DETECTION OF QRS COMPLEX AND CLASSIFICATION OF ELECTROCARDIOGRAM SIGNALS USING COMPUTATIONAL INTELLIGENT ALGORITHMS

2021
MASTER THESIS
ELECTRICAL-ELECTRONICS ENGINEERING

MARWAH MUWAFAQ KADHIM AL-MOZANI

THESIS ADVISORS
Assoc.Prof.Dr. Mustafa Burak TÜRKÖZ
Assist.Prof.Dr. Eftâl ŞEHIRLI

DETECTION OF QRS COMPLEX AND CLASSIFICATION OF ELECTROCARDIOGRAM SIGNALS USING COMPUTATIONAL INTELLIGENT ALGORITHMS

Marwah Muwafaq Kadhim AL-MOZANI

T.C.

Karabuk University

Institute of Graduate Programs

Department of Electrical-Electronics Engineering

Prepared as

Master Thesis

Thesis Advisors

Assoc.Prof.Dr. Mustafa Burak TÜRKÖZ

Assist.Prof.Dr. Eftâl ŞEHIRLI

KARABUK

June 2021

I certify that in my opinion the thesis submitted by Marwah Muwafaq Kadhim AL-MOZANI titled "DETECTION OF QRS COMPLEX AND CLASSIFICATION OF ELECTROCARDIOGRAM SIGNALS USING COMPUTATIONAL INTELLIGENT ALGORITHMS" is fully adequate in scope and in quality as a thesis for the degree of Master of Science.

Assoc.Prof.Dr. Mustafa Burak TÜRKÖZ .....
Thesis Advisor, Department of Electrical-Electronics Engineering

Assist.Prof.Dr. Eftâl ŞEHIRLI .....
Thesis Co-Advisor, Department of Medical Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Electrical-Electronics Engineering as a Master of Science thesis. June 14, 2021

Examining Committee Members (Institutions)	Signature
Chairman : Assist.Prof.Dr. Emrah IRMAK (AAKU)	.....
Member : Assoc.Prof.Dr. Mustafa Burak TÜRKÖZ (KBU)	.....
Member : Assist.Prof.Dr. Eftâl ŞEHIRLI (KBU)	.....
Member : Assist.Prof.Dr. Abdullah Talha SÖZER (KBU)	.....
Member : Assist.Prof.Dr. Ahmet Reşit KAVSAOĞLU (KBU)	.....

The degree of Master of Science by the thesis submitted is approved by the Administrative Board of the Institute of Graduate Programs, Karabuk University.

Prof. Dr. Hasan SOLMAZ .....
Director of the Institute of Graduate Programs

"I declare that all the information within this thesis has been gathered and presented in accordance with academic regulations and ethical principles and I have according to the requirements of these regulations and principles cited all those which do not originate in this work as well."

Marwah Muwafaq Kadhim AL-MOZANI

ABSTRACT

M. Sc. Thesis

DETECTION OF QRS COMPLEX AND CLASSIFICATION OF ELECTROCARDIOGRAM SIGNALS USING COMPUTATIONAL INTELLIGENT ALGORITHMS

Marwah Muwafaq Kadhim AL-MOZANI

Karabük University
Institute of Graduate Programs
The Department of Electrical-Electronics Engineering

Thesis Advisors:
Assoc.Prof.Dr. Mustafa Burak TÜRKÖZ
Assist.Prof.Dr. Eftâl ŞEHIRLI
June 2021, 76 pages

Electrocardiogram (ECG) is a bioelectric signal that records the electrical activity of the heart versus time, and it is frequently used as a diagnostic instrument for evaluating the functions of the heart. In this thesis, the ECG signals were analysed and classified using a new and highly accurate algorithm in MATLAB software. The analysis process comprises four stages: preprocessing, QRS complex detection, extraction of QRS features, and data classification. 823 ECG signals from an arrhythmia database created for research by the MIT-BIH used in this thesis. The dataset was split into two classes as healthy and non-healthy. The QRS complex was detected using the K-means clustering method and tracking local extrema points. The average sensitivity, specificity, and accuracy of detection of QRS points were respectively computed as 97.04%, 92.68%, and 95.81% for R peak value, 91.30%, 95.22%, and 94.02% for Q

peak value and 91.65%, 95.05%, and 93.95% for S peak value. Training and test datasets have been created based on 5-fold cross-validation. Decision Tree (DT), Random Forest (RF), and Linear Discriminant Analysis (LDA) algorithms have been utilized to classify the ECG signals. The average sensitivity, specificity, and accuracy of classification of ECG signals are respectively computed as 89.9%, 90.0% and 90.0% for DT, 96.0%, 85.4% and 90.5% for RF, 84.5%, 74.8% and 79.6% for LDA. The highest accuracy and Matthews Correlation Coefficient (MCC) were recorded for RF, while the lowest value of accuracy and MCC were recorded for LDA.

Key Words : Electrocardiogram, QRS complex, Machine Learning, Signal Processing.

Science Code : 90524

ÖZET

Yüksek Lisans Tezi

HESAPLAMALI ZEKİ ALGORİTMALAR KULLANILARAK QRS YAPISININ TESPİTİ VE ELEKTROKARDİYOGRAM SİNYALLERİNİN SINIFLANDIRILMASI

Marwah Muwafaq Kadhim AL-MOZANI

Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Elektrik-Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanları:
Doç. Dr. Mustafa Burak TÜRKÖZ
Dr. Öğr. Üyesi Eftâl ŞEHIRLI
Haziran 2021, 76 sayfa

Elektrokardiyogram (EKG), kalbin elektriksel aktivitesini zamana karşı kaydeden bir biyoelektrik sinyaldir ve sıklıkla kalbin işlevlerini değerlendirmek için bir tanı aracı olarak kullanılmaktadır. Bu tezde, MATLAB yazılımında geliştirilen yeni ve yüksek doğrulukla çalışan bir algoritma ile EKG sinyalleri analiz edilmiş ve sınıflandırılmıştır. Analiz işlemleri, ön işleme, QRS yapısı tespiti, QRS yapısının özelliklerinin çıkarılması ve veri sınıflandırması olarak dört aşamadan oluşmaktadır. MIT-BIH tarafından araştırma yapmak için oluşturulan aritmi veri tabanından 823 adet EKG sinyali bu tezde kullanılmıştır. Veri seti, sağlıklı ve sağlıklı olmayan olarak iki sınıfa ayrılmıştır. QRS yapısı, K-ortalamalar kümeleme metodu kullanarak ve yerel ekstrem noktaları izleyerek tespit edilmiştir. QRS noktalarının ortalama duyarlılık, özgüllük ve saptama doğruluğu sırasıyla R tepe değeri için %97.04, %92.68 ve %95.81, Q tepe

değeri için %91.30, %95.22 ve %94.02 ve S tepe değeri için% 91.65,% 95.05 ve %93.95 olarak hesaplanmıştır. Eğitim ve test setleri 5 çapraz doğrulama ile oluşturulmuştur. EKG sinyallerini sınıflandırmak için Karar Ağacı (DT), Rastgele Orman (RF) ve Lineer Diskriminant Analizi (LDA) algoritmaları kullanılmıştır. EKG sinyallerinin ortalama duyarlılığı, özgüllüğü ve doğruluğu sırasıyla DT için %89.9, %90.0 ve %90.0, RF için %96.0, %85.4 ve %90.5, LDA için %84.5, %74.8 ve %79.6 olarak hesaplanmıştır. En yüksek doğruluk ve Matthews Korelasyon Katsayısı (MCC) RF için elde edilirken, en düşük doğruluk ve MCC değeri LDA için elde edilmiştir.

Anahtar Kelimeler : Elektrokardiyogram, QRS kompleksi, Makine Öğrenimi, Sinyal İşleme.

Bilim Kodu : 90524

ACKNOWLEDGMENT

I would like to express my sincere gratitude and appreciation to my research supervisor Assoc. Prof. Dr. Mustafa Burak TÜRKÖZ. He has been a tremendous mentor and guides for me. Also, I am extremely grateful to my co-supervisor, Assist. Prof. Dr. Eftâl ŞEHIRLI for his guidance, support, and patience towards the completion of this work. He provided me with precious ideas, advice, and suggestions to complete this research work. I would like to thank him for encouraging my research.

Special thanks to my husband, my daughters, my mother, and my father for supporting me every time.

CONTENTS

Page

APPROVAL…ii
ABSTRACT…iv
ÖZET…vi
ACKNOWLEDGMENT…viii
CONTENTS…ix
LIST OF FIGURES …xii
LIST OF TABLES …xiii
SYMBOLS AND ABBREVITIONS INDEX …xiv

PART 1 …1
INTRODUCTION …1
1.1 SIGNAL PROCESSING AND FILTERS …3
1.2 MACHINE LEARNING (ML) …4
1.2.1 Supervised Learning …4
1.2.2 Unsupervised Learning…5
1.3 PROBLEM STATEMENT …5
1.4 THESIS OBJECTIVES…5

PART 2 …6
ECG SIGNAL …6
2.1 ECG RECORDING…6
2.2 DISEASES DIAGNOSED BY ECG …8
2.2.1 Normal Sinus Rhythm (NSR)…8
2.2.2 Atrial Premature Beats (APB) …8
2.2.3 Atrial Flutter (AFL) …9
2.2.4 Atrial Fibrillation (AFIB) …10
2.2.5 Wolff-Parkinson-White (WPW) Syndrome …10
2.2.6 Premature Ventricular contraction (VPC) …11

Page

2.2.7 Left Bundle Branch Block (LBBB) 12
2.2.8 Right Bundle Branch Block (RBBB) 13
2.2.9 PR Interval 14

PART 3 16
LITERATURE REVIEW 16

PART 4 23
MATERIALS 23
4.1 MIT-BIH ECG DATABASE 23
4.2 COMPUTER PROPERTIES 24
4.3. MATLAB 24
4.3.1 Signal Processing Toolbox 25

PART 5 26
METHODS 26
5.1 SIGNAL PREPROCESSING 27
5.1.1 Zero Phase Filter 27
5.1.2 Moving Average (MA) Filter 28
5.2 QRS COMPLEX DETECTION 29
5.2.1 K-Means Clustering Algorithm 29
5.2.2 Detection of R- Points Using K-Means Clustering Method 30
5.2.3 Detection of Q and S Points 31
5.2.4 Feature Extraction 32
5.3 MACHINE LEARNING 33
5.3.1 Linear Discriminant Analysis (LDA) 34
5.3.2 Random Forests (RF) 36
5.3.3 Decision Trees (DT) 36
5.4 CROSS-VALIDATION 37
5.5 CONFUSION MATRIX 38

Page

PART 6 42
RESULTS AND DISCUSSION 42

PART 7 53
CONCLUSION 53
REFERENCES 54
RESUME 60

LIST OF FIGURES

Page

Figure 1.1. Cardiac electrical conduction system [1]. 2
Figure 2.1. (a) the human body, (b) heart anatomy, and (c) ECG graph. 7
Figure 2.2. ECG graph of NSR [8]. 8
Figure 2.3. ECG graph of APB [9]. 9
Figure 2.4. ECG graph of AFL [9]. 9
Figure 2.5. ECG graph of AFIB [9]. 10
Figure 2.6. ECG graph of WPW [10]. 11
Figure 2.7. ECG graph of VPC [9]. 12
Figure 2.8. ECG graph of LBBB [13]. 13
Figure 2.9. ECG graph of RBBB [14]. 14
Figure 2.10. ECG graph of PR Interval [15]. 15
Figure 5.1. Typical ECG signal processing flowchart. 26
Figure 5.2. Block diagram of a moving average filter [34]. 28
Figure 5.3. Illustration of Q points [29]. 31
Figure 5.4. Illustration of S points [29]. 32
Figure 5.5. LDA transformation and the optimal linear decision boundary [49]. 35
Figure 5.6. The decision tree [54]. 37
Figure 5.7. Illustration of the 5-fold cross-validation procedure [60]. 38
Figure 6.1. (a) raw ECG, (b) filtered ECG, (c) R-peak detection, (d) S-peak detection, and (e) Q-peak detection. 43
Figure 6.2. ECG signals whose QRS complex are marked. (a) NSR, (b) AFIB, (c) AFL, (d) APB, (e) LBBBB, (f) RBBBB, (g) PVC, and (h) WPW. 47
Figure 6.3. ROC curves. 52

LIST OF TABLES

Page

Table 3.1. An annual number of articles on QRS detection using machine learning was published in the web of science over the 1988- 2020 period. 17
Table 3.2. QRS complex detection proposed in the previous studies. 20
Table 4.1. MIT-BIH database of ECG signals. 23
Table 4.2. Properties of the used computer. 24
Table 5.1. The standard confusion matrix M. 39
Table 6.1. QRS complex segmentation comparison. 48
Table 6.2. Performance metrics of classification of ECG signals 50

SYMBOLS AND ABBREVITIONS INDEX

ABBREVITIONS

ECG	: Electrocardiogram
ICD	: Implantable Cardioverter and Defibrillator
LSM	: Least Mean Square Algorithm
ML	: Machine Learning
NSR	: Normal Sinus Rhythm
AFL	: Atrial Flutter
APB	: Atrial Premature Beats
AFIB	: Atrial Fibrillation
WPW	: Wolff-Parkinson-White
PVC	: Premature Ventricular contraction
LBBB	: Left Bundle Branch Block
RBBB	: Right Bundle Branch Block
HBR	: Heartbeat Rate
DFA	: Deterministic Finite Automata
DWT	: Discrete Wavelet Transform
CVDs	: Cardiovascular Diseases
DCM	: Dilated Cardiomyopathy
HCM	: Hypertrophic Cardiomyopathy
MI	: Myocardial Infarction
SFS	: Sequential Forward Selection
KNN	: K-Nearest Neighbor
FBGA	: Field-Programmable Gate Arrays
FIR	: Finite Impulse Response
STFT	: Short Time Fourier Transform
HR	: Heart Rate
HRV	: Heart Rate Variability

DER : Detection Error Rate

GUIs : Graphical User Interfaces

DSP : Digital Signal Processing

IIR : Infinite Impulse Response

FFT : Fast Fourier Transformation

DCT : Discrete Cosine Transformations

MA : Moving Average

GLCM : Grey-Level Co-Occurrence Matrix

LDA : Linear Discriminant Analysis

RF : Random Forests

DT : Decision Trees

TP : True Positive

FP : False Positive

FN : False Negative

TN : True Negative

MCC : Matthews Correlation Coefficient

AUC : Area Under Curve

ROC : Receiver Operating Characteristic

PART 1

INTRODUCTION

The automatic detection and diagnosis of heart disease have received significant attention from researchers and medical professionals. The use of computers in this process requires high accuracy as it relates to human life. Computer devices and modern equipment have become widely used in medical institutions, as they improve the quality of diagnosis and reduce the error rate to a minimum. The presence of good software helps doctors in medical institutions. It has become an urgent need, especially these days.

The human heart consisted of four partitions that work efficiently and continuously for life. The pericardial intake of muscles is controlled by an electrical current that flows through the heart accurately and precisely. The electrical current that triggers each heart attack is generated in the cardiac system (the frontal node or atrial sinus node), placed at the top of the right upper heart chamber (right atrium). The rate at which the heart discharges the electrical current determines the heart rate. This proportion is impacted by nerve impulses and certain hormone levels in the circulation system.

In adults, at the rest time, the heart beats ranged from 60 to 100 beats per minute. However, the lower rates may be normal in young people, particularly those who are physically appropriate. The heart pulse generally differs in response to exercise and stimuli such as pain and anger. The heart rhythm is abnormal only when the heart rate is fast, slow, irregular or when electrical impulses move in abnormal pathways.

The electrical current streams from the atrial sinus node first over the right atrium and then over the left atrium, making the muscles of these chambers shrinkage and pump blood from the atrium towards the lower heart chambers (ventricles).

The electric flow arrives at the ventricular atrial node, placed at the lower part of the wall between the atria near the ventricles. The ventricular atrial node delivers the only electrical connection between the atria and ventricles. The atria are isolated from the ventricles by non-electrical tissue at other sites. The ventricular atrium node delays the transmission of the current. Subsequently passing through the ventricular atrial node, the electric current travels downwards to the swarm, a group of fibers that divide into a left branch of the left ventricle and a right branch of the right ventricle. The electrical current then spreads in an orderly manner on the ventricles' surface, from bottom to top, leading the ventricle to begin to shrinkage, removing blood from the heart, as illustrated in Figure 1.1.

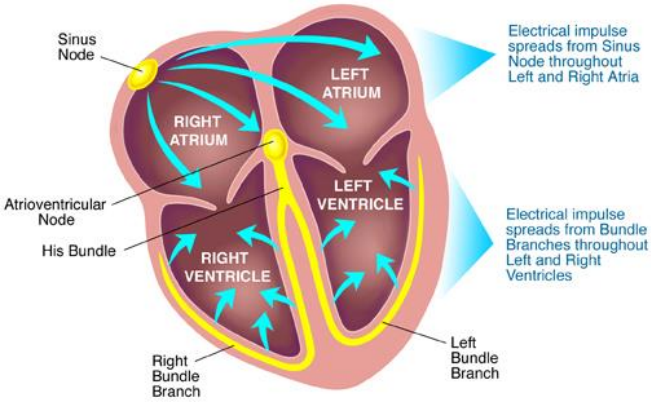


Figure 1.1. Cardiac electrical conduction system [1].

An arrhythmia describes an irregular heartbeat [2]. Arrhythmia appears in the form of heart failure, tachycardia, fluttering or fibrillation, premature cardiac constriction causing early heart attacks, symptoms that may accompany this condition: dizziness, shortness of breath, sometimes heart palpitations, and sometimes heart failure. The problem with arrhythmia may make the patient more likely to develop serious medical conditions such as stroke or cardiac arrest. Arrhythmia may not require any medical intervention to treat it at times. However, the doctor will perform the necessary medical procedures to treat the problem if there are clear and significant signs, such as dizziness, chest pain, loss of consciousness, or if the problem increases the likelihood of heart failure, stroke, or sudden cardiac death arrest.

A doctor will need to diagnose an arrhythmia needs to know the symptoms the patient has, as well as medical history and physical examination. Then, it performs several specialized tests to detect the disease, the most important of which are:

- Electrocardiogram (ECG).
- Holter (It is a battery-powered medical device that measures heart activity such as pulse rate and regularity. The Holter device is used continuously for 24 hours and, during that time, records the heart rate).
- Echocardiogram (A heart examination using ultrasound technology is an echocardiogram of non-invasive medical imaging examinations of the heart chambers. During the examination, the heart muscle's strength and the shape of the heart valves are recognized. The heart's echo is sipped through the chest by placing the tip of the probe that broadcasts ultrasound waves in a sectional way to the skin at the chest wall between two ribs to obtain imaging of the heart from several sides).
- Implantable heart rate recorder (It is a small device powered by a battery placed in the chest to monitor an irregular heartbeat and detect an irregular heartbeat. The implantable cardioverter and defibrillator (ICD) can deliver electrical shocks via one or more wires connected to the heart to repair abnormalities in the heart rhythm).

1.1 SIGNAL PROCESSING AND FILTERS

Signal processing is manipulated signals to extract meaningful information from the raw data, information extraction based on the relationship of various signals, or generating a substitute signal depiction [3]. Regularly the processing mechanisms are specified by a series of mathematical expressions through qualitative rules are applied. Currently, there are many motivations for signal processing; however, the most common include the following:

- Predicting future signal values to anticipate source behaviour.
- Information extraction by rendering more useful or apparent forms.
- Elimination of unwanted signal elements that may corrupt the signal quality.

Filtration is calculated and integrated into the ECG signals for noise elimination. An adaptive filter, for example, is a system with built-in scalability self-aligning functionality [4]. Spectral overlaps exist in most situations between signal and noise; if the present band is detected or differs over a defined period, the coefficient must be changed or not defined in advance. The adaptive filtering system eliminates much of the noise in the ECG signal. For example, for noise reduction, an adaptive self-tuning filter can be added. These filters are applied using the least mean square algorithm (LSM). The LSM algorithm can be integrated using various MATLAB tools that present a graphical and interacting computation environment that enables users to view results by applying the two phases of an adaptive filter on a noisy ECG signal.

1.2 MACHINE LEARNING (ML)

ML is perhaps one of the fastest-growing computer science areas, with numerous applications. ML can be used to play advanced and complex computations [5]. By having access to large data sets, ML utilizes the supplied data to learn from the past and make predictions on future data based on the presented data. The data consist of two or more classes and is trained based on feature vectors. Each feature in the feature vectors provides one dimension, thus having feature vectors with length n creates an n-dimensional feature space. To train a classification model, the learning method depends on if the feature vectors have been labelled or not and can be split into supervised and unsupervised learning.

1.2.1 Supervised Learning

Supervised learning is performed when the actual state, also known as the label, for known data set. Consequently, the data is applied to train a classifier by dividing the different data sets based on their actual state. When new data that is not labelled is presented to the classifier, it will be based on various metrics, provide the state with the highest score. The metrics used to estimate which state the new data belongs to vary from model to model [5].

1.2.2 Unsupervised Learning

Unsupervised learning is the method where only the data is known. Hence, the actual state, or label, of each of the classes is unknown. Unsupervised learning is done by clustering methods or associations, where the likelihood of the samples is used to group the samples into classes [6].

1.3 PROBLEM STATEMENT

The underlying complexities and the mechanical and clinical interaction of arrhythmias often create difficulties in diagnosing the medical doctors and primary care practitioners, who produce regular misdiagnoses and cross-classification with the aid of visual criteria. After a comprehensive survey, it is stated that although many filtering procedures have been tested in recent years, these procedures excessively worsen the quality of the ECG signal causing low accuracy and elevated detection error rate. Therefore, there is an urgent necessity of utilizing filters to locate the QRS segmentation complex for remote monitoring and processing of ECG signal and classification of arrhythmias [7].

1.4 THESIS OBJECTIVES

The specific objectives of this thesis are to fabricate new and highly accurate algorithms for ECG classification as normal and abnormal using MATLAB software. The analysis process comprises four stages: digital filtering (preprocessing step), QRS complex detection, extraction of QRS features (processing step), and data classification (postprocessing step). This program will help the medical staff diagnose, treat, and prevent heart problems and reduce mortality. The possibility of extracting the desired features from it to help the doctor form a deeper perception of a classification disease, as a program that can overcome these difficulties enables the doctor to read more correctly and medical signals. These improves overall diagnostic performance and reliability.

PART 2

ECG SIGNAL

ECG device is frequently employed to estimate the electrical activity of the heart and it is used as a routine practice prior to the patient go through surgery or identifies the reasons or illnesses that cause experiencing chest pain or the occurrence of pulses. The first ECG device appeared in 1903, designed by the scientist Einthoven, and he named it the heart painter. It is noteworthy that he took his concept of the galvanometer, which is involved in estimating the electric current's intensity. This sketch was massive in size and is not ideal in manufacture. Still, with the progress of the times and the introduction of permanent upgrades to this tool, it suited small and user-friendly and precise in displaying the outcome.

2.1 ECG RECORDING

To record the electrical activity of the heart, ten electrodes installed on the chest area of the human body. The electrical activity of the heart is recorded in the form of waves drawn on a graph. The patterns of these different waves express the electrical activity of the heart. The doctor diagnoses the patient's condition based on the results of this test. It should be noted that some types of heart disorders may not appear during this type of procedure, so the doctor may use other types of planning devices that help plan the electrical activity of the heart during different times of the day.

The ECG is measured at 25 mm/sec (5 large squares/sec), and the electrical energy is adjusted so that 1 mV = 10 mm (2 large squares) in the vertical direction. Hence, each small 1-mm square is 0.04 sec in time and 0.10 mV in voltage. Five fundamental waves exist on the screen of the ECG device (as shown in Figure 2.1), each symbolizing a particular point inside the heart, and these waves are:

- P wave (atrial depolarization) (The first wave appeared on the device's screen, as it reveals depolarization due to the passage of electric current in the left and right atria).
- Q wave (depolarization due to the passage of current through the septum between the two ventricles).
- P wave (atrial depolarization) (It is the first wave appeared on the screen of device, as it reveals depolarization due to the passage of electric current in the left and right atria).
- R wave (This wave is the greatest in terms of current strength, and it suggests depolarization of the left and right ventricles).
- S wave (It shows depolarization of the walls of the left and right ventricles).
- T wave (Represents ventricular repolarization. Generally, the T wave shows a positive deflection. This occurs because the last cells to depolarize are sited in the subepicardial region of the ventricles and these cells have shorter action abilities than observed in the subendocardial areas of the ventricular wall).

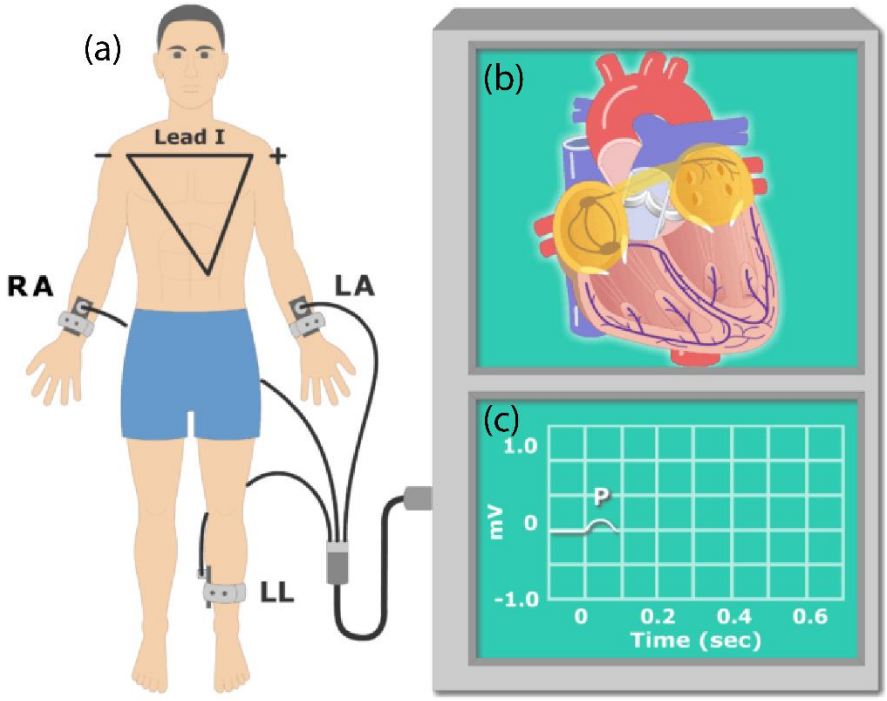


Figure 2.1. (a) the human body, (b) heart anatomy, and (c) ECG graph.

2.2 DISEASES DIAGNOSED BY ECG

Based on the location and figure of P, Q, R, S, and T peaks, various diseases can be diagnosed from the ECG graph include:

2.2.1 Normal Sinus Rhythm (NSR)

NSR comes from the SA node and is 50 to 100 beats per minute (Figure 2.2). When the incidence is below the low limit, sinus tachycardia is above the upper bound, and the rhythm is termed sinus bradycardia. Rest, while external stimuli in physical or intellectual stimulation are absent, the heart rhythm is ultimately normal. The continuing variance causes these minor changes in the cardiac velocity in the equilibrium between both component parameters of the autonomy of the nervous system, which affect the firing rate of the node of SA [8].

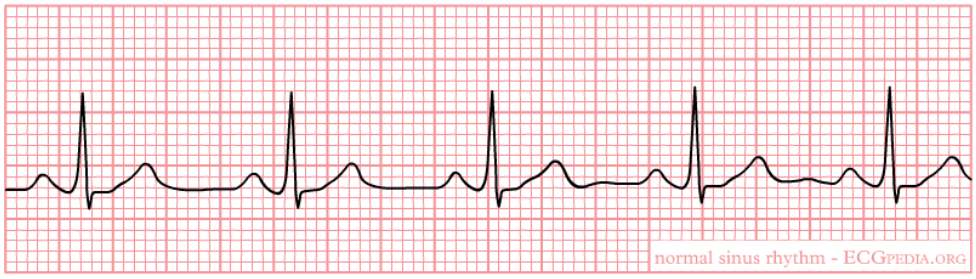


Figure 2.2. ECG graph of NSR [8].

2.2.2 Atrial Premature Beats (APB)

APB is a premature atrial depolarization that occurs at a premature atria activation, normally beyond sinus node from a site as seen in Figure 2.3 and can also be referred to as atrial or supraventricular extrasystoles. Premature depolarizations from the auricular node or its bundle are called premature auricular beats.

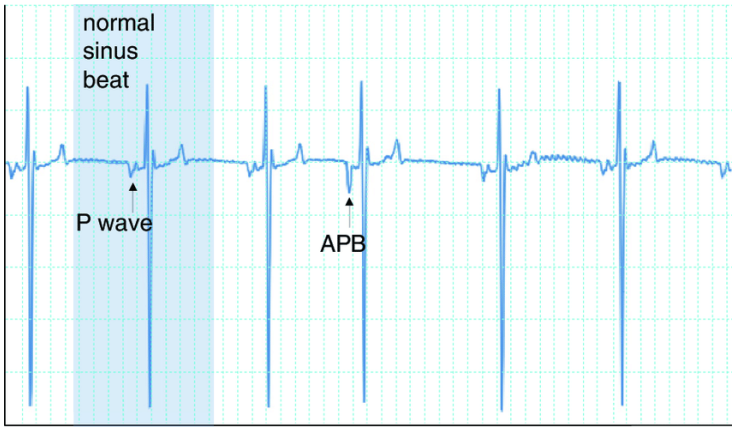


Figure 2.3. ECG graph of APB [9].

2.2.3 Atrial Flutter (AFL)

AFL is a normal, macro-re-entered arrhythmia historically defined as super ventricular tachycardia with 240–320 beats per minute of an auricular scale (bpm). AFL shall be described as similar, consecutive repeat P waves of the same morphology without clear diastolic iso-electrical interval (Figure. 2.4). For the classification of arrhythmia as AFL, at least four such P waves are needed. This can be normal or abnormal in ventricular rhythm.

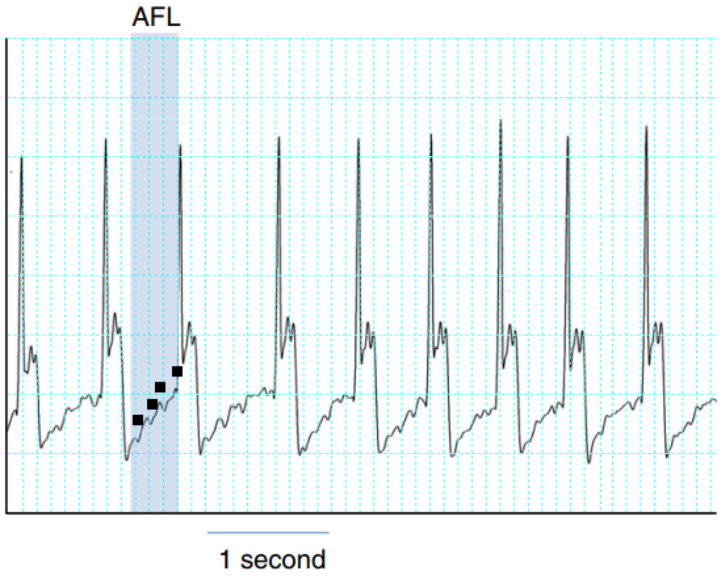


Figure 2.4. ECG graph of AFL [9].

2.2.4 Atrial Fibrillation (AFIB)

AFIB is an irregular and often quick heart rate that can boost the threat of strokes, heart failure, and other heart-related complications. AFIB is identified as a sequence of continuous atrial complexes that can no longer be recognized from one alternative as P waves, and where consecutive complexes vary in the peak-peak interval, height, and intrinsic shape, with the variant being neither progressive nor repetitive (Figure 2.5). A minimum of four such complexes is needed for the categorization of arrhythmia as AFIB [9].

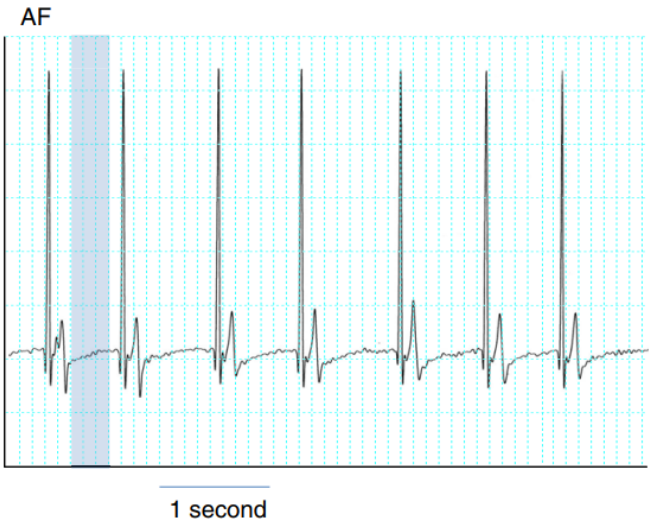


Figure 2.5. ECG graph of AFIB [9].

2.2.5 Wolff-Parkinson-White (WPW) Syndrome

WPW is very common, leading to abnormally accelerated heart pumping for times. The explanation is an additional cardiac electrical relation. This heart condition begins at birth (congenital), but symptoms will only develop later in life. Many cases of otherwise stable young adults are diagnosed. ECG characteristics of WPW in sinus rhythm include PR <120m interval, Delta wave – sluggish initial QRS slurry rise, QRS extension <110m, ST and wave discordant shifts – i.e., pseudo infarction pattern can be seen in up to 70 per cent of patients in the opposite direction of the main part of the QRS complex – due to negative drift delta waves in the lower/anterior system (mimicking posterior infarction) [10], as illustrated in Figure 2.6.

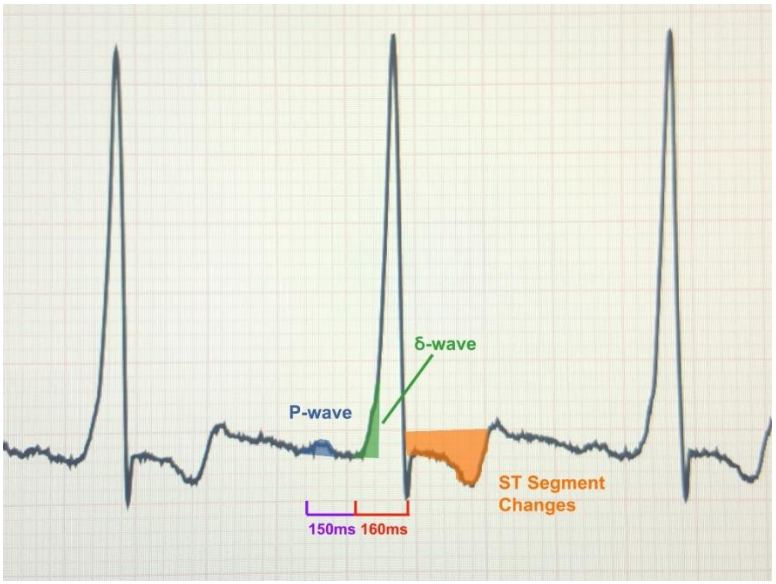


Figure 2.6. ECG graph of WPW [10].

2.2.6 Premature Ventricular contraction (VPC)

VPC are classified as the electrical complex ventricular system (the entire electrical occurrence: QRS, RS, QRST, or RST), which is dissimilar in form (voltage and/or length, i.e., height and/or width) and premature to the previous ventricular complex (NVPB) (Figure. 2.7). The ventricular complex usually is immediately followed by the VPB. Still, when it happens late in the heart cycle, the P wave unrelated to the VPB will precede the VPB (giving rise to a seemingly brief PR interval). In addition, premature atrial beats will lead to an aberrant (i.e., the bundle branch block) to ventricular stimulation, but arrhythmias should not be categorized as VPBs [9].

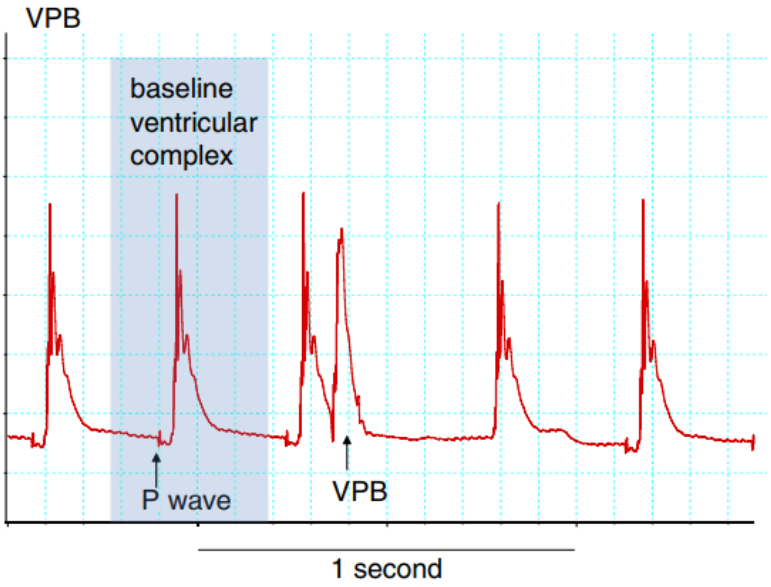


Figure 2.7. ECG graph of VPC [9].

2.2.7 Left Bundle Branch Block (LBBB)

A typical (ECG) abnormality in the LBBB, which is seen in patients who have disrupted natural cardiac conduction in both the front and back left pillars [11] of His-Purkinje system (Figure 2.8). LBBB differentials provide intraventricular lead time, leading to identical ECG outcomes and may not have a V6 R wave. A paceful pulse may also be misunderstood with LBBB, but there is often no R wave in V6, typically with pacer spikes. An incomplete LBBB is considered an LBBB pattern with QRS length below 120 ms. Without super-ventricular stimulus, ventricular rhythm (PVC running) can seem very close to the LBBB and will, in certain rare cases, be unmistakable from LBBB [11, 12].

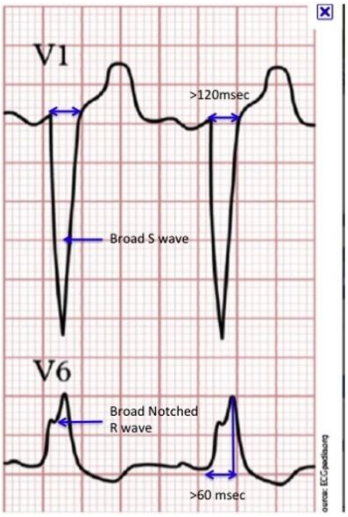


Figure 2.8. ECG graph of LBBB [13].

2.2.8 Right Bundle Branch Block (RBBB)

RBBB refers to an electrical conductivity machine heart block in the RBBB (Figure 2.9). The right ventricle is not triggered directly by pulses that pass through the RBBB during the RBBB. But typically, the left bundle branch also activates the left ventricle. Via the myocardial ventricle to the right ventricle, these impulses can migrate and depolarize the right ventricle. The QRS complex is seen as being expanded as it leads through the myocardial tone slower than the lead through the His-Purkinje fibre bundle. The QRS complex often shows an additional deflection reflecting fast left-ventricular depolarization followed by slower right-ventricle growth.

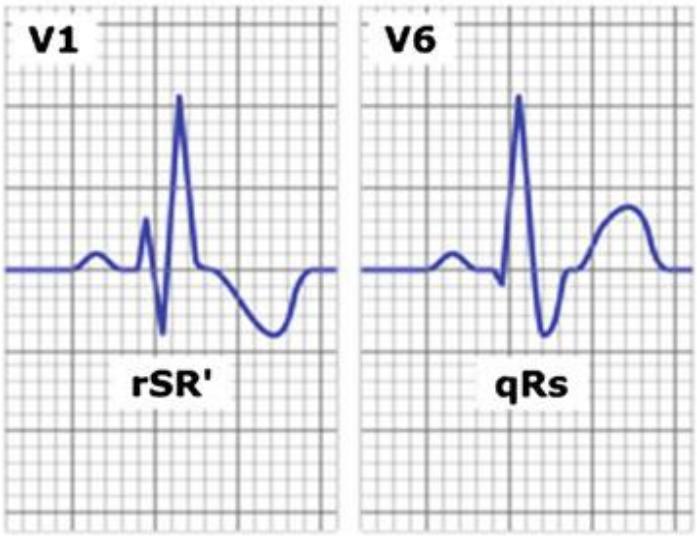


Figure 2.9. ECG graph of RBBB [14].

2.2.9 PR Interval

The PR interval is measured from the start of the P wave to the beginning of the QRS complex. The drive through the AV node is mirrored.

- Usually, the length of PR is between 120 – 200 ms (three to five small squares).
- The first-grade heart block is said to be present if the PR interval is > 200 ms.
- Pre-excitement syndrome, including the Wolff– Parkinson–White or Lown–Ganong–Levine syndrome, can be correlated with PR interval < 120 ms.

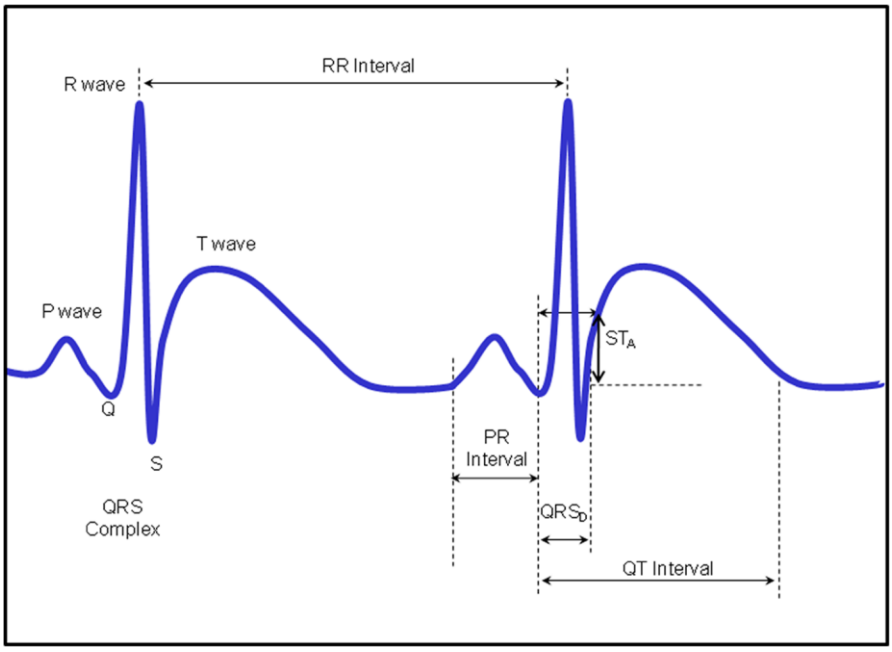


Figure 2.10. ECG graph of PR Interval [15].

PART 3

LITERATURE REVIEW

To achieve an accurate and quick diagnosis, automatic ECG analysis algorithms play a crucial role, beginning with QRS detection. QRS complex detection is reviewed as the first step in ECG analysis. In this case, the QRS complex is one of the most significant waveforms in ECG investigation. The QRS complex is represented by a combination of three graphical deflections witnessed in a typical ECG analysis. Moreover, it the central and most compelling section of the illustration. It is viewed as the main spike as considered in an ECG line. It displays the electric movement of the heart during a ventricular contraction; its occurrence and shape provide details about the heart's status. It serves as a framework for automating cardiovascular identification based on its form. The ECG analytical algorithms consider this to be the basis for the cardiac cycle classification.

In this case, QRS detection offers the basis for automating the ECG analysis algorithms [16]. For decades, the QRS complete detection algorithm has been analyzed, resulting in various hardware and software techniques for completing the same. There are currently numerous QRS detection algorithms – motivation for this variation is the need to enhance the process as each algorithm seeks to be better and faster than others.

QRS length is among the critical features of the QRS complex, which is applied in the assessment and organization of ECG signals. The QRS field is the length over and beyond S and Q nodes above the isoelectric (ISO) axis. The R-R interval is the spectrum of the two Heartbeat rate (HBR) and QRS complexes subsequent to it. The PR interval subsequently demonstrated the duration between the initiation of atrial depolarization and the start of ventricular growth while allowing for the manifestation of systole. In this relation, the R-wave is the R-wave amplifier, the largest R-wave node. The R-T interval is between crests, describing the T-wave and QRS complex

successive crests. It can be considered as the time interval between the ventricular depolarization peak and consecutive ventricular polarization peak.

ECG compression techniques have sought to minimize ECG the created dataset size and maintain all scientifically relevant characteristics like the P-wave, QRS complex and T-wave, as previously adopted. The compressed signal allows for better efficiency by programmers for the real-time processing of rhythmic classification networks. Many ECG compression techniques were suggested in the literature [17-19]. These methods were divided into processes of direct extraction, transformation, and parameters [20]. The latest extraction function algorithm presents morphologic and statistical features derived from regular and irregular ECG signals to finish this paragraph. Added to the extraction capabilities, the compressed shape of the signal enables us to verify the classification efficiency improvements during a classification stage.

The number of research papers with the keywords 'QRS detection using machine learning' available in the journals indexed by a web of science from 1988 to 2020 is tabulated in Table 3.1. The rapidly increasing number of scientific articles proves how engaging the machine learning field is in mainstream researchers.

Table 3.1. An annual number of articles on QRS detection using machine learning was published in the web of science over the 1988- 2020 period.

YEAR	**NUMBER of ARTICLES**
1988 - 2006	0
2007	3
2008	1
2009	2
2010	3
2011	4
2012	2
2013	4

YEAR	NUMBER of ARTICLES
2014	5
2015	7
2016	7
2017	8
2018	20
2019	22
2020	29

According to Chazal [21], 178 structures were generated from a QRS Complex – representing ECG beats. Subsequently, transformation, the features were in a summation of 229 transformed features. Techniques for evaluating feature in this context were achieved by reviewing various ECG based literature. In a comparable simulation, 30 features for a neural network were obtained employing the backpropagation training algorithm [22]. Subsequently, the two models discussed in this section demonstrate that the more inputs, the more challenging the classification network structure. As mentioned at the beginning of this research, it was stated that there might be challenges in computing the QRS Complex. For example, the classification speed reduces significantly in a normal personal computer that is cannot complete the research. Managing this challenge involves essential and crucial features from ECG waveforms that must be familiarized from the new study.

Hamdi et al. suggested a novel method for medical assessment and clinical evaluation assistance of the ECG signals based on the deterministic finite automata (DFA) and some requirements. This research confirms that regular grammar is helpful in the extraction of QRS complex and clarification of ECG signals. The DFA applied to represent a normalized QRS complex as a sequence of negative and positive peaks. A QRS is deemed a set of adjacent peaks that fulfill specific standard deviation criteria and duration. The suggested approach in this study is employed to numerous sorts of ECG signals gathered from the standard MIT-BIH arrhythmia database. Numerous metrics are evaluated, including QRS intervals, RR distances and peak amplitudes. Regular grammar and DFA demonstrated functional for ECG signal diagnosis. The

proposed method offered a sensitivity rate of 99.74% and the positive predictivity rate of 99.86% [23].

Adam et al. utilized a new discrete wavelet transform (DWT) approach fused with nonlinear features for automated characterization of cardiovascular diseases (CVDs). ECG signals of normal, and dilated cardiomyopathy (DCM), hypertrophic cardiomyopathy (HCM), and myocardial infarction (MI) are exposed to five levels of DWT. The relative wavelet of four nonlinear features such as fuzzy entropy, sample entropy, fractal dimension, and signal energy is obtained from the DWT coefficients. These features are supplied to the sequential forward selection (SFS) method and then ranked utilizing the ReliefF process. The suggested protocol in this study reached maximum classification accuracy of 99.27%, sensitivity of 99.74%, and specificity of 98.08% with K-nearest neighbour (kNN) classifier using 15 features ranked by the ReliefF method [24].

Zhang et al. introduced an adaptive threshold algorithm in ECG signal feature extraction using Kalman filtering theory. Low computational cost, low storage requirement, and fast response feature are attained using two sets of adaptive threshold systems in various requirements. As evidence of the hypothesis, the suggested algorithm in this study is validated in Matlab and implemented on field-programmable gate arrays (FPGA) utilizing the MIT-BIH database. The results showed the proposed algorithm uses a low resource of FPGA and displays 99.30 % detection sensitivity and 99.31 % positive prediction on average, respectively. The suggested algorithm can quickly adjust various individuals with the self-adjusting system in satisfying detection accuracy [25].

Yakut et al. enhanced extraction of QRS complex method having low complexity is suggested. This approach comprises two steps as preprocessing and decision making. To assess the proposed method's implementation, it was examined utilizing the ECG recordings (about 1.3 million beats) taken from the five standard databases as MIT-BIH Arrhythmia, Fantasia, MIT-BIH Noise Stress Test QT and European ST-T. In this study, 1296137 beats of 272 cases were studied for QRS detection, and the average sensitivity was achieved as 99,60%, while the average positive predictivity, +P was

presented as 99,77%. The impact of the suggested method is that the training, selection, setting, and prediction processes are not necessary while determining the essential parameters [26].

Shaik et al. suggested using an adaptive threshold method on spectrogram computed using Short Time Fourier Transform (STFT) for QRS complex detection in ECG signal. The proposed algorithm comprises preprocessing the raw ECG signal to wipe out the power-line interference, computing the STFT, using adaptive thresholding technique, and identifying QRS peaks. The proposed algorithm provides good outcomes in terms of sensitivity of 99.56%, specificity of 99.52% and QRS detection error rate of 0.93% achieved against the MIT-BIH Arrhythmia database [27].

Berwal et al. proposed a high-performance QRS complex detector appropriate for medical devices. In this study, a Biorthogonal wavelet filter bank with fourth-level decomposition is first employed to get the denoised ECG signals. The soft threshold method is employed to obtain the QRS complex peaks by the QRS complex peak detector block. The proposed model has been tested for its robustness on multiple datasets. Sensitivity of 99.31%, positive predictivity of 99.19%, and the Detection Error Rate (DER) of 1.49% shown by the suggested design make it appropriate for QRS complex detectors utilized in wearable healthcare devices [28].

Table 3.2. QRS complex detection proposed in the previous studies.

AUTHORS (YEAR)	**SENSITIVITY**	**SPECIFICITY**	**POSITIVE PREDICTED VALUE)**	**METHOD**
Sehirli and Turan (2021) [29]	94.75%	95.96%	91.86%	Consecutive difference method, k-means clustering, tracking local

AUTHORS (YEAR)	SENSITIVITY	SPECIFICITY	POSITIVE PREDICTED VALUE)	METHOD
				extreme points.
Shaik et al. (2015) [27]	99.56%	99.52%	----	Utilize an adaptive threshold technique on spectrogram computed using STFT.
Uchaipichat et al. (2010) [30]	99.10%	99.60%	----	STFT
Darrington (2006) [31]	99.00%	99.20%	----	Fine-to-coarse algorithm
Hamdi, Abdallah et al. (2018) [23]	99.74%	----	99.86%	Deterministic finite automata (DFA)
Adam, Oh et al. (2018) [24]	99.74%	98.08%.	----	Discrete wavelet transform (DWT)
Zhang, Yu et al. (2020) [25]	99.30 %	----	99.31 %	A Kalman filtering based adaptive threshold algorithm

AUTHORS (YEAR)	SENSITIVITY	SPECIFICITY	POSITIVE PREDICTED VALUE)	METHOD
Yakut and Bolat (2018) [26]	99,60%	----	99,77%	Dynamic threshold process
Berwal, Kumar et al. (2018) [28]	99.31%	----	99.19%	Soft threshold method

PART 4

MATERIALS

4.1 MIT-BIH ECG DATABASE

The MIT-BIH Database was the first set of standard test materials that generally available to evaluate arrhythmia detection. Since 1980, this database has been used to research cardiac dynamics at about 500 sites worldwide [32]. This study used 823 normal ECG signals and abnormal ECG signals with different heart diseases such as APB, WPW, AFL, AFIB, VPC, LBBB, and RBBB. Numerous of these files were chosen because the rhythm, QRS morphology variation, or signal quality could be supposed to present considerable difficulty to arrhythmia detector; these records have earned significant notoriety among database users [33].

Table 4.1. MIT-BIH database of ECG signals.

				5-FOLD CROSS-VALIDATION			
NO	**CLASS**	**FRAGMETS NUMBER**	**PATIENTS NUMBER**	**TRAINING SET**	**TEST SET**	**TRAINING SET**	**TEST SET**
1	NSR	283	23	255	28	252	12
2	APB	66	9	60	6	54	2
3	AFL	20	3	18	2	18	18
4	AFIB	135	6	122	13	117	4
5	WPW	21	1	19	2	18	16
6	PVC	133	14	120	13	117	10
7	LBBB	103	3	93	10	90	13
8	RBBB	62	3	56	6	54	8
	Sum	823	62	743	80	720	83

4.2 COMPUTER PROPERTIES

Properties of the used computer throughout this thesis are recorded in Table 4.2.

Table 4.2. Properties of the used computer.

PRODUCT	PROPERTY
Centra processing unit (CPU)	Core i7 2.6 GHz
Read access memory (RAM)	8.00 GB
Hard Disc	256 GB
Cache memory	4.00 MB
Display card	4.00 GB
Monitor	15 inches full HD
Operating system	Windows 10 (64-bit)

4.3. MATLAB

MATLAB is an efficient, embedded, and easy-to-use technology computation environment. It offers a single immersive interface that incorporates computer computing, simulation, and scripting to engineers, scientists, and other technical professionals. MATLAB contains a family of toolboxes and application-specific. MATLAB encourages the building of its reusable tools. One of its key strengths. MATLAB code can conveniently build personalized special features and programs. Biomedical engineers use MATLAB to research, design, and produce embedded medical device algorithms and systems. MATLAB provides various benefits relative to other conventional numerical approaches, for example:

- It makes coding in an extremely high language fast and simple.
- The architectures generated for the dataset require limited attention arrays.
- A fast exploration and simple debugging graphical user interfaces (GUIs).
- There are high-quality services to graphics and envision.
- MATLAB M files on a wide variety of systems are completely portable.
- The framework can be supplemented with toolboxes, providing specialized signal processing installations, for example.

In comparison, MATLAB is a modern software application and problem-solving environment. MATLAB has advanced data set architectures and provides built-in debugging and profiling capabilities. These aspects make MATLAB a perfect language for teaching and a vital instrument for studying and solving practical problems.

4.3.1 Signal Processing Toolbox

The Signal Processing Toolkit offers a rich and customizable structure for analogue and digital signal processing (DSP). GUIs support interactive architectures and research, and progressing algorithm creation supports command line functions. The toolbox is suitable for signal analysis and the development of a DSP algorithm. The algorithms used for industrial signal processing have been selected and applied for optimal performance and computational reliability. The key features are M-file routines, written in the MATLAB language, and provide access to source code and algorithms. In addition, MATLAB's open framework philosophy and the toolboxes allow you to alter or apply experiments to existing functions. The main features of the toolbox for signal processing are:
A complete range of signal models and linear device models.

- Design of filter analogy methods.
- Optical filter architecture, analysis, and application tools for FIR and infinite pulse response (IIR).
- The transformations used most, including fast Fourier transformation (FFT) and discrete cosine transformations (DCT).
- Spectrum estimation and processing approach for statistical signal.

PART 5

METHODS

The flowchart diagram of the proposed QRS detection method is shown in Figure 5.1. It contains four stages which are digital filtering (preprocessing step), QRS complex detection (processing step), extraction of QRS features, and data classification (post-processing step).

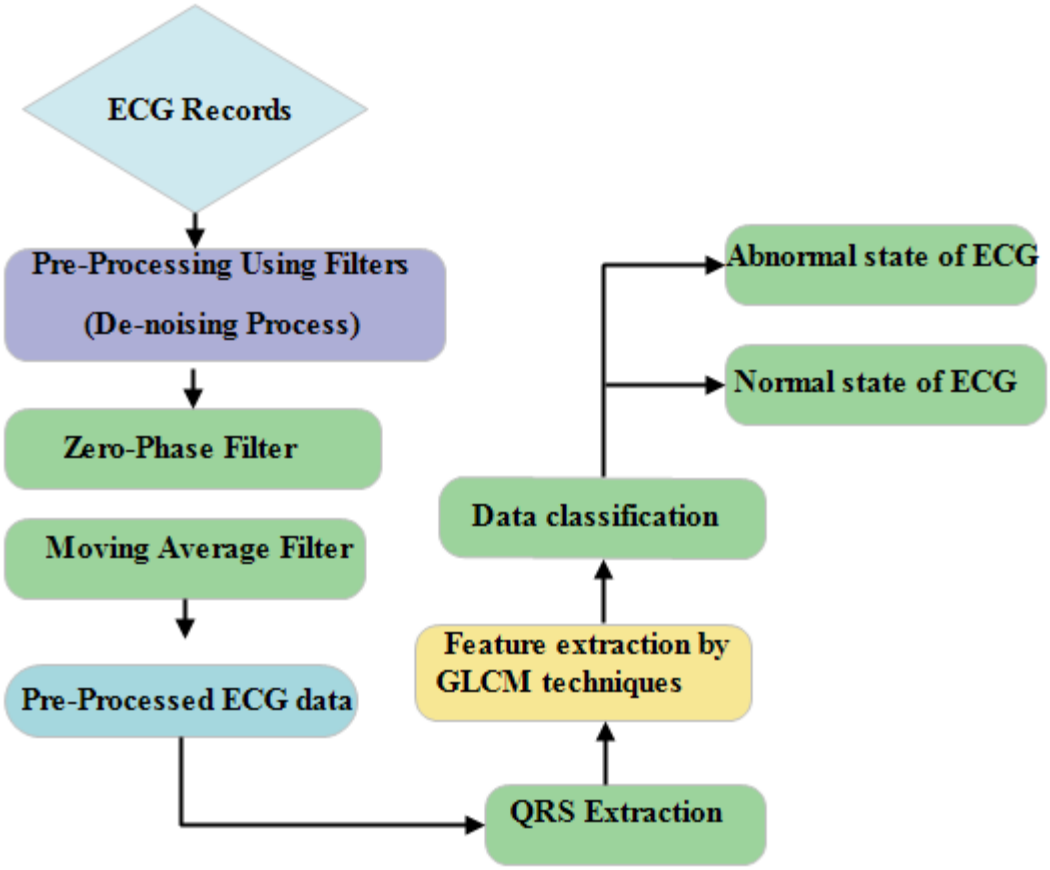


Figure 5.1. Typical ECG signal processing flowchart.

The MIT-BIH ECG annotation files were uploaded and read using the Matlab software. The uploaded files passed via three processing steps: preprocessing, processing, and post-processing to achieve the thesis objectively.

5.1 SIGNAL PREPROCESSING

A non-noise signal from ECG data from several sources can be considered the first form of exploration during this study. Digital signal processing technologies give a versatile and efficient approach. In addition, these techniques may be used to detect critical signal events in sources such as ECG data to increase the accuracy of bio-signals.

Digital filters are used for various signal processing tasks, including noise de-noise, waveform modulation, signal softening, and signal recovery. The filters are divided into two types.

5.1.1 Zero Phase Filter

A non-phase filter is a particular situation in which a linear phase filter is α=0.

$$y = filtfilt\ (b, a, x) \tag{5.1}$$

$$y = filtfilt\ (sos, g, x) \tag{5.2}$$

$$y = filtfilt\ (d, a, x) \tag{5.3}$$

$y = filtfilt\ (b, a, x)$ executes digital zero-step filtering, both forward and reverse, by processing the input for the data set generated, x. Once the generated dataset is filtered on, filtfilt switches the filtered series and reverses it through the buffer. The outcome has the following features:

- Distortion of the zero steps.
- A filter transfer feature equal to the initial filter transfer feature squared magnitude.

- A filter order doubles the order defined by b and a for the filter.

By balancing initial conditions, filtfilt minimizes begin-up and ending transients. y = zero-phase $filtfilt\ (sos, g, x)$filters the input of the generated dataset,x, with the second-order filter (biquad) seen with a matrix of sos and scale of g values.

y = zero-phase $filtfilt\ (d, a, x)$ filters the data set created,x, by a digital filter,d. Use designfilt to produce d based on the specification for frequency response.

5.1.2 Moving Average (MA) Filter

Owing to its conceptual flexibility and ease of use, the MA filter is perhaps one of the most used FIR filters. As seen in the diagram below (Figure. 5.2), the filter requires no multiplication, only additions, and a delay line, making it supremely well-suited for many low-power, computational-sensitive devices.

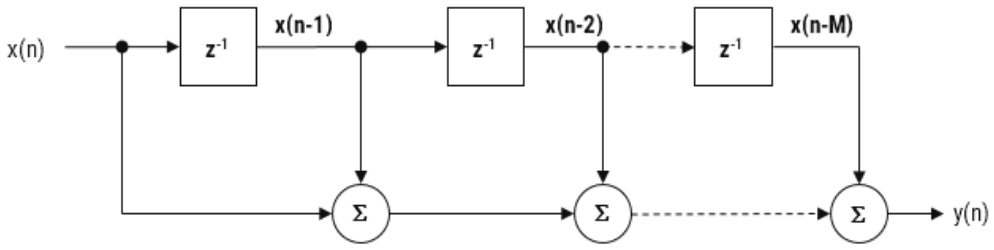


Figure 5.2. Block diagram of a moving average filter [34].

However, considering its simplicity, the moveable average filter is suitable for random noise reduction while maintaining an acute phase response, making it a flexible compound for smart sensor signals. For example, the following can be defined: a moving average L longitude filter for an input signal x(n):

$$y(n) = \frac{1}{L} \sum_{k=0}^{L-1} x(n-k)\ \ for\ \ n = 0,1,2,3, \ldots \quad (5.4)$$

If a basic thumb law states, the noise decline is proportional to the square root of the total number of points.

5.2 QRS COMPLEX DETECTION

The second part of the method is to prepare QRS complex detection on the signals.

5.2.1 K-Means Clustering Algorithm

The K-means technique is based on an output index minimization defined as a sum of the square distances between all the cluster areas and the cluster center. The K-means clustering algorithm can be used if a certain number of samples are supplied and clustered into K numbers of the clusters. It is based on the output indices minimization, defined as the sum of square distances between all points inside a cluster domain and the cluster center. The following are the separate procedural steps in the K-means clustering algorithm:

Phase 1: select K cluster centers in the beginning Z1(1), Z2(1) … Zn (1). The first K samples from the provided instances set X, Zl+1(1), for l = 1, 2, …, K − 1, are typically picked.

Phase 2: Disperse samples X, using the following relation in the kth iterative process, among the K cluster domains.

$$X \epsilon S_j(k) \text{ if } \left\|X - Z_j(k)\right\| < \|X - Z_i(k)\| \quad (5.5)$$

For all I = 1, 2, ..., K, I = j, where Sj(k) indicates the set of samples with a cluster core of Zj (k).

Phase 3: measure the new cluster centers Zj(k+1), j=1, 2., K from the product of step 2, to reduce the number of squared distances to the new cluster center from all points in Sj(k). In other words, the new Zj(k+1) center is determined to mean performance:

$$J_j = \sum X \epsilon S_j(k) \left\|X - Z_j(k+1)\right\|^2, \quad \text{j = 1, 2,..,k is minimized.} \quad (5.6)$$

The Zj(k+1) that minimizes these indices of results is the Sj (k) average. The new cluster core is therefore provided:

$$Z_j(k+1) = \frac{1}{N_j} \sum X \epsilon S_j(k) X, \quad j = 1,2, \ldots\ldots, K \tag{5.7}$$

Where Nj is Sj sample number where (k), the name "K-means" derives from the way clusters are modified sequentially.

Phase 4: if Zj(k+1) = Zj(k) = 1, 2..., K, the experiment is over, the algorithm has been converged. The comportment of the K-means clustering algorithm depends on the number of cluster centers indicated, the selection of the initial cluster, the order in which the sample is obtained, and the geometrical characteristics of the data set [35].

5.2.2 Detection of R- Points Using K-Means Clustering Method

Massive amplitude variations describe points within a QRS complex. The slope value will represent this attribute. This analysis is then used to derive the R- points using the slope knowledge and procedures.

- Measure the absolute pitch, calculate the pitch between adjacent two points from the input ECG (n).

$$\text{Slope (k)} = data\ (k+1) - data\ (k) \tag{5.8}$$

$$\text{abs}_{\text{slop}} = |\text{slope}| \tag{5.9}$$

- Cluster the divided data segment's absolute slope value by K-means, and K value = 2.

The amplitude value represents the changes in amplitude directly. The pitch map shows the characteristics of the R-points. There are very high positive pathways and very low negative paths within the R-points. The absolute slope value increases the slope characteristic of the R-points. The values are arranged in two groups using the K-means approach based on absolute slope. The K-means clustering algorithm may be

used if the number of samples is specified and clustered into clusters k. This is focused on minimizing the index of efficiency, defined as the amount of the square distances to the cluster centre from all points in the cluster domain [35, 36]. R-points have more features in general in a comparatively short period. Two cluster centers are easily selected by the K-means algorithm, one with a higher pitch and a lower pitch.

5.2.3 Detection of Q and S Points

With the detection of R peaks, the Q, S points can be detected using tracking local extrema points. The definition of Q points shown in Figure 5.3 is utilized. Q points of QRS complex are identified by scanning left through all R peaks. When a point at the left direction of R peak is lower than two points before and after it, this point is held in Q list [29].

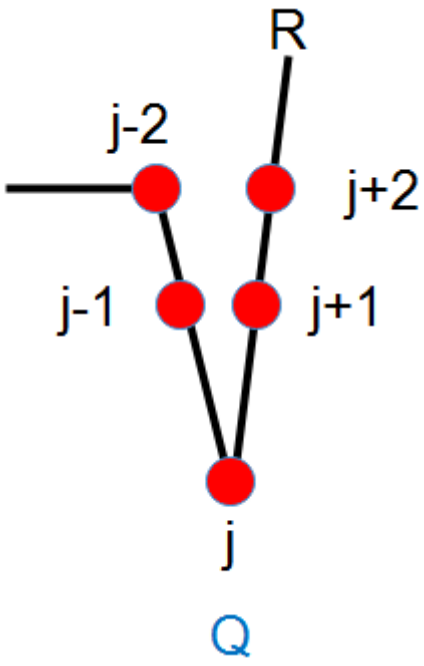


Figure 5.3. Illustration of Q points [29].

In the same way as identifying Q points, S points of QRS complex are identified by scanning right through all R peaks. When a point in the right direction of the R peak is lower than two points before and after it, this point is held in the S list [29], as shown in Figure 5.4.

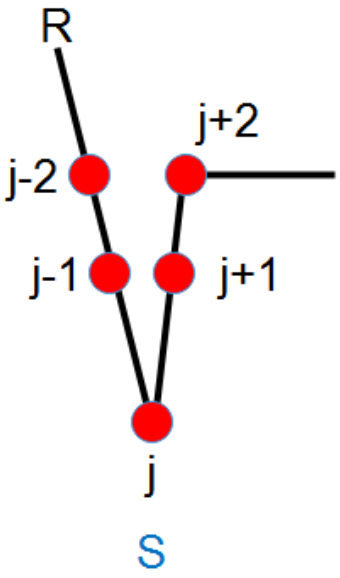


Figure 5.4. Illustration of S points [29].

5.2.4 Feature Extraction

Owing to the dynamic feature of cardiac arrhythmias and non-stationary, the exact ECG extraction features are extremely significant. Grey-Level Co-Occurrence Matrix (GLCM) is introduced by Haralick et al. [37]. Its exceptional statistical extraction power has been commonly used for texture classification in health care and disease ratings [38-40].

The use of probability for co-occurrence using GLCM to derive different grey texture characteristics. It is defined as a two-dimensional histogram of grey levels divided by a fixed spatial ratio in a pair of pixels. The GLCM of an image is determined by the use of a displacement vector (d), defined by its radius β and orientation (a) [41]. For an image, Haralick extracted 7 GLCM texture features representing the following: Homogeneity, contrast, angular second moment, dissimilarity, entropy, correlation, and variance[42-44].

$$\text{Homogeneity (HOM)} = \sum_{i,j=0}^{N-1} i \frac{P_{i,j}}{1+(i-j)^2} \tag{5.10}$$

$$\text{Contrast (CON)} = \sum_{i,j=0}^{N-1} iP_{i,j}(i-j)^2 \tag{5.11}$$

$$\text{Dissimilarity (DIS)} = \sum_{i,j=0}^{N-1} iP_{i,j|i-j|} \tag{5.12}$$

$$\text{Entropy (ENT)} = \sum_{i,j=0}^{N-1} iP_{i,j}(-\ln P_{i,j}) \tag{5.13}$$

$$\text{Variance (VAR)} = \frac{\sum_{i,j=0}^{N-1}(P_{i,j}-\mu)^2}{N-1} \tag{5.14}$$

$$\text{Angular Second Moment (ASM)} = \sum_{i,j=0}^{N-1} iP_{i,j}^2 \tag{5.15}$$

$$\text{Correlation (COR)} = \sum_{i,j=0}^{N-1} ijP_{i,j} - \mu_1\mu_2 \tag{5.16}$$

$$\mu_1 = \sum_{i=0}^{N-1} i \sum_{j=0}^{N-1} P_{i,j} \tag{5.17}$$

$$\mu_2 = \sum_{i=0}^{N-1} j \sum_{j=0}^{N-1} P_{i,j} \tag{5.18}$$

$$\sigma_1^2 = \sum_{i=9}^{N-1}(i-\mu)^2 \sum_{j=0}^{N-1} P_{i,j} \tag{5.19}$$

HOM tests the lack of heterogeneity or local similitude in the scene. Small grey-coloured variations in pair components suggest high HOM values. CON is a calculation of the amount of local pixel value difference between neighbouring pixels. It is strong in areas with big local differences and is the other way around HOM. DIS is CON-like and HOM-like reciprocal. If the place has a high CON, it is high. ENT is a disorder calculation in an image. If the picture is uniform or heterogeneous, the ENT is smaller. It is the other way around ASM. VAR is elevated if the grey level in the local area is a broad standard deviation. ASM and uniformity are the uniformity dimensions and the repeat pixel pair. It is strong when GLCM is homogeneous locally; it is similar to Heter. COR is a grey-level estimate of linear image dependencies. High COR values suggest that the grey pixel pairs are linear [45].

5.3 MACHINE LEARNING

Machine learning and artificial intelligence branch enable machines to work with intelligent software cleverly. The methods of statistical learning are the foundation of

intelligent applications for the development of machine intelligence. Because machine learning algorithms need the dataset to be created, the discipline must be connected to the database discipline. Also, there are well-known concepts such as data discovery and pattern recognition, Information Discovery from Data (KDD) [46].

Machine learning algorithms are often categorized as supervised or unsupervised.

- Supervised machine learning algorithms can apply what has been realized in the past to new data utilizing labelled models to assume future trials. Starting from the analysis of a recognized training dataset, the learning algorithm creates an inferred function to create expectations about the output values. Consequently, the system is capable of giving targets for any new input after sufficient training. The learning algorithm can also evaluate its output with the correct, expected output and locate errors to adjust the model accordingly.
- In contrast, unsupervised machine learning algorithms are utilized when the information employed to train is neither classified nor labelled. Unsupervised learning studies how systems can propose a function to describe a hidden structure from unlabelled data. The system does not figure out the correct output, but it investigates the data and can draw inferences from datasets to describe invisible structures from unlabelled data.

To analyse and classify the obtained results in this thesis, three algorithms have been used in this study, as described below:

5.3.1 Linear Discriminant Analysis (LDA)

In 1936, Fisher used a technique of discrimination using a dimension reduction to distinguish objects into distinct groups [47-49]. The LDA is an intuitive approach that defines class assignment by determining the linear transformation of data into feature space, optimizing the ratio of its variance between classes to their variance within the class, and obtaining the largest class separation, as seen in Figure. 5.5.

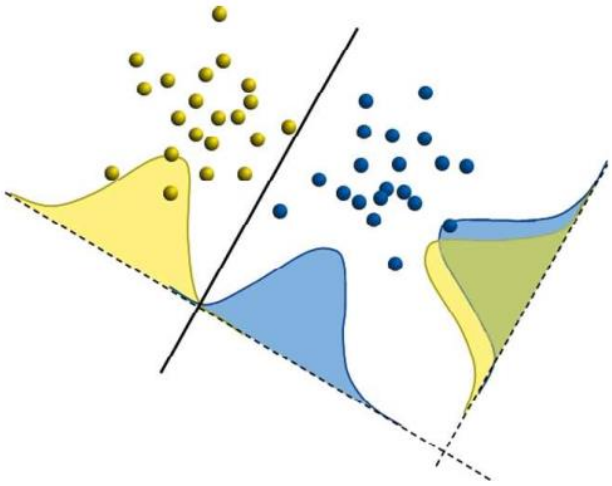


Figure 5.5. LDA transformation and the optimal linear decision boundary [49].

The result is a linear boundary for determination, identical to that defined by maximum probability discrimination, which is outstanding (in Bayesian sense) when the assumptions underlying multivariate normality are met, and matrices with equal covariance are met [50]. The maximal class division in the two-class case can be seen when the coefficient parameter, w, for defining the linear transformation is used as follows:

$$\mathrm{W} = \sum^{-1}(\mu_1 - \mu_0) \tag{5.20}$$

Where μk is the typical matrix of covariance and where μk is the mean class k vector. Usually, we estimate §1 with the combined estimate of covariance, SW, provided by only restricted details.

$$S_W = \sum_{K=1}^{N} S_K \ , \ S_K = \ \sum_{i=1}^{N_K} (X_{Ki} - \bar{X}_K)(X_{Ki} - \bar{X}_K)^T \tag{5.21}$$

In the above equations, **x***ki* and **x***k* denote the *i*th training sample of class *k*, and the corresponding class means.

5.3.2 Random Forests (RF)

In 1995, Tin Kam Ho of Bell Labs suggested the RF judgment conception for the first time [51, 52]. Leo Breiman used the more general term RF to characterize the overall method, later generalized and formalized this model [53]. In [53] Breiman showed that RFs are highly efficient. Still, they also cope easily with many problems, complicating and influencing other classification methodologies used throughout different application domains. In specific, the random forests do not require simplifying assumptions concerning data and error distribution models.

$$PE^* \leq \frac{\bar{\rho}(1-S^2)}{S^2} \tag{5.22}$$

$$S = 1 - 2\,.\,PE^*_{tree} \tag{5.23}$$

Where ρ denotes the mean tree prediction association, s stands for tree power, and PE*tree is the predicted tree classification widespread error.

5.3.3 Decision Trees (DT)

It is a supervised learning method used to solve classification complications mostly. It works for both categorical and continuous input and output variables. The model learns simple decision instructions inferred from its data features and then expects the value of a target variable. In other words, the population or sample is split into two or more homogeneous sets (or sub-populations) based on the most significant differentiator in input variables.

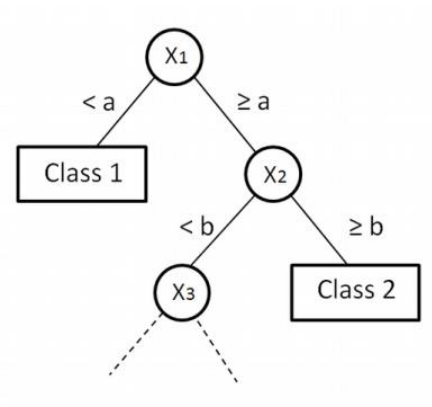


Figure 5.6. The decision tree [54].

Figure 5.6. has a typical scheme for the DT. The judgment nodes are the X variables. each node has an attribute, variables a and b are the limits of the features that separate the decision into three tree paths.

Nominal or interactive. The class variables are the tree leaves, so the entity under analysis can be labelled [54] in equation 5.21. H(S) is an entropy that measures the quantity of variability of the data in the dataset S [55-57].

$$\mathrm{H(S)} = \sum_{x \in X} -p(x)\ \log_2 p(x) \tag{5.24}$$

In equation 5.24, IG(A) is information gain that measures the entropy's difference after dividing set S on attribute A [55-57].

$$\mathrm{IG\ (S, A)} = \mathrm{H(S)} - \sum_{t \epsilon T} P(t)H(t) = H(S) - H(S|A) \tag{5.25}$$

The decision tree identifies the most significant variable and its value that gives the best homogeneous sets of the population [57-59].

5.4 CROSS-VALIDATION

It introduces the most common method for model calculation and model selection in machine learning practice: k-fold cross-validation. The term cross-validation is referred to the train/test holdout method as a cross-validation technique. However, it

might make more sense to think of cross-validation as a crossing overtraining and validation stages in successive rounds. Here, the core idea behind cross-validation is that each sample in our dataset has the chance of being tested. k-fold cross-validation is a particular case of cross-validation where iterate over a dataset set k times. In each round, the dataset was divided into K parts: one part is used for validation. The remaining k − 1 part is combined into a training subset for model evaluation, as shown in Figure 5.7, which illustrates the process of 5-fold cross-validation [60].

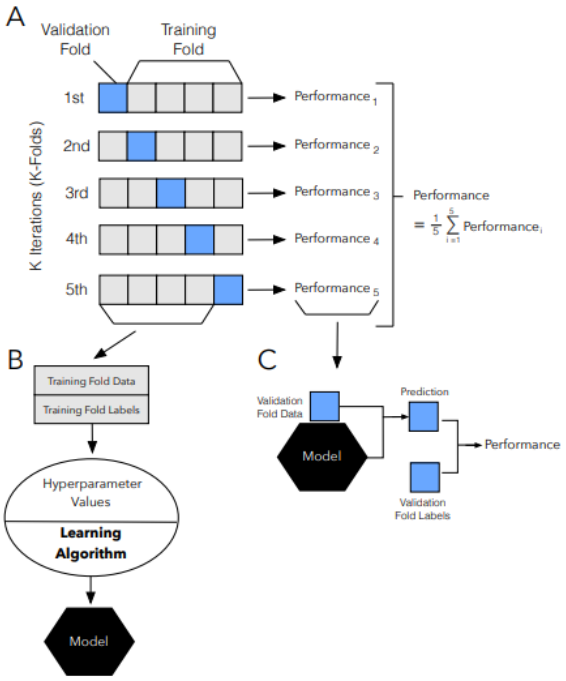


Figure 5.7. Illustration of the 5-fold cross-validation procedure [60].

5.5 CONFUSION MATRIX

A machine learning role that involves a solution for a binary classification problem is where our study developed. The data collection representing the task consists of N+ examples in one class, positive and negative in the other. For example, the stable people are typically branded negatively in a biomedical case-control report, while the positive mark is commonly assigned to the patients. The positive class correlates to the pathological phenotype, given two phenotypes. For example, in various stages of a tumour, this rating is essential.

The classification model predicts each data instance, attributing its expected mark to each sample (positive or negative). Thus, each sample is dropped in one of the following four cases at the end of the classification procedure:

- If the program says positive and the experiments say positive, it's true positive (TP).
- If the program says positive and the experiments say negative, it's false positive (FP).
- If the program says negative and the experiments say positive, it's false negative (FN).
- If the program says negative and the experiments say negative, it's true negative (TN).

The consequence of the classification task can be represented entirely in a 5.1. table labelled uncertainty matrix $M = \frac{\text{TP FP}}{\text{FN TN}}$ extended in Table 5.1.

Table 5.1. The standard confusion matrix M.

	PREDICTED POSITIVE	**PREDICTED NEGATIVE**
Actual positive	TP	FN
Actual negative	FP	TN

TP and TN are the correct predictions, while FN and FP are the incorrect predictions.

If M= $\begin{pmatrix} n^+ & 0 \\ 0 & n^- \end{pmatrix}$, the grading is perfect.

Since it would be time-intensive to examine the entire four types of uncertainty matrix individually, statistics implemented some applicable statistical rates that could instantly explain the consistency of a forecast.

Sensitivity refers to the test's capability to record ill patients who do have the condition correctly. In the example of a medical test used to detect the condition, the sensitivity (sometimes also named the detection rate in a clinical setting) is the proportion of

people who test positive for the disease among those who have the disease. Mathematically, this can be expressed as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+FN}} \tag{5.26}$$

Specificity refers to the test's capability to reject healthy patients without a condition correctly. Specificity of a test is the proportion of those who truly do not have the disease who test negative for the disease. Mathematically, this can also be written as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{5.27}$$

Many researchers think that accuracy computation is the normal approach for global metrics of M inputs or higher. In particular, precision refers to the ratio of correctly forecast instances to all instances in the dataset:

$$\text{Accuracy} = \frac{TP + TN}{n^+ + n^-} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.28}$$

(worst accuracy: 0; Best accuracy: 1.)

For any confusion matrix M and ranges, by definition, precision is specified at the real Unit Interval [0,1]; the best value of 1,00 is M=$\begin{pmatrix} n^+ & 0 \\ 0 & n^- \end{pmatrix}$ and the worst value of 0,00 is perfected M= $\begin{pmatrix} 0 & n^+ \\ n^- & 0 \end{pmatrix}$ [61].

The Matthews correlation coefficient (MCC) is the contingency matrix determined in the Pearson product-moment coefficient of correlation [61, 62] between real and projected value as an alternate measure not influenced by the unbalanced data set problem. MCC reads as follows about the entries of M:

$$\text{MCC} = \frac{TP.TN - FP.FN}{\sqrt{(TP+FP).(TP+FN).(TN+FP).(TN+FN)}} \tag{5.29}$$

(worst value: -1; best value: +1)

MCC is the only high score binary classification rate achieved if the binary predictor can reliably forecast most positive data instances and most negative data instances [63, 64]. In the case of perfect misclassification or perfect classification, the spectrum is [−1, +1], the extreme values – 1 and +1, respectively.

PART 6

RESULTS AND DISCUSSION

The MIT-BIH database consists of eight cases, normal ECG signals (NSR) and abnormal ECG signals such as (AFIB), (AFL), (APB), (LBBB), (RBBB), (PVC), and (WPW)). All records in. mat format is available. These files are read and processed using MATLAB software. This thesis section discusses an algorithm built in the ECG single-lead signal to identify and classify ECG signals as normal and abnormal QRS complexes.

The obtained results of each step of the algorithm were illustrated in Figure 6.1.

- Step a: The patient acquires a raw ECG signal. Then, the crude ECG signal is uploaded and plotted using Matlab software (Figure 6.1(a)).
- Step b: The crude ECG signal is frequently influenced by such disruptions as contact with the power line, external electromagnetic fields, random body movements or respiration [65]. In this step, the ECG signals are preprocessed for noise removal using the zero-phase and moving average filters (Figure 6.1(b)).
- Step c: The K-means clusters method identifies the R peaks value (Figure 6.1(c)).
- Step d and e: Based on the tracking local extrema points, the Q and S peaks were detected, whereas the Q peak located on the left side of R peak (step d) and S peak situated on the right side of R-peak (step e).

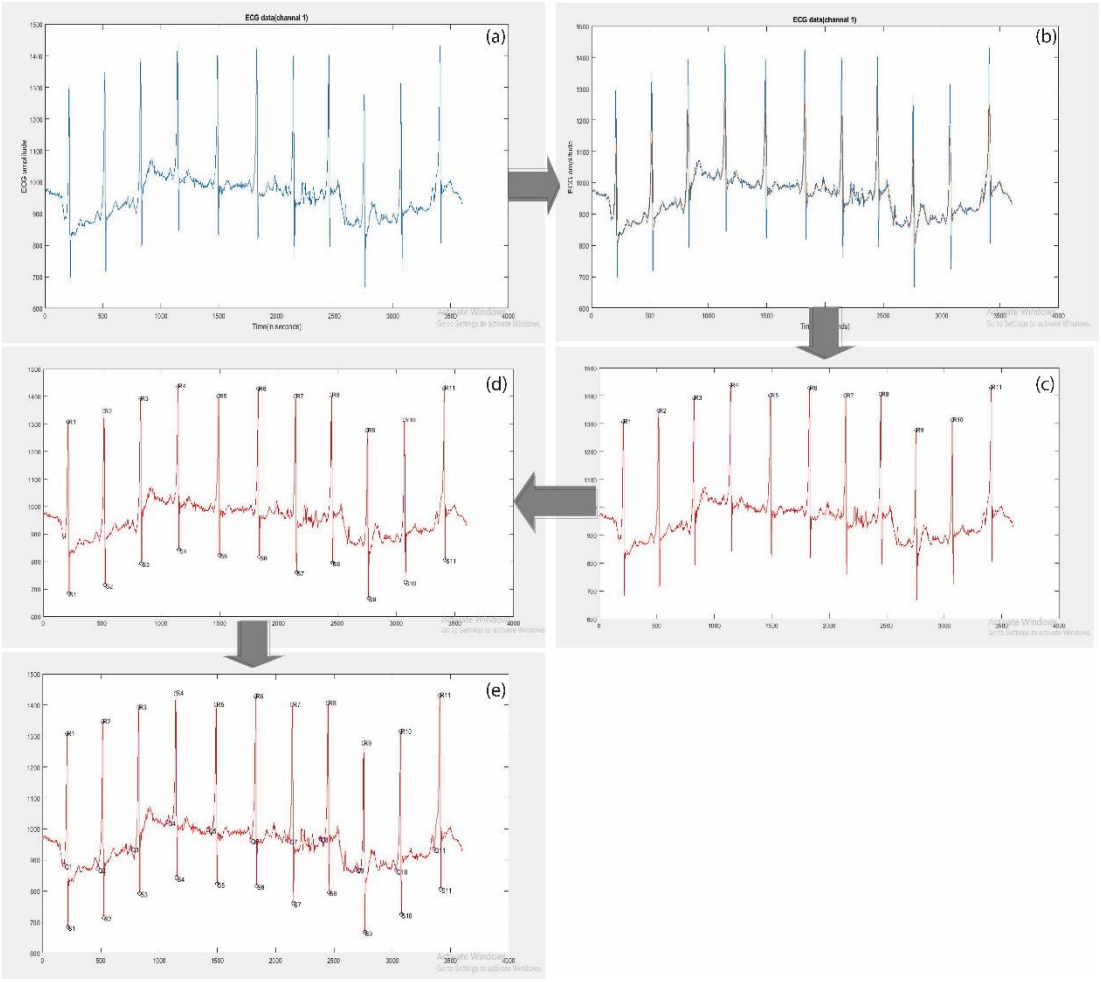
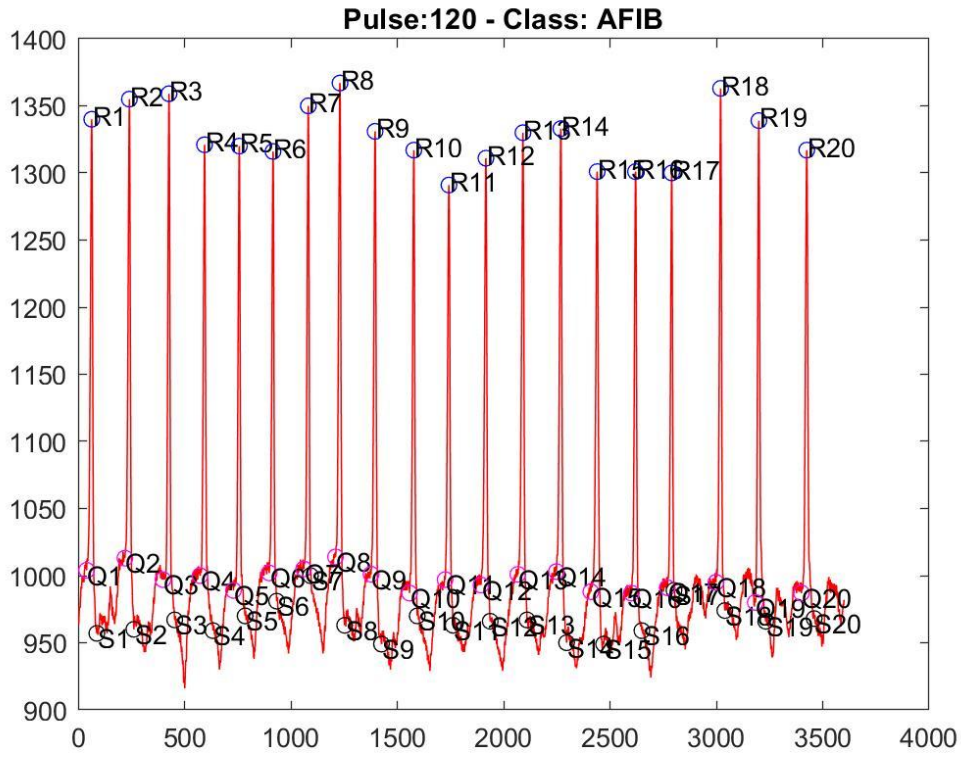


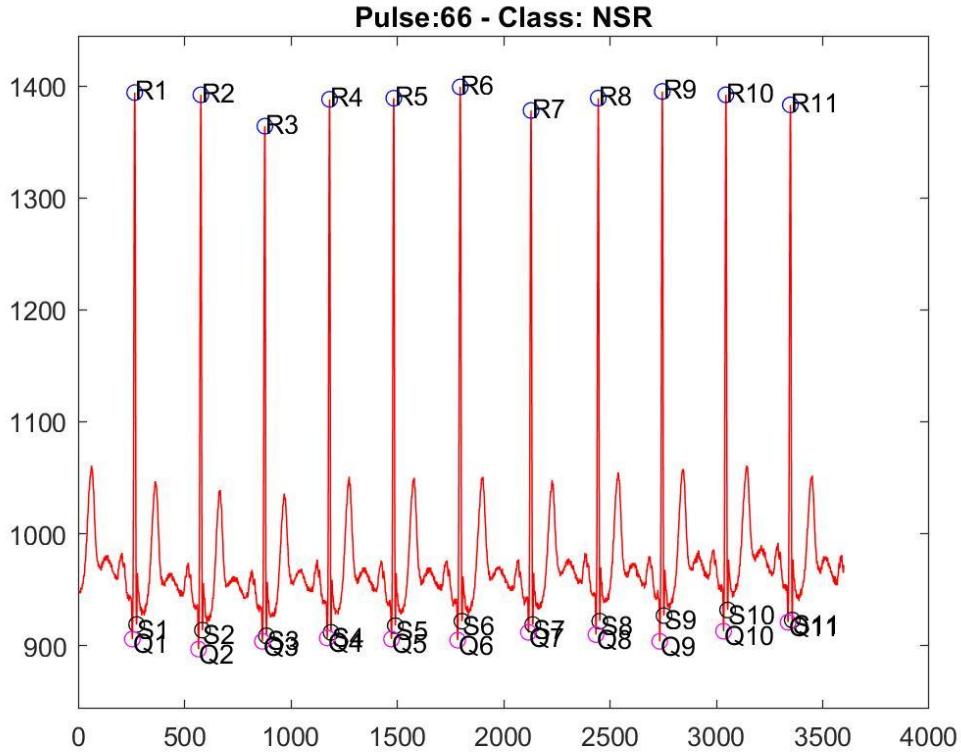
Figure 6.1. (a) raw ECG, (b) filtered ECG, (c) R-peak detection, (d) S-peak detection, and (e) Q-peak detection.

One sample ECG signal for each class is plotted. The QRS complex peaks are marked, and the pulse rate was measured as shown in Figure 6.2.

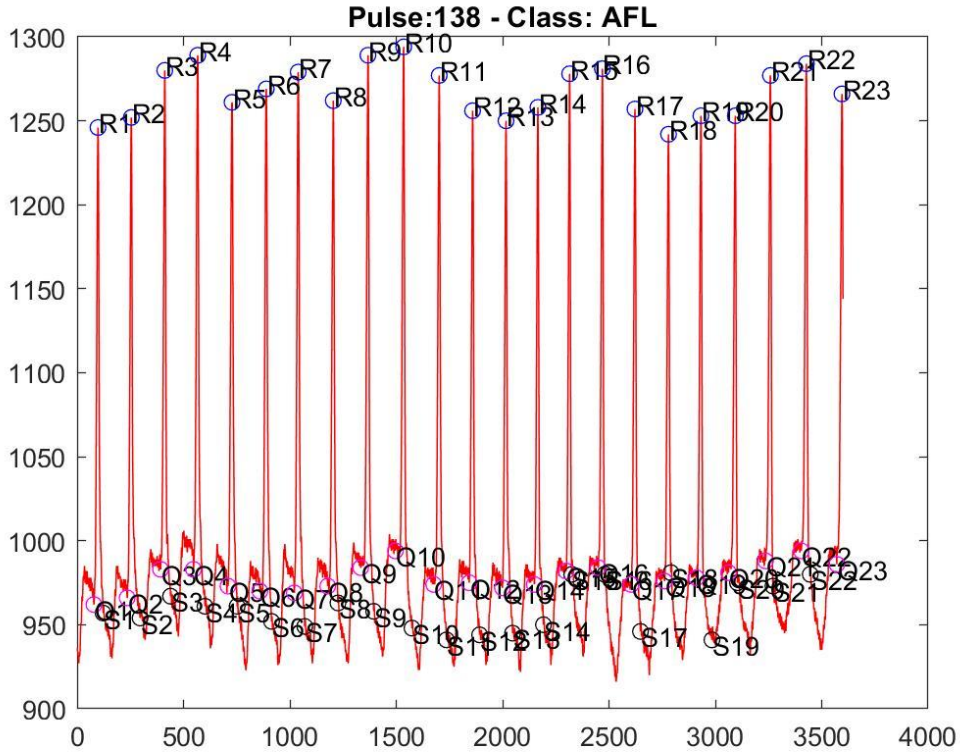
(a)



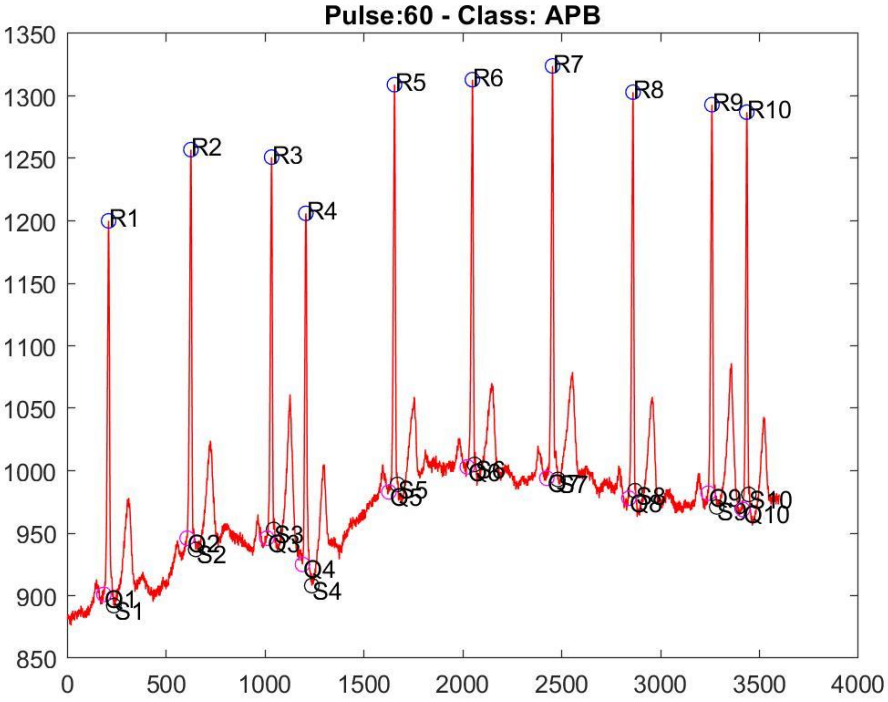
(b)



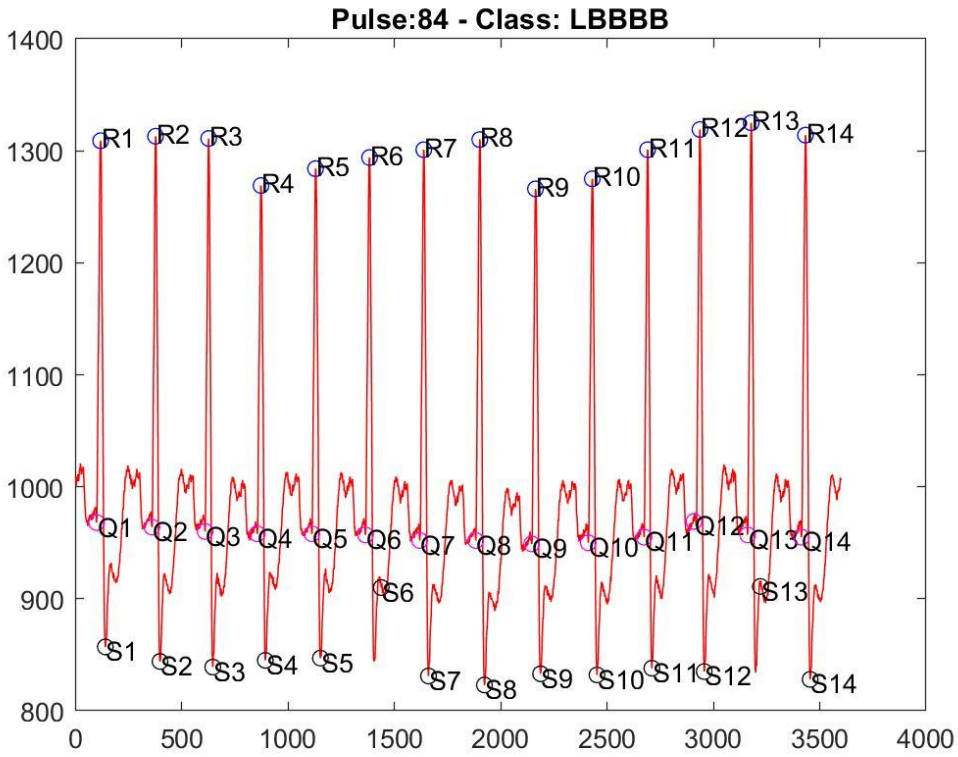
(c)


Pulse:138 - Class: AFL
1300
1250
1200
1150
1100
1050
1000
950
900
0
500
1000
1500
2000
2500
3000
3500
4000

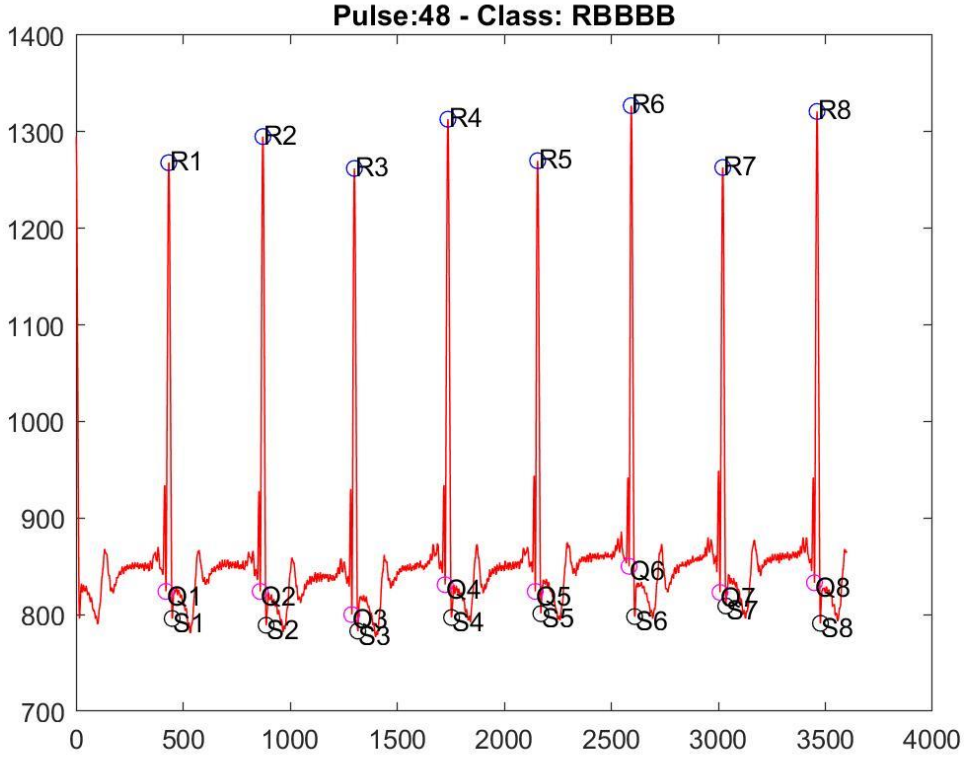
(d)


Pulse:60 - Class: APB
1350
1300
1250
1200
1150
1100
1050
1000
950
900
850
0
500
1000
1500
2000
2500
3000
3500
4000

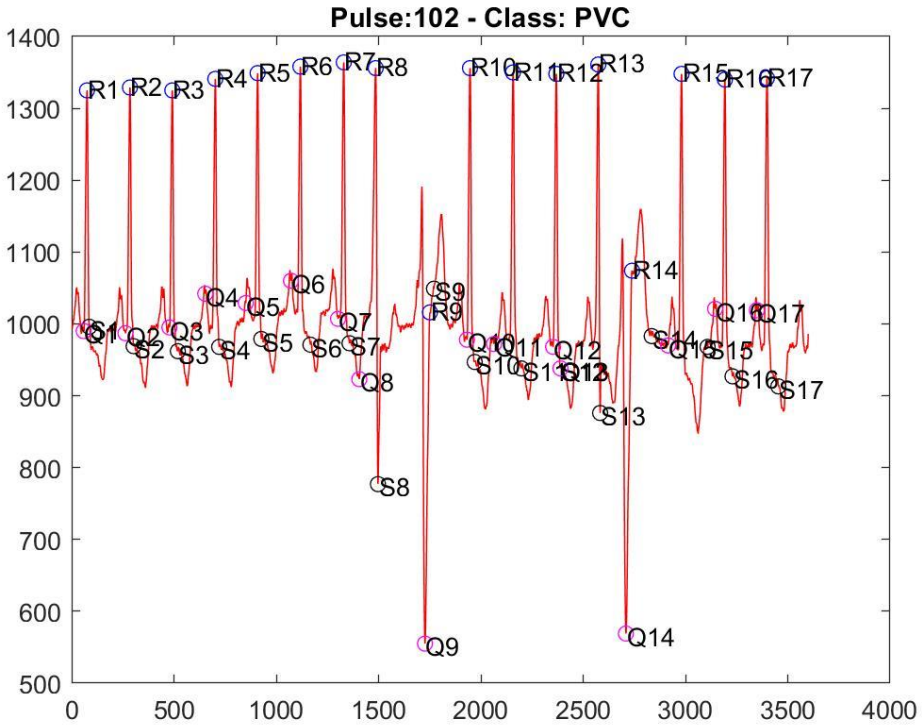
(e)


Pulse:84 - Class: LBBBB
1400
1300
1200
1100
1000
900
800
0
500
1000
1500
2000
2500
3000
3500
4000
R1 R2 R3 R4 R5 R6 R7 R8 R9 R10 R11 R12 R13 R14
Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8 Q9 Q10 Q11 Q12 Q13 Q14
S1 S2 S3 S4 S5 S6 S7 S8 S9 S10 S11 S12 S13 S14

(f)


Pulse:48 - Class: RBBBB
1400
1300
1200
1100
1000
900
800
700
0
500
1000
1500
2000
2500
3000
3500
4000
R1 R2 R3 R4 R5 R6 R7 R8
Q1 Q2 Q3 Q4 Q5 Q6 Q7 Q8
S1 S2 S3 S4 S5 S6 S7 S8

(g)



(h)

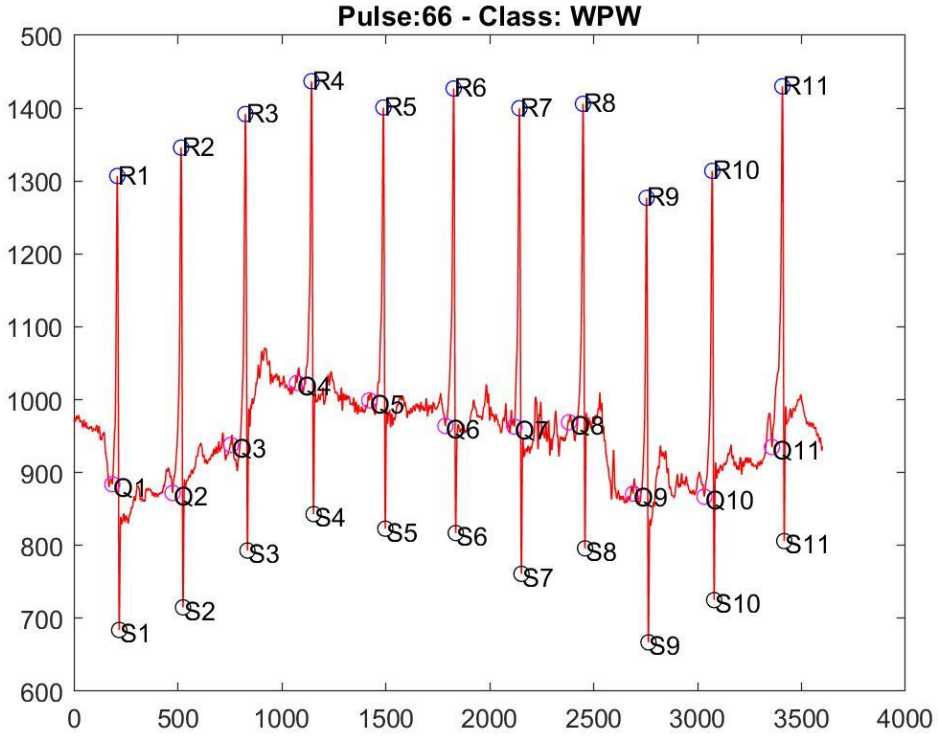


Figure 6.2. ECG signals whose QRS complex are marked. (a) NSR, (b) AFIB, (c) AFL, (d) APB, (e) LBBBB, (f) RBBBB, (g) PVC, and (h) WPW.

The average of sensitivity, specificity, and accuracy of detection of QRS points are respectively computed as 97.04%, 92.68%, and 95.81% for (R), 91.30%, 95.22% and 94.02% for (Q) and 91.65%, 95.05% and 93.95% for (S). The sensitivity, specificity, and accuracy of QRS complex segmentation are summarized and compared with other studies as presented in Table 6.1.

In this thesis, the K- mean clustering method and tracking local extrema points are proposed to detect the QRS complex segmentation. The obtained results were comparable to other works which investigated the same topic. In Table 6.1, the results of the proposed method achieved 93.33% of sensitivity, 94.31% of specificity, and 94.59% of accuracy, which are nearly on par with other popular methods. The proposed process takes 0.42 seconds on average in a full-automatedly way. Comparison based on execution time could not be made because it was not shared in the literature [29]. This thesis deduced that K- means clustering method and tracking local extrema points have the potential to improve the performance in QRS complex detection.

Table 6.1. QRS complex segmentation comparison.

AUTHORS (YEAR)	**DATABASE**	**METHOD**	**RESULT**
Sehirli and Turan (2021) [29]	MIT-BIH	Consecutive difference method, k-means clustering, tracking local extreme points.	95.57% accuracy, 94.75% sensitivity, 95.96% specificity.
Shaik et al. (2015) [27]	MIT-BIH	Utilize an adaptive threshold technique on spectrogram computed using STFT.	99.56% sensitivity, 99.52% specificity.
Uchaipichat et al. (2010) [30]	MIT-BIH	STFT	99.10% sensitivity,

AUTHORS (YEAR)	DATABASE	METHOD	RESULT
			99.60% specificity.
Merino et al. (2015) [66]	European ST-T, Fantasia DB, MIT-BIH	Envelopment filter and K-means	99.77% accuracy, 99.92% sensitivity.
Subramanian (2017) [67]	MIT-BIH	Multiwavelet transform	93.35% accuracy, 98.5% sensitivity
Curtin et al. (2018) [68]	MIT-BIH	Windowing algorithm	94.3% accuracy, 96% sensitivity
Darrington (2006) [31]	MIT-BIH	Fine-to-coarse algorithm	99.00% sensitivity, 99.20% specificity.
The proposed method	MIT-BIH	k-means clustering and tracking local extrema points	93.33% sensitivity, 94.31% specificity, 94.59% accuracy,

A hybrid model based on machine learning models is developed to classify ECG as normal and abnormal signals. Three algorithms (DT, RF, and LDA) have been used to classify ECG signals. With the 5-fold cross-validation, performance metrics ECG signals are calculated and presented in Table 6.2. Overall sensitivity, specificity, and accuracy of classification of ECG signals are respectively computed as 89.9%, 90.0% and 90.0% for Decision tree, 96.0%, 85.4% and 90.5% for random forest, 84.5%,

74.8% and 79.6% for LDA. The highest value of sensitivity is recorded for RF, while the lowest value is recorded for LDA. The highest specificity value was recorded for DT, while the lowest value of specificity was recorded for LDA. The highest accuracy value was recorded for RF, while the lowest value of accuracy was recorded for LDA. This study concluded that the performance metrics features obtained from the RF model were the highest compared to other models, which later confirmed using AUC-ROC graph.

Table 6.2. Performance metrics of classification of ECG signals.

DT										
	TP	**TN**	**FP**	**FN**	**Train Set**	**Test Set**	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**	**MCC**
Cross Val1	109	90	13	14	340	226	88.618	87.379	88.053	0.759
Cross Val2	109	95	10	12	340	226	90.083	90.476	90.265	0.805
Cross Val3	99	105	9	13	340	226	88.393	92.105	90.265	0.806
Cross Val4	106	101	11	8	340	226	92.982	90.179	91.593	0.832
Cross Val5	106	97	11	12	340	226	89.831	89.815	89.823	0.796
Total	529	488	54	59	1700	1130	89.966	90.037	90.000	0.800
RF										
	TP	**TN**	**FP**	**FN**	**Train Set**	**Test Set**	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**	**MCC**
Cross Val1	106	98	14	8	340	226	92..982	87.500	90.265	0.806
Cross Val2	106	100	18	3	340	226	97.222	84.746	90.708	0.822
Cross Val3	105	105	14	2	340	226	98.131	88.235	92.920	0.864
Cross Val4	97	103	18	8	340	226	92.381	85.124	88.496	0.773
Cross Val5	112	92	21	1	340	226	99.115	81.416	90.265	0.818
Total	525	498	85	22	1700	1130	95.978	85.420	90.531	0.816

LDA										
	TP	**TN**	**FP**	**FN**	**Train Set**	**Test Set**	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**	**MCC**
Cross Val1	98	85	17	26	340	226	79.032	83.333	80.973	0.621
Cross Val2	97	79	38	12	340	226	88.991	67.521	77.876	0.576
Cross Val3	94	88	24	20	340	226	82.456	78.571	80.531	0.611
Cross Val4	87	86	32	21	340	226	80.556	72.881	76.549	0.535
Cross Val5	102	84	31	9	340	226	91.892	73.043	82.301	0.660
Total	478	422	142	88	1700	1130	84.452	74.823	79.646	0.596

AUC (Area under the ROC Curve) - ROC (receiver operating characteristic) curve is a performance measurement for the classification problems at various threshold settings. ROC is a possibility curve, and AUC represents the degree or measure of separability. It tells how much the model is capable of distinguishing between classes. The higher the AUC, the better the model is at predicting. By analogy, the Higher the AUC, the better the model determines patients with the disease and no disease.

The AUC of DT, RF, and LDA models was 9000, 9070, and 7964. The AUC for RF was significantly higher than other models that matched performance metrics calculations (Table 6.2).

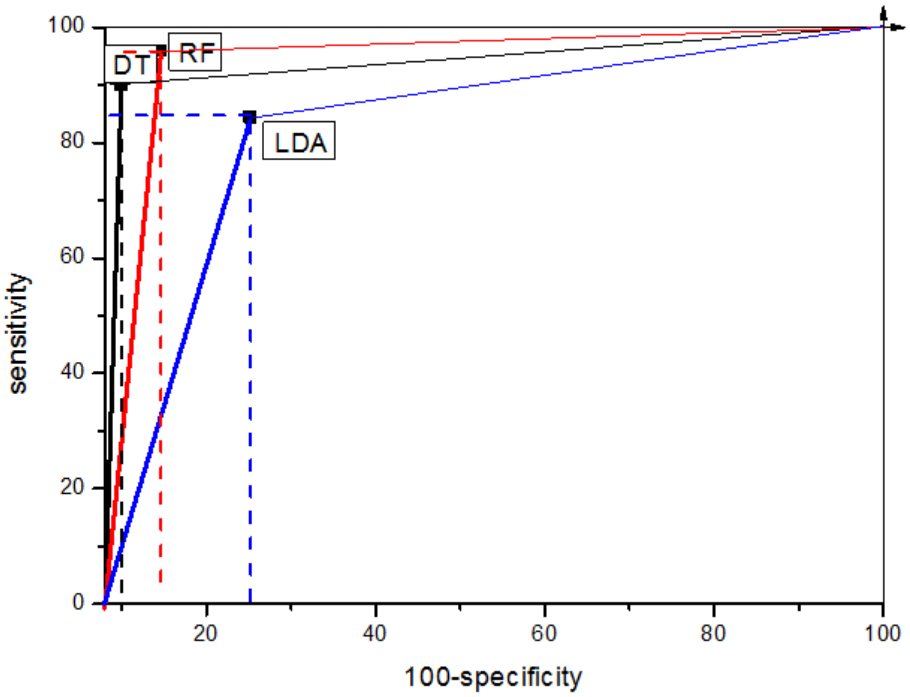


Figure 6.3. ROC curves.

PART 7

CONCLUSION

The critical stage in ECG diagnosis identifies and measures different waves that compose the entire ECG loop, effectively extracting and finding suitable structures and algorithms for the ECG signal classification. To this end, various approaches have been validated for ECG signal processing and interpretation. The features selected were suited for diagnosing eight cardiac conditions and increased the machine learning performance, simplicity, and recognition rate:

- The ECG signals were successfully uploaded and smoothed using the zero-phase filter and moving average filter.
- The QRS complex segmentation was successfully detected using the K-means clustering algorithm and tracking local extrema points.
- Performance metrics of classification of ECG signals were achieved using DT, RF, and LDA models.
- The values of accuracy and MCC calculated using RF were significantly higher than others.
- All performance metrics features obtained from the LDA model were the lowest compared to other models.
- The AUC of DT, RF, and LDA models was 9000, 9070, and 7964, respectively. The AUC for RF was significantly higher than other models.
- Based on the obtained results, the classification of ECG signals using the RF model was the best compared to other models.

REFERENCES

1. A.R. Dehaghani, "Uncertainty Quantification and Reduction in Cardiac Electrophysiological Imaging", Thesis. ***Rochester Institute of Technology,*** Rochester, New York ,(2015).

2. R.J. Martis, M.M.R. Krishnan, C. Chakraborty, S. Pal, D. Sarkar, K. Mandana, A.K. Ray, "Automated screening of arrhythmia using wavelet based machine learning techniques", ***Journal of medical systems***, 36(2): 677-688 (2012).

3. E.N. Bruce, "Biomedical signal processing and signal modeling", ***Wiley New York***, (2001).

4. H. Gholam-Hosseini, H. Nazeran, K.J. Reynolds, "ECG noise cancellation using digital filters", ***Proceedings of the 2nd International Conference on Bioelectromagnetism,*** Melbourne, VIC, Australia, 151-152 (1998).

5. S. Shalev-Shwartz, S. Ben-David, "Understanding machine learning: From theory to algorithms", ***Cambridge university press,*** (2014).

6. L. Sathyapriya, L. Murali, T. Manigandan, "Analysis and detection R-peak detection using Modified Pan-Tompkins algorithm", 2014 ***IEEE International Conference on Advanced Communications, Control and Computing Technologies,*** Ramanathapuram, India, 483-487 (2014).

7. A. Sheetal, H. Singh, A. Kaur, "QRS detection of ECG signal using hybrid derivative and MaMeMi filter by effectively eliminating the baseline wander", ***Analog Integrated Circuits and Signal Processing,*** 98(1):1-9 (2019).

8. L. Sörnmo, P. Laguna, "Bioelectrical signal processing in cardiac and neurological applications", ***Academic Press***, (2005).

9. M.J. Curtis, J.C. Hancox, A. Farkas, C.L. Wainwright, C.L. Stables, D.A. Saint, H. Clements-Jewery, P.D. Lambiase, G.E. Billman, M.J. Janse, "The Lambeth Conventions (II): guidelines for the study of animal and human ventricular and supraventricular arrhythmias", ***Pharmacology & therapeutics,*** 139(2): 213-248 (2013).

10. S. Assaf, C. Libby, "Asymptomatic Wolff-Parkinson-White Syndrome: Incidental EKG", ***Journal of Education and Teaching in Emergency Medicine***, 2(3), (2017).

11. Scherbak D, Hicks GJ. "Left Bundle Branch Block". ***In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing***, (2020). PMID: 29489192.

12. M.H. Nikoo, A. Aslani, M.V. Jorat, LBBB: state-of-the-art criteria", ***International Cardiovascular Research Journal***, 7(2): 39–40 (2013).

13. V. Kumar, R. Venkataraman, W. Aljaroudi, J. Osorio, J. Heo, A.E. Iskandrian, F.G. Hage, "Implications of left bundle branch block in patient treatment", ***The American Journal of Cardiology***, 111(2) : 291-300 (2013).

14. S. Bhattacharyya, U. Snekhalatha, "Classification of right bundle branch block and left bundle branch block cardiac arrhythmias based on ECG analysis", ***Artificial Intelligence and Evolutionary Algorithms in Engineering Systems, Springer***, 331-341 (2015).

15. M.P. Bonomini, P.D. Arini, G.E. Gonzalez, B. Buchholz, M.E. Valentinuzzi, "The allometric model in chronic myocardial infarction", ***Theoretical Biology and Medical Modelling***, 9(1) : 1-13(2012).

16. B.-U. Kohler, C. Hennig, R. Orglmeister, "The principles of software QRS detection", ***IEEE Engineering in Medicine and biology Magazine,*** 21(1): 42-57(2002).

17. P.S. Hamilton, W.J. Tompkins, "Compression of the ambulatory ECG by average beat subtraction and residual differencing", ***IEEE Transactions on Biomedical Engineering,*** 38(3): 253-259 (1991).

18. S.M. Jalaleddine, C.G. Hutchens, R.D. Strattan, W.A. Coberly, "ECG data compression techniques-a unified approach", ***IEEE transactions on Biomedical Engineering,*** 37(4) : 329-343(1990).

19. H. Al-Nashash, A dynamic Fourier series for the compression of ECG using FFT and adaptive coefficient estimation", ***Medical engineering & physics,*** 17(3): 197-203(1995).

20. G. Nave, A. Cohen, "ECG compression using long-term prediction", ***IEEE transactions on Biomedical Engineering,*** 40(9) : 877-885 (1993).

21. P. de Chazal, Automatic classification of the Frank lead electrocardiogram", Thesis, ***University of New South Wales,*** Kensington, Australia, (2000).

22. L.C. Pretorius, C. Nel, Feature extraction from ECG for classification by artificial neural networks", ***Proceedings Fifth Annual IEEE Symposium on Computer-Based Medical Systems,*** Durham, NC, USA**,** (1992).

23. S. Hamdi, A.B. Abdallah, M.H. Bedoui, "A robust QRS complex detection using regular grammar and deterministic automata", ***Biomedical Signal Processing and Control,*** 40: 263-274 (2018).

24. M. Adam, S.L. Oh, V.K. Sudarshan, J.E. Koh, Y. Hagiwara, J.H. Tan, R. San Tan, U.R. Acharya, "Automated characterization of cardiovascular diseases using

relative wavelet nonlinear features extracted from ECG signals", ***Computer methods and programs in biomedicine,*** 161:133-143 (2018).

25. Z. Zhang, Q. Yu, Q. Zhang, N. Ning, J. Li, "A kalman filtering based adaptive threshold algorithm for qrs complex detection", ***Biomedical Signal Processing and Control,*** 58: 101827 (2020).

26. Ö. Yakut, E.D. Bolat, "An improved QRS complex detection method having low computational load", ***Biomedical Signal Processing and Control,*** 42: 230-241 (2018).

27. B.S. Shaik, G. Naganjaneyulu, T. Chandrasheker, A. Narasimhadhan, "A method for QRS delineation based on STFT using adaptive threshold", ***Procedia Computer Science,*** 54: 646-653 (2015).

28. D. Berwal, A. Kumar, Y. Kumar, "Design of high performance QRS complex detector for wearable healthcare devices using biorthogonal spline wavelet transform", ***ISA transactions,*** 81: 222-230 (2018).

29. E. Sehirli, M.K. Turan, " A Novel Method for Segmentation of QRS Complex on ECG Signals and Classify Cardiovascular Diseases via a Hybrid Model Based on Machine Learning", ***International Journal of Intelligent Systems and Applications in Engineering,*** 9(1):12-21(2021).

30. N. Uchaipichat, S. Inban, "Development of QRS detection using short-time fourier transform based technique", ***IJCA Journal,*** 7-10(2010).

31. J. Darrington, "Towards real time QRS detection: A fast method using minimal pre-processing", ***Biomedical Signal Processing and Control,*** 1(2): 169-176 (2006).

32. Z.F.M. Apandi, R. Ikeura, S. Hayakawa, "Arrhythmia detection using MIT-BIH dataset: A review", ***International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA),*** Kuching, Malaysia***,*** 1-5 (2018).

33. G.Q. Gao, "Computerised detection and classification of five cardiac conditions", Thesis, ***Auckland University of Technology***, Auckland, New Zealand, (2003).

34. Advanced Solutions Nederland B.V., "The moving average filter", ***https://www.advsolned.com/the-moving-average-filter/***, (2018).

35. S. Mehta, D. Shete, N. Lingayat, V. Chouhan, "K-means algorithm for the detection and delineation of QRS-complexes in Electrocardiogram", ***Irbm,*** 31(1): 48-54 (2010).

36. Y. Xia, J. Han, K. Wang, "Quick detection of QRS complexes and R-waves using a wavelet transform and K-means clustering", ***Bio-medical materials and engineering,*** 26(s1): S1059-S1065 (2015).

37. R.M. Haralick, K. Shanmugam, I.H. Dinstein, "Textural features for image classification", ***IEEE Transactions on systems, man, and cybernetics,*** (6): 610-621 (1973).

38. J. Rahebi, F. Hardalaç, "Retinal blood vessel segmentation with neural network by using gray-level co-occurrence matrix-based features", ***Journal of medical system,*** 38(8): 85 (2014).

39. H. Yu, J. Scalera, M. Khalid, A.-S. Touret, N. Bloch, B. Li, M.M. Qureshi, J.A. Soto, S.W. Anderson, "Texture analysis as a radiomic marker for differentiating renal tumors", ***Abdominal Radiology,*** 42(10): 2470-2478 (2017).

40. W. Sun, N. Zeng, Y. He, "Morphological arrhythmia automated diagnosis method using gray-level co-occurrence matrix enhanced convolutional neural network", ***IEEE Access,*** 7: 67123-67129 (2019).

41. D. Gadkari, "Image quality analysis using GLCM", Thesis, ***University of Central Florida***, Florida, USA, (2004).

42. S.P. Aware, "Image Retrieval Using Co-Occurrence Matrix & Texton Co-Occurrence Matrix For High Performance", ***International Journal of Advances in Engineering & Technology,*** 5(2): 280 (2013).

43. T.A. Pham, "Optimization of texture feature extraction algorithm", Thesis, ***Delft University of Technology***, Delft, Netherlands, (2010).

44. S.K.A.Z.F. Jabr, ECG Heart diseases Diagnosis in Three Cases (Normal, Bradycardia, Tachycardia) by Using GLCM and Fuzzy Logic, ***International Journal of Innovative Engineering and Emerging Technology***, 2(4), (2016).

45. J. Zhou, R.Y. Guo, M. Sun, T.T. Di, S. Wang, J. Zhai, Z. Zhao, "The Effects of GLCM parameters on LAI estimation using texture values from Quickbird Satellite Imagery", ***Scientific Reports,*** 7(1): 1-12 (2017).

46. M. Mohammed, M.B. Khan, E.B.M. Bashier, "Machine learning: algorithms and applications", ***Crc Press,*** (2016).

47. G.J. McLachlan, "Discriminant analysis and statistical pattern recognition", ***John Wiley & Sons,*** (2004).

48. R.A. Fisher, "The use of multiple measurements in taxonomic problems", ***Annals of eugenics,*** 7(2): 179-188 (1936).

49. T.D. Lemmond, A.O. Hatch, B.Y. Chen, D. Knapp, L. Hiller, M. Mugge, W.G. Hanley, "Discriminant Random Forests", ***DMIN***, 55-61 (2008).

50. K. Mardia, J. Kent, J. Bibby, "Multivariate Analysis", chapitre 2, ***San Diego: Academic Press***, (1992).

51. T.K. Ho, Random decision forests, “Proceedings of 3rd international conference on document analysis and recognition”, ***IEEE***, 278-282 (1995).

52. T.K. Ho, “The random subspace method for constructing decision forests”, ***IEEE transactions on pattern analysis and machine intelligence,*** 20(8): 832-844 (1998).

53. L. Breiman, “Random Forests Machine Learning”, ***Machine learning***, 45(1): 5-32 (2001).

54. N. dos Santos Bastos, D.F. Adamatti, C.Z. Billa, “Decision Tree to Analyses EEG Signal: A Case Study Using Spatial Activities”, ***Latin American Workshop on Computational Neuroscience, Springer***, 159-169 (2017).

55. J.R. Quinlan, “Induction of decision trees”, ***Machine learning,*** 1(1): 81-106 (1986).

56. A.J. Taggart, A.M. DeSimone, J.S. Shih, M.E. Filloux, W.G. “Fairbrother, Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo”, ***Nature structural & molecular biology,*** 19(7): 719-721 (2012).

57. S.R. Tithi, A. Aktar, F. Aleem, “Machine Learning Approach for ECG Analysis and predicting different heart diseases”, Thesis, ***BRAC University***, Dhaka, Bangladesh, (2018).

58. S.L. Salzberg, “C4. 5: Programs for machine learning” ***j. ross quinlan. morgan kaufmann publishers, inc.***, ***Springer,*** 1993: 235-240 (1994).

59. T. Hastie, R. Tibshirani, J. Friedman, “Unsupervised learning, The elements of statistical learning”, ***Springer,*** 485-585 (2009).

60. S. Raschka, “Model evaluation, model selection, and algorithm selection in machine learning”, ***arXiv preprint arXiv***:1811.12808 (2018).

61. D. Chicco, G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation”, ***BMC genomics,*** 21(1): 6 (2020).

62. D.M. Powers, Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, ***arXiv preprint arXiv***:2010.16061 (2020).

63. G. Jurman, S. Riccadonna, C. Furlanello, “A comparison of MCC and CEN error measures in multi-class prediction”, ***PLoS One,*** 7(8) e41882 (2012).

64. D. Chicco, “Ten quick tips for machine learning in computational biology”, ***BioData mining,*** 10: 35 (2017).

65. M. AlMahamdy, H.B. Riley, “Performance study of different denoising methods for ECG signals”, ***Procedia Computer Science***, 37: 325-332 (2014).

66. M. Merino, I.M. Gómez, A.J. Molina, “Envelopment filter and K-means for the detection of QRS waveforms in electrocardiogram”, ***Medical engineering & physics,*** 37(6): 605-609 (2015).

67. B. Subramanian, ECG signal classification and parameter estimation using multiwavelet transform”, ***Biomed. Res***, 28: 3187–3193 (2017).

68. A.E. Curtin, K.V. Burns, A.J. Bank, T.I. Netoff, “QRS complex detection and measurement algorithms for multichannel ECGs in cardiac resynchronization therapy patients”, ***IEEE journal of translational engineering in health and medicine,*** 6: 1-11(2018).

RESUME

Marwah ALMOZANI completed her elementary education in Basra (IRAQ). She completed high school education in ALaraeej High School (for girls) in 2010. She graduated from Iraq University College, Faculty of Engineering, Department of Computer Engineering (IRAQ) in 2015. She started the master program at Karabük University, Department of Electrical-Electronics Engineering, in 2019.