sheng, W. (2012). Human gesture recognition through a Kinect sensor. 2012 IEEE International Conference on Robotics and Biomimetics, ROBIO 2012 - Conference Digest, 1379–1384. http://doi.org/10.1109/ROBIO.2012.6491161

[12] Wang, Q., Kurillo, G., Ofli, F., & Bajcsy, R. (2015). Unsupervised Temporal Segmentation of Repetitive Human Actions Based on Kinematic Modeling and Frequency Analysis. Proceedings - 2015 International Conference on 3D Vision, 3DV 2015, (1111965), 562–570. http://doi.org/10.1109/3DV.2015.69

[13] Heinland, D., Ronfard, R., & Boyer, E. (2011). A survey of vision-based methods for action representation, segmentation and recognition. Computer Vision and Image Understanding, 115(2), 224–241. http://doi.org/10.1016/j.cviu.2010.10.002

[14] Bandera, J. P., Marfil, R., Romero-Garcés, A., & Voilmy, D. (2018). A new paradigm for autonomous human motion description and evaluation: Application to the Get Up & Go test use case. Pattern Recognition Letters. http://doi.org/10.1016/j.patrec.2018.02.003

[15] Fod, A., Matarić, M. J., & Jenkins, O. C. (2002). Automated derivation of primitives for movement classification. Autonomous Robots, 12(1), 39–54. http://doi.org/10.1023/A:1013254724861

[16] M. Raptis, D. Kirovski, and H. Hoppe. Real-time classifi- cation of dance gestures from skeleton animation. In Processings of Eurographics/ACM SIGGRAPH Symposium on Computer Animation (SCA), pages 147–156, 2011.

[17] J. Sung, C. Ponce, B. Selman, and A. Saxena. Unstructured human activity detection from rgbd images. In Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pages 842–849, 2012.

[18] C. Lu and N. J. Ferrier. Repetitive motion analysis: segmentation and event classification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(2):258–263, 2004.

[19] Ozcan, C., Sen, B., Nar, F., "Sparsity-driven despeckling for SAR images", IEEE Geosci. Remote Sens. Lett., 3(1): 115-119, Jan. 2016.

Keogh, E., Chu, S., Hart, D., & Pazzani, M. (2001). An online algorithm for segmenting time series. In Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on (pp. 289-296). IEEE.

Understanding effects of hyper-parameters on learning: A comparative analysis

Y.Y. BAYDİLLİ[1] and Ü. ATİLA[1]

[1] Department of Computer Engineering, Karabuk University, Karabuk/Turkey, yusufbaydilli@karabuk.edu.tr
[1] Department of Computer Engineering, Karabuk University, Karabuk/Turkey, umitatila@karabuk.edu.tr

***Abstract* - In this study, we analyzed the hyper-parameters which are frequently used in deep learning methods on a generated DNN. On the Fashion-MNIST dataset, we had chance to interpret the evolution of the model to the end as a result of tests performed on a low epoch number. At the end of the study, we reached a success rate of about 90 percent on the test data and showed that the selected hyper-parameters by created model were the most accurate.**

***Keywords* – classification, deep learning, hyper-parameter optimization.**

I. INTRODUCTION

DEEP architects can be created easily and complex problems can be solved with high accuracy, due to the increased processing speed with today's technology. Although deep architects have provided great benefits, two new problems have emerged that must deal by the researchers; time spent to train the model, and the hyper-parameter choice. In fact, since the selection of hyper-parameters is an element directly affecting the training time, this can be said that the most important problem is the hyper-parameter selection itself.

Although several definitions have been made for the concept of DL (Deep Learning), non-linear structures can be generally solved using three or more hidden layers [1]. During the construction of these structures, two types of hyper-parameters are selected and model parameters are trained. Problems such as under-fitting and over-fitting that may be encountered during training are directly related to these hyper-parameters [2]. This makes the right hyper-parameter selection a must.

Since the hyper-parameter selection is an optimization problem in itself, it is possible to make improvement with various optimization techniques. However, the high number of hyper-parameters and working with too many combinations do not provide a definite solution for the time consumed and the correct model. Apart from the optimization techniques developed, there is a trial/error-based method called *babysitting*, also [3]. This technique behaves like a *brute-force* and is quite beneficial in that it in the sense of observing how the constructed model evolves, even though it does not produce positive results for the time spent.

In this study, we tried to construct a successful *DNN* (Deep Neural Network) model from scratch by using a problem that can be solved 90-95% success rate by *CNN* (Convolutional Neural Network) in order to examine the positive or negative effects of hyper-parameter selection. While trying to reach the best result, on the other hand, we were able to analyze the effects of hyper-parameters to learning process.

II. DATA

Using a term like *"Hello World of Deep Learning"* for MNIST dataset could not be wrong. However, this dataset, created by LeCun et al. in 1998, is now become easily resolvable with even simple CNN and MLP structures. Therefore, researchers have begun to perform benchmark operations with datasets such as Fashion MNIST [4].

This dataset consists of *28 x 28* grayscale images just like MNIST. There are *70,000* images (*60,000* for train, *10,000* for validation) that are labeled in *10* different categories. So, we aimed to analyze the effects of hyper-parameters on deep learning architects by trying to achieve the highest score with a DNN on this dataset, which can be solved with CNN-based architects at high success rates.

III. ARCHITECTURE AND TEST

Hyper-parameters can be defined as factors provide model optimization. These factors can be examined in two groups as model-based and train-based [5].

A. Model-Based Hyper-parameters

1) Number of hidden layers and units

The *DEEP* term brings first to mind multi-layered and multi-noded complex architectures. *Deep Learning* is also thought to have taken its power from this complexity. This is both right and wrong hermeneutics. Increasing the number of parameters that can be trained, may increase the learning as well as it decreases. Since numbers of hidden layers and number of hidden nodes are the quantities that affect the number of parameters to be trained (weight, bias), the most optimal model, i.e. *"the most complex model with least parameters"*, must be established both in terms of time and learning.

The *"rule of thumb"* methods are used in many topics as well as in selecting hidden layers and hidden units. Some of these rules [6]:

- # of hidden neurons should be between size and input.
- # of hidden neurons should be *2/3* size of *input + output* layer.

Also, we took consider *"growing/pruning"* method [7], and finally, we tried to build our model with an algorithm that offers a *pyramid-like* architecture, based on the increasing and

decreasing of the number of neurons, as learning rate increased. The dynamics of the algorithm we used are shown in Algorithm 1:

Algorithm 1 creating layers for MLP

```
procedure GROWING_PRUNING (n)
    n = 1
    add layer (2^n)
    for 2^n < # of input units do
        # of hidden units = 2^n
        if test accuracy increased then
            n = n + 1
        else add layer (2^n)
        end if
    end for
    for 2^n > # of output units do
        # of hidden units = 2^n
        if test accuracy increased then
            n = n − 1
        else add layer (2^n)
            continue
        end if
    end for
end procedure
```

In the algorithm, we increased the number of neurons in the first hidden layer as learning continues and until it reaches size of input layer, then, we tried to improve learning rate by adding new layers and decreasing number of neurons. Process continued until learning rate reaches to size of output layer. So, there are two main stop criteria; size of *input/output* layer and *learning rate*. While performing our tests, we used the Keras framework with TensorFlow backend. Each epoch took as long as about *6* secs with GeForce GTX 1050 TI GPU. For $epoch = 5$, we had a chance to analyze each hyper-parameter for in a short time as *30* secs.

All this *trial/error* process performed for hidden units and layers can be seen in Table 1.

Table 1: Searching for best model.

		Layer 1		Layer 2		Layer 3		Layer4	
n	2^n	acc	loss	acc	loss	acc	loss	acc	loss
1	2	0.663	0.852	-	-	-	-	-	-
2	4	0.761	0.567	-	-	-	-	-	-
3	8	0.787	0.494	-	-	-	-	-	-
4	16	0.794	0.476	-	-	0.788	0.479	0.792	0.486
5	32	0.797	0.470	-	-	**0.801**	**0.447**	-	-
6	64	0.797	0.468	0.795	0.462	0.796	0.451	-	-
7	128	0.797	0.467	**0.801**	**0.453**	0.797	0.453	-	-
8	256	0.798	0.467	0.800	0.457	-	-	-	-
9	512	**0.798**	**0.466**	0.798	0.460	-	-	-	-

acc : test accuracy
loss : test loss

We have seen that after the tests performed, the second layer does not develop after *128* nodes and after *32* nodes for the third one. We did not add a new layer to model due to there is not observed any contribution to test accuracy of additional *16* nodes. The final version of our model seen in Figure 1.

2) *Weight initialization*

Weight initialization is a significant agent on reducing the time spent on learning [8], [9]. When we examined the accuracy and loss values obtained from the tests performed, although no significant difference occurred, we decided to continue benchmark process with He Normal [10], in which the most improvement was observed (Table 2).

We have also observed that *zero* or *one* weight initialization leads to local minimum problem.

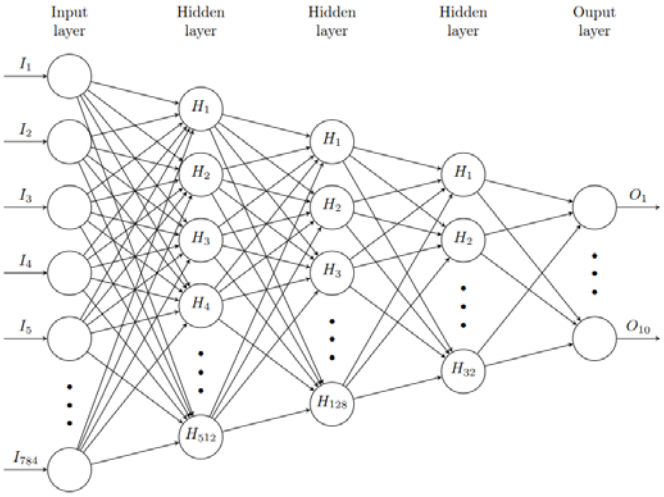


Figure 1: Constructed model.

Table 2: Weight initialization methods.

Initialization	acc	loss
zeros	0.100	2.303
ones	0.100	14.506
random normal	0.800	0.470
random uniform	0.803	0.469
truncated normal	0.791	0.465
orthogonal	0.800	0.455
Lecun normal	0.805	0.453
Lecun uniform	0.806	0.462
Glorot normal	0.796	0.482
Glorot uniform	0.809	0.466
He normal	**0.810**	**0.448**
He uniform	0.807	0.452

3) *Activation function*

The activation functions, which transfers the values from the previous layer to the next layer, and the success rates on the test values are given in the following Table 3.

Table 3: Activation functions.

Activation Function	acc	loss
elu	0.839	0.394
selu	**0.840**	**0.394**
soft plus	0.806	0.4701
soft sign	0.822	0.440
relu	0.832	0.393
tanh	0.824	0.407
sigmoid	0.605	1.024
hard sigmoid	0.506	1.256
linear	0.805	0.465

It has been observed that *sigmoid* and *hard-sigmoid* functions gave the worst accuracies due to the *"gradient vanishing"* problem [11]. On the other hand, *selu* was one step ahead of the *"linear unit"* family, which produced the highest accuracies [12].

B. Train-Based Hyper-parameters

1) Optimizers

Optimizer functions provide to update the parameters and minimize the objective function. In this respect, it is one of the most important hyper-parameters for learning. It was seen that the best values were obtained with the *adamax* optimizer as a result of the tests that we performed with the most used optimizer functions among the researchers [13].

Table 4: Optimization functions.

Optimizer	acc	loss
sgd	0.837	0.399
rmsprop	0.857	0.400
adagrad	0.864	0.344
adadelta	0.852	0.391
adam	0.858	0.379
adamax	**0.869**	**0.346**
nadam	0.850	0.382

2) Batch size

The batch size approach is a method that improves both learning speed and learning rate, which proposes of learning all the training data in small pieces instead of learning in one go [14]–[16].

Table 5: Batch sizes.

Batch size	acc	loss
None	0.867	0.367
512	0.871	0.355
256	0.870	0.344
128	**0.873**	**0.391**
64	0.863	0.379
32	0.872	0.346
16	0.873	0.382

After performed tests for various batch size values, no improvement was observed after *128*.

3) Epoch

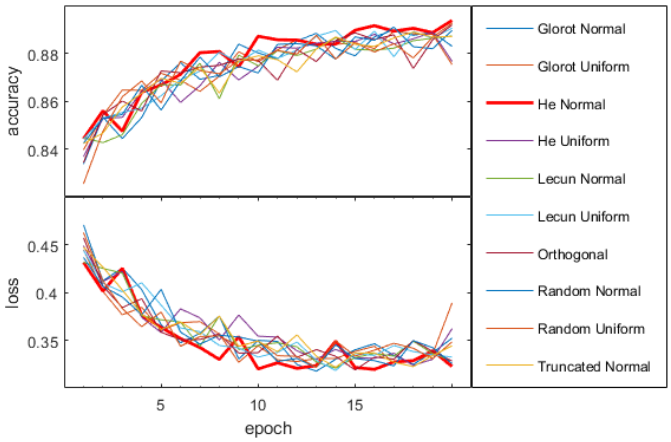
The epoch number is defined as going through forward and backward once over all parameters [17]. An over-selected epoch number may increase the accuracy of the training data, but, stops the contribution after a step on the test data. This leads to the *over-fitting* problem [18], [19]. In our model, we have not seen any improvements after *20* epochs, briefly.

Table 6: Epoch number.

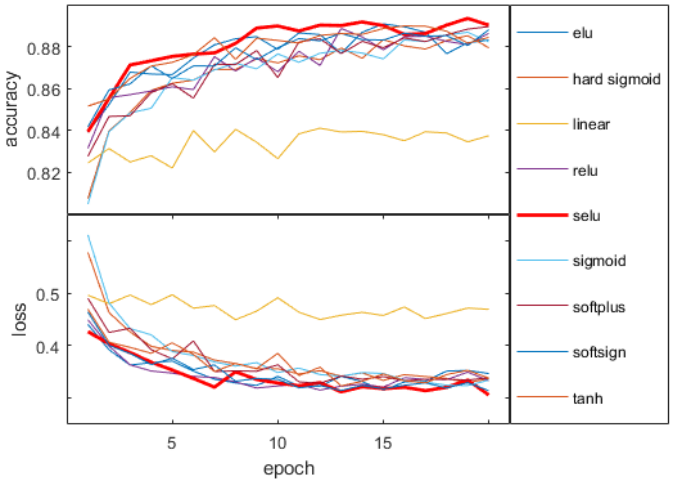
Epoch	acc	loss
5	0.873	0.391
10	0.881	0.327
20	**0.895**	**0.344**
30	0.886	0.406
50	0.890	0.494

IV. RESULTS AND DISCUSSION

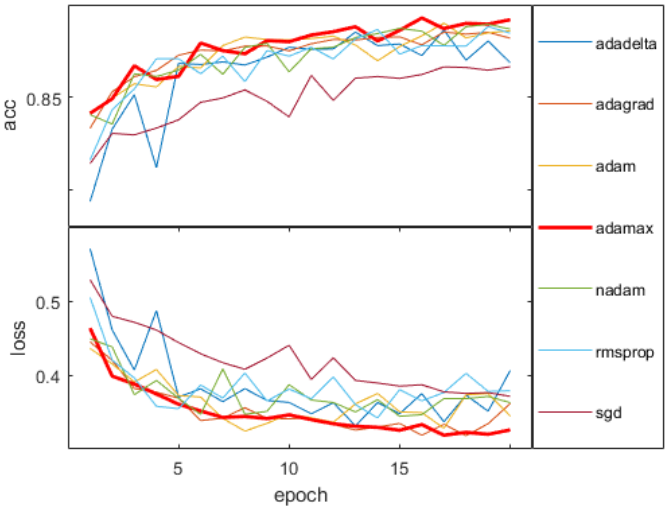
Right hyper-parameter selection is a process that directly affects the learning and prediction processes. Performing with the wrong hyper-parameters, learning cannot be realized at the desired rates, also, learning process can be unnecessarily prolonged due to problems such as *under-fitting* and *over-fitting*.



(a) weight initializers.



(b) activation functions.



(c) optimizers.

Figure 2: Accuracy and loss progress for $epoch = 20$.

One of the most important quantities of hyper-parameters is the number of hidden layer and number of hidden units. The

low number of this quantity may ensure low learning rates, while the high numbers may lead to long training times of the model.

So, in order to overcome this problem, we suggested an algorithm that offers a pyramid-like architecture and considering the *"rule of thumb"* and *"pruning/growing"*, which are the most widely used methods for deciding the number of hidden layers and number of hidden units. Thanks to this algorithm, we were able to examine in detail that how using more or less layers and/or neurons affected the model, in the tests performed for $epoch = 5$. Eventually, we were able to carry prediction values from 75% levels to 80% in a short time.

In the second phase of the study, we again proposed a kind of *"winner takes all"* model in order to select the best hyper-parameters as soon as possible. In this model, we implemented our trainings on the created model by testing both the model-based and train-based hyper-parameters, which affects the feed-forward and backpropagation processes, as the best to be fetched for $epoch = 5$. At the end of the training, we observed that the most contribution to learning rates was earned by activation function, optimizers and finally, epoch number.

At the end of the study, we were able to estimate this classification problem, which can be solved 90% - 95% accuracy with CNN architecture, with 90% accuracy with DNN construction. When we tried to insert all hyper-parameters one by one for $epoch = 20$ in the last model we created, we observed that the selected hyper-parameters for $epoch = 5$ are the most contributing hyper-parameters to learning process (Figure 2). Hence, it can be said that getting proper results in choosing the hyper-parameters are realized at the beginning of the learning. It would not be wrong interpreting this situation as *"As the twig is bent, so grows the tree"*.

In this study, we performed our tests for hyper-parameters, which are highly effective in learning, but, we think that performing tests with low epoch values using such as the method we proposed, would provide great benefits in obtaining the most accurate hyper-parameters that affects the learning positively and helps to overcome problems like over-fitting.

We also think that it is important to examine how the train and test accuracy has been evolved by selection of *sub-quantities* (L2 penalty, dropout rate, etc.) in hyper-parameters. Furthermore, hyper-parameter searching for different deep learning models (CNN, RNN, etc.) in low epochs is among the issues that need to be studied in the future to see how it would make difference in the rest of the training.

REFERENCES

[1] L. Deng and D. Yu, "Deep Learning: Methods and applications," Now Publishers, Boston, MA, US, Tech Report MSR-TR-2014-21, 2014.

[2] M. Hushchyn, "Model selection & hyperparameters tuning," presented at the Cyclotron Seminar ALICE/NA61/SPbSU, Saint Petersburg State University, Sankt-Peterburg, Russia, 2017.

[3] A. Karpathy, "CS231n Convolutional Neural Networks for Visual Recognition (1)," 2017. [Online]. Available: http://cs231n.github.io/neural-networks-1/. [Accessed: 15-Mar-2018].

[4] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," *CoRR*, Aug. 2017.

[5] T. Gao, "Hyperparameter optimization for deep learning," presented at the UNC CS Deep Learning Journal Club, University of North Carolina at Chapel Hill, NC, USA, 2016.

[6] M. Kirk, *Thoughtful Machine Learning with Python: A Test-Driven Approach*, 1 edition. Boston, MA, US: O'Reilly Media, 2017.

[7] M. Thoma, "Analysis and optimization of convolutional neural network architectures," M. Sc. Thesis, Karlshure Institute of Technology, Karlsruhe, Germany, 2017.

[8] S. K. Kumar, "On weight initialization in deep neural networks," *CoRR*, Apr. 2017.

[9] S. Koturwar and S. Merchant, "Weight initialization of deep neural networks (DNNs) using data statistics," *CoRR*, Oct. 2017.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," presented at the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1026–1034.

[11] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, USA, 2011, vol. 15, pp. 315–323.

[12] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-normalizing neural networks," presented at the NIPS 2017: The Thirty-first Annual Conference on Neural Information Processing Systems, Long Beach, CA, USA, 2017.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," presented at the 3rd International Conference on Learning Representations, San Diego, USA, 2014.

[14] S. L. Smith, P.-J. Kindermans, C. Ying, and Q. V. Le, "Don't decay the learning rate, increase the batch size," presented at the 6th International Conference on Learning Representations, Vancouver, Canada, 2017.

[15] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," presented at the 5th International Conference on Learning Representations, Toulon, France, 2016.

[16] D. Mishkin, N. Sergievskiy, and J. Matas, "Systematic evaluation of CNN advances on the ImageNet," *Comput. Vis. Image Underst.*, vol. 161, no. C, pp. 11–19, Aug. 2017.

[17] A. Karpathy, "CS231n Convolutional Neural Networks for Visual Recognition (3)," 2017. [Online]. Available: http://cs231n.github.io/neural-networks-3/. [Accessed: 16-Mar-2018].

[18] L. Prechelt, "Early Stopping — But When?," in *Neural Networks: Tricks of the Trade*, Berlin, Heidelberg: Springer, 2012, pp. 53–67.

[19] R. Caruana, S. Lawrence, and L. Giles, "Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping," in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, Cambridge, MA, USA, 2000, pp. 381–387.

Artificial Bee Colony Algorithm for The Linear Ordering Problem

Emrullah SONUÇ[1]

[1] Karabuk University, Karabuk/Turkey, esonuc@karabuk.edu.tr

Abstract **- Linear Ordering Problem is an NP-hard combinatorial optimization problem. Several metaheuristic algorithms (Tabu Search, Memetic Algorithm, Variable Neighborhood Search, Simulated Annealing, Scatter Search, Greedy randomized adaptive search procedure) present for the linear ordering problem in literature for finding high quality solutions. This paper presents an Artificial Bee Colony algorithm for solving the linear ordering problem. The results are compared between the other implementations of metaheuristics and Artificial Bee Colony algorithm can produce good solutions for the linear ordering problem instances.**

Keywords **– linear ordering problem, artificial bee colony, metaheuristics, combinatorial optimization.**

I. INTRODUCTION

Linear Ordering Problem (LOP) is an NP-hard problem [1], [2] . In a real world, there are a lot of field of application for the LOP. Some of these fields are economy, sociology, graph theory, archaeology, and task scheduling. Due to this fact, many algorithms are presented for solving the LOP. We can divide these algorithms into two categories as exact methods and metaheuristic algorithms. Among the exact solution methods, a cutting plane algorithm is proposed by Grötschel et al. [3] and a combined interior point/cutting plane algorithm is proposed by Mitchell and Borchers [4]. However, when the size of the problem increases, problem of calculation time arises for these exact methods. Therefore, metaheuristic algorithms have been applied for the LOP. The metaheuristic algorithms for solving the LOP in the literature are as follows:

- Tabu Search [5] (1999).
- Memetic Algorithm [6] (2004).
- Variable Neighborhood Search [7] (2006).
- Simulated Annealing [8] (2007).
- Scatter Search. [9] (2001).
- Greedy randomized adaptive search procedure [9] (2001).

In this paper, Artificial Bee Colony algorithm that is a very popular algorithm in recent years, is proposed for solving the LOP. This paper is organized as follows. In Section II, mathematical formulation and definition of the LOP problem is introduced. Artificial Bee Colony algorithm is described in Section III. Section IV gives details about computational experiments and results. Finally, Section VI states some conclusions.

II. LINEAR ORDERING PROBLEM

Given a matrix $C = \{c_{ij}\}_{nxn}$, the LOP consists of finding a permutation π of the rows and columns of *C*. The aim of the LOP is to maximize sum of the numbers above the main diagonal as it seen in Fig. 1. The LOP can be formulated as follows:

$$cost_C(\pi) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} c_{\pi(i)\pi(j)} \quad (1)$$

where $\pi(i)$ is the index of the column (and row) in position *i* in the permutation. This representation of the LOP is also known as the triangulation problem of input-output matrices.

Sample instance is shown in Fig. 1. The permutation π = (1,2,3,4,5) and its objective value $cost_C(\pi)$ = 179.

	1	2	3	4	5
1	0	13	7	14	26
2	11	0	21	4	17
3	9	6	0	16	33
4	8	10	18	0	28
5	23	12	29	21	0

Figure 1: Sample instance for the LOP (π).

The optimal solution for sample instance is given by π^* = (1,4,2,3,5) (see Fig. 2.) and its objective value $cost_C(\pi^*)$ = 187.

	1	4	2	3	5
1	0	14	13	7	26
4	8	0	10	18	28
2	11	4	0	21	17
3	9	16	6	0	33
5	23	21	12	29	0

Figure 2: Sample instance for the LOP (π^*).

III. ARTIFICIAL BEE COLONY FOR THE LOP

Artificial Bee Colony (ABC) algorithm is a popular evolutionary algorithm. The ABC algorithm is an effective algorithm when comparing to other algorithms in many studies in the literature. The ABC algorithm is developed by Dervis Karaboga [10] in 2005. The ABC algorithm is inspired by honey bees and contains three groups of honey bees namely employed, onlooker and scout bees. The number of employed bees is equal to the number of onlooker bees and this group of bees is called colony. Pseudocode of the ABC algorithm is as follows:

Algorithm 1: Artificial Bee Colony Algorithm

Parameters: NP, FN, NMC, limit
// *NMC* represents the number of maximum cycles
// *NP* is the number of the colony size
// *FN* is the food number that is *NP/2*
// *limit* is maximum number of trials
Initialization;
Set *iteration* = 1;
while *iteration* ≤ *NMC* **do**
  The employed bees phase;
  The onlooker bees phase;
  The scout bees phase;
  Memorize the best solution achieved so far;
  Set *iteration* = *iteration* + *1*;

Figure 3: Artificial Bee Colony Algorithm.

Employed bees search for new permutation to achieve better solutions for the LOP within the neighborhood of the permutation in their memory. Employed bees share their permutation with onlooker bees and then onlooker bees choose their permutations with some probabilities. Employed bees whose solutions cannot be improved through a predetermined number of trials, this is a parameter of the ABC algorithm called "limit", become scouts and their solutions are abandoned.

Metaheuristic algorithms like the ABC algorithm start from an initial solution randomly and then try to improve current solution using some neighborhood mechanism. For the LOP, insertion is an effective and best performing for neighborhood mechanism. For that reason, insertion mechanism is used in this paper. Illustration of insertion(7,3) neighborhood mechanism is shown in Fig. 4.

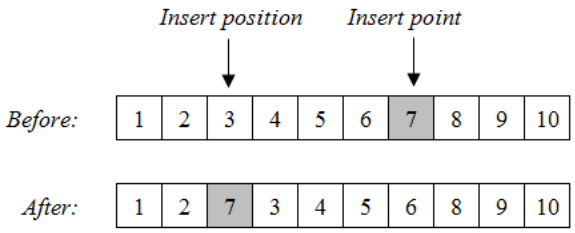


Figure 4: Insertion neighborhood mechanism.

IV. COMPUTATIONAL RESULTS

Computational tests were carried out on several test instances from taken [11]. Before starting the computational tests, parameter selection has been done for the ABC algorithm. Three different tests are determined for solving the LOP using the ABC algorithm. For the parameter *NP*, Karaboga and Basturk [12] pointed out that increment of the NP value does not improve the performance of the ABC algorithm. So that *NP* is chosen as 40 for all runs. *NMC* is 10^5, 10^6 and 10^7 for three different tests, respectively. The value of the parameter *limit* was determined by running sample instance 10 times with the predetermined values of *limit*. After the experiments, the most proper value for the limit is *50n*.

In Fig. 5, for the test instance which is dimension value n = 111, the effect of the limit parameter is shown.

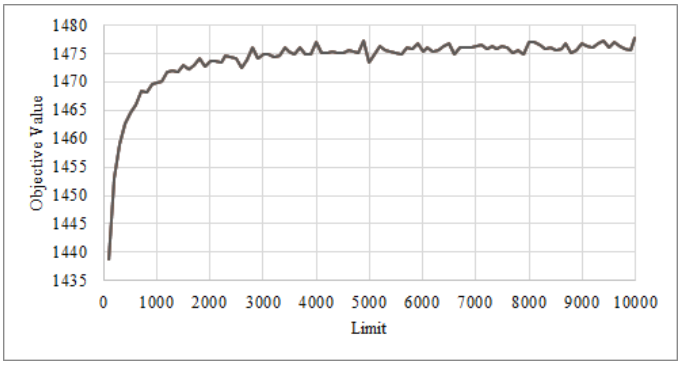


Figure 5: Effect of the parameter *limit*.

The experiments are conducted on Intel Xeon CPU with 2.40 GHZ, 32GB Ram. Each experiment test runs 10 times with random seeds and the best, the worst, the mean, the median and standard deviations are reported in the tables.

Table 1, Table 2 and Table 3 shows the experimental results for *NMC* = 10^5, *NMC* = 10^6, *NMC* = 10^7 respectively.

Table 1: Results of the ABC algorithm on test instances (*NMC* = 10^5).

Instance	Best	Worst	Mean	Median	SD
N-atp111	1478	1472	1475.3	1476	1.73
N-atp134	1774	1765	1768.5	1768.5	2.87
N-atp163	2046	2034	2039.6	2039	3.95
N-atp24	172	172	172	172	0.00
N-atp48	482	481	481.3	481	0.46
N-atp66	758	754	756.1	756	1.14
N-atp76	928	924	925.8	926	0.98
N-EX1	448	445	446.3	446	0.78
N-EX2	441	437	438.2	438	1.33
N-EX3	437	434	435.7	436	0.78
N-EX4	390	387	388.2	388	0.98
N-EX5	405	402	403.8	404	0.98
N-EX6	394	393	393.4	393	0.49

Table 2: Results of the ABC algorithm on test instances (*NMC* = 10^6).

Instance	Best	Worst	Mean	Median	SD
N-atp111	1482	1479	1480.5	1480.5	1.20
N-atp134	1775	1770	1772.6	1773	1.50
N-atp163	2045	2039	2042.5	2042.5	1.75
N-atp24	172	172	172	172	0.00
N-atp48	483	482	482.1	482	0.30
N-atp66	759	757	757.7	758	0.64
N-atp76	929	927	927.9	928	0.70
N-EX1	448	447	447.9	448	0.30
N-EX2	441	438	439.6	440	0.80
N-EX3	437	436	436.9	437	0.30
N-EX4	390	389	389.7	390	0.46
N-EX5	405	404	404.8	405	0.40
N-EX6	395	394	394.3	394	0.46

Table 3: Results of the ABC algorithm on test instances (*NMC* = 10^7).

Instance	Best	Worst	Mean	Median	SD
N-atp111	1482	1479	1480.5	1480.5	1.20
N-atp134	1775	1770	1772.6	1773	1.50
N-atp163	2045	2039	2042.5	2042.5	1.75
N-atp24	172	172	172	172	0.00
N-atp48	483	482	482.1	482	0.30
N-atp66	759	757	757.7	758	0.64
N-atp76	929	927	927.9	928	0.70
N-EX1	448	447	447.9	448	0.30
N-EX2	441	438	439.6	440	0.80
N-EX3	437	436	436.9	437	0.30
N-EX4	390	389	389.7	390	0.46
N-EX5	405	404	404.8	405	0.40
N-EX6	395	394	394.3	394	0.46

In [11], some experiments were presented in a Dual Intel Xeon at 3.06 GHz with 3.2 GB of RAM. There was a two different time limits: 10 seconds and 10 minutes. There is no information about how many times experimental tests run. Even if the comparison of the ABC algorithm is not completely make sense because of the differences of hardware and other specifications, comparison of the best results of three test instances for algorithms are shown in Table 4. The best results in the table are shown as bold.

Table 4: Comparison of experimental results between the ABC algorithm and other algorithms.

Algorithm	N-atp111	N-atp134	N-atp163
TS[5]	**1495**	1788	2066
MA [6]	**1495**	**1796**	**2073**
VNS [7]	1492	1784	2072
SA [8]	1473	1767	2035
SS [9]	1489	1790	2067
GRASP [9]	1488	1783	2055
ABC	1485	1780	2051

According to results in Table 4, Memetic algorithm performs better solutions for all three instances. The ABC algorithm is the second worst performing after Simulated Annealing algorithm. The possible reason of these results can be parameter selections, neighborhood mechanism etc.

V. CONCLUSION

In this paper, the ABC algorithm is proposed for solving the LOP. Results are compared with the other metaheuristic algorithms for the LOP. In terms of quality solutions, the ABC algorithm can produce good solutions for the LOP instances. In future, the ABC algorithm can be parallelized on CPU or GPU for getting better solutions. Also there was a many variants of the ABC algorithm better than the classic ABC algorithm. These variants can use for solving the LOP.

REFERENCES

[1] H. B. Chenery and T. Watanabe, "International comparisons of the structure of production," Econom. J. Econom. Soc., pp. 487–521, 1958.
[2] H. R. Lewis, "Computers and Intractability. A Guide to the Theory of NP-Completeness," 1983.
[3] M. Grötschel, M. Jünger, and G. Reinelt, "A cutting plane algorithm for the linear ordering problem," Oper. Res., vol. 32, no. 6, pp. 1195–1220, 1984.
[4] J. E. Mitchell and B. Borchers, "Solving linear ordering problems with a combined interior point/simplex cutting plane algorithm," in High performance optimization, Springer, 2000, pp. 349–366.
[5] M. Laguna, R. Marti, and V. Campos, "Intensification and diversification with elite tabu search solutions for the linear ordering problem," Comput. Oper. Res., vol. 26, no. 12, pp. 1217–1230, 1999.
[6] T. Schiavinotto and T. Stützle, "The linear ordering problem: Instances, search space analysis and algorithms," J. Math. Model. Algorithms, vol. 3, no. 4, pp. 367–402, 2004.
[7] C. G. Garcia, D. Pérez-Brito, V. Campos, and R. Martí, "Variable neighborhood search for the linear ordering problem," Comput. Oper. Res., vol. 33, no. 12, pp. 3549–3565, 2006.
[8] I. Charon and O. Hudry, "A survey on the linear ordering problem for weighted or unweighted tournaments," 4OR, vol. 5, no. 1, pp. 5–60, 2007.
[9] V. Campos, F. Glover, M. Laguna, and R. Martí, "An experimental evaluation of a scatter search for the linear ordering problem," J. Glob. Optim., vol. 21, no. 4, pp. 397–414, 2001.
[10] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," J. Glob. Optim., vol. 39, no. 3, pp. 459–471, 2007.
[11] "Linear Ordering Problem." [Online]. Available: http://www.optsicom.es/lolib/. [Accessed: 02-Apr-2018].
[12] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," Appl. Soft Comput., vol. 8, no. 1, pp. 687–697, 2008.

Systematic Literature Review on Security Vulnerabilities and Attack Methods in Web Services

F.SABAZ [1], Y.ÇELİK [2]

[1] Karabuk University, Karabuk /Turkey, furkansabaz@karabuk.edu.tr
[2] Karabuk University, Karabuk /Turkey, yukselcelik@karabuk.edu.tr

***Abstract* - The using and importance of web services that flexibly meet the need for communication regardless of parameters such as platform, operating system are increasing day by day. In a simple sense, Web services carry a lot of data because they provide communication. With the increasing usage of web services and evolving technology, various methods have been developed to conceal data, to provide security and to prevent access by third parties in web services. At the same time, attack and injection methods for web service servers or web services have been developed. Some of these methods exploit the fact that the web services are XML-based. For example, XML injection, XPath (which stands for XML path language) injection are some of them. However, perhaps the most common type of attack are DOS and XDOS attacks. The purpose of this study is to gather the reasons of web service attacks, the precautions to be taken against the attack, the solutions for the exploits.**

***Keywords* - Web Service, Security, Attack, Injection, Exploit, Spoofing**

I. INTRODUCTION

Web services are the new generation of web applications. Web services perform their tasks over the web, and these operations take place via the standard HTTP or HTTPS protocol [1,3]. An important feature of web services is that internal data and interfaces are shared with other users or programs. As with traditional APIs, a web application can use multiple web services shared among other applications, or even between companies [2]. Working and communicating all these processes independently from platform, operating system, programming languages is another reason to make this technology attractive. Web Services (WS) has become an important part of the Web because of its attractive features such as user friendly structure, platform independence and XML / SOAP support [1]. Web services have features that can fulfill almost all companies' simple requests (for example, sending details of a product to be purchased) or relatively complex requests (such as checking and realizing travel bookings) [2].

There are 4 technologies on the basis of web services [1].

- eXtensible Markup Language (XML)
- Simple Object Access Protocol (SOAP)
- Web Services Description Language (WSDL)
- Universal Description, Discovery, and Integration (UDDI)

XML is typically used for (hierarchical) data definition. SOAP is XML-based, used for sending and receiving packets with certain standards. WSDL can be easily understood from the description, as well as for service description and to determine the response to the request (data types, messages to be accepted). UDDI is used for publishing a written web service, opening a service and listing available web services [2,3].

Web services can have security vulnerabilities because they connect several platforms and it does some processing with user data. These explanations may relate to XML, SOAP, WSDL, or UDDI [4,5,6]. Especially in complex systems, since the output of a web service is also the input of another web service, it may be easier to manipulate data and values in some way and the danger can reach another dimensions [4]. Some of the exploits and attack types are as follows;

- Injection,
- Denial Of Service (DOS),
- Spoffing,
- Man in the middle attack.

Injection has certain types. These are XML Injection, XPath injection, Sql injection because web services are xml based. Dos attacks, which are basically service stop logic, are known as XDOS. The reasons for XDOS include recursive payloads and oversize payloads. DOS and XDOS are not intended to access system information. It is often intended to stop services and make them unserviceable [4,6].

Spoofing is generally a process of accessing the system as if it were a user. In spoofing, SOAP messages are often sent to the web service server, but the destination address of the SOAP message is redirected to any web server client by the person. In fact, DOS attacks are taking place because of the large amount of these packages. Spoofing has types such as TCP, IP and DNS spoofing. Man in the middle attack is usually the process

REF	Attack Type	Method	Purpose of the paper
[6]	SQL Injection & DDOS	The SOA platform enables the consolidation and collaboration of a number of web services applications served over a network. The security wall of the XML messaging framework designed for SOA-based systems is transparent. SOAP messages can be viewed or other types of attacks can be performed by bypassing HTTP and SMTP protocols. Passing a firewall brings security problems that can interfere with enterprise networks and cause DDOS attacks. Especially for DDOS and Sql Injection attacks, a framework called 'The Integrated Application and Protocol Framework' is developed. Within this framework, there are anti-attack measures implemented via UDDI protocol, WSDL protocol and SOAP.	Exploit, attack methods.
[9]	SQL Injection & XPATH Injection	A tool called CIVS-WS (Command Injection Vulnerability Scanner for Web Services) is developed to detect attacks on Web services. It is aimed to detect similar attacks by teaching query types to the system	Detection of attacks.
[10]	SQL Injection &XPATH Injection	A comparison was made between web security scanners for injection methods. SQL Injection and XPath Injection have been observed to be large in number and rate.	Security vulnerability.
[11]	XSS & SQL Injection	A hybrid method has been developed. In this method, two parts, static and dynamic analysis, were developed. It is noted that in static analysis, attention will be given during coding. Dynamic analysis is a measure that can be taken after the application is published.	Detection of attacks and exploit.
[12]	XSS	In the study, theoretical information is given about XSS vulnerabilities. It has been indicated that it is possible to prevent XSS code from running without session permission by opening security tag to soap messages and adding tags such as userName and role. It is indicated that there are many standard and programming languages to detect injection. As in many methods, it has specified 2 phases as static and dynamic phase. In addition, an attempt was made to penetrate web services by developing a script. The analysis data of the injection process is included in the study.	Security vulnerability, prevention of hacking attacks and penetration test
[13]	General security vulnerabilities based on SOAP message security	With the developing SOA technology and universal applications of SOA technology, heterogeneous platform-based security vulnerabilities are on the rise. Previously, two different security vulnerabilities for Apache Axis platforms were specified after .NET and their solutions were specified. A Web services security module based on the .NET and Axis2 platform has been proposed. For example, in the module developed for .NET, 4 different filters are defined for requests and sent data to web services. This filter is divided into 2 in itself. Each filter is assigned a task. This structure is named WSE Security Framework.	Security vulnerability, and prevention of hacking attacks
[14]	XPATH Injection, XML Injection, SOAP attacks, DDos,	Because Web services want to exchange information between companies and within the company, security vulnerabilities in web services can constitute serious business losses. Security issues have been investigated in three stages. XML, SOAP and discovery (web service). SOAP messages contain XML Injection-like injection methods and configurations for the .NET platform to prevent some of attacks. AJAX technology, which is often used in web applications, has been shown to be able to play session and cookie informations in particular, making use of the script code.	Attack type and prevent attacks
[15]	DDOS	The aim of the paper is on cloud systems and vulnerabilities in other web applications such as web services, especially DDOS attacks. In the study, the ways to prevent hacking attacks on web services at the application layer are shown.	Attack type and prevent attacks
[16]	SQL Injection, XSS	Applications have been proposed to detect various exploits. It has been shown that exploits can be detected through applications such as Acunetix, Netsparker, Firefuzzer	Exploit
[17]	XML-SOA-Based Vulnerabilities	Peer-to-peer architectures can be vulnerabilities of web services developed for the architects. A framework called OGSA-P2P has been developed for this purpose.	Prevent attacks
[18]	DDOS	There are 2 types of DDOS in the study; Protocol Deviation attacks and vulnerability exploitation are explained. In Protocol Deviation attacks, a single packet may cause the system to crash. The most familiar technique is the Ping of Death method. It is based on the method of exploiting the network of the attacked system by exploiting bandwidth, memory and all other processing resources in the source exploitation attack. An xml authentication based step system called CheckWay Gateway is developed to prevent attack. In the installed system, a gateway is established between the Web Service server and the client. The SOAP Request message from the client to the server must pass through this gateway authentication in particular.	Prevent attacks.
[19]	DOS-DDOS	In this study, DDOS attacks are divided into 2 sections as Flood and Semantic attacks. Flood attacks are defined as the fact that the service does not respond to other packages by sending regular packets to the web service server in a very serious amount. Semantic attacks are attacks that an attacker sends with many WS-Security headers containing nested encrypted items. The service will not respond to other packages again because the semantic attack requires a serious resource and time to respond to the request. To prevent these two situations, the author developed a method called Client Puzzle. The server asks the client to resolve some puzzles in moderate to hard terms before starting to make a difficult or expensive operation in terms of incoming and calculating. The purpose here is to make a distinction between the person who makes the request and the person who is the bot.	Exploit detection
[20]	DDOS	In the study, various DDOS attack experiments were performed on Java Metro, Apache Axis, and .NET platforms. Experiments have been done through exploits such as nested xml expressions, WSDL Flooding, and heavily cryptographic operations. Analyzes of attacks against platforms, especially memory and processor consumption have been made. Simple Network Management Protocol (SNMP) and network interface card are used to obtain analysis data. Management data of a network managed by SNMP is stored in the management information base called MIB (Management Information Base).	Attack Scenarios

[21]	SQL Injection, XPath Injection	SQL Injection, and XPath Injection have been prevented as much as possible by the encoder, they must be re-tested through various third-party software. these programs are as follows; • HP WebInspect, • IBM Rational AppScan, • Acunetix Web Vulnerability Scanner The program that gives the best result for XPath injection and SQL Injection seems to be HP WebInspect. The study also includes reports on whether the attacker is able to obtain information that should be hidden, such as the user name, password, etc., carried through the web services, and can view the file paths of the server. It has been determined that the program that gives the best result regarding these 2 explanations is Acunetix Web Vulnerability Scanner.	Automatic detection of vulnerabilities
[22]	DDOS, XDOS	The system designed in this study has 2 modes. Normal and attack mode. If the system is not in normal mode, it does not perform a validation request. If the system is taken into attack mode, the service requests to the server must be verified before processing. If any of the requests sent for the attack are validated, only one of them will be processed. Thus, the server does not waste resources for hack attacks done to the resources. In addition to this, the server is also spending resources for these validations. The resulting requests must be verified. An attacker can still consume the victim's system resources by sending a large number of requests that force verification and authentication. In order to prevent this, a mechanism called operations provider has been established. The operation of the operation provider directs the incoming authentication requests through the services in the name of the authenticator.	Prevent hack attacks
[23]	SQL Injection, XPath Injection	A tool called CIVS-WS (Command Injection Vulnerability Scanner for Web Services) has been developed to take precautions against SQL and XPath injection attacks. This tool is integrated into the system in 5 steps. One of the most important of these steps is the training of the system. For this, a workload comprising various parameters of the injection methods is created. The workload is then applied to the system to train the system. After training system injection methods, an attack load based on a large set of SQL Injection and XPath Injection attacks is created (in other words, a set of web service calls containing malicious parameters). Subsequently, attack load is applied to the trained system. The system is run in order to detect security vulnerabilities at the point of detection of attacks that are intended to injection in the system with normal inquiries during the detection of attacks. In addition, the author compared this application with some other anti hacking softwares and found that his own CIVS-WS tool gave better results.	Detection of attacks and prevent hack attacks
[24]	CDATA Field exploit, DDOS attacks, XML Injection, SOAP attacks, Mitnick attack	The study mentioned the explanations of a web services and other web applications. The Mitnick attack was first carried out by Kevin Mitnick in the U.S.A. The method utilizes the vulnerabilites detected in the three-way handshake method in TCP. Assuming that there is a TCP connection between two devices named H1 and H2, the attacker tries to act like one of the devices by interrupting the connection between H1 and H2. If it behaves like H1 here, it will perform Syn / Flood attack on H1 device. It also creates several TCP packets to find the TCP sequence number H2 generates for H1 and sends it to H2. The attacker takes the IP of H1 by spoofing and acts like it and sends the SYN packet to H2. It is the first step of the handshake method. The goal is to try to keep the TCP connection between the attacker and the H2 uninterrupted. Because H1 is exposed to flood attack, he will want to terminate the connection. H2 responds to H1 with a SYN / ACK message, but H1 can not send an RST packet to terminate the connection because it is initially under SYN flood attack. In this case, according to the TCP sequence number and handshake operation, it is assumed that a secure connection is established between the H2 device and the H1, H2 is not aware that it is actually connected to a malicious person. The **CDATA FIELD vulnerability** is due to the fact that web services are XML-based. CDATA allows characters such as '<' and '>' to run in the field. So the script codes can run and the CDATA FIELD vulnerability occurs in the web services. <![CDATA[<]]> **SCRIPTCODES** <![CDATA[>]]> alert('YOU HAVE BEEN HACKED') <![CDATA[<]]> /**SCRIPTCODES** <![CDATA[>]]> Script code can be run in the fields written in bold font. There are various methods and subclasses of DDOS attacks in Web services. When trying to parse too many xml tags nested inside the server, attacker proof the server resources. XML Injection attacks can also be done in web services. DOM documents, on the other hand, have a more complex and intelligent structure, but there are also missing aspects of the DOM structure. Because attackers can not resist other types of attacks, including DoS attacks, when they can send extremely complex but well-documented XML documents (DOM attacks). The system is forced to create large sized objects in memory and the memory of system is exhausted. Another example of an attack for Web services is attacks based on SOAP vulnerabilities. If a hacker creates SOAP messages with very complex SOAP headers (SOAP headers), SOAP Header Attacks that can be used in WS DoS attacks can be performed. Another example of SOAP Attacks (also used in WS DoS) are attacks that can occur if an attacker sends repetitive SOAP messages to over-force the WS. The attack type is called SOAP Replay attack.	Automatic detection of vulnerabilities.

between the hardware and the user to capture and inspect packets. Similar to the spoofing method. [7,8].

II. APPROACH REVIEW OF LITERATURE

In this study, a list of studies about security vulnerabilities for web services and their prevention in the literature has been made. when preparing this paper;

- Study prepared on web service security until now and total number of these studies,
- Studies, are focused on which types of attacks for web services,
- Recommended solution methods for web service security
- The security of the web service and other web technologies are connected to each other.

These titles are considered in this study.

IV. RESULTS OBTAINED

Looking at the previous study on web services, the importance of web services is clear. With this importance, it is understood that the vulnerabilities in web services sometimes cause serious problems. Because Web services are separate XML based from other web applications, unlike classic web applications, there are various types of exploits and attacks such as XML Injection, XPath Injection, SOAP attacks. They need to be well analyzed and used in this way. It is very important that these analyzes are done well. Otherwise, hackers can see data that transfered with the web services or block communication of web services because many important information and data are carried through web service.

III. CONCLUSIONS

There are basically two kinds of attacks in the literature when we look at the studies prepared on web service security. These are DDOS and various injection techniques. In DDOS studies, the aim is usually to distinguish whether the request to the server is a real user or a request sent from malicious software. Several methods have been developed for this. By solving a small and easy puzzle to the user, it can be understood whether the owner is a real person or not [19]. In addition to, by monitoring the general traffic of the network with SNMP, it can taken precaution more safely by detectin that the system is under attack depending on various parameters[20].

SQL Injection, which is the most known of injection methods, is a common injection method not only in web services but also in all other web applications. In order to prevent SQL Injection, it is generally seen in the literature that the values coming from the client are analyzed well and the characters have to be taken carefully from the client. It can be done by the encoders, previously developed frameworks and programs can be also used[6, 9]. For example, applications such as Acunetix have been found to be useful in terms of determining what kind of exploits are detected on which pages and uploading some suggestions before they are uploaded as a project [16].

Since Web services are XML-based, problems and exploits related to XML can occurse in web applications. Attacks with XML Injection and nested xml data are the most common attacks. Especially when attackers are attacking the server to proof its resources, the data in the xml file begins to unfold after validating the source of the XML data, such as in DDOS attacks. Otherwise, attempting to unravel each of the incoming data seriously lost time and resources to the server.

Other types of attacks, such as XSS attacks, are also made through web services. In XSS attacks, cookies and session information are usually run through forwarding or scripting codes. Then the attackers can access the personal information of users in the web applications. One of the most effective ways of preventing XSS exploit is to provide the role and session information to the user after adding a security tag in the SOAP messages, so that the script code can be executed according to the given roles. In this case, execution of source unknown scripts can be prevented[12].

REFERENCES

1. Andrews, M., Whittaker, J., A., "How to Break Web Software: Functional and Security Testing of Web Applications and Web Services", Addison-Wesley Professional, New York, 240 (2006).

2. Jønvik, T., E., Hartvigsen, A., M., van Thanh, D., "Ubiquitous Access to Personalised Services", Mobile and Wireless Communications, 301-308, 106:(2003).

3. Yin, Y., Zou, B., "The Analysis and Research of OPC XML-DA Server", International Conference on Future Energy, Environment, and Materials, (16): 1535-1540 (2012).

4. Mouli, V., R., Jevitha, K., "Web Services Attacks and Security-A Systematic Literature Review", 6th International Conference On Advances In Computing & Communications, 6-8 (2016).

5. Zhou, B., Shi, Q., Yang, Q., Yang, P., Yu, Y., "Measuring Web Service Security in the Era of Internet of Things", Computers & Electrical Engineering, (2017).

6. Elçi, A., "Web Services Security: Attacks and Solutions", Doğu Akdeniz Üniverstesi, (2009).

7. Templeton, S., J., Levitt, K., E., "Detecting Spoofed Packets", DISCEX, (2003).

8. "Web Service Spoofing", http://www.ws-attacks.org/WS-Addressing_spoofing

9. Antunes N., Laranjeiro N., Vieira M., Madeira H., "Effective Detection of SQL/XPath Injection Vulnerabilities in Web Services", 2009 IEEE International Conference on Services Computing, 260-267 (2009).

10. Vieira M., Antunes N., Madeira H., "Using Web Security Scanners to Detect Vulnerabilities in Web Services",

IEEE/IFIP Intl Conf. on Dependable Systems and Networks, (2009).

11. Sarıman G., Küçüksille E., "Web Servislerinin Yazılım Güvenlik Testleri için Önerilen Hibrit Yaklaşım", SDU International Journal of Technological Science, 1-14(2016).

12. Salas M. I. P., Martins E., "Security Testing Methodology for Vulnerabilities Detection of XSS in Web Services and WS-Security", Electronic Notes in Theoretical Computer Science, 133-154(2014).

13. Yue H., Tao X., "Web Services Security Problem in Service-oriented Architecture", International Conference on Applied Physics and Industrial Engineering, 1635-1641(2012)

14. Stamos A, The Next Generation of Vulnerable Enterprise Applications, www.isecpartners.com

15. Carlin A., Hammoudeh M., Aldabbas O., Defence for Distributed Denial of Service Attacks in Cloud Computing, The International Conference on Advanced Wireless, Information, and Communication Technologies, 490-497(2015).

16. Saleh A. Z. M., Rozali A. N., Buja A. G., Jalil K. A., Ali F. H. M., Rahman T. F. A., "A Method for Web Application Vulnerabilities Detection by Using Boyer-Moore String Matching Algorithm", Information Systems International Conference, 112-121(2015).

17. Cristescu M. P., Stoica E. A., Ciovica L. V., "Web Services Specific Security Standards", International Economic Conference, 597 – 602(2014)

18. Gruschka N., Luttenberger N., "Protecting Web Services from DoS Attacks by SOAP Message Validation", Christian-Albrechts-University of Kiel

19. Suriadi S., Stebila D., Andrew C. and Liu H., "Defending Web Services Against Denial of Service Attacks Using Client Puzzles", IEEE International Conference on, (2011).

20. Suriadi S., Clark A. and Schmidt D., "Validating Denial of Service Vulnerabilities in Web Services", International Conference on Network and System Security, 175-182(2010).

21. Vieira M., "Using Web Security Scanners to Detect Vulnerabilities in Web Services", Departamento De Engenharıa Informátıca, (2009).

22. Ye X., "Countering DDoS and XDoS Attacks against Web Services", International Conference on Embedded and Ubiquitous Computing, 346-352(2008).

23. Antunes N., Vieira M., Laranjeiro N., Madeira H., Effective Detection of SQL/XPath Injection Vulnerabilities in Web Services, IEEE International Conference on Services Computing, 260-267(2009).

24. Vorobiev A., Han J. H. J., "Security Attack Ontology for Web Services", Semantics, Knowledge and Grid, Second International Conference, 42-48(2006)

Road Extraction Techniques from Remote Sensing Images: A Review

I. KAHRAMAN[1], I. R. KARAS[1]

[1]Department of Computer Engineering, Karabuk, Turkey

Abstract **- The importance of analysis high resolution satellite imagery plays an important research topic for geographical information analysis of cities. Geospatial data plays an important role in important issues such as governmental, industrial, research topics on traffic management, road monitoring, GNSS navigation, and map updating. In this study, road detection from satellite imagery methods are classified as classification-based, knowledge-based, mathematical morphology and dynamic programming. In the beginning, the road structures including feature and model are analyzed. Then, the advantages and disadvantages of road detection methods are evaluated and summarizez their accuracy and performance based on road detection principles. Therefore, in order to obtain remarkable results for road detection, it is better to use more than one method. In after days, performing a complex road extraction from a satellite image is still a necessary and important research topic.**

Keywords **- Remote Sensing Image, Classification, Road Extraction, Road Feature**

I. INTRODUCTION

After release of enhencement in land observation satellites, many technologies applied to the remote sensing images with image processing developed rapidly. The main purpose of the RS implementations is to gather meaningful and semantic information and clarify interested targets to understand the whole image. Road network extraction from a RS image has many complex difficulties but it is worth to study. Roads are the fundamental part of transportation, traffic management, city planning, GPS navigation and map updating. Due to rapid changes on road network in urban transportation system, it is required to extract road network for instant road mapping. Various road detecting approachs have been presented for road network extraction from high resolution satellite images.

II. ROAD MODELS AND FEATURES

Road extraction from RS images, image characteristics of road features have difficulties according to spectral and spatial resolution, weather conditions, sensor type, light change and soil characteristics. Practically, extracting a road network based on structural model is very complex. Therefore, analysis of road characteristics and road modeling are very important. Geometric features, photogrammetric features, topological features, functionality and texture features are the main road features.

Different road sections in the images have different properties for road extraction from an image. The geometric properties of a path are directly related to the path shape. Photometric features are more representative of the gray level and color of the path. Topological and functional properties are relatively easy, but it is difficult to use these properties in real applications.

III. ROAD EXTRACTION ALGORITHMS

Due to numerious applications that applies diffirent road extraction methods, it is not easy to classify the techniques in detail. High number of research studies show that, most of the methods for extracting road segments consist of klowledge-based, classification-based, morphology-based and dynamic programming. These methods are briefly described in the following section.

A. Classification-Based Methods

Methods for classification-based techniques generally use geometric properties, photometric properties, and texture properties for road sections. The accuracy for classification is unsatisfactory due to misclassification among road and road-like objects such as roads and building structures, parking lots, area blocks and water areas.

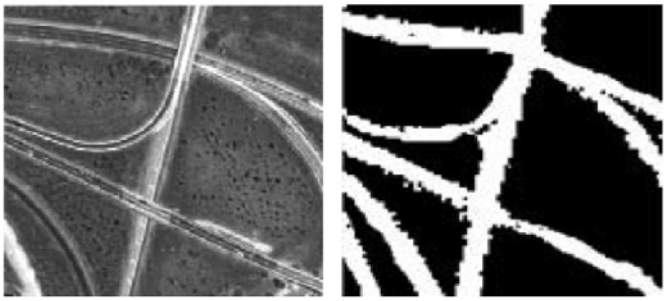
Labeled samples are trained for supervised classification methods. For the accuracy of these methods depend on selected features and labeled samples. In general, three types of supervised classification method is described as follows.

(1) Artificial Neural Networks Classification Method

The first studies were based on spectral and contextual information of image pixels for direct classification of back propagation neural networks and advanced model. Tu-Ko (2003) presented a powerful approach to the extraction of the roadway major line, where the neural network is trained with edge and spectral features. The results can have good results throughout the system, even with many segments that are not roadside. The back propagation neural network method presented tested the network structure by experimenting with different parameters and obtaining the optimal input vector [5]. The process of determining optimal input parameters, network

structure, and termination conditions in education is quite troublesome.

A method have been implemented with back-propagation network for road detection. As a first step, spectral information was used for road detection. After that, as an input texture parameters for gray level co-occurrence matrix (GLCM), contrast, energy, entropy, and homogeneity were calculated obtained from the input image. For the purpose of optimizing the system performance and evaluate the contribution of texture parameters to path determination, determinated texture parameters were associated with spectral information. The output road map is demonstrated in Figure 1. [6].



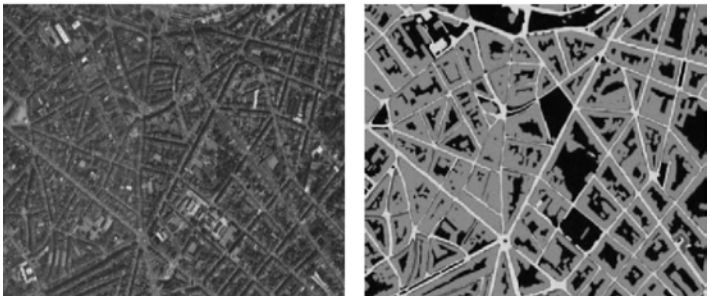
(a) (b)

Figure 1. Road map, (a) input data, (b) output road map

Since, there are some disadvantages of the back propagation neural network methods such as; slow convergence, required more training samples, the performance is getting slow by increasing the number of classes, and it is easy to encounter over-fitting. Therefore, many newly enhenced neural network models is proposed to derive the road sections from RS images. For instance, radial floor nerve network function, fuzzy neural networks, spiking neural network, and hybrid neural network [4].

(2) Support Vector Machines Classification Method

Risk minimization and generalization accuracy are the benefitial aspects for object detection using SVM classification methods in high resolution RS images. Hence, determining kernel functions, selection of domain and training samples constitute complexity using SVM methods.

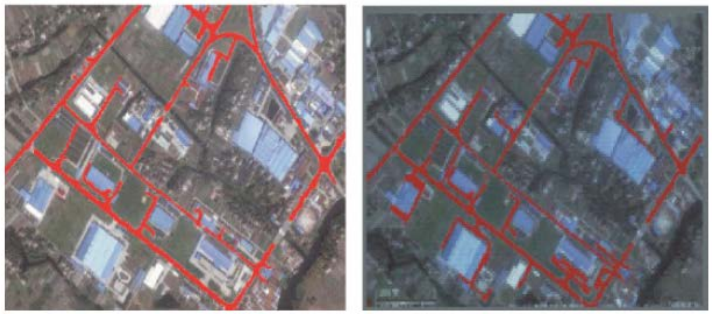


(a) (b)

Figure 2. Extraction of road network. (a) Original image. (b) Extracted road network.

(3) Markov Random Field (MRF) classification methods

A Markov Random Field (MRF) is a graphical model of a joint probability distribution. Due to spatial correlation between all pixels in an digital image, it can be anaylzed effectually by defining the conditional probability distribution function and texture statistical properties. MRF is widely used in areas such as image segmentation, edge detection, restoration and reconstraction. In can be seen from the studies and image below that relationship complexity between extracted objects in a satellite image makes it hard to model accurately by using just MRF model [15]. So, the hybrid model should be used to achieve remarkable results for complex object detectşon from satellite images.



(a) (b)

Figure 3. Two method for road extraction. (a) MRF-based detection. (b) SVM and FCM-based detection.

B. Knowledge Based Methods

It is hard to extract the roads from satellite images using only local spectrum and texture features. Since the structure of the road lane that makes identification of the eigenvector difficult, the data can not be entered directly into the classifier. For this reason, parameter models such as energy function can be used to work on the maximum value of the energy function. Common parameter models usually derive some structural elements from each other in relation to each other, and finally detect that it is specific enough to perform object detection [15]. Although the knowledge-based methods have been

implemented to derive from RS images, it has some disadvantages such as over-subtraction, occlusions and shadowing.

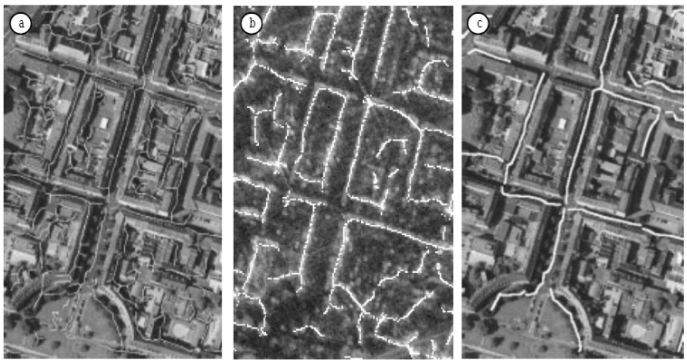


Figure 4: The segmented lines, i.e. road candidates, from the visual image (a) must be accompanied by parallel lines as hint for buildingsin the SAR image (b) to verify the road hypothesis (c) [14]

C. Mathematical Morphology Methods

Mathematical morphology takes a great deal of interest in academic areas, such as computer vision, image processing, pattern recognition, and other topics. Since 1980s, researchers proposed many methods based on mathematical morphology for road detection. Hence, image segmentation processes have been applied by mathematical morphology methods combined with different convenient methods.

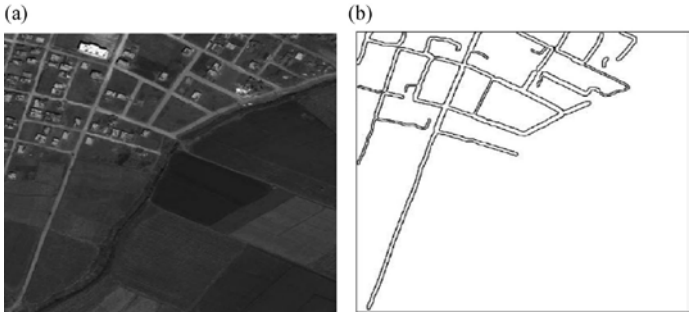


Figure 5. Road detection by Cem algorithm. (a) Ikonos image. (b) Road network detected.

Mathematical morphology methods reveal certain advantages that are commonly used in deriving from satellite images. The image segmentation outcome are considerably influenced by the selection of the structural elements by shape and size. Since the components of the construction elements, it is hard to use only the mathematical morphology method in order to obtain remarkable accuracy and significant extraction outcomes.

D. Dynamic Programming And Grouping

Dynamic programming is a metod that implements a mathematical method to perform the decision process. Ordinarily, the path must be given as a parameter model and expressed as a cost function. Dynamic programming is considered as a computation tool to specify the best path between the seed points. An algorithm for automatic road detecting from an satellite image with dynamic programming and Kalman filter for path monitoring is proposed. In spite of missing edge and occlusion of cars and bridges, it can detect roads simultaneously and in real time. But, there are particular limitations in this method for many assumptions as much as the previous information available in the monitoring operation. Movaghati et al. (2010) introduced a extracting model using the Extent Kalman filter and the Particle filter. These algorithms use a clustering algorithm without considering road barriers. This algorithm may follow all path intersections, but the result depends on the parameters that are set in the modüle [12].

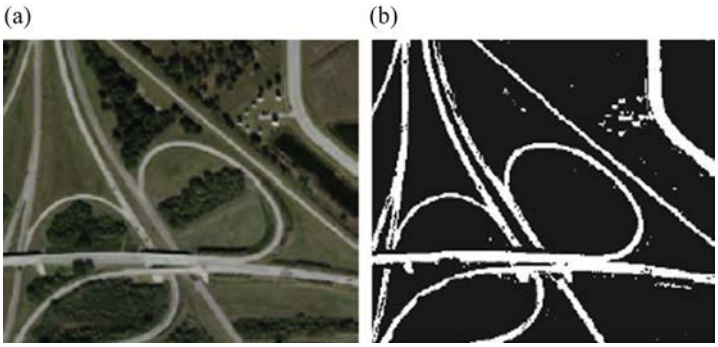


Figure 6. Road and road network extraction results. (a) Input image. (b) Output image.

IV. CONCLUSION

As explained above, each algorithm has advantages and disadvantages according to the run time, accuracy and complexity. It can be detected roads from satellite images using only one algorithm. However, the accuracy will not be in remarkable accuracy, unless using combination of various methods. Additionally, The main problem for road detection from satellite images is to determine features of the road. A linear line is defined for road sections in most of the presented methods. After incrising advanced technology, it is easy to reach high resolution images. So many detailed road features will emerge including lots of noise. These roads should be defined fully. For this reason, many researchers concern about proper road model and how to detect roads in remarkable accuracy and quickly.

REFERENCES

[1] Vosselman, G., Knecht, J., 1995. Road tracing by profile matching and Kalman filtering. In: Gruen, A., Baltsavias, E., Henricsson, O. (Eds.), Workshop on Automatic Extraction of Manmade Objects from Aerial and Space Images, Birkhauser, Berlin, pp. 265e274.

[2] Baumgartner, A., Steger, C., Mayer, H., et al., 1999. Automatic road extraction based on multi-scale, grouping, and context. Photogrammetric Engineering and Remote Sensing 65 (7), 777e785.

[3] Herumuti, D., Uchimura, K., Koutaki, G., et al., 2013. Urban road extraction based on hough transform and region growing. In: The 19th KoreaeJapan Joint Workshop on Frontiers of Computer Vision, Incheon, 2013.

[4] George, J., Mary, L., Riyas, K.S., 2013. Vehicle detection and classification form acoustic signal using ANN and KNN. In: 2013 International Conference on Control Communication and Computing, Beijing, 2013.

[5] Mokhtarzade, M., Valadanzoej, M.J., 2007. Road detection from high-resolution satellite imagery using artificial neural networks. International Journal of Applied Earth Observation and Geoinformation 9 (1), 32e40.

[6] Kirthika, A., Mookambiga, A., 2011. Automated road network extraction using artificial neural network. In: IEEEInternational Conference on Recent Trends in Information Technology, Chennai, 2011.

[7] Zhu, C., Shi, W., Pesaresi, M., et al., 2005. The recognition of road network from high-resolution satellite remotely sensed data using image morphological characteristics. International Journal of Remote Sensing 26 (24), 5493e5508.

[8] Wang, Y., Zheng, Q., 1998. Recognition of roads and bridges in SAR Images. Pattern Recognition 31 (7), 953e962.

[9] Hu, J., Razdan, A., Femiani, J.C., et al., 2007. Road network extraction and intersection detection from aerial images by tracking road footprints. IEEE Transactions on Geoscience and Remote Sensing 45 (12), 4144e4157.

[10] Zhang, D.B., 2007. Research on Interactive Road Extraction Method from High-resolution RS Image (PhD thesis). Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Scienee, Xi'an.

[11] Barzohar, M., Cooper, D.B., 1996. Automatic finding of main roads in aerial images by using geometric-stochastic models and estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (7), 707e721.

[12] Movaghati, S., Moghaddamjoo, A., Tavakoli, A., 2010. Road extraction from satellite images using particle filtering and extended Kalman filtering. IEEE Transactions on Geoscience and Remote Sensing 48 (7), 2807e2817.

[13] Das, S., Mirnalinee, T.T., Varghese, K., 2011. Use of salient features for the design of a multistage framework to extract roads from high-resolution multispectral satellite images. IEEE Transactions on Geoscience and Remote Sensing 49 (10), 3906e3931.

[14] Tönjes, Ralf & Growe, Stefan. (2002). Knowledge Based Road Extraction from Multisensor Imagery.

[15] Wixing Wang, Nan Yang, Yi Zhang, Fengping Wang, Ting Cao, Patrik Eklund, A review of road extraction from remote sensing images, Journal of Traffic and Transportation Engineering, Volume 3, Issue 3, June 2016, Pages 271-282

Delaunay Triangulation and Its Applications

M. AKSİN[1], E. DEMİRAL[2] and İ. RAKIP KARAŞ[3]

[1] Karabuk University, Karabuk/Turkey, mustafaaksin@karabuk.edu.tr
[2]Karabuk University, Karabuk/Turkey, emrullahdemiral@karabuk.edu.tr
[3]Karabuk University, Karabuk/Turkey, ismail.karas@karabuk.edu.tr

Abstract **- Data collection, data retention and analysis are becoming more and more important every day because of the fact that technology is involved almost all our life. The processing and analysis of the data can become more difficult with increasing precision.**

In three-dimensional surface modeling, due to the increase in sensitivity and data size depending on the surface state and extent of the surface, collecting and processing the data may become difficult. As a solution to this situation, we can see that Computational Geometry is used extensively. Computational Geometry derives intermediate interpolations by taking the start and end data as references instead of keeping each data separately. In this way, it is possible to model by determining intermediate values based on the mentioned reference points.

There are various methods in Computational Geometry such as intersection detection, point position and triangulation. According to the needs, a solution way can be produced by various geometric computations. In our study, "Delaunay Triangulation" which is the most used type of triangulation methods will be examined.

Keywords **- Delaunay Triangulation, Surface Modeling, Computational Geometry.**

I. INTRODUCTION

THE importance of developing technology in human life is increasing day by day. Every day, technology is spreading to more areas. Existing technologies are further developed and, accordingly, the need for data acquisition and data processing is also increasing. New needs arise, such as larger capacity memory units, more powerful processors, faster data transmission.

Small size items that we can use in our home can be made with 3D printers. Robot technology can now be designed to be close to reality. Historic buildings, museums, uninhabited cities and many other places can be seen and traveled through a computer program while sitting at home. 3D applications are developed with technology make our lives easier. In addition, technological requirements are increasing.

In some cases, these requirements can be met by installing systems that are much stronger than normal, but these needs may become unobtainable as the scope of the data expands. In these cases, generalization of the data may be the solution. This method facilitates the processing and management of the data.

II. MODELING METHODS

Surface modeling currently used in 3D applications is facilitated in terms of visualization, but it is still not possible to obtain at the desired level, store and analyze data in a topological sense. Although it is possible to model smooth surfaces or periodic rough surfaces, the situation is different especially on irregularly distributed surfaces.

When we think of a three-dimensional irregular surface, there are many complicated relationships between many parameters for modeling this area on a computer. One of the most common problems in cartography is surface digitization in order to work on irregular surfaces. It is possible only to digitize a map surface for a certain area and model it in a computer environment, although this is not possible exactly to everywhere. Usually the area, region, island, parcel etc. each area is estimated within itself by making measurements on the basis. Relation with other regions and areas is not determined. In this case, measurements are made in the relevant region, reference points are determined and intermediate values such as position and height according to these reference points are determined by interpolation method. In this way, a whole surface modeling is done.

Since it is not possible to model all real world surfaces in a one-to-one manner, the digitized model, which can be regarded as the primary model, is obtained by primarily structuring and generalizing spatial data. In the model generalization, it is the basic aim to prepare the data in the density and structure in accordance with the secondary models to be presented to the user by visualization and to obtain the cartographic model [1].

Various methods are used to generalize topological data so that it can be compiled as much as possible. Various geometric calculations can be made that vary according to the structure of the data to be processed, the requirements of the analyzes to be performed, and the data to be obtained. Here are some of them [2,3]:

a. Geometric Search
b. Concave Wall Creation
c. Separation by triangles
d. Finding proximity between objects
e. Finding the intersection

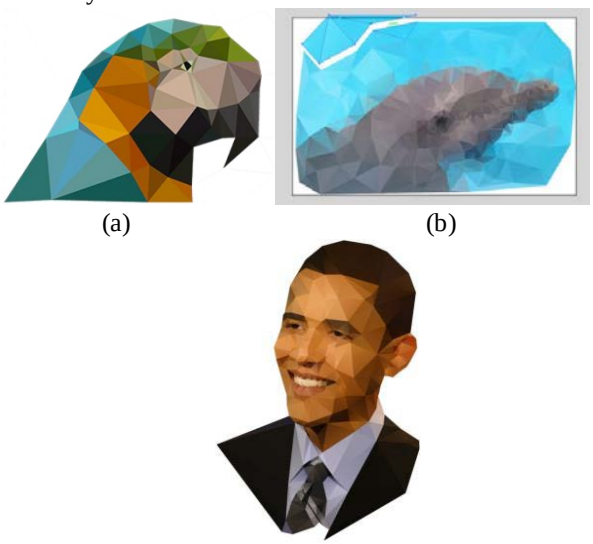
Within these methods, "Separation by triangles" method is frequently used within the scope of geographical information systems. Among the triangulation methods, "Delaunay triangulation" is the most commonly used triangulation method in surface modeling.

III. DELAUNAY TRIANGULATION

The main method used to model the surfaces is the "Delaunay Triangulation" method. In Turkish it is known as "Delaunay Üçgenlemesi". This method developed by Boris Delaunay in

1934.

The Delaunay triangulation relies on the reference points on a 2D or 3D surface, linking these points. It is a method that provides modeling by generating intermediate values by interpolation method. It is also used in different areas. However, map engineers seem to be used extensively in the models they need.



(a) (b)

(c)

Figure 1: Delaunay Triangulation applied pictures

We can list the important features of the Delaunay Triangulation as follows.

- It makes sense. It is definite and only one result.
- Independent from initial reference point.
- The resulting triangles are as equilateral triangles as possible. Due to the formation of very narrow angled triangles, a linear connection between the distant and non-directly related points is prevented [4].
- There is no other empty spot in the periphery of the diagram.
- The convex frame of the dataset is contained within the cluster. The convex frame of a point set is the smallest polygon that takes in that point set.
- The circle of reference points is located in the right triangle formed by the pair of points closest to each other.
- It forms the edge of a triangle with the right piece, which connects each point with another point closest to itself [5].

Delaunay Triangulation occupies a very important place in computational geometry [6]. It is necessary to define the Voronoi diagram in order to understand this triangular model which is so important.

IV. VORONOI DIAGRAM

The Voronoi diagram is also referred to in the literature as the Dirichlet, Thiessen, or Wigner-Seithz diagrams. Any point of the set of endpoints located in the plane is called the "Voronoi polygon (Field)" instead of the geometric plane of the points located closer to the other points in the cluster. The union of the Voronoi polygons of all points in the cluster forms the "Voronoi Diagram" of that cluster [2].

In the Voronoi diagram, the edges of the Voronoi polygon from the lines separating the points from the areas of the other points separate the point inside the area from the other neighboring points. The lines joining the reference points form the middle struts of the edges of the Voronoi Polygon.

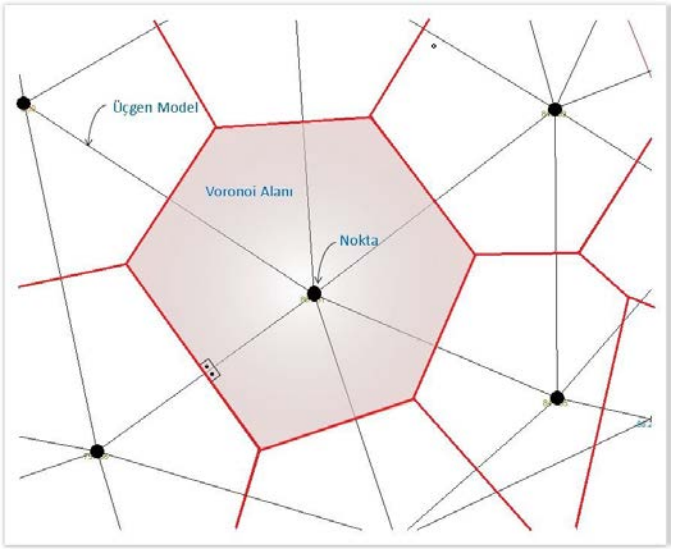
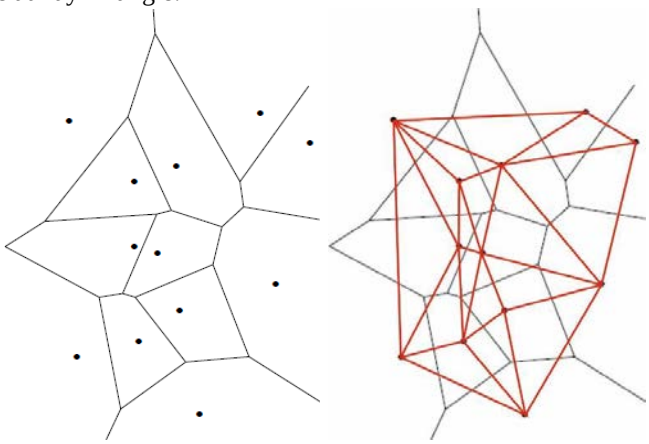


Figure 2: Voronoi polygon [7]

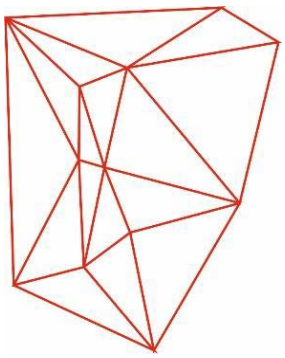
A. *Voronoi Polygon*

The Voronoi Polygon belonging to one point distinguishes that point from the other points adjacent to it. The other points closest to a point are defined as the points adjacent to that point. The middle struts that pass through the midpoints of the lines connecting the point and neighboring points are forming the edges of the polygon [7].

After the reference points are determined, the Voronoi diagram can be extracted. Voronoi diagram is a definite structure used for nearest point problems. In the Voronoi diagram, the Delaunay triangulation can be obtained as a result of connecting the reference points with each other reference points in their neighboring areas. In other words, the Voronoi diagram formed by the Voronoi polygons forms the basis of the Delaunay Triangle.



(a) (b)



(c)

Figure 3: Obtained samples of Delaunay triangle from Voronoi diagram

First of all, Voronoi Polygons are obtained by taking the midpoint sheds of the determined reference points. Then the triangulation is completed by obtaining the connections by the Delaunay triangulation method of these reference points.

V. The Importance Of Delaunay Triangulation In Geographical Information Systems

The modeling of generalization in Geographical Information Systems is a complex task because it depends on a wide variety of parameters. In addition to the roads, it is required to perform network analysis with parameters such as water network, sewer line, electricity lines. Generalization methods are gaining importance in order to make all these structures together and to make analysis such as shortest path analysis on the whole.

Bildirici and Selvi, in their studies on the generalization methods used in the Geographic Information Systems, the importance of triangulating and analyzing the generalization methods used in Geographical Information Systems with their strengths and weaknesses;

"In their work, four methods that have been assessed as feasible from a large number of previously published methods have been analyzed with their strengths and weaknesses. As a result of the study, it is planned to apply the triangulation method which is evaluated as the best in terms of applicability with macro support in a GIS software in the near future in order to create path axes. As a result, the following suggestions were made:

"It is for these reasons that field-line transformations are needed to be automated. As a result of the study, it is planned to apply the triangulation method which is evaluated as the best in terms of applicability with macro support in a GIS software in the near future in order to create path axes "[8].

Delaunay Triangulation, which has an important place in analysis methods in geographical information systems, is a preferred method of triangulation in terms of definite and single outcome.

VI. Conclusion

The size of the data to be analyzed is the factors that influence the application of the method of factors such as the number of reference points. It is important that a stable algorithm is used for the calculation-based software to work properly. The fact that Delaunay triangulation is the only result can be seen as the reason why the preference is high in this respect. However, it cannot be expected that the software will be successful in practice alone. Although the equipment has an important place here, it is expected that it may be insufficient depending on the size of the data to be processed.

Since the earth has an irregular surface, data entry at every point may be correct in terms of the health of the measurement, but it is not possible to apply it in practice. It is possible to perform modeling by making measurements at more points and producing intermediate interpolations as much as possible. When the applications are examined, independent measurements are made and models are made. Depending on the variation of the data processing capacity according to the number of points, the speed of the software can be seen as a disadvantage in terms of achieving healthy results.

In addition to this disadvantage, the more data entry is made, the higher the modeling accuracy. But besides this, data can be a serious problem in storage. If the same area is considered to have more than one network, a large-area modeling of all these networks, such as surface, road, waterline, electricity network, will require high capacity storage space. Apart from this, the data processing speed will also decrease accordingly. In addition to speed, the capacity problem can be considered as a disadvantage.

Although these are disadvantages, Yanalak M. in his work on topological surface modeling suggests that "Delaunay triangulation is the only meaning and that triangles are the closest triangles to the triangles, so it would be a good choice to use these triangles as a base in studies to be done on the network" [2,3].

As a result, "Delaunay Triangulation" alone is not enough. However, within the topological surface modeling methods, we can say that it is the most suitable modeling method in terms of applicability and efficiency.

Acknowledgment

This study was supported by Research Fund of the Karabuk University. Project Number: "KBÜBAP /17 - DR - 437".

References

[1] Uçar D., Bildirici İ.Ö., Uluğtekin N., Coğrafi Bilgi Sistemlerinde Model Genelleştirmesi Kavramı ve Geometri ile İlişkisi, Coğrafi Bilgi Sistemleri ve Jeodezik Ağlar Çalıştayı, SÜ, Konya, 24-26 Eylül 2003, s.94-103.

[2] Yanalak M., Sayısal Arazi Modellerinden Hacim Hesaplarında En Uygun Enterpolasyon Yönteminin Araştırılması, İTÜ FBE, Doktora Tezi, 1997.

[3] Yanalak, M., Yüzey Modellemede Üçgenleme Yöntemleri, Harita Dergisi, Harita Genel Komutanlığı, Sayı:126, 2001.

[4] Watson D.F., Philip G.M., Systematic Triangulations, Computer Vision, Graphics and Image Processing, 26, 217-233, 1984.

[5] Worboys M.F., GIS: A Computing Perspective, Taylor&Francis Ltd, 2000.

[6] Güdükbay U., Sinop A.K., Hesaplamaya Dayalı Geometri (Computational Geometry) (in Turkish). Türkiye Bilişim Ansiklopedisi, pp. 445–448, Papatya Yayıncılık, 2006.

[7] Netcad Yazılım A.Ş (2015). Voronoi, Available: http://portal.netcad.com.tr/display/HELP/Voronoi (Access: march 2018) Bildirici İ.Ö., Selvi H.Z., Model Genelleştirmesinde Geometri Değişimlerinden Alan-Çizgi Dönüşüm Yöntemleri, TMMOB Harita ve Kadastro Mühendisleri Odası 10. Türkiye Harita Bilimsel ve Teknik Kurultayı, Mart 2005, Ankara.

Hand Gesture Recognition with One-Shot-Learning

E. ŞEKER[1] and O. FINDIK[2]

[1] Karabuk University, Karabuk/Turkey, esmaseker@karabuk.edu.tr
[2]Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

***Abstract* – In this paper, one-shot-learning gesture recognition methods are reviewed and an approach of hand gesture recognition using one-shot-learning is proposed. This approach aims to recognize new categories of gestures from a single video clip of each gesture. The gestures are generally related to a particular task, for instance, hand signals used by divers, finger codes to represent numerals, etc. In this study, both RGB and depth images are utilized for a given dataset. A rich dataset, namely the ChaLearn Gesture Dataset (CGD2011), are employed. The dataset is divided into 20 different files which include 940 videos in total. Although training the system with only one example is difficult, depth and RGB images provide many new possibilities. We used the standard deviation of the depth images of a gesture and motion history image (MHI) method. Also, two dimensional fast fourier transform (2D FFT) is used to reduce the effect of camera shift. It is seen that FFT has no distinct effect on the image quality. Then, we compare image templates based on the correlation coefficients and Levenshtein, Mahalanobis, Frobenius distance measures. The Levenshtein distance measure is more suitable to match image templates compared to other distance measures. It is observed that MHI method gives better hand gesture recognition accuracy about one-shot-learning.**

***Keywords* – One-shot-learning, Gesture Recognition, ChaLearn Dataset.**

I. INTRODUCTION

Various approaches to human computer interaction (HCI) have been proposed in the last few decades. HCI is an interdisciplinary field of study that deals with the design, development, evaluation and implementation of interactive technologies which uses diversified interfaces [1,2]. These interfaces can interact with the software as well as with various hardware components. Motion-based operations allow the interaction to be performed directly by body movements without being tied to any device.

Human gestures are very important to use real-time machines wisely and effortlessly. But, due to constraints, it is determined that a single method is not sufficient for the gesture recognition. In that case, existing methods are separated into particular categories. For example, a space-time template approach uses a motion energy image (MEI) and a motion history image (MHI) to preserve space and time information of motion [3,4]. The model-based approach employs the 2D and 3D models of the body to define the limbs of the human body. This approach is significantly based on the motion estimation [5,6].

Hidden Markov Model (HMM) is widely accepted approach for gesture recognition. There are various similarities between recognition of gesture and voice. Therefore, the HMM approach used for voice recognition is also used for gesture recognition. The gestures have characteristic meanings like voices, so that the gestures can vary according to space, time, and social factors. Finally, regularities in gesture performance are similar to syntactic rules of voice recognition. Consequently, such approaches require a large amount of data to identify all the gestures and a dataset has to contain related information, e. g. the number of samples, the input field, and the number of gestures [7,8,9,10].

The conditional distance approach gives the distance between two gestures. In this approach, the level building algorithm, also known as the segmentation and recognition matching algorithm, is used to affect multiple connected gestures. The level building algorithm has dynamic programming implementation. This approach depends on a distance function that compares two gestures. Hence, the conditional distance function is employed in the level building algorithm. The experimental results of the approach show 82% success over the 179 the classes in the dataset. It also displays that these results are comparable to the most advanced methods in the last few decades [11,12,13].

In this paper, a new approach is proposed to separate different gestures from a single video. Gestures are represented by image templates that can be used to compare and match gestures. Depth images are used in these image templates. The gestures do not require prior knowledge of any motion estimation or impression. The ChaLearn dataset (CGD2011) is divided into training dataset and test dataset. Although there is only one gesture in a video within the training dataset, more than one can be found in a video within the test dataset. Therefore, the test phase is separated into several steps. Firstly, different gestures within the video are separated and feature vector of each gesture is obtained. Then, existing feature vector is compared to all the feature vectors in the training dataset and three types of distance measures, i.e. Levenshtein, Mahalanobis, Frobenius, are used to find the best match.

II. PROPOSED APPROACH

Many methods used for gesture recognition have difficulties to identify new gestures. In addition, these methods require a lot of training dataset [14]. The proposed method allows us to recognize new gestures and it provides training with an only

example using RGB and depth images. The gestures used in the method represent particular tasks. For example, marks for deaf, hand marks used by divers, hand marks used for guide vehicles and aircrafts.

A. Gesture Recognition Dataset

Nowadays, there are many gesture recognition datasets, such as American Sign Language (ASL) dataset, Naval Air Training and Operating Procedures Standardization (NATOPS) aircraft handling signals database, Keck gesture dataset, Cambridge Hand Gesture (CHG) dataset, etc. [15]. ASL is a large dataset with many different mark videos. This dataset is also an important source for sign language recognition and human activity analysis. It was implemented using a test dataset of 206 video sequences belonging to 108 different expressions and a training dataset of 999 video sequences belonging to 992 different expressions [16].

NATOPS aircraft handling signals database is a database used to track body and hand gestures together. It uses a stereo camera to collect 3D images. Unlike previous gesture databases, this database requires knowledge about body gestures and hand gestures to distinguish. 24 body and hand gestures are employed. These gestures are repeated 20 times and 400 samples are generated with 2.34 seconds each. The videos are recorded in a closed environment with patience lighting [17].

The Keck gesture dataset is employed to recognize military gestures in difficult conditions. It consists of 14 different gesture classes. The images are collected using a fixed and color camera for this dataset. Each gesture is performed by 3 people, so that all gestures in the dataset are repeated 3 times. As a result, the training, set contains 126 videos sequences, while there are 168 video sequences in the test set. The videos in this set are taken from a moving camera. In addition, these videos have background complexity [18]. The CHG dataset is employed for hand segmentation and gesture recognition in various lighting conditions. It only contains image sequences. There are 900 sequences of 9 hand gesture classes in the dataset. All the sequences are recorded with a fixed camera. The images are obtained by moving the hands right or left on the table with different poses. The entire dataset is divided into five groups containing image sequences taken under different illumination conditions [19].

The data in the ChaLearn dataset is recorded with a Microsoft Kinect™ camera providing both a RGB image and depth image obtained with an infrared sensor. This dataset focuses on the upper body gestures of a user looking at the camera. Thus, it offers an unprecedented large dataset on arms and hands. Unlike other datasets, the dataset provides the opportunity to conduct research from the semiotic, linguistic, sociological, psychological and aesthetic points of view. It completes the sign language recognition datasets by accessing more variety of gesture types. The spatial resolution of the camera limits depth information to resolve correctly finger poses. But the RGB model provides sufficient resolution for such studies. Therefore, this dataset has both a normal video set and a 3D video set [20].

The ChaLearn dataset is divided into 9 categories. These categories are (1) activity (like writing, answering the phone), (2) pantomime (gesture performed to mimic actions), (3) dance, (4) body language, (5) gesticulations, (6) illustrators (like Korean gestures), (7) emblems (like symbolic or political gestures), (8) sign language, (9) signals (like diving signals) [20,21]. In the proposed approach, ChaLearn dataset is used. Some depth frames of the training video samples of the dataset are shown in *Fig. 1*. There are 20 files in the dataset with 47 rgb and 47 depth videos in each file. The training videos are 10 rgb videos and 10 depth videos in a file. The remaining videos in the file are test videos. We employed only depth videos in our study.

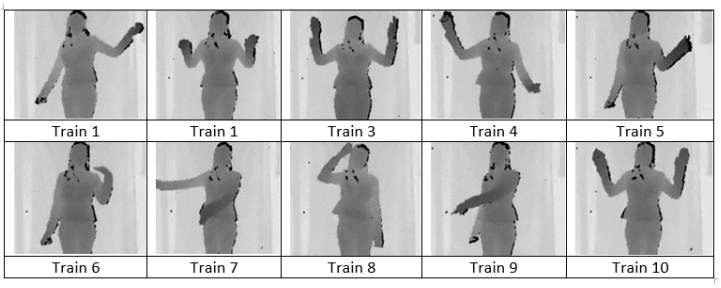


Figure 1: Depth frames from the training dataset in the first file.

B. Feature Extraction

We convert the depth data in the dataset to grayscale which presents two advantages. Firstly, the grayscale gives the pixel density information in proportion to the distance of the object to the camera. Secondly, it provides to use the grayscale threshold on each pixel of the image to separate the foreground from the background. The Otsu method is used for the threshold level. It is a method of determining a threshold that converts a grayscale image to a binary image. While this method is used, it is assumed that the image is composed of two color classes, foreground and background as shown in *Fig. 2*. Then, the intraclass variance value of these color classes is calculated for all threshold values. When the intraclass variance value is the minimum, the variance value between the classes becomes maximum. The maximum variance value between the classes ensures fast and accurate results [22,23].

The proposed method focuses on four operations. These operations are (1) standard deviation (STD), (2) motion history image (MHI), (3) 2 dimensional fast fourier transform (2D FFT), (4) some distance measurements to calculate the distance between two images (Levenshtein, Mahalanobis, Frobenius).

The videos in the dataset are divided into frames. Each of these frames consists of 240x320 matrices. The 4 steps are required to find the STD in a video. These steps are as follows: (1) All frames in the video are summed up and divided by the total number of frames, so the average matrix is found. (2) Each frame in the video is subtracted from the average matrix. (3) The resulting matrices are squared and summed up. (4) The result is divided by the total number of frames in the video. Thus, the gesture of a person in the video can be monitored by the STD. *Figure 3* shows the STD of some videos in the training dataset. For the i^{th} gesture at a video consisting of N frames, the STD σ_i^2 of pixel (x,y) across the frames is given by,

$$\sigma_i^2 = \frac{\sum(I_{xy} - \overline{I_{xy}})}{N} \quad (1)$$

Here I_{xy} is the pixel value of the location *(x,y)* of a frame in a video, where $x = 1,2,3,...,m$ and $y = 1,2,3,...,n$ and N is total frames in the video. $\overline{I_{xy}}$ is average matrix.

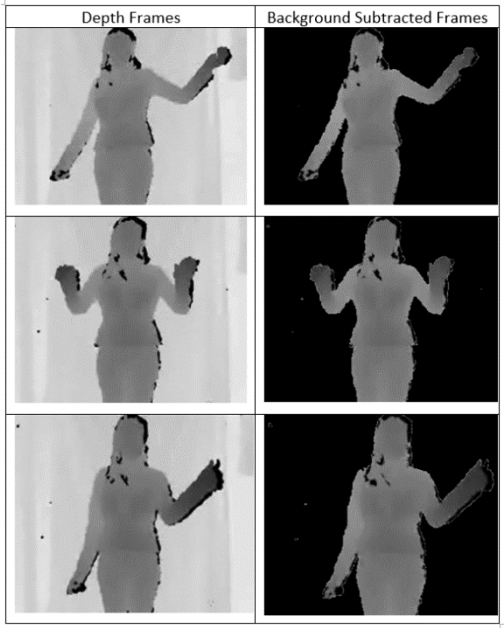


Figure 2: Otsu threshold used to depth frames.

Since the STD does not indicate the flow direction of the gesture as shown in *Fig. 3*, the temporary position information of the gesture disappears. For this reason, we use MHI in the proposed approach with two main advantages. Firstly, the method is not sensitive to noise like missing parts, shadows, etc. Thereby, the dominant gesture information is preserved. It can also be applied with cheap cameras and low power CPUs in the low light areas. As shown in *Fig. 4*, unlike the STD method, it preserves gesture direction information and is used to distinguish directions (e.g. moving hand from up to down and from down to up). The MHI $H_T(x,y,t)$ is computed from an update function $\Psi(x,y,z)$, as given in [24,25]

$$H_T(x,y,t) = \begin{cases} T, & \Psi(x,y,z) = 1 \\ max(0, H_T(x,y,t) - \delta), & otherwise \end{cases} \quad (2)$$

where *(x,y,t)* is the spatial coordinates *(x,y)* of an image pixel at time t (in terms of image frame number), the duration T determines the temporal extent of the gesture in terms of frames, and δ is the decay parameter. $\Psi(x,y,z)$ is defined as

$$\Psi(x,y,z) = \begin{cases} 1, & D(x,y,t) \\ 0, & otherwise \end{cases} \quad (3)$$

where $D(x,y,t)$ is a binary image comprising pixel intensity differences of frames separated by temporal distance Δ, i.e.,

$$D(x,y,z) = | I(x,y,t) - I(x,y,t \pm \Delta)| \quad (4)$$

and $I(x,y,t)$ is the intensity value of pixel with coordinates *(x,y)* at the t^{th} frame of the image sequence.

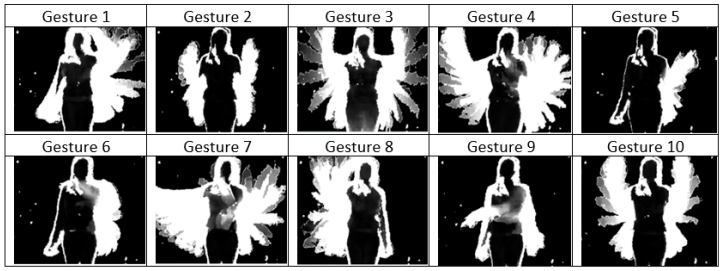


Figure 3: STD applied videos in the training dataset.

A gesture in dataset always returns to its initial position, but the subjects of the frames may shift with greatly affecting the classification results. For this reason, 2D FFT is used before the STD method and after the MHI method. In the study, the camera can be fixed in the training samples but not in the test samples. This procedure makes it difficult to match training and test samples. The 2D FFT transforms the depth data into the spectral domain and suppresses the time resolution before using the STD method. Thereby, the effect of the cameras is reduced and the matching between training and test samples becomes easier. For a frame of size $M \times N$ the 2D FFT is given by,

$$F_{uv}(t) = \frac{1}{MxN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I_{xy}(t) e^{-2i\pi((ux/M)+(vy/N))} \quad (5)$$

where $I_{xy}(t)$ is the value of the *(x,y)* pixel of the image given in the spatial domain for the frame at time t where $t = 1,2,...,T$. The exponential term is the basis function corresponding to each point $F_{uv}(t)$ in the Fourier space [26]. The feature obtained by this process provides theoretically more accurate gesture recognition.

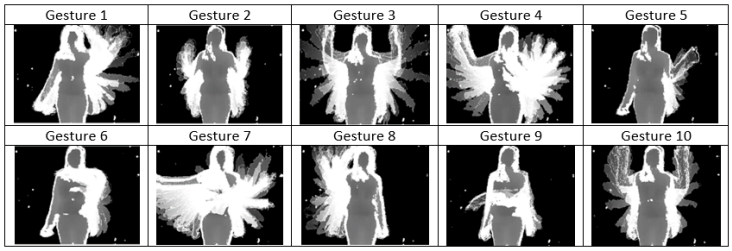


Figure 4: MHI applied videos in the training dataset.

III. EXPERIMENTAL RESULTS

There are 9 categories in the dataset as mentioned in section *A*. We particularly studied hand and arms gestures. Depth videos were used in the approach. There are 20 files in the dataset with 47 videos in each file which consists of 10 training videos and 37 test videos. Each video in the training dataset focuses on only one gesture. The gesture always returns to its initial position. More than one gesture can be found in a video in the test dataset. For this reason, different gestures must be separated from each other as a first step. Fortunately, as shown

in *Fig. 5*, all gestures in a video return to initial position. This makes it easier to separate the different gestures within the video. The separating different gestures within a test video is given by,

$$\begin{aligned} P_1 &= \left(I_{11}(t) - I_{11}(1)\right)^2 + \left(I_{12}(t) - I_{12}(1)\right)^2 \\ &\quad + \left(I_{13}(t) - I_{13}(1)\right)^2 + \cdots \\ &\quad + \left(I_{1n}(t) - I_{1n}(1)\right)^2 \\ P_2 &= \left(I_{21}(t) - I_{21}(1)\right)^2 + \left(I_{22}(t) - I_{22}(1)\right)^2 \\ &\quad + \left(I_{23}(t) - I_{23}(1)\right)^2 + \cdots \\ &\quad + \left(I_{2n}(t) - I_{2n}(1)\right)^2 \end{aligned}$$

… … …

$$\begin{aligned} P_m &= \left(I_{m1}(t) - I_{m1}(1)\right)^2 + \left(I_{m2}(t) - I_{m2}(1)\right)^2 \\ &\quad + \left(I_{m3}(t) - I_{m3}(1)\right)^2 + \cdots \\ &\quad + \left(I_{mn}(t) - I_{mn}(1)\right)^2 \end{aligned}$$

(6)

$$P_s = \sum_{k=1}^{m} P_k \qquad (7)$$

where $I_{xy}(t)$ is the pixel value of position (x,y) at time t. $t = 1,2,3,..,T$. P_s is sum of the square differences between the current frame and the reference frame. After the different gestures in the test videos are separated, the gestures in the test and training videos are compared and the similar gestures are matched. This comparison has been performed with some distance measurements.

Three different distance measurements were used in the approach. These are Levenshtein, Mahalanobis, and Frobenius. The Levenshtein distance measurement gives the most accurate match in the comparison between the gestures in the test and training videos. It is used to rank the similarity between two sequences. It basically performs operations such as insertion, deletion, or substitution between entities such as two arrays, two words, two sentences [27,28]. In this study, this distance measurement method is used to compare two images. If the image from the training dataset is V and the image from the test dataset is Y, then L (V, Y) expression becomes minimum, so the images are similar.

In the study, 9 feature sets were created using three basic operations. Thereby, 9 methods are proposed. These methods are: (1) Otsu threshold was applied to the images in the dataset. Different gestures are separated in the video within the test dataset and the training and test samples are compared only with the STD. (2) The background was subtracted from the images in the dataset, so that we focus only on the subjects in the images. The action boundaries were determined to separate the images in the test dataset. Lastly, the MHI was applied to compare the training samples with the test samples. (3) The shifting of the camera can cause to wrong classification. For this reason, 2D FFT was applied before comparing the gestures in the training and test videos with STD. (4) After using the MHI, 2D FFT was applied to prevent wrong classification. (5) The Levenshtein distance measurement was used in conjunction with the third method to match gestures. (6) The Levenshtein distance measurement was also used with the fourth method to match gestures. (7) Three basic operations, which are known as 2D FFT, STD, MHI, have been used together to make a more accurate comparison. The Levenshtein distance measurement was used to match gestures. In the methods (8) and (9), three basic operations were also used. Unlike the other methods, Mahalanobis and Frobenius distance measurements were used respectively to get more accurate results in matching. The methods used in the approach and percentages of gesture recognition are displayed in Table 1.

Each file in the dataset has different has different gestures. Therefore, the results vary from file to file. Since some files contain finger gestures that are difficult to recognize, the gesture recognition percentage of these files are relatively low. Also, some gestures can be in different positions relative to the body, this causes misclassification. The success of the proposed approach depends on the correct separation of the gestures in the test videos and in any video. Making an expected gesture of the person causes a wrong matching of the gestures.

All methods within the approach have been examined. The most accurate classification was made in method 7. In this method, three basic operations were combined. In addition, 2D FFT was applied twice against camera movements. Unlike other distance measurements, the most successful distance measurement was Levenshtein. Second best method was observed to be method 6 (2D FFT after using MHI) which was not sensitive to noise. Also, it gave motion direction information with preventing incorrect gesture matching.

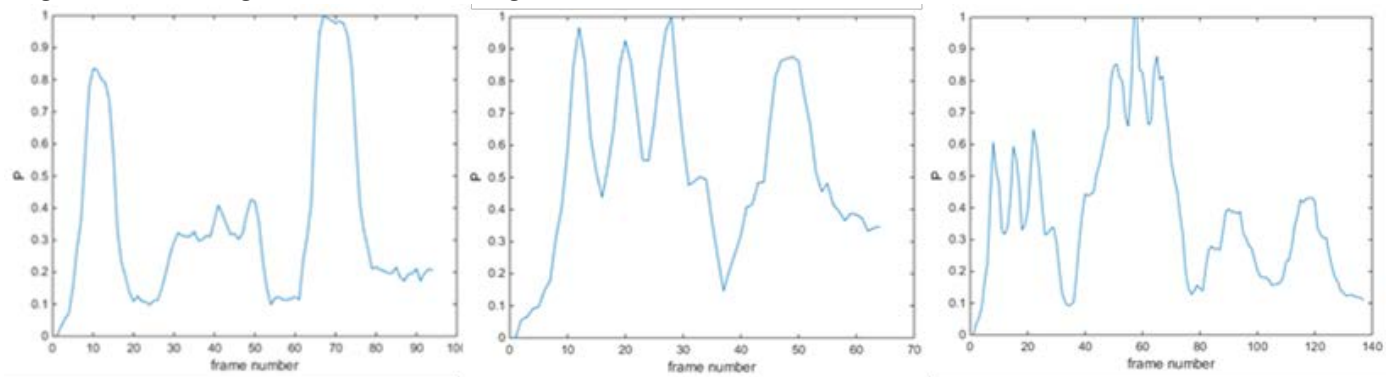


Figure 5: Action boundary for more than one gesture in the videos.

Table 1: Summary of the proposed methods.

Methods	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Depth data	√	√	√	√	√	√	√	√	√
Background subtraction	√	√	√	√	√	√	√	√	√
STD	√								
MHI		√							
2D FFT before using STD			√		√		√	√	√
2D FFT after using MHI				√		√	√	√	√
Intelligent classification	√	√	√	√	√	√	√	√	√
Frobenius distance measurement								√	
Mahalanobis distance measurement									√
Levenshtein distance measurement					√	√	√		
Percentages of gesture recognition	50.22	53.69	51.68	56.64	55.53	63.55	76.15	41.70	52.20

IV. CONCLUSION

The proposed approach aims to recognize hand gestures with one-shot learning. For this study, the ChaLearn Dataset 2011, which is a very rich but complex dataset having a lot of diversity and classification, is used. In the approach, firstly, all the different gestures in the test videos are separated. Secondly, the background is subtracted from the image using Otsu threshold. Lastly, various combinations are applied to gestures with 3 basic operations known as 2D FFT, STD and MHI. Some distance measurements are used to match gestures. The Levenshtein distance measurement gives the most accurate gesture matching result. Intelligent classification is applied for matching gestures in the study. As a result of all these operations, it has been observed that the use of three basic operations together in the recognition of hand gestures with one-shot learning gives more accurate results. There are various difficulties in recognizing finger gestures in the proposed method. Therefore, we aim to improve the recognition of the finger gestures in further studies.

REFERENCES

[1] N. C. Kılıboz, U. Güdükbay, "A hand gesture recognition technique for human-computer interaction", Journal of Visual Communication and Image Representation, vol. 28, pp. 97-104, April 2015.

[2] S. S. Rautaray, A. Agrawal, "Vision based hand gesture recognition for human computer interaction: survey", Artificial Intelligence Review, vol. 43, pp. 1-54, January 2015.

[3] M. A. R. Ahad, J. K. Tan, H. S. Kim, S. Ishikawa, "Human activity recognition: various paradigms", in *Internat. Conf. on Control Automation and Systems (ICCAS) 2008*, pp. 1896-1901.

[4] A. F. Bobick, J. W. Davis. (2001, March). The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Machine Intell. (23)3*. pp. 257-267.

[5] M. A. R. Ahad, (2011). *Computer vision and action recognition: a guide for image processing and computer vision community for action understanding*. Atlantis Ambient and Pervasive Intelligence. Atlantis Press.

[6] H. Zhou, D. Tao, Y. Yuan, X. Li, "Object trajectory clustering via tensor analysis", in *16th IEEE Internat. Conf. on Image Processing (ICIP) 2009*, pp. 1925-1928.

[7] S. Belgacem, C. Chatelain, T. Paquet, "Gesture sequence recognition with one shot learned CRF/HMM", Image and Vision Computing, vol. 61, pp. 12-21, May 2017.

[8] P. Premaratne, S. Yang, P. Vial, Z. Ifthikar, "Centroid tracking based dynamic hand gesture recognition using discrete Hidden Markov Models", Neurocomputing, vol. 228, pp. 79-83, March 2017.

[9] J. Beh, D. K. Han, R. Durasiwami, H. Ko, "Hidden markov model on a unit hypersphere space for gesture trajectory recognition", Pattern Recognition Letters, vol. 36, pp. 144-153, January 2014.

[10] X. Gong, L. Han, J. Wang, M. Ran, "Recognition and simulation of parachute action based on continuous hidden markov model", *Chinese Automation Congress (CAC) 2017*.

[11] R. Krishnan, S. Sarkar, "Conditional distance based matching for one-shot gesture recognition", Pattern Recognition, vol. 48, pp. 1302-1314, April 2015.

[12] P. Hong, M. Turk, T. Huang, "Gesture modeling and recognition using finite state machines", in *IEEE Conference on Automatic Face and Gesture Recognition 2000*, pp. 410-415.

[13] H.-II Suk, B.-Kee Sin, S.-Whan Lee, "Hand gesture recognition based on dynamic Bayesian network framework", Pattern Recognition, vol. 43, pp. 3059-3072, September 2010.

[14] H. Cheng, Z. Dai, Z. Liu, Y. Zhao, "An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition", Pattern Recognition, volç55, pp. 137-147, July 2016.

[15] S. Ruffieux, D. Lalanne, E. Mugellini, O. A. Khaled. (2014). *A survey of datasets for human gesture recognition*. From book: Human-Computer Interaction. Advanced Interaction Modalities and Techniques. vol. 8511, pp. 337-348.

[16] V. Athitsos, C. Neidle S. Sclaroff, J. Nash, "The American Sign Language Lexicon Video Dataset", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008.

[17] Y. Song, D. Demirdjian, R. Davis, "Tracking body and hands for gesture recognition: NATOPS aircraft handling signals database", in *IEEE International Conference on Automatic Face & Gesture Recognition and Workshops (FG2011)*. 2011.

[18] Z. Jiang, Z. Lin, L. S. Davis. (2012, March). Recognizing human actions by learning and matching shape-motion prototype trees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 34.

[19] L. Baraldi, F. Paci, G. Serra, L. Benini, R. Cucchiara, "Gesture recognition using wearable vision sensors to enhance visitors' museum", IEEE Sensors Journal, vol. 15, pp. 2705-2714, March 2015.

[20] J. Wan, S. Z. Li, Y. Zhao, I. Guyon, S. Escalera, "ChaLearn looking at people RGB-D isolated and continuous datasets for gesture recognition", in *IEEE Conference Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2016.

[21] U. Mahbub, H. Imtiaz, T. Roy, M. S. Rahman, M. A. R. Ahad, "A template matching approach of one-shot-learning gesture recognition", Pattern Recognition Letters, vol. 34, pp, 1780-1788, November 2013.

[22] C. Yu, C. Dian-ren, L. Yang, C. Lei, "Otsu's thresholding method based on gray level-gradient two-dimensional histogram", *2nd International Asia*

Conference on Informatics in Control, Automation and Robotics (CAR), April 2010.
[23] P. P. Vijay, N. C. Patil, "Gray scale image segmentation using Otsu thresholding optimal approach", Journal for Research, vol. 02, July 2016.
[24] A. Bobick, J. Davis, "An appearance-based representation of action", in *Proc. 1996 International Conference on Pattern Recognition (ICPR'96)*, vol. 7270, pp. 307-312.
[25] M. A. R. Ahad, J. K. Tan, H. Kim, S. Ishikawa, "Motion history image: its variants and applications", Machine Vision and Applications, vol. 23, pp. 255-281, March 2012.
[26] R. C. Gonzalez, R. E. Woods. (2001). *Digital Image Processing*. Addison-Wesley Longman Publishing Co. Inc.
[27] V. I. Levenshtein. (1965). *Binary codes capable of correcting deletions, insertions, and reversals*. Soviet Physics Doklady. vol. 163, pp. 845-848.
[28] S. Konstantinidis, "Computing the edit distance of a regular language", Information and Computation, vol 205, pp. 1307-1316, September 2007.

Diacritic Restoration of Turkish Tweets with word2vec

Z.OZER[1], I.OZER[2] and O.FINDIK[3]

[1] Karabuk University, Karabuk/Turkey, zeynep.ozer@outlook.com
[2]Karabuk University, Karabuk/Turkey, ilyas.ozer@outlook.com
[3]Karabuk University, Karabuk/Turkey, oguzfindik@karabuk.edu.tr

Abstract **- Social media platforms such as Twitter have grown at a tremendous pace in recent years and have become an important source of data providing information countless field. This situation was of interest to researchers and many studies on machine learning and natural language processing were conducted on social media data. However, the language used in social media contains a very high amount of noisy data than the formal writing language. In this article, we present a study on diacritic restoration which is one of the important difficulties of social media text normalization in order to reduce the noise problem. Diacritic is a set of marks used to change the sound values of letters and is used on many languages besides Turkish. We suggest a 3-step model for this study to overcome the top of the diacritic restoration problem. In the first stage, a candidate word producer produces possible word forms, in the second stage the language validator chooses the correct word forms and at the final word2vec is used to create vector representations of the words and make the most appropriate word choice by using cosine similarities. The proposed method was tested on both synthetic and real data sets, and we achieved a relative error reduction of 37.8% in our data sets compared to the previous study with an average of 94.5% performance.**

Keywords **- Text Mining, Diacritics restoration, Twitter, Tweet Normalization.**

I. INTRODUCTION

In the past few years, the use of social networks such as Twitter and Facebook has grown very rapidly all over the world and researchers have become very interested so that social networks have become a platform where people share their views on every field and the data obtained from social networks began to used in a wide area ranging from traffic anomaly detection [1] to crime estimation [2]. Nevertheless, the spelling language used in social networks varies considerably from the formal ones, and the presence of very high amounts of noisy data has seriously complicate researchers' works. For this reason, efforts to normalize social network data before use have gained momentum recently [3-4-5].

This article focuses on the Diacritic Restoration (DR) problem, which is common in social networks. Diacritic is a set of marks that are added to the letters and change their phonetic, and they are used in many languages such as Turkish, French, Greek, Hungarian and Spanish. The increased of mobile devices usage caused the use of the American Standard Code for Information Interchange (ASCII) equivalents to be widespread due to reasons such as inappropriate keyboard layouts on these devices, and diacritic letters being secondary characters. Turkish contains seven diacritic characters consisting of (ç, ı, l, ğ, ö, ş, ü) also the ASCII equivalents of these characters are (c, i, l, g, o, s, u) respectively, as shown in Table 1. Diacritic Restoration, which is called as deacification or diacritization, is to write correct way the words which is written in partly or completely ASCII form. The main problem here is that the ASCII equivalents of diacritic characters are valid letters, used in the Turkish alphabet. Thus, during diacritic restoration, there are two different possibilities for each of these letters, the ascii and non-ascii form In this case, "n" is the number of diacritic characters at the word, so 2^n different possible surface forms can be written for each word, and in most cases more than one surface form is a legal word used in Turkish. For example, "cok" has 2 diacritic letters, so it has 2^2 surface forms and 3 of these surface forms are correct Turkish words ("çok" - very, "çök" - "cok" - in colloquial language: horn). Therefore, to obtain the correct word after diacritization, it is not enough to just look at the current token. Rather, they should be evaluated together with neighboring words. On the other hand, the vast majority of diacritization studies for Turkish are based either on character-level modeling or on a decision based on the frequencies of possible roots passing through the corpus, and contex-based studies are extremely limited.

In this study, we propose a 3-stage model to solve the problem of DR. In the proposed model, possible surface forms are produced by a candidate word generator first. In the second stage, the morphological analyzer checks whether the surface forms produced are legal Turkish words. We do not prefer the lexicon lookup method at this stage because of Turkish is a very powerful example that the morphologically rich languages, it can produce an infinite number of words by introducing new additions to the roots. Furthermore, if only one surface form is produced at this stage that correct Turkish word, deasciication process is completed. However, if more than one correct Turkish dictionary is produced, it must be determined which one is the most appropriate word. In the last stage, we learn vector representations in high-dimensional vector space using a compilation with word2vec tool, and we choose the most

appropriate word by evaluating the semantic relation of word to neighboring words using cosine similarity. The proposed method provides a fairly high performance compared to the state of the art techniques.

Table 1: List of Turkish diacritic characters and their ASCII counterparts.

Turkish	ç	ğ	ı	ö	ş	ü
	Ç	Ğ	İ	Ö	Ş	Ü
ASCII	c	g	i	o	s	u
	C	G	I	O	S	U

II. RELATED WORKS

In recent years, a very high amount of mistakes has been made in the spelling language used in social networks, depending on the reasons such as mobile device usage and user habits. The use of missing diacritic markers is also one of the largest parts of typing mistakes [5]. It is possible to define the diacritic restoration process as the automatic addition of these marks when the diacritic marks are missing or not at all.

In the study on DR [6], a language independent method based on the letter level learning mechanism, which does not require any additional tagging tools, has been proposed especially for languages with few resources. Similarly, in [7], a method with a learning mechanism at the letter level is presented.

The method they propose is suitable for languages that do not have big dictionaries and generalizations are possible. In [8], a method based on statistical machine translation instead of the letter-level learning mechanism has been used for the restoration of diacritic texts in Algeria. In addition to these studies, DR studies have been conducted for languages such as Croatian [9], Vietnamese [10], Romanian [11] and Arabic [12].

In the first study on Turkish DR, [13] a hidden Markov model with a character base was constructed using a 18 million word assembly. System performance is tested using different n-gram language models and a 4-gram model provides a significant error reduction compared to the 3-gram model.

Unlike many other models, Zemberek [14], which is one of the most commonly used tools in Turkish studies, produces more than one candidate word for each token, ranking these words according to the frequency in the corpus of their roots. In the study of Emacs Turkish Mode [15], which was inspired by the research in [13], decision lists were constructed by using Greedy Prepend Algorithm, which is a kind of decision list algorithm with 1 million words. In the study for Turkish [5], a language independent model was proposed using Conditional Random Fields (CRFs) and language validator, and the proposed model was tested on two separate data sets, synthetic and real. For each test, 97.06% and 95.43 accuracy were achieved, respectively. In the study on social media text normalization [4], [5], the recommended method was used. In this study [16], deasciification performance and the effect on information retrieval were discussed.

III. WORD2VEC

Word2vec is a shallow neural network [17] which input a corpus and output a vector set. Neural networks are used very successfully in many areas [18-19-20-21]. Word2vec predicts words based on context with two different neural models: Continuous bag of words (CBOW) and Skip-Gram.

While the CBOW model predicts the current word based on the context, the Skip-gram model, on the contrary, tries to guess the other words around this word based on the current word. The CBOW model compares the word and output and corrects the representation of the word, depending on the back propagation of the error gradient. In fact, CBOW tries to maximize the following equation (1):

$$\frac{1}{V}\sum_{t=1}^{v}\log p(m_t \mid m_{t-\frac{c}{2}}...m_{t+\frac{c}{2}}) \quad (1)$$

On the other hand, Skip-Gram search the prediction of the context given a word and it tries to maximize the following equation (2):

$$\frac{1}{V}\sum_{t=1}^{V}\sum_{j=t-c,\, j\neq t}^{t+c}\log p(m_j \mid m_t) \quad (2)$$

Once the feature vector for each word is obtained, the similarity between the two words is calculated using cosine similarity. Let $a(x_1, y_1)$ and $b(x_2, y_2)$ be 2 points given in two-dimensional space, the cosine similarity between these two points can be written as (3):

$$\cos\theta = \cos(a,b) = \frac{a \cdot b}{\|a\|\|b\|} = \frac{x_1x_2 + y_1y_2}{\sqrt{x_1^2 + x_2^2} \times \sqrt{y_1^2 + y_2^2}} \quad (3)$$

On the other hand, if we increase the size, we can show the vectors a and b as $a(a_1, a_2, a_3...a_n)$ and $b(b_1, b_2, b_3...b_n)$. In this case, we can rewrite (3) as follows (4):

$$\cos\theta = \cos(a,b) = \frac{\sum_1^n (a_i \times b_i)}{\sqrt{\sum_1^n a_i^2} \times \sqrt{\sum_1^n b_i^2}} \quad (4)$$

Here, cos Θ is in the range [0,1] and 0 means no semantic relation between two words, 1 means that the word has the same meaning.

IV. THE PROPOSED MODEL

Since the ASCII equivalents of the diacritic characters used in the Turkish alphabet are also legal letters, so more than one valid Turkish word can occur after the diacritic restoration. This creates an ambiguity about which of the legal words produced is the correct word. For this reason, it is not enough to model only the relevant token at the character level, or to look at the roots of the obtained words at the corpus, in order to determine the correct word. On the other hand, most of the diacritic

restoration tools developed for Turkish, in which deasciificaiton by making probabilistic modeling at character level.

If we will evaluate this problem on a real tweet on Twitter: There are two diacritic words in the sentence "Hakem bunu sut olarak degerlendirdi". There is only one correct word in the form of "değerlendirdi" (it is regarded) for the "degerlendirdi" token and there are 2 valid words which is "süt"(milk) and "şut" (shot) for the "sut" token. In this case, there are two possible sentences can be written:

Hakem bunu süt olarak değerlendirdi.
(The referee regarded it as a milk.)
Hakem bunu şut olarak değerlendirdi.
(The referee regarded it as a shot.)

The above sentence is obviously a "şut" of the correct when assessed semantically. In the our proposed model, firstly probabilistic modeling at the character level is making and if value above the threshold, is send to the LV layer. At the LV level, morphological analyzes of candidate words are made to determine whether they are legal Turkish words. Using a lexicon lookup method for languages that are quite rich in terms of morphology, such as Turkish, is not very efficient due to the size of possible surface forms, so morphological analyzer as LV is preferred in this study. If there is only one valid word obtained after morphological analysis, the diacritic restoration process is completed.

On the other hand, if more than one valid Turkish word is obtained, it is necessary to make a semantic evaluation. For this purpose, by using a corpus, the feature vector of each word in the corpus is determined by the word2vec tool. After this step, the harmony of the word with the other words is calculated using the cosine similarity. Here, "a" is the candidate word and "b" is the other words in the related sentence, "v" is the number of the other words and "p" total similarity rate, cosine similarity is calculated as (5):

$$p = \sum_{t=1}^{p} \cos(a, b_t) = \frac{\sum_1^n (a_i \times b_{ti})}{\sqrt{\sum_1^n a_1^2} \times \sqrt{\sum_1^n b_{ti}^2}} \quad (5)$$

For each candidate word, the total similarity value, which is the other words, is calculated and the word with the highest similarity value is considered as the result of the diacritic restoration process. Figure 1 shows the diacritic restoration process in detail.

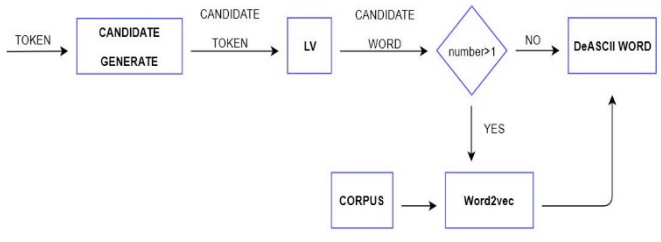


Figure 1: the diacritic restoration process

V. Experimental Setup and Results

In this section, we provide information on the data sets and evaluation methodology that we used first. We are discussing the results we achieved later.

A. Dataset & Experimental Evaluation

A synthetic data set was created primarily for the diacritization task. Synthetic dataset creation is a simple task, unlike the diacritization process. All diaritic characters in the word are replaced by ASCII equivalents to form the synthetic data set. The most important feature of the data set we create is to select words that constitute more than one valid Turkish word after diacritic restoration in order to observe the ambiguity problem. In this way, the performance of the semantic evaluation process of words can be better observed. For this purpose, using Twitter REST API, we collected 6,146,742 tweets between 10.08.2016 and 25.04.2017 in Ankara city limits. We choose randomly 1000 tweets from these data to use synthetic data sets in accordance with the above-mentione rules. On the other hand, we used tweets that were directly cast in ASCII form for our actual data set. Similarly, there are 1000 tweets in our data set. Another important issue related to the data set is the need for a large corpus so that the words can be evaluated semantically. For this purpose, it was used as a corpus of the Hürriyet newspaper archives [22] and The Turkish Wikipedia dump (TRWiki-20150121-pages-meta-current) data set also used in the study in [16]. This data set has a total of 129,288,786 words and 53,375,687 of them contain at least one diacritic character. No matter how large the corpus is, all surface forms may not be found in the corpus because of the presence of too many surface forms of each word in jointed language such as Turkish. In this case, when calculating the cosine similarity, we calculate the similarity based on the longest root of the related word. In addition, due to the fact that there are too many misspelling words or social media specific words in social networks, we check words with LV, which is the same sentence in diacritic ones and we only make similarity calculations with correct Turkish words.

In order to compare the performance of the proposed model, similar to the studies in [5-16], we based on the ratio between the number of fully correct converted words and the total number of words (6).

$$Acc = \frac{\#ofcorr \cdot diacritizedwords}{\#ofwords} \quad (6)$$

B. Results and discussion

In this section, firstly we provide information about our system performance by testing the synthetic and real data sets that we have created, with the previous DR tools and comparing the results we obtained. In this context, we compare our system performance with those of [5-14-15]. It is included in a tool [23] of the study in [5]. We also compare the results from Zemberek 2 since the Zemberek library does not include the deascii tool in its final version.

As seen in Table 2, we obtained the best score skip-gram

model with 94.2% and 94.8% for both real and synthetic datasets, respectively. For the skip-gram model, an average of 94.5% performance was achieved in both sets of data. In addition, in [5] and [15] achieved similar results for tools, while Zemberek's performance was somewhat lower in both sets of data than for other tools.

Table 2: DR evaluation results with previous work.

System	Synthetic	Real
Adalı	86,6%	82,0%
Yüret	88,3%	82,6%
Zemeberek	76,8%	75,3%
Cbow	94,1%	93,3%
Skip-gram	94,8%	94,2%

We also tested four different models to determine the relationship between the restored diacritic word and the other words in the same sentence.

We first applied the cbow and skip-gram model using the diacritic word and the first word that comes after it and as seen in Table 3 we obtained 85.7% and 84.3%, 86.2% and 85.9% results for synthetic and real data set for cbow, respectively, and skip-gram for models using single words. These results provide similar results with previous studies. On the other hand, when we evaluate the relation between all other words in the related sentence, the results increased markedly, reaching 94.1% and 93.3% for cbow and 94.8% and 94.2% for skip-gram, respectively.

Table 3: DR evaluation results with previous work

	Sentetik	Gerçek
$Cbow_{one}$	85,7%	84,3%
$Cbow_{all}$	94,1%	93,3%
$Skip-gram_{one}$	86,2%	85,9%
$Skip-gram_{all}$	94,8%	94,2%

VI. CONCLUSION

In this article, we proposed a DR model based on Word2vec and tested the proposed model with both the previous work and the whole sentence and one word adjacent to the restored word. In synthetic and real data sets, we achieved the best results with a skip-gram model based on the whole sentence. On the other hand, in evaluations based on one word, we obtained similar results to other studies in the literature. In the entire sentence model of the Skip-gram, we achieved a relative error reduction of 37.8% in our data sets compared to the previous study with an average of 94.5% performance. However, in the model we propose, the collection is quite foreground. The development of the collection will affect the performance of the system positively.

REFERENCES

[1] Giridhar, P., Amin, M. T., Abdelzaher, T., Wang, D., Kaplan, L., George, J., & Ganti, R. (2016). ClariSense+: An enhanced traffic anomaly explanation service using social network feeds. Pervasive and Mobile Computing, 33, 140-155.

[2] Wang, X., Gerber, M. S., & Brown, D. E. (2012, April). Automatic crime prediction using events extracted from twitter posts. In International conference on social computing, behavioral-cultural modeling, and prediction (pp. 231-238). Springer, Berlin, Heidelberg.

[3] Saloot, M. A., Idris, N., & Mahmud, R. (2014). An architecture for Malay Tweet normalization. Information Processing & Management, 50(5), 621-633.

[4] ERYİĞİT, G., & TORUNOĞLU-SELAMET, D. İ. L. A. R. A. (2017). Social media text normalization for Turkish. Natural Language Engineering, 23(6), 835-875.

[5] Adali, K., & Eryiğit, G. (2014). Vowel and diacritic restoration for social media texts. In Proceedings of the 5th Workshop on Language Analysis for Social Media (LASM) (pp. 53-61).

[6] Mihalcea, R., & Nastase, V. (2002, August). Letter level learning for language independent diacritics restoration. In proceedings of the 6th conference on Natural language learning-Volume 20 (pp. 1-7). Association for Computational Linguistics.

[7] Mihalcea, R. F. (2002, February). Diacritics restoration: Learning from letters versus learning from words. In International Conference on Intelligent Text Processing and Computational Linguistics (pp. 339-348). Springer, Berlin, Heidelberg.

[8] Harrat, S., Abbas, M., Meftouh, K., & Smaili, K. (2013, August). Diacritics restoration for Arabic dialect texts. In INTERSPEECH(pp. 1429-1433).

[9] Šantić, N., Šnajder, J., & Bašić, B. D. (2009). Automatic diacritics restoration in croatian texts. INFuture2009: Digital Resources and Knowledge Sharing, 309-318.

[10] Do, T. N. D., Nguyen, D. B., Mac, D. K., & Tran, D. D. (2013, August). Machine translation approach for vietnamese diacritic restoration. In Asian Language Processing (IALP), 2013 International Conference on (pp. 103-106). IEEE.

[11] Grozea, C. (2012, September). Experiments and results with diacritics restoration in Romanian. In International Conference on Text, Speech and Dialogue (pp. 199-206). Springer, Berlin, Heidelberg.

[12] Azmi, A. M., & Almajed, R. S. (2015). A survey of automatic Arabic diacritization techniques. Natural Language Engineering, 21(3), 477-495.

[13] Tür, 2000 Tür G. A statistical information extraction system for Turkish, Bilkent University (2000) Ph.D. thesis

[14] Akın, A. A., & Akın, M. D. (2007). Zemberek, an open source nlp framework for turkic languages. Structure, 10, 1-5.

[15] Yuret, D., & De La Maza, M. (2006, November). The greedy prepend algorithm for decision list induction. In International Symposium on Computer and Information Sciences (pp. 37-46). Springer, Berlin, Heidelberg.

[16] Arslan, A. (2016). DeASCIIfication approach to handle diacritics in Turkish information retrieval. Information Processing & Management, 52(2), 326-339.

[17] Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR, arXiv; 2013. p. 1301-3781.

[18] Turkson, Richard Fiifi, et al. "Artificial neural network applications in the calibration of spark-ignition engines: An overview." Engineering science and technology, an international journal19.3 (2016): 1346-1359.

[19] Eichie, J. O., Oyedum, O. D., Ajewole, M. O., & Aibinu, A. M. (2017). Artificial Neural Network model for the determination of GSM Rxlevel from atmospheric parameters. Engineering Science and Technology, an International Journal, 20(2), 795-804.

[20] Ozer, I., Ozer, Z., & Findik, O. (2018). Noise robust sound event classification with convolutional neural network. Neurocomputing, 272, 505-512.

[21] [21] Ozer, I., Ozer, Z., & Findik, O. (2017). Lanczos kernel based spectrogram image features for sound classification. Procedia Computer Science, 111, 137-144.

[22] http://gurmezin.com/derlemtr-projesi/

[23] Eryiğit, G. (2014). ITU Turkish NLP web service. In Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (pp. 1-4).

Performance Evaluation of Bicycle Sharing System in Urban Transportation

Ö. BATTAL[1] and Z. ÖZER[2]

[1] Karabuk University, Karabuk/Turkey, ozlembattal@karabuk.edu.tr
[2]Karabuk University, Karabuk/Turkey, zeynepbayram@karabuk.edu.tr

Abstract - With the increase in population density in cities, the use of motor vehicles is increasing. In addition, traffic congestion, causes some issues such as air pollution noise pollution that affect both human health and urban life negatively. For the elimination of the disadvantageous of the motor vehicle usage people have to be encouraged to transport with non-motorized vehicle s such as bicycle. In this study, views about cycling were taken by online survey for the purpose of popularize the usage of bicycle in Karabuk University Campus and campus transportation.
According to the obtained data, 30% of the participants were female and 70% were male and they have the biggest share in the age distribution, with 80% and 15-25 age range. Looking at the educational status, 78% of them gave undergraduate education and 13% gave a postgraduate education. 15% of the participants were working and 85% were students. In the scope of the study, bicycle usage of the people was examined. Accordingly, 43% of the participants have a bicycle. It is also seen that bicycle use is 61% when it is not bicycle owner. 84%, the highest rate in the questionnaire assessed participants of the purpose of using the bike has to have a positive effect on health. Among the factors limiting the use of bicycles, the rate of 71%, which was the highest rate, was the lack of bicycle routes and 65% was shared with motor vehicles.

I. INTRODUCTION

Survey results obtained in order to determine the relationship between user responses and free bicycle usage were tested with the state of the art classification algorithms using a 10-fold cross validation in the Matlab environment. Here, we have prepared a data set consisting of a total of 46 features and 1 output, including free bicycle usage output and other survey question feature vectors. As shown in Table 1, the highest classification performance was obtained with Quadratic SVM with 93%. On the other hand, since the size of our data set is still at a limited level, it is seen that other classification algorithms give very close results. However, when classification performance is evaluated, it is seen that there is a correlation between feature vector and free bicycle use preference. Similarly, using Quadratic SVM, we achieved a 71% performance when we tested the classification performance of gender output and other questionnaire feature vectors.

Table 1: The results of various classification algorithms.

Classifier Type	Prediction Speed	Memory Usage	Interpretability	Accuracy
Decision Tree	Fast	Small	Easy	90.0%
Linear SVM	Binary: Fast Multiclass: Medium	Medium	Easy	91.0%
Quadratic SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	93.0%
Cubic SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	91.0%
Fine Gaussian SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	92.0%
Medium Gaussian SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	91.0%
Coarse Gaussian SVM	Binary: Fast Multiclass: Slow	Binary: Medium Multiclass: Large	Hard	91.0%

This study reflects the thoughts about the use of bicycles, it is intended to shed light on the present and future transportation planning. Bicycle routes and safe parking areas should be organized by the local government in order to spread the use of bicycles and transportation. It is also important and necessary that individuals be motivated to use public transport vehicles and / or bicycles, to integrate public transport with bicycles, if possible, to encourage students to use bicycles, especially from the primary school level. In addition, drivers involved in urban traffic, the greatest risk for the spread of bicycle culture, should be made aware of bicycle users.

REFERENCES

[1] Mert, K. ve Öcalir E.V. (2010) Konya'da Bisiklet Ulaşımı: Planlama ve Uygulama Süreçlerinin Karşılaştırılması, METU JFA.
[2] Öncü E. Çakan C. Candan S. (2003) Trafik Güvenliğinin ve Kullanımının Artırılması için Bisikletlilere Yönelik Düzenlemeler: Konya Bisiklet Planı.
[3] Özalp, M., Öcalir, E.V. (2008) Türkiye'deki Kentiçi Ulaşım Planlaması Çalışmalarının Değerlendirilmesi, ODTÜ Mimarlık Fakültesi Dergisi, c: 25, n: 2; 71-97.
[4] Planners Collaborative INC (2008) Massachusetts Bicycle Plan EOT (Massachusetts Executive Office of Transportation), USA.
[5] *Akay, A. (2006) Ulaşımda Bisikletin Yeri ve Ankara Bilkent Koridorunda Bisiklet Yolu Önerisi, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Ankara; 19-20.*

Parallel Artificial Atom Algorithm for Large Scale Travelling Salesman Problem

A. N. ALTINTAŞ TANKÜL[1] and B. SELÇUK[2]

[1] Karabuk University, Karabuk/Turkey, aysenuraltintas@karabuk.edu.tr
[1]Karabuk University, Karabuk/Turkey, bselcuk@karabuk.edu.tr

Abstract **- Optimization algorithms yield acceptable results in the shortest time, even if they cannot always guarantee the best end result in the given problem. There are classical mathematical methods and meta-heuristic methods that have become very popular for solving optimization problems. Meta-heuristic algorithms can be categorized in many type such as physics based, social based, biological based, chemistry based, sport based, swarm based, mathematics based and also hybrid based. In this study, the Artificial Atom Algorithm (A^3) is applied in parallel to solve the Traveling Salesman Problem (TSP). A^3 is chemistry-based technique that is improved by inspired the compounding process of atoms and the application of parallel A^3 is particularly easy and promises significant gains in performance especially for large scale TSP. TSP is one of the route planning problems that finds the lowest cost path of visiting all the cities on the giving map and returns to starting point, it was aimed to plan the best route. The performance of algorithm in terms of the city number, the route distance and the calculation time of this route will be examined. An interface will be designed to implement the application and observe the experimental results.**

Keywords **- Artificial Atom Algorithm. Traveling Salesman Problem. Ant Colony Optimization. Optimization Algorithm. Parallel Programming**

I. INTRODUCTION

One of the most basic principles in our world is to search for the optima, optimal state. The atoms in physics try to create bonds to minimize the energy of their electrons [1]. The same is true for the optimal biology of survival, which, together with biological evolution, leads to better adaptation of species to their environment [1]. As long as mankind exists, many fields are striving for perfection. He wants to achieve maximum happiness with minimum effort. In our economy, profits and sales should be maximized and costs should be as low as possible. For this reason, optimization is one of the oldest in science, including everyday life.

Optimization in engineering is to achieve the best possible result under given conditions. Most decisions taken in engineering are aimed at minimizing effort and maximizing profits. Effort or gain can often be expressed as a function of some design variables. Hence optimization is the process of finding the conditions that give the maximum or minimum value of a function [2]. Minimization as an example; cost, distance, cross length, weight, processing time, material, energy consumption, number of objects, maximization; profit, value, output, return, efficiency, utility, productivity, capacity, number of objects can be shown.

Generally, optimization algorithms can be divided into two basic classes: deterministic and probabilistic algorithms. Deterministic algorithms are often used if there is a clear relationship between possible solutions and benefits of a given problem [2]. Fixed parameters are used to solve mathematical functions, the data are known in advance and you know exactly what happens when you start the system. If the relationship between a solution candidate and its "conformity" is not so obvious or very complex, or if the dimensions of the search area are too high, resolving a problem deterministically can cause the search field to be numbered extensively, even for relatively small problems.

In this point where the deterministic algorithms are insufficient, probabilistic algorithms are introduced. The probabilistic algorithms, which have become one of the most important research areas in optimization, have a chance component [1]. Defined parameters take random values and are used to solve nonlinear problems [3]. The solution achieves an acceptable result, even if it is not at the best possible level, in a shorter working time. They are sometimes called heuristic and sometimes meta-heuristic. Heuristic used in optimization are functions that help to decide which of the possible solutions will be examined later [1]. Meta- heuristic is a higher-level procedural or heuristic method designed to find, generate or select an heuristic method that can provide a sufficiently good solution to the optimization problem, especially with inadequate or incomplete knowledge or limited computing capacity [4].

There are many different meta-heuristic optimization algorithms. Some of these are Ant Colony Optimization Algorithm(ACO) [5], Particle Swarm Optimization Algorithm [6], Artificial Immune System Algorithm [7], Bacteria Foraging Algorithm [8], Genetic Algorithm [9], League Championship Algorithm [10], Imperialist Competitive Algorithm, Simulated Annealing [11] and Artificial Atom Algorithm [12].

The Traveling Salesman Problem (TSP) is a combinatorial optimization problem that requires an exponential time at N [13] and it is impossible to solve larger problems with classical mathematical techniques so heuristic algorithms are more suitable for this problem. Many of optimization algorithms have been applied to TSP and compared their efficiency.

Although optimization algorithms are faster than deterministic algorithms, as the size of problems increases, the working time also increases. Parallel programming may be preferred at this point instead of serial calculations performed by a single processor. In parallel programming, it is aimed to run the algorithm faster with the help of more than one

processor. Parallel programming has several forms of implementation.

In this study, the A^3 is applied to TSP in parallel and compared with ACO that is found in the literature. Firstly, the description of TSP, A^3 and ACO are given and then the implementation of these algorithms on TSP in parallel is explained.

II. TRAVELLING SALESMAN PROBLEM

The Traveling Salesman Problem (TSP) is a combinatorial optimization problem. In TSP, the salesperson has to visit each of the cities and return to the city where he started. The purpose of the problem is to reduce the total length of the traveling salesperson travel. This problem is known to be difficult by the NP and cannot be solved exactly in polynomial time. As the number of nodes increases, the time spent solving the problem increases exponentially. The exponential increase of the solution time with increasing number of nodes in the TSP is shown in Table 1.

Table 1: Evaluation of Hamilton Cycles (see [14])

Düğüm Sayısı	Döngü Sayısı (n-1)!	Gerekli Zaman
12	39.916.800	0.004 saniye
13	479.001.600	0.05 saniye
14	6.227.020.800	1 saniye
15	87.178.291.200	9 saniye
16	1.307.647.368.000	2 dakika
17	2.1 * 1013	35 dakika
18	3.6 * 1014	10 saat
19	6.4 * 1015	7.5 gün
20	1.2 * 1017	140 gün
21	2.4 * 1018	7.5 yıl
22	5.1 * 1019	160 yıl
23	1.1 * 1021	3.500 yıl
24	2.6 * 1022	82.000 yıl
25	6.2 * 1023	2 milyon yıl

The integer linear programming model of the navigator problem is shown below [15]:

$$
\begin{aligned}
&\min \sum_{i=0}^{n} \sum_{j\neq i, j=0}^{n} c_{ij} x_{ij} \\
&0 \leq x_{ij} \leq 1 && i,j = 0,\dots n \\
&x_{ij} \text{ integer} && i,j = 0,\dots n \\
&\sum_{i=0, i\neq j}^{n} x_{ij} = 1 && j = 0,\dots,n \\
&\sum_{j=0, j\neq i}^{n} x_{ij} = 1 && i = 1,\dots,n \\
&u_i - u_j + n x_{ij} \leq n-1 && 1 \leq i \neq j \leq n.
\end{aligned}
\tag{1}
$$

If the starting city is given in the problem, the number of possible Hamiltonian roads is equal to (n-1) city's displacement, i.e. (n-1)!. So the exact solution would be to try all the permutations and use the brute force search to select the lowest cost way. This method guarantees the best solution for a small number of cities, but not even for 20 cities. Although the solution of the problem is simple, scanning the entire solution space is not a feasible approach to achieve the result [7]. So using heuristic and meta- heuristic techniques to solve TSP problems is a more efficient way.

III. ANT COLONY OPTIMIZATION ALGORITHM

ACO is a meta-heuristic method developed by inspiration from the sense of direction of the ants and logic to find the sources of food. The route that minimizes the distance between the nests of real ants and the points where they collect food is based on the detection of pheromone chemistry they secrete [14]. Real ants identify the route they will follow when they leave their nest to search for food, from the pheromone secreted by previous ants on the road [16]. Adaptive behavior of real ants can be seen in Figure 1. Pheromone is a chemical secreted from the legs of the ants, staying on the road for a certain period of time and evaporating over time.

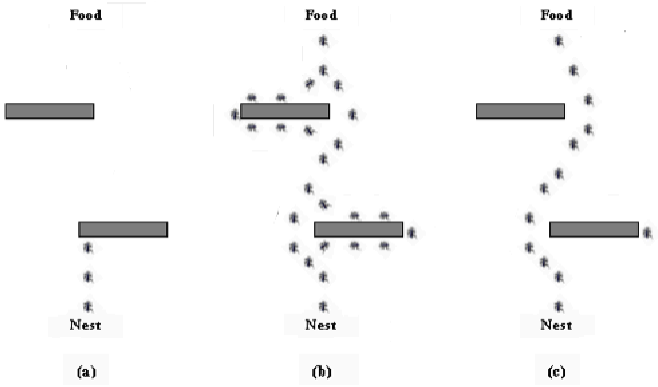


Figure 1: Adaptive behavior of real ants a) Start to search food b) Ants first search food source with equal probability when there are no pheromone on the roads, c) After some time, majority of ants choose shortest path

In the beginning, the ants randomly search for food. The ants that uses shorter path to return in a shorter time leave more pheromone on the road. As the search for food proceeds, the amount of pheromone on short distances will be concentrated, and the rate of pheromone evaporation will decrease in parallel with the shortness of the time. In the same way, the rate of preference of the longer-distance path by the ants due to the lesser initially pheromones on roads decreases and after a while the pheromone on the road is completely evaporated and no ant will use them [14]. This process has been modeled as ACO Algorithm that is shown at Figure 2.

```
procedure ACO
    initialise pheromone trails;
    while (termination condition not satisfied) do
        construct candidate conformations;
        perform local search;
        update pheromone values;
    end
end
```

Figure 2: Basic structure of Ant Colony Optimization Algorithm for combinatorial optimization problems

ACO algorithms are very useful for optimization problems, because they examine numerous solutions at each step of the algorithm. Parallelized ACO algorithm techniques even speed

up the search. For parallelization of ACO there are different approaches. These approaches can be categorized in 3 different titles [17].

1) Parallelization of ant colonies [18]
2) Parallelization of ants [19]
3) Parallelization of evaluation of solution elements

In this paper, parallel ants approach has been applied on TSP. Each ant/slave, are assigned to a separate processor to create the solution. It is the responsibility of the master processor to receive input, to randomly place ants, to update the pheromone, and to generate the communication overhead [17].

IV. ARTIFICIAL ATOM ALGORITHM

A. *Definition of A^3*

The artificial atom algorithm (A^3) that has been improved recently by Karcı [20] is a meta-heuristic algorithm based on chemical compounds based on the process of compound formation of atoms and giving successful results in different optimization problems. The formation process of chemical compounds can be imitated for the optimization of scientific and engineering problems. The ionic bond and covalent bond chemical processes are imitated separately from each other. These two processes constitute the basic two operations of the algorithm (Figure 4). The problem can be thought of as an atom, each parameter will be considered as an electron, Figure 3 shows a representation of it.

Figure 3: Representation of Atom, E is an electron

Every atom is a random solution of the problem at the beginning. The process starts with more than one atom, and these solutions are called the Atom Set. Each of the decision variables is called the electron can be thought as gene, the series of decision variables is the atom as chromosome, and the matrices composed of the atoms are called the atom set as population in Genetic Algorithm.

Create random Atom Set (A_0)
Compute individual effect of each electron for each atom in the current Atom Set.
$i \leftarrow 0$
Do the following until stopping criterion met
  Apply Covalent Bond to A_i.
    // $B \leftarrow CovalentBond(A_i - 1)$
  Apply Ionic Bond to A_i.
    // $A_i + 1 \leftarrow IonicBond(B)$
  Compute effects of electrons in IR.
  Compute objective function value for each atom.
  $i \leftarrow i + 1$

Figure 4: Artificial Atom Algorithm (see [20])

After a random Atom Set is created in the method, the value of each atom's objective function and the effect of each electron on the objective function are calculated. The electron values, which have great influence on the solution, are at the atom's initial indices, the electron values that have less effect on the solution are at the end of the atom.

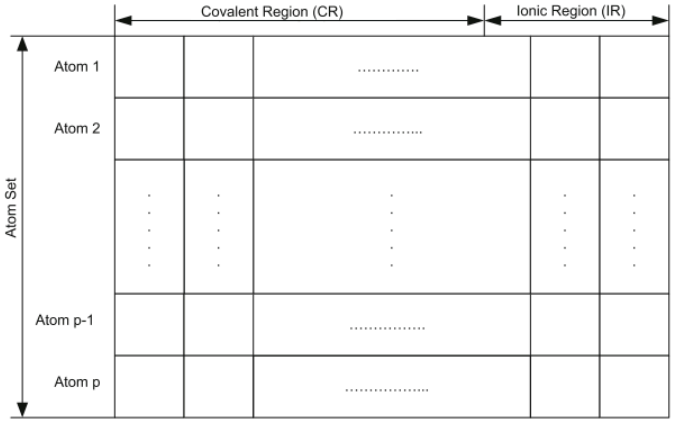


Figure 5: Representation of Atom Set, Covalent Region (CR) and Ionic Region (IR) (see [20])

There are two regions of atom, Covalent Region (CR) and Ionic Region (IR) that is seen in Figure 5. The electron values near the beginning of the atom that is CR are copied between the matched atoms depending on the magnitude of the effect values by the covalent bond operator (Figure 6) [20]. In other words, it is tried to increase the number of parameter values having a positive effect for the solution in the atom set. Figure 7 shows the representation of covalent bond operation.

$k \leftarrow 1,2,\ldots,\beta n$ // $k \leq \beta n$
  if $E[A_j(k)]$ is better than $E[A_r(k)]$
    Copy value of $A_j(k)$ to $A_r(k)$
  else
    Copy value of $A_r(k)$ to $A_j(k)$

Figure 6: Covalent Bond Operation

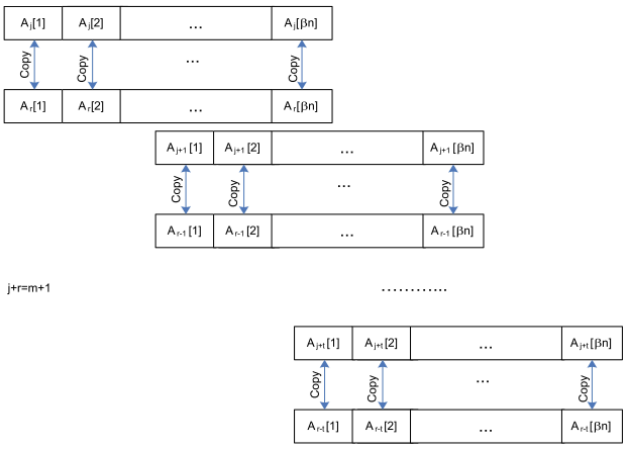


Figure 7: Representation of Covalent Bond Operation (see [20])

Then, the values of the electrons at the end of the atom that is IR are renewed randomly to approximate the solution, while

the ionic bond operator (Figure 8) removes the values with negative effect on the solution from the atom [20].

$$k \leftarrow \beta n + 1, \beta n + 2, \dots, n \;//\; \beta \text{ is covalent rate}, (1-\beta)n = \alpha n$$
$$A_r[k] \leftarrow L_k + \eta * (U_k - L_k)$$
$$//L_k \text{ is lower bound for } k^{th} \text{ attribute}$$
$$//U_k \text{ is upper bound for } k^{th} \text{ attribute}$$

Figure 8: Ionic Bond Operation

The objective function value and electron effects of the new state of atoms are calculated. This process continues until the end condition is met.

While none of the algorithms can examine the effect of the parameter values on the solution individually, the Artificial Atom Algorithm calculates the result of each decision variable. It evaluates the effect individually. For this reason, unlike other heuristic algorithms, the result in each iteration approaches the optimal value or maintains the state. In other heuristic algorithms, the decision variables are renewed by random methods, so the result can be approached and departed optimally.

B. Application of TSP with A^3

A^3 has been applied to different problems such reinforcement learning [21], group elevator control optimization [22], data clustering analysis [23] and preparation the optimum daily nutrition plan [24]. TSP is also designed with A^3 for small-scale problems [12] by Yıldırım and Karcı that forms the basis of this work.

When performing the TSP application, some changes made at original A^3 design according to structure of the problem, shown in the Figure 9,10,11 and 12 in [12]. First, to avoid repeat visits to a previously visited city, atom sets is generated randomly with a permutation method in the construction of the atomic set. When applying A^3 operators in the same way, some techniques have been developed to protect the unrepeated and ordered form of the atomic set. Flowchart of A^3 is in Figure 9.

After calculating the objective function values of the atoms and the effect values of the electrons that depends on the distance to the next city, the ionic bond operator for TSP was applied before the covalent bond operator.

When the ionic bond operator is applied, it is necessary to calculate the effect values of the atoms since new atoms are introduced randomly into the atom cluster (Figure 10). For this reason, before the covalent bond operator, the effect values of the electrons on the objective function are recalculated. Then the covalent bond operator is applied. In the covalent bond operation, the atoms are matched in pairs, the effect values of the electrons on the objective function are compared (see [12]). The electron with the smaller effect value is located on the other atom and is named as the electron *a*. The next electron from *a* is called electron *b*. The electron with the greater impact value is identified as *c* and its next index is called the electron *d*. Once these definitions have been made, the 2-opt method is applied for electrons with larger effect values (Figure 11).

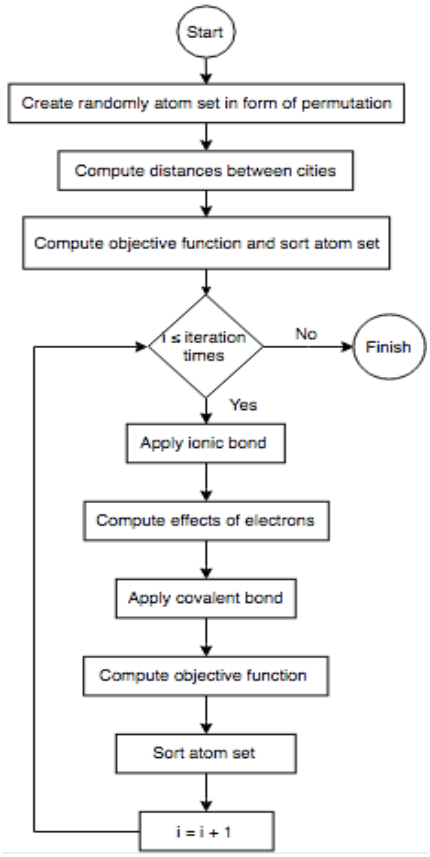


Figure 9: Flowchart A^3 for TSP an

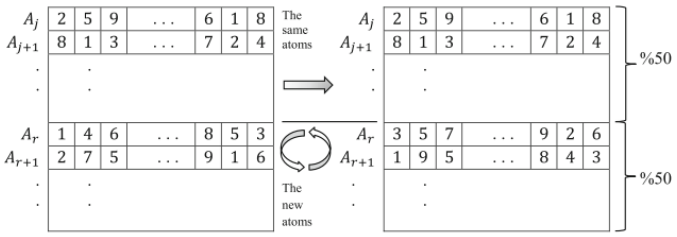
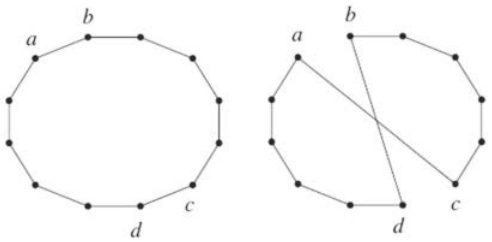


Figure 10: Example of ionic bond operation for TSP [12]



(a) (b)

Figure 11: Covalent bond operation for a larger value electron of a paired atom (a) Initial tour (b) The tour after applied the method [12]

There are 5 different cases when applying covalent bond operation (see [12]). Suppose that A_j and A_r are paired atom with n electron, $E[A_j(k)]$ is the effect of kth electron of atom A_j and $E[A_r(k)]$ is the effect of kth electron of atom A_r such that $E[A_j(k)] < E[A_r(k)]$.

The cases of covalent bond are [12]:

$If\ k < l,\ A_r[k+1:l] = A_r[l:-1:k+1]$
$If\ l \leq k;$
(a) $if\ l = 1\ and\ k = n-1, A_r[k+1\ l] = A_r[l\ k+1]$
(b) $If\ l \neq 1\ and\ k = n-1,\ A_r[k+1:-1:l] = A_r[l:\ k+1]$
(c) $If\ l = 1\ and\ k \neq n-1, A_r[k:-1:l+1] = A_r[l+1:\ k]$
(d) $If\ l \neq 1\ and\ k \neq n-1,\ A_r[k+1:-1:l] = A_r[l:\ k+1]$

After the covalent bond operation, the objective function values are calculated for the atom set and the atomic set is sorted. The number of iterations is checked and the A^3 operators continue to be applied until the number of iterations is reached.

When applying A^3 for comparison problems, both the atomic group is sorted according to the objective function values and each atom is sorted according to the electron effect values. The order of the electrons for TSP, on the other hand, is the order of visits of the cities. For this reason, aligning the electrons according to the effect values is not suitable for the structure of the TSP because the candidate solution will completely disintegrate. Thus, in the form of A^3 adapted for TSP, the atoms are not ordered according to electron impact values (Figure 12).

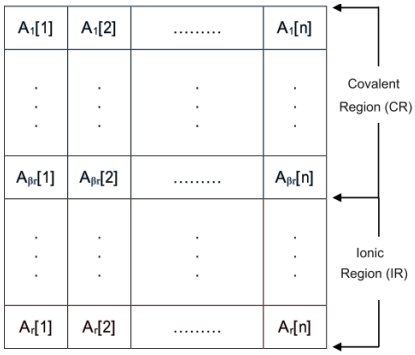


Figure 12: Representation of Atom Set, Covalent Region (CR) and Ionic Region (IR) for TSP [12]

C. *Parallel Application of A^3*

There are three different type of parallelism [25]. In bit-level parallelism, acceleration in computer architecture was achieved by doubling the computer word size, the amount of information that can be processed per cycle in processor. Increasing the word size reduces the number of instructions that the processor must execute to perform an operation on variables that are larger than the length of the word [25].

The other parallelism is instruction-level parallelism that reordered the instructions in the program, combined them into groups and without changing the result, executed these combined instructions in parallel [25]. Parallelism at the instruction-level has 2 basic approaches as Pipelining and Superscalar computing. Pipelining can try to isolate a process or a directive for a processor to different interconnected threads and process different threads in parallel. Superscalar computing also means that a processor can calculate the same phases in parallel in more than one direction in a single cycle.

Task parallelism refers to the simultaneous execution of different threads on different or the same data, in systems containing more than one processing unit. The fundamental difference in parallelism at the data level is that it cannot be fully scaled according to the data and resources available. The total process can only be scaled relative to the number of independent tasks.

In this paper, task parallelism is used for parallel A^3 application and parallelization is done on multi-core CPU. In the atom set, for each atom generation of atoms, computation of effects and ionic bond operations are discrete tasks. For covalent bond operation, each pair of atoms is processed independently. Each process performs the same functions and communicates with a "Master" Process but do not with each other. So, the application of task parallelism that works on different data is the most appropriate method for the program.

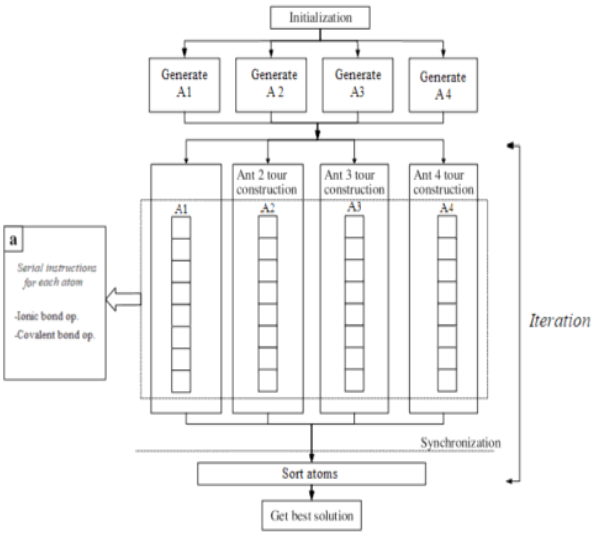


Figure 13: Representation of parallel structure of A^3

Parallel Generate Atoms: Each atom in atom set is a permutation of given cities that represent different data segment. That's why parallelization of generate atom operation can be done. Figure 13 shows the representation of parallel structure of A^3.

Parallel Ionic Bond: This operation changes IR of atom set randomly with new atoms. It is similar with generate atoms and parallelization is also same with generate atom operation.

Covalent Bond: First half of atoms in CR of atom set paired with the atoms in other half of the region and for each pair; the electrons were compared in term of the effects of electrons on the objective function. Covalent bond operation on each pair is independent from other pairs so that can be done in parallel discretely.

V. EXPERIMENTAL SETUP AND RESULT

For parallel applications hardware specification is important for the speed of the algorithm. CPU parallelization method is used in this project so the processor affects the performance. The processor used on this project is Intel Core i7-6700K that's specification is given in Table 2.

Table 2: Processor Specifications

Product	Core i7-6700K
Brand	Intel
# CPU Cores	4
Frequency (GHz)	4 GHz
Cache (MB)	8
Threads/Core	8
Frequency (GHz)	4 GHz
Turbo Frequency (GHz)	4.2 GHz
Cache (MB)	8
Instructions Width (bits)	64
# Memory Channels	2
Memory Bandwidth (GB/Sec)	34.1GB/s

ACO and A^3 algorithms are applied for different size of TSP with different number of ant/atom values.

For serial ACO, 20, 50, 200, 500 and 622 size TSP are used for 125, 250 and 10 ants and for parallel ACO, 200, 500 size TSP are used for 125 ants.

For serial A^3 20, 50,200, 500, 436, 436, 622, 711, 1083 size TSP used for 500, 100 atoms and for parallel A^3 200, 500, 436, 436, 622, 711, 1083 size TSP used for 500, 100 atoms.

Table 3: Different TSP problems with different specifications

Algorithm	**Number of City**	**Number of Ant/Atom**
Serial ACO	20	125 ants
	50	125 ants 250 ants
	200	125 ants
	500	125 ants
Serial A^3	20	500 atoms
	50	500 atoms
	200	500 atoms
	500	500 atoms
	436	500 atoms
	436	100 atoms
	622	100 atoms
	711	100 atoms
	1083	100 atoms
Parallel ACO	200	125 ants
	500	125 ants
Parallel A^3	200	500 atoms
	500	500 atoms
	436	500 atoms
	436	100 atoms
	622	100 atoms
	711	100 atoms
	1083	100 atoms

Table 3 shows experimented problems and for these problems what values have been used for algorithm specifications to see how they affect performance and compare results.

Table 4: Test results of ACO algorithm for different

Number of City	Tour Length	Time	Number of Ant/Atom
20	6954,7	0:8	125 ants
50	19010,5	1:40	125 ants
50	23771,9	2:23	250 ants
200	98860,23	01:08:27	125 ants
500	260074,28	16:15:39	125 ants

Table 4 shows that running time for ACO algorithm for the TSP increases exponentially with the number of cities. On the other hand, if the number of iterations is fixed and the number of ants is doubled the execution time will be doubled.

Table 5: Test results of Parallel ACO algorithm

Number of City	Tour Length	Time	Number of Ant/Atom
20	6037,4	0:4	125 ants
50	18724,35	00:45	125 ants
200	96759,93	27:45	125 ants
500	254093,31	2:24:00	125 ants

Parallel ACO has about 2,5 times better execution time on 4-core processor that can be seen by comparison the values on Table 4 and Table 5.

Table 6: Test results of A^3

Number of City	Tour Length	Time	Number of Ant/Atom
20	6187,6	00:01	500 atoms
50	16307,82	0:18	500 atoms
200	92609,8	7:50	500 atoms
500	262542,54	43:58	500 atoms
436	73163	32:00	500 atoms
436	75030,6	6:18	100 atoms
622	135054,6	14:42	100 atoms
711	205540,4	17:15	100 atoms
1083	289030	40:46	100 atoms

The running time of A^3 is also increasing exponentially with the number of city in TPS and it is proportional with the number of atoms.

Table 7: Test results of Parallel A^3

Number of City	Tour Length	Time	Number of Ant/Atom
20	5134,8	00:01	500 atoms
50	17730,38	00:02	500 atoms
200	93683,30	01:53	500 atoms
500	253110,9	10:57	500 atoms
436	71765,87	8:18	500 atoms
436	74806	1:44	100 atoms
622	134898,3	3:59	100 atoms
711	213714,4	4:31	100 atoms
1083	257306,5	10:25	100 atoms

Implementing A^3 as parallel gives about 4 times better

execution time on 4-core processor that can be seen by comparison the values on Table 6 and Table 7. Figure 14 shows the analysis of running time A^3 and Parallel A^3 for different size TSP and different number of atom.

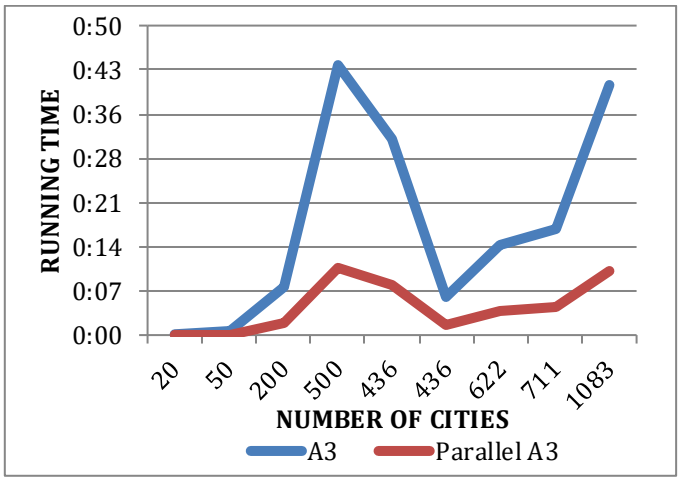


Figure 14: Running time comparison for A^3 and Parallel A^3

In Figure 14, 500 atoms are used first 5 values of number of city and 100 atoms are used for last 4 values on the graph. Increasing number of atoms is also increasing the running time but gives better tour length.

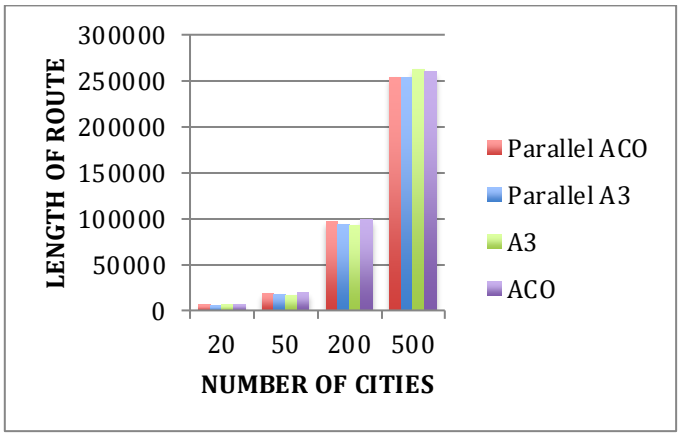


Figure 15: Tour length comparison for ACO, A^3, Parallel ACO and Parallel A^3

The result for comparison of tour length can be seen at Figure 15. The A^3 gives better results than ACO in general even though the difference is not too much. This result is same for parallel and serial application of the algorithms.

Even if the running time of ACO and A^3 is increasing exponentially, the increase in the ACO is much more than A^3.

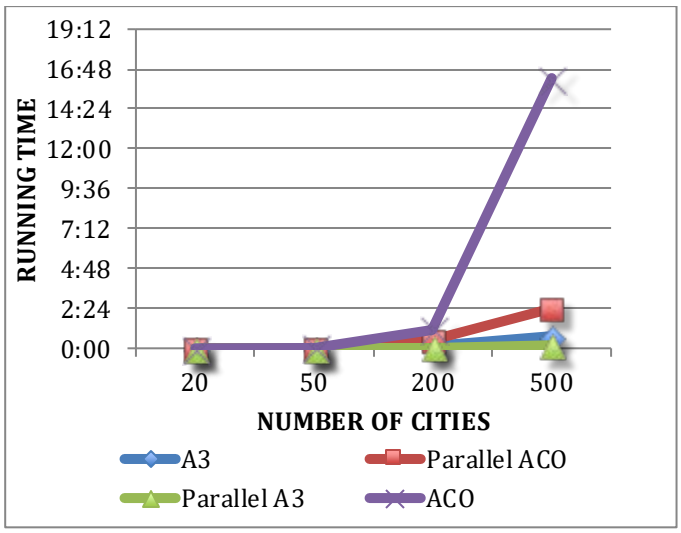


Figure 16: Running time comparison of ACO, A^3, Parallel ACO and Parallel A^3

VI. CONCLUSION

In this study, we implemented the Artificial Atom Algorithm parallel, which is natural-inspired meta-heuristic method based on compound formation process of atoms, on large scale Traveling Salesman Problem.

The proposed method was applied on two small scales and 6 large scales TSP with different number of atom. The parallel application of the A^3 was accomplished by paralleling ionic and covalent processes on each atom and atom pair in the atom set. For the 4-core CPU parallelization, this method works 4 times faster than original algorithm that shows how effective the proposed method works. The number of atoms gives a direct proportional effect on the method.

Parallel A^3 was also compared with Parallel Ant Colony Optimization Algorithm in terms of tour length and running time. Parallel A^3 has better results in average for tour length and much better results for running time than Parallel ACO. The efficiency in parallelization of A^3 is higher than ACO since A^3's running time increased 4 times while ACO's increased 2.5 times according to obtained results.

In future works, A^3 can be applied with CUDA (Compute Unified Device Architecture), which is an architecture and technology offered as an add-on to NVIDIA's C programming language for the GPU to achieve better runtime performance.

REFERENCES

[1] T. Weise, "Global Optimization Algorithms–Theory and Application," *URL http//www. it-weise. de, Abrufdatum*, vol. 1, p. 820, 2009.

[2] A. Astolfi, *Optimization: An introduction*, no. September. 2006.

[3] M. Canayaz, "Cırcır Böceği Algoritması: Yeni Bir Meta-Sezgisel Yaklaşım Ve Uygulamaları," 2015.

[4] Z. Michalewicz and D. B. Fogel, *How to Solve It: Modern Heuristics*, Second, Re., no. December. Springer, 2004.

[5] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," *Theor. Comput. Sci.*, vol. 344, no. 2–3, pp. 243–278, 2005.

[6] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Neural Networks, 1995. Proceedings., IEEE Int. Conf.*, vol. 4, pp. 1942–1948 vol.4, 1995.

[7] D. Dasgupta, "Guest editorial: Special issue on artificial immune systems," *IEEE Trans. Evol. Comput.*, vol. 6, no. 3, pp. 225–226, 2002.

[8] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization and control," *Control Syst. IEEE*, vol. 22, no. 3, pp. 52–67, 2002.

[9] J.-Y. Potvin, "Genetic Algorithms," *Ann. Oper. Res.*, vol. 63, pp. 339–370, 1996.

[10] A. Husseinzadeh Kashan, "League Championship Algorithm (LCA): An algorithm for global optimization inspired by sport championships," *Appl. Soft Comput. J.*, vol. 16, pp. 171–200, 2014.

[11] D. Henderson, S. Jacobson, and A. Johnson, *THE THEORY AND PRACTICE OF SIMULATED ANNEALING*. 2003.

[12] A. E. Yildirim and A. Karci, "Applications of artificial atom algorithm to small-scale traveling salesman problems," *Soft Comput.*, 2017.

[13] J. Grefenstette, R. Gopal, B. Rosmaita, and D. Van Gucht, "Genetic Algorithms for the Travelling Salesman Problem," *In Proceedings of the first International Conference on Genetic Algorithms and their Applications*. pp. 160–168, 1985.

[14] S. Kuzu, O. Önay, U. Şen, M. Tunçer, B. F. Yıldırım, and T. Keskintürk, "Gezgin Satıcı Problemlerinin Metasezgiseller ile Çözümü," *Istanbul Univ. J. Sch. Bus.*, vol. 43, no. 1, pp. 1–27, 2014.

[15] C. H. Papadimitriou and K. Steiglitz, "Combinatorial optimization: Algorithms and complexity," *IEEE Trans. Acoust.*, vol. 32, no. 6, pp. 1258–1259, 1984.

[16] R. Matai, S. Singh, and M. Lal, "Traveling Salesman Problem: an Overview of Applications, Formulations, and Solution Approaches," *Travel. Salesm. Probl. Theory Appl.*, 2010.

[17] H. Liu, P. Li, and Y. Wen, "Parallel Ant colony optimization algorithm," *Autom. 2006. WCICA 2006. …*, vol. 1, no. 2, pp. 77–82, 2006.

[18] M. Manfrin, M. Birattari, T. Stützle, and M. Dorigo, "Parallel Ant Colony Optimization for the Traveling Salesman Problem," *Interface*, vol. 4150, no. 9, pp. 224–234, 2006.

[19] S. Tsutsui and N. Fujimoto, "Parallel ant colony optimization algorithm on a multi-core processor," *Swarm Intell.*, pp. 488–495, 2010.

[20] A. Karcı, "A new meta--heuristic algorithm based on chemical process: Atom algorithm," *Proc. 1st Int. Eurasian Conf. Math. Sci. Appl. PriÇtine, Kosova*, 2012.

[21] A. KARADOĞAN and A. KARCI, "Takviyeli Öğrenme İçin Yapay Atom Algoritması (A3) Kullanımı," 2014.

[22] A. Erdogan Yildirim and A. Karci, "Group elevator control optimization using artificial atom algorithm," *IDAP 2017 - Int. Artif. Intell. Data Process. Symp.*, no. September 2017, 2017.

[23] M. Demir, "Veri Kümeleme de Yapay Atom Algoritması ve Cırcır Böceği Algoritmasının Karşılaştırılmalı Analizi," vol. 2016, no. November, 2016.

[24] E. Cifci, B. Acma, and B. Selcuk, "Yapay atom algoritmasi kullanarak bireye Özgü Öğün programli beslenme Çizelgesi hazirlanmasi," *2017 Int. Artif. Intell. Data Process. Symp.*, no. November, pp. 1–4, 2017.

[25] Go, "Parallel Computing."

ICATCES 2018

INTERNATIONAL CONFERENCE ON ADVANCED TECHNOLOGIES, COMPUTER ENGINEERING AND SCIENCE

INDEXIVE
SMART CONFERENCE INDEXING SERVICE

All papers have been published with INDEXIVE.

WWW.ICATCES.ORG

ISBN: 978-605-9554-18-3

KARABÜK ÜNİVERSİTESİ 2007

KONGRE KOORDİNASYON BİRİMİ

KARDEMİR

Ziraat Bankası