DETECTION OF PLANT DISEASES USING IMAGE PROCESSING WITH MACHINE LEARNING

2023
MASTER THESIS
COMPUTER ENGINEERING

Raghad Mula ALYAS

Thesis Advisor
Assist. Prof. Dr. Nehad T.A RAMAHA

DETECTION OF PLANT DISEASES USING IMAGE PROCESSING WITH MACHINE LEARNING

Raghad Mula ALYAS

Thesis Advisor
Assist. Prof. Dr. Nehad T.A RAMAHA

T.C.
Karabuk University
Institute of Graduate Programs
Department of Computer Engineering
Prepared as
Master Thesis

KARABUK
February 2023

I certify that in my opinion the thesis submitted by Raghad Mula ALYAS titled "DETECTION OF PLANT DISEASES USING IMAGE PROCESSING WITH MACHINE LEARNING" is fully adequate in scope and in quality as a thesis for the degree of Master of Computer Engineering.

Assist. Prof. Dr. Nehad T.A RAMAHA ....

Thesis Advisor, Department of Computer Engineering

This thesis is accepted by the examining committee with a unanimous vote in the Department of Computer Engineering as a Master of Science thesis. February 27,2023

Examining Committee Members (Institutions)		Signature
Chairman	: Assist. Prof. Dr. Zafer ALBAYRAK (SUBU)	....
Member	: Assist. Prof. Dr. İsa AVCI (KBÜ)	....
Member	: Assist. Prof. Dr. Nehad T.A RAMAHA (KBÜ)	....

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

Prof. Dr. Muslum KUZU ....

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."

Raghad Mula ALYAS

ABSTRACT

M. Sc. Thesis

DETECTION OF PLANT DISEASES USING IMAGE PROCESSING WITH MACHINE LEARNING

Raghad Mula ALYAS

Karabuk University
Institute of Graduate Programs
The Department of Computer Engineering

Thesis Advisor:
Assist. Prof. Dr. Nehad T.A RAMAHA

February 2023, 71 pages

The quantity and quality of the plants generated by the crop may be significantly impacted by insects and diseases that affect agricultural plants. Plant diseases and insects that feed on plants may be found via the analysis of digital photos. Recent fundamental developments have made deep learning far more effective than previous methods. These developments in digital image processing are noteworthy. Serious researchers are pursuing deep learning and machine learning as artificial intelligence to detect diseases and pests. However, previous research work in the field has not reached satisfactory results. Therefore, this research project suggested a new model using image processing, feature extraction, and machine learning to identify five distinct plant diseases: Anthracnose, Bacterial, Citrus canker, Powdery mildew, and Grey mold. This study uses two classifiers: K-mean for clustering and multi-SVM for classification. These classifiers were used to evaluate the best feature models, and then we compared those models to the detailed features of the used, pre-trained disease classification models. Comparisons were made between the efficacy of the three binary-based approaches: one against all, one against one, and DAGSVM. According to our research findings, the "one against the other" and DAG techniques are superior to the alternatives, where we were able to obtain an accuracy of the linear kernel with 500 iterations is 98.38%

Keywords : Support vector machine (SVM), K-mean, Plant disease, Feature extraction.

Science code : 92431

ÖZET

Yüksek Lisans Tezi

BİTKİ HASTALIKLARININ MAKİNE ÖĞRENMESİ YOLUYLA GÖRÜNTÜ İŞLEME ÜZERİNDEN TESPİTİ

Raghad Mula ALYAS

Karabük Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı:
Dr. Öğr. Üyesi Nehad T.A RAMAHA
Şubat 2023, 71 sayfa

Rekolte bitkilerinin miktarı ve kalitesi böceklerden ve tarım bitkilerini tutan hastalıklardan anlamlı derecede etkilenebilir. Bitki hastalıkları ve bitkilerden beslenen böcekler dijital fotoğrafların analizleriyle bulunabilir. Yakın zamanlarda gerçekleşen temel bazı gelişmeler önceki yöntemlere kıyasla derin öğrenmeyi çok daha etkin kılmıştır. Dijital görüntü işleme alanındaki bu gelişmeler dikkate değer. Ciddi araştırmacılar hastalıkları ve zararlıları tespit etmek için derin öğrenme ve yapay zeka olarak makine öğrenmesini takip etmektedirler. Ancak, bu alanda önceki araştırmalar tatmin edici sonuçlar vermemiştir. Bu yüzden, bu araştırma projesinde beş farklı bitki hastalığının tanımlanmasında görüntü işleme, özellik ekstraksiyonu ve makine öğrenmesini kullanan yeni bir model önerilmektedir: Antraknoz, Bakteriyel, Turunçgiller Kanseri, Küllenme ve Kurşuni Küf Hastalığı. Bu çalışmada iki ayırıcı kullanılmıştır: kümeleme için K-ortalama ve sınıflandırma için multi-SVM. Bu ayırıcılar en iyi özellik modellerinin değerlendirilmesi için kullanıldı, sonra bu modelleri kullanılan eğitim öncesi sınıflandırma modellerinin ayrıntılı özellikleriyle karşılaştırdık. Üç adet ikili temelli yaklaşımın etkinlikleri arasında karşılaştırmalar yapıldı: hepsine karşı biri, birine karşı diğeri ve DAGSVM. Bizim araştırma bulgularımıza göre, 500 tekrarla elde edilen lineer çekirdeğin doğruluğu %98.38 olduğundan "birine karşı diğeri" ve DAG teknikleri diğer alternatiflere göre üstündür.

Anahtar kelimeler : Destek vektör makinesi (SVM); K-ortalama, Bitki hastalığı, Özellik ekstraksiyonu.

Bilim Kodu : 92431

ACKNOWLEDGMENT

First and foremost, I would like to thank my supervisor Assist. Prof. Dr. Nehad T.A RAMAHA , for his guidance, encouragement, and expertise throughout this process. He has been instrumental in shaping my research and helping me navigate the challenges that arose along the way.

I am grateful to the faculty former member at Karabuk University, Assist. Prof. Dr. Adnan Saher M. AL-AJEELI, for his valuable input on my research. his guidance helped me to develop my research skills to complete this study.

I want to thank my family and friends for their unconditional love, unwavering support, and encouragement throughout my academic journey. Without their support, I would not have been able to persevere through the challenges of this research.

Especially I would like to dedicate this thesis to my mother, who passed away before I could complete this research. She was a constant source of support and motivation, and her unwavering belief in me was the driving force that kept me going even when times were tough. She instilled in me the importance of education and encouraged me to pursue my academic goals.

Although my mother is no longer with us, her legacy lives on in my academic achievements, and I know she would be proud of what I have accomplished. I wish she could be here today to share in this moment with me.

Thank you, Mom, for your unwavering love and support. This thesis is dedicated to you, and I hope to make you proud.

I would like to express my sincere gratitude to Karabük Üniversitesi for providing me with the opportunity to pursue my academic goals. My time at the university has been a truly transformative experience, and I am proud to be a part of this community.

CONTENTS

Page

<u>**Page**</u>

LIST OF FIGURES

Page

LIST OF TABLES

Page

ABBREVITIONS

RGB	: Red green blue
ANN	: Artificial neural network
AWS	: Amazon Web Services
BIV	: Blue image strength values
BMP	: Windows bitmap
BPNN	: Back Propagation Neural Network
CCM	: color co-occurrence method
CIELAB	: International Commission on Illumination LAB
CNN	: Convolutional neural network
DNN	: Deep Neural Network
FC	: Fuzzy curve
FNBC	: Fuzzy Naïve Bayes Classifier
FS	: Fuzzy surface
GAN	: Generative adversarial network
GLCM	: Gray Level Co-occurrence Matrix
Gif	: graphics interchange format
GIV	: Green image strength values
GPU	: Graphics processing unit
GUI	: Graphical user interface
HSI	: Hue, saturation, and intensity
HSI	: Hue Saturation Intensity
HSV	: Hue, Saturation, and Value
JPG	: Joint Photographic Experts Group
KNN	: K-Nearest Neighbors
LDA	: Linear discriminant analysis
MCPDA	: Modified Color Recovery Detection Algorithm

NN	: Neural network
OCR	: optical character recognition
PCA	: Principal component analysis
PNN	: Probabilistic neural network
RBF	: Radial basis function
R-CNN	: Region-based Convolutional Neural Networks
RF	: Random forest
RIV	: Red image strength values
PNG	: Portable Network
RP	: Recall-precision
SD	: Standard deviation
SIFT	: symbolism, imagery, figurative language, tone and theme
SWIR	: short-wave infrared
3D	: Three dimension
VNIR	: visible to near infrared
VA	: Variances deviation
VGG	: Visual Geometry Group
YVMV	: Yellow Vein Mosaic Vi

PART 1

INTRODUCTION

1.1. OVERVIEW

A plant disease diagnosis takes knowledge of both science and art. Diagnoses, also known as identifying signs and symptoms, are made visually and need correct judgment and scientific principles. For research, instruction, and diagnosis, it is essential to have photographic images of plant diseases and the symbols often used to help people comprehend plant diseases. These digital images may be combined by pathologists using digital imaging transmission techniques, which can subsequently be used to diagnose plant diseases [1], [2].

Many organisms, including fungi, bacteria, viruses, and parasites, are responsible for many plant illnesses; figure (1.1) displays examples of infected plants with fungi and bacterial diseases. Saprophytes are often called necrotrophic or heterotrophic organisms. In contrast, autotrophic organisms are usually referred to as parasites. Nevertheless, both may be infectious agents or germs. Parasites and saprophytes are types of creatures that live within others living objects. In comparison, facultative organisms can adapt their behavior to the surrounding conditions [1],[3].

A. *Citrus canker* disease

B. *Powdery mildew* disease

C. *Grey mold* disease

D. *Anthracnose* disease

Figure 1.1. Sample images of infected plants with multiple diseases [4,5].

The damage to a plant's leaves that results from a disease is frequently thought to be the first indication that the sickness will spread to other plants; here are a few leaf ailments:

- Leaf spots: Leaf spots result from the fungus adhering to a leaf. It starts as a little brown patch on the leaf's skin. It then spreads throughout the entire plant, seriously harming it.
- Leaf Rust: The lower leaves of plants with this fungus, typically seen on mature plants, rust.
- Powdery mildew: A fungus infects plants, causing white, powdery leaf holes.
- Downy mildew: Affected plants (oomycete organisms) produce white or grey

letters in the veins of their leaves, which facilitate crop harvesting.

- The mosaic virus infects plants and leaves them with white, yellow, or green (light or dark) dots.
- Yellow Leaves: Low or excessive watering causes moisture stress, which results in the yellowing of the leaves.
- Leaf curl: This illness causes leaves to twist, curl, and change color. Both mildew and a virus are the culprits.

Several diseases might induce symptoms, the primary diagnostic tools for field ailments. These signs show that external systems are involved in potential host-pathogen interactions. Consequently, several significant decisions have been made on safe practices, plant development, and processing [6].

One of the topics that scientists are interested in researching is the automatic diagnosis and treatment of sickness. This internal context aids the diagnosis and classification of diverse plant diseases. To better understand the several diverse computer applications that may be used in integrated spaces, the following publications were studied as part of the literature research.

India currently produces the most grapes per capita globally, but there is room for growth. Grape exports from India account for 1.54 % of the world's total grape exports and have a total value of 48,505 US dollars and 53,910 metric tonnes. Agriculture employs more than 70% of the working population [7], [8], [9]. The fruit grapes are prevalent in India. The leaf diseases powdery mildew, downy mildew, and anthracnose, among others, are severe for grapes. The grape suffers significantly as a result of this. The disease has affected the plant's stems, fruits, and leaves. Because it is too early, identifying leaf spot infections in the agricultural sector provides a significant issue [10], [11].

Applications for various industries, including quality control in manufacturing, medical imaging, agricultural remote sensing, and others, have been created employing digital image processing techniques. Image processing has led to the development of efficient methods for crop identification, gauging soil quality, calculating yield, and other digital applications utilized in agriculture. One agricultural processing technique is used in digital image applications to identify plant diseases [12].

Depending on where they manifest on the plant, diseases may be categorized as viral, viral, fungal, etc. Anthracnose, downy mildew, and powdery mildew are a few fungal diseases that may affect grapevines in India [11], [13]. The proposed project's primary focus is image processing to gather information and classify fungal illnesses such as powdery mildew and downy mildew. This research aims to demonstrate the feasibility of an automated diagnosis of grape leaf diseases. Because the system can identify and distinguish between grape leaf diseases automatically, quickly, properly, and for less money than conventional approaches.

1.2. MOTIVATION

Environmentally friendly, economical, and more efficient techniques are gaining in popularity. Direct visual examination and field observations have historically been the primary methods for diagnosing plant diseases. Still, both techniques have considerable disadvantages, notably the high labor costs associated with finding and treating them. In smart agriculture, automated plant disease detection devices have been employed to save labor costs and boost output. Only instrumental therapy can identify illness, despite all these processes distinguishing between healthy and diseased plants [14],[15]. Each of these systems widely used this strategy. With the huge potential of image processing technology, automated plant disease detection may be made simpler. Its use in innovative agricultural projects leads to improved earnings, lower spraying, more effective growth methods, and higher yields. This automated plant disease detection might be implemented using image processing technology. There are several more applications for which it may be helpful, such as identifying plant species or evaluating the development stage of plants [16], [17]. Applications like this are becoming more significant as interest in intelligent agriculture grows. Therefore, this thesis addresses some significant challenges that still need to be resolved in this sector.

1.3. PROBLEM STATEMENT

Exposure of farms to agricultural pests affects the process of sustainable agriculture, and this is due to negative consequences for the agricultural economy of countries. Detecting agricultural diseases requires two critical factors: experience and time, where the expert in agricultural pests visits the site in the field to detect the plants and take samples from the leaves of the infected plants to study them or compare them with diseases that were previously detected. Peasants are usually from the financially poor class and do not have scientific expertise that enables them to diagnose diseases. Calling the expert to the farm site is costly, and experts may not be present in some regions. Therefore, machine learning techniques play an essential role in facilitating this problematic task by training the machine on the expertise of specialists to be able to work instead of the expert. Thus, the farmer can use innovative programs that do not require experience, and these programs detect disease and establish a diagnosis to treat it. The proposed system allows the farmer to take a picture of the leaves of the affected plants and puts them in the system. After that, the system will analyze the picture, extract its features, and then compare those features with the diseases the experts have classified. This way will overcome the time and cost issues. In the proposed work in this research, we studied previous works in this field and found that some did not reach satisfactory results because they used classical methods in programming project algorithms. Therefore, we have done a project that relies on Multi-SVM and other image processing techniques to increase the accuracy of extracting features from the images. Therefore, we reached higher accuracy of the program's work to diagnose plant disease efficiently.

1.4. OBJECTIVES

This research's primary goal is to identify the most efficient way to find various diseases that could be present in plant leaves. This stage includes finding the best technique for image processing, feature extraction, and machine learning. Using this as a springboard, we may specifically emphasize the following objectives:

- To identify significant and often utilized components from the corpus of prior research used to diagnose ailments in plant leaves.
- To compare the best features obtained using in-depth features produced using pre-trained deep classification approaches for sickness classification with the best features from the available literature.

1.5. RESEARCH CHALLENGES

To monitor and enhance the health of plants and fruits, growers and farmers may utilize the vital information on leaf health and production provided by both pre-harvest and post-harvest disease detection technology. Based on the range of various photographic techniques, we chose a group of images of plants. By analyzing the past literature, we could pinpoint limitations discovered while building the data set for pictures of ill plants. These limitations are as follows:

- Documents are photographed in controlled settings and away from the field. Automation technologies are still in their infancy.
- Limited applications to manage massive numbers of plants in the field in real-time.
- The testing is conducted in controlled environments rather than in the field.

1.6. CONTRIBUTION

The main topics of this thesis are developing a technology for plant disease detection and creating a series of tactics to improve performance in identifying hazardous and ill plants. To prove the points made in this thesis, the problems that must be overcome to create image-processing technology for diverse applications are discussed in this thesis. These worries consist of the main contributions made by this thesis are listed below:

- A thorough analysis of plant segmentation-based techniques was drawn from the corpus of previous research, including how well-suited specific algorithms are to particular growth conditions.
- Create an image processing algorithm that can identify plant diseases at different stages of plant growth and in various environmental conditions using HSV shape and color space characteristics analysis.

PART 2

LITERATURE REVIEW

The many definitions of disease-specific diagnosis that may be made using photo imaging technology are explained in this section.

A proper diagnosis may assist in preventing crop losses as well as declines in agricultural output. It takes substantial time to recover, labor, and effort on preexisting issues. Image imaging is utilized in diagnosing illnesses as a direct result of this. Diagnosing a disease involves several processes, including capturing photographs, editing, categorizing, and developing and shrinking certain traits. The editor described the potential use of leaf photos for identifying plant issues.

The many methods described for differentiating between various immune system components are highlighted. In addition to the role that disease-causing bacteria play in the process, the studies also explored a range of other additional elements and techniques for extracting the characteristics of a damaged leaf. A precise diagnosis of the issue and the condition, which may be carried out via imaging if required, are prerequisites for effective therapy. The group leader emphasized how differentiating immune system components may be done in various ways.

In addition to the role that disease-causing bacteria play in the process, the studies also explored a range of other additional elements and techniques for extracting the characteristics of a damaged leaf. Neural network (NN) techniques may be used well for plant pathologies, such as autoimmune features, backpropagation algorithms, support vector machines, etc. We can identify and categorize various disorders using imaging technologies.

2.1. DETECTION USING SVM

Support Vector Machine (SVM) is a supervised machine learning method that may be applied to classification and regression problems. Although we also examine regression difficulties, its application is better for categorization.

A system that can identify pests based on images of leaf damage was presented by Rani et al. [18]. After contrast enhancement, image segmentation was carried out using the k-means algorithm. An SVM classifier is used to produce identification after the features of this image have been extracted. The proportion of regions affected was calculated.

Dhaware et al. [19] found numerous techniques to determine whether plants are affected. Preprocessing involves converting RGB to HSV for each image. For segmentation, background subtraction with clusters is employed. Energy, homogeneity, correlation, and other concepts. Using the SVM method, images may be categorized. The results were accurate despite the limited sample size. Infected leaves were not categorized according to illness.

SVMs are used by U. Mokhtar et al. [20] to identify tomato leaf diseases. To perform image preprocessing, they used morphological methods with leaf image separation, image scaling, and background subtraction. Color and texture feature. The Gabor filter eliminates texture.SVM is applied to categorization.

MATLAB R2017a was used to construct an SVM classifier system to classify leaf diseases such as Alternaria, Cercospora Leaf Spot, and Bacterial Blight by Zaw, Ko Ko [21]. This study transforms RGB to HIS (Hue Saturation Intensity). Gray Level Co-occurrence Matrix (GLCM) extracts damaged area features using k-means clustering in segmentation. The median filter removes noise before feature extraction.

Finally, a support vector machine determines leaf illness categorization and accuracy (SVM). The system's best accuracy is 83%.

Classification can be done on both structured and unstructured data. The act of classifying involves dividing up a certain piece of data into various categories. Making an educated guess as to the class of the supplied data points is the first step in the procedure. The classes may also be called goals, labels, or categories in everyday speech. It is used to analyze images of ill leaves, which helps identify the many forms of diseased leaves. Figure (1.2) illustrates the classification method found here [22].

Image processing was used by Madiwalar et al.[23]. Scientists used color and texture to diagnose plant diseases. Their algorithms ran on 110 RGB pictures. The image was classified using GLCM features, RGB and YCbCr averages (A group of color spaces designed for use in digital still and moving images), and a Gabor-convoluted image. The researchers categorized the leaves using SVM. Color and Gabor filter properties differentiate anthracnose and leaf spot. Using all the gathered information, they achieved 83.3% accuracy.

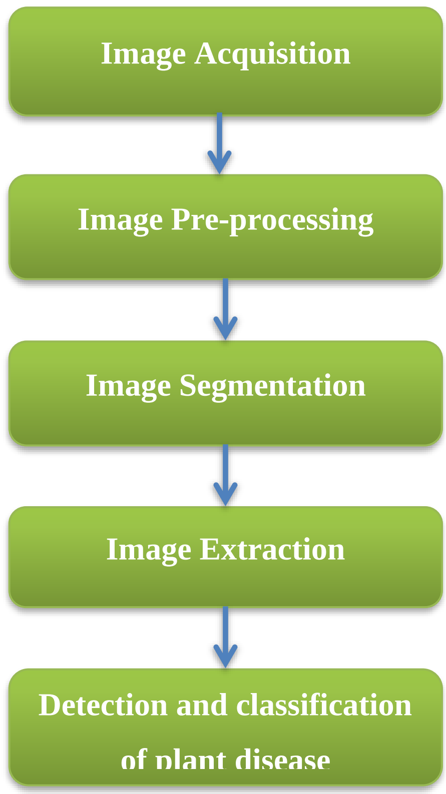

Figure 1.2. Plant disease detection and identification block diagram.

Dandawate and Kokare [24] created a technique for automatically detecting diseases in soybean plants. The color space for the image was altered from RGB to HSV, or hue, saturation, and value. Color-based and cluster-based techniques were used for segmentation. Using the shape of the leaf as a guide, the SIFT method was utilized to identify the plant species.

Citrus illness was successfully identified by Pydipati et al. [25] using color texture data and discriminant analysis. Additionally, they tested if the statistical classification methods and the hue, saturation, and intensity (HSI) color attributes might help identify the damaged leaves using the color co-occurrence method (CCM). With this method, they achieved an accuracy of more than 0.95.

The illnesses that can harm tomato plants were divided into different categories by Hlaing and Zaw [26] using a variety of texture and color traits. To learn more about the texture, they employed a method known as the Scale Invariant Feature Transform or SIFT. The characteristics' size, placement, and shape were all described in detail in this information. The RGB channel was also used to retrieve the color information.

Agrawal et al. [27] Grape leaf diseases Black Rot, Esca, and Leaf Blight are accurately classified using LAB and HSI color model features. [6] Developed a method to identify diseases in Tea plants. Three different types of diseases with less in features are detected using SVM classifiers. The developed method classified the diseases with an accuracy of 90 %.

2.2. DETECTION USING KNN

Umapathy et al. [28], In machine learning, the K-Nearest Neighbors algorithm has been used to identify patterns, calculate statistics, and categorize data. A survey on identifying plant diseases was conducted using the KNN classifier. An algorithm for sugarcane disease detection was suggested. The process of feature extraction is carried out using image processing techniques. A 95 % success rate was attained in diagnosing the leaf scorch disease in sugarcane leaves.

Kaur, S.et al. [29] Used for soybean culture to detect three different diseases Downy Mildew, Frog eye, and Septories leaf blight. They reported an average classification accuracy of approximately 90 % using a big dataset.

Parikh, Aditya, et al. [30] Developed a technique for determining the cotton plant disease's severity and identifying it. Grey Mildew Disease could be recognized with an accuracy rating of 82.5% using 40 images.

Suresha et al. [31] provided a method for identifying illnesses in paddy leaves. In this method, the images of paddy leaf diseases were captured using a camera while the paddy leaves were in paddy fields. After converting the RGB images to the HSV color space, Otsu segmentation was used to locate the region of interest in the image. The K-Nearest Neighbor algorithm was used to extract and classify geometric parameters such as area, perimeter, minor axis length, and major axis length to reach an accuracy of 68.1 %.

A technique for the illness diagnosis of guava leaf was devised by Thilagavathi et al. [32]. It was necessary to perform some preprocessing, such as scaling and image enhancement. The sick area is then brought to light by using the region growth procedure. The CIELAB color space (demonstrates the three-dimensional, quantifiable link between colors.) is applied here, while the SIFT algorithm is employed for feature extraction.

Image processing is the process of turning a physical image into its digital equivalent and then going through a set of steps to extract useful information from it [33]; when implementing specific specified signal processing algorithms, the image processing system treats all images as 2D signals instead of 3D ones [34].

Sandhu et al. [35] identified plant diseases using back propagation neural network (BPNN) and digital image processing. By examining leaf photos, researchers have developed techniques for diagnosing plant diseases. Otsu's thresholding, boundary detection, and spot detection techniques were used to segment diseased leaf sections. They classified plant diseases based on color, texture, morphology, edges, etc. Plant pathogens are classified and found using BPNN.

Using KNN and neural network approaches, Kurale, Neha G can identify and categorize plant leaf diseases.[36] the dataset of sick plant leaves is considered. This dataset includes leaves from healthy plants and leaves afflicted by early blight, late blight, and black rot. They have determined and quantified how much of the leaf is changed by their findings based on their categorization.

Moghadam, Peyman, et al. [37] used hyperspectral imaging to find plant illnesses. This investigation used visible-to-near-infrared (VNIR) and short-wave infrared (SWIR) spectra. Leaf segmentation is accomplished using spectral k-means clustering. They developed a technique for taking grids out of hyperspectral images. Researchers achieved 83 % and 93 % accuracy with the full spectrum for VNIR vegetation indicators. The proposed solution is excessively expensive because a 324-band hyperspectral camera is needed**.**

Sharath D. M. et al. [38] created the Pomegranate Blight detection method to identify bacterial blight in the plant. Grab-cut segmentation was used. Image edges were extracted using a Canny edge detector. Researchers devised a way to measure fruit contamination.

Garima Shrestha et al. [39] used a convolutional neural network. The authors categorized 12 plant diseases accurately at 88.80%. Experiments used 3000 high-resolution RGB photos. Three parts do convolution and pooling. Compute-intensiveness suffers. The model's F1 score of 0.12 is low since it made many inaccurate negative predictions.

2.3. DETECTION OF USING ANN

A proposed study for plant disease recognition using the feed forward back propagation algorithm was performed well by Al Bashish et al. [40]. they examined early scorch, late scorch, cottony mould, and microscopic whitening disorders through an ANN classifier for the detection of plant diseases

Arivazhagan et al. [41] break the illness process into four steps: deploying the color transformation structure for RGB input photographs and eliminating and masking green pixels. Segmentation and texture statistics simplify this approach. The classifier ANN enhances sickness categorization by extracting characteristics. On 500 database leaves, test the algorithm's robustness. Kulkarni and Patil6 used image processing and neural networks to diagnose plant illnesses.

Anjomshoae and Rahi [42] used a template-based approach to study and research rubber tree leaf overlap detection. The Scale Invariant Feature Transform (SIFT) extracted significant information from the data.

Vibhute and Bhode [43] surveyed numerous image-processing methods in agricultural applications. The main emphasis has been remote sensing, hyper-spectral imaging, fuzzy logic, neural networks, evolutionary algorithms, wavelet, PCA, etc. Image processing has been studied as a potential tool for automatically classifying fruits and weeds.

Kumari et al. [43], Imaging, segmentation, feature extraction, and classification were used to detect disease. K-mean aggregation yielded affected-population features. Image segmentation. Collect variance, correlation, power, homogeneity, mean, and SD. Mean, SD, and VA gathered. The disease classifier used group characteristics. This study used NN classification. Bacterial leaf spot and cotton leaf disease obtained 90% target localization accuracy.

To categorize the sickness the plant is suffering from, Thushara et al. [44] claim that the numerous characteristics of the plant leaf, such as intensity, color, and size, are retrieved.

Li, Yang, et al. [45] presented a memory-based categorization technique Combined CNN and GAN (GAN). Since CNN learns commonalities between input paired data, it needs little raw data to generate good classification results. GAN abstracts images from completed projects for future work. The standard CNN model forgets during continuous work (pest and plant classification).

In their study, Rastogi et al. [46] suggested a two-phase method for classifying plants based on leaf traits. The first step includes preprocessing leaf images, feature extraction, training, and classification using artificial neural networks. In phase two, leaf illnesses are classified using K-Means, feature extraction, and ANN.

2.4. DETECTION USING NAIVE BAYES

Mondal et al. [47] developed a method for recognizing the yellow mosaic virus using high-resolution leaf pictures. After being transformed from RGB to grayscale, input photos undergo morphological operations. Mean, SD, entropy, homogeneity, mode, and median are retrieved. Naive Bayes classified the disorders.

Sethy, Prabira Kumar, et al. [48] picked rice and apple leaf illnesses (speckled deciduous disease, yellow-leaf disease, round spot diseases, and mosaic diseases) as their research subjects. The apple leaf spot image was used to extract color, texture, and shape. The BP neural network model, which had an average recognition accuracy of 92.6 %, was used to categorize diseases.

Nababan, Marlince, et al. [49] created an intelligence-based application with an input-output plan to diagnose oil palm plant illness using Naïve Bayes. Based on the research result of the Bayes method with known symptoms, a diagnosis of oil palm plant disease could be made.

The yellow vein mosaic virus (YVMV) disease in okra leaves was identified and categorized using image processing, K-means, and a Naive Bayesian classifier by Mondal, Dhiman, et al.[50] .

79 images of okra leaves, some of which were ill and others not, were used to test the proposed approach. The input leaf images are categorized into four groups based on the severity of the YVMV infection.
Wahyuni Eka [51] used the Naive Bayes Classifier to determine the findings of the disease classification. Additionally, forward-chaining search strategies were used to conduct the test. The accuracy of the forward chaining approach was 90%, compared

to the accuracy of the FNBC method, which was 88%.

Ding Min et al. [52] enhanced linear discriminant analysis (LDA) by combining a high-dimensional texture index with a simple Bayesian classification model. 91.56 % of both procedures recognize and 98.44 % of leaves. The RP curve demonstrates the redesigned blade classification approach's advantage for high-dimensional indicator classification.

2.5. DETECTION USING FUZZY

Zhang, Yan-Cheng, et al. [53] suggested fuzzy curves (FC) and surfaces (FS) for the selection of disease image features in cotton leaves. This inquiry is in two stages. They begin by using FC to quickly and automatically separate only a limited number of important features from a bigger collection of original features. Second, utilize FS to pinpoint features that depend on the most crucial characteristics.

The automated plant species detection is done using the leaf shape descriptor created by Salve et al. [54]. This facilitates the taxonomic categorization process and aids in the automatic classification of plants.

Sabrol, Hiteshwari, and Satish Kumar [55] employed image processing and soft computing to identify plant diseases automatically. Five tomato diseases were assessed using natural images. Tomato leaf curl, fungus late blight, fungal septoria leaf spot, bacterium canker, and one unscathed image from each group (healthy). 180 training and test images are included. Color moments, histogram, and coherence vector attributes were transformed to CIE XYZ (a device-invariant representation of color).
Using fuzzy entropy and the PNN classifier, Majid et al. [56] developed a smartphone application for identifying paddy plant diseases. Paddy illnesses lower rice output and revenue. Fuzzy entropy extracts paddy illnesses from digital pictures of rice leaves. Cross-validation measures how well conclusions from statistical analysis generalize to different datasets. 91.46 % of paddy ailments can be correctly identified.

Sutha, et [52] use fuzzy to classify plant leaf diseases. The fuzzy membership function evaluates vertex structure interactions to identify plant disease. A test and database image can be differentiated using extracted metrics like skewness, mean, and deviation. The method's accuracy is 93%.

2.6. DETECTION USING DEEP LEARNING

Srinidhi, V. V., Apoorva Sahay, and K. Deeba [57] demonstrated that the Deep Convolutional Neural Network models EfficientNet and DenseNet could correctly identify apple plant diseases from images of apple plant leaves. They presented models based on EfficientNetB7 and DenseNet that offer 99.8% and 99.75% accuracy, respectively, and overcome known convolutional neural network limitations.

Swaminathan et al. [58] discovered 29 diseases in 7 plants using 35779 photos from Kaggle's Huges DP Plant-Village dataset. The original image is transformed to HSV colors for training and classification, and a thresholded masked image is sent to the suggested model. Training the model on Google Colab obtained 94.96% accuracy for all plant disease classes (Tesla-T4 GPU).

Hassan, Sk Mahmudul, et al. . [59] originally described shallow VGG using RF and Xgboost. They made use of tomatoes, potatoes, and maize. Their shallow VGG model with Xgboost outperforms deep learning methods in accuracy and precision. The best VGG combination for corn, potatoes, and tomatoes is Xgboost. The accuracy of shallow VGG with Xgboost is 94.22 %, 97.36%, and 93.14 %, respectively.

Rezende, Vanessa, et al. [60] classified 20 plant diseases affecting 10 species using image processing. This method used modified VGG architectures and ImageNet weights to scale, resize, and pick the images. A study compared the accuracy, precision, recall, and F1-score categorization factors.

Nagasubramanian, Koushik, et al. [61] model has a 0.87 infected class F1 score and a classification accuracy of 95.73%. They deduced the training model using a saliency map and visualizing classification-sensitive pixels. The saliency map display determined the sensitivity of wavelength classification. 733 nm was identified as the most sensitive wavelength by saliency map visualization.

The three fundamental neural network architectures that Kumar and Sumit [62] explored were the Fully CNN, Single Shot Multibook Detector, and Deep Neural Network (DNN). Convolution Neural Networks can be used to leverage AI-based Deep Learning to handle this complicated challenge, as evidenced by their validation accuracy of 94.6%

Islam, Md. Tariqul [63] employed image processing and a CNN model to train the dataset. With this method, they attained an accuracy rate of 94.29 %.

A LeNet-based model was developed by Wallelign et al. [64] to categorize soybean plant diseases. 12,673 leaf image samples, including pictures of healthy leaves, were available in the PlantVillage database. The photographs were taken under unplanned circumstances. The deployed model's classification accuracy was 99.32%, demonstrating CNN's ability to classify plant illnesses.

PART 3

MACHINE LEARNING IN MATLAB

MATLAB makes the engine that builds up one's knowledge much simpler. MATLAB is the ideal environment for spreading engine knowledge into your information analytics because it provides tackles and objectives for conducting greater data and applications that make engine knowledge available. Thanks to MATLAB, engineers and information scientists have fast access to prebuilt purposes, Wade toolboxes, and specific applications for organizing, reversing, and grouping. MATLAB allows using of the following:

- Making an exact classical that best detentions the prognosis control of your information using classical modification and discount strategies.
- Assimilate engine Copies of knowledge can be found in initiative schemes, bunches, and smokes, and board duplicates can be embedded into entrenched real-time hardware.
- Obtain a group with instinctive encryption capabilities for embedded device analytics.
- Facilitate the integration of workflows ranging from information analytics to placement

3.1. MACHINE LEARNING

People naturally pick things up via experience, and machine learning teaches robots to do the same. Instead of relying on an equation as a model, machine learning algorithms use computer techniques to "learn" information directly from data. This makes models unnecessary for machine learning techniques. In response to a rise in

The algorithms can enhance their performance dynamically by the number of samples available for learning.

In machine learning, two techniques are supervised and unsupervised, as shown in figure (3.1). Unsupervised learning looks for hidden patterns or intrinsic structures in input data. In contrast, supervised learning includes training a model using data with known inputs and outputs to predict future outputs [65].

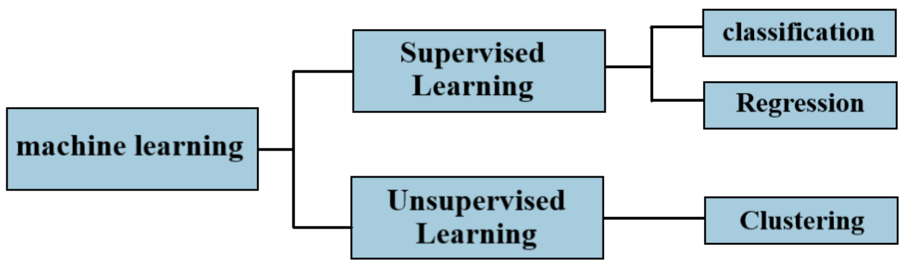

Figure 3.1. Machine learning techniques.

Supervised machine learning aims to build a model that can make predictions in the face of uncertainty by using data as a foundation for such predictions. When employing supervised learning, an algorithm may be trained by giving it a known input data set and known responses to the data (output). As a result, the program may predict how the model will respond to incoming data. By combining classification and regression approaches, supervised learning can create predictive models.

Without human supervision, learning reveals hidden patterns and internal data structures. This technique allows inferences from datasets that simply contain input data and no tagged responses. The unsupervised learning approach that is most frequently utilized is clustering. It is used in exploratory data analysis to find hidden groupings and patterns in the data. Clustering applications include the study of gene sequences, consumer behavior research, and object identification.

3.1.1. Supervised Learning

One key distinction between supervised and unsupervised learning is the use of labeled datasets to train algorithms with high accuracy in data classification and prediction. When new data is added to a model during cross-validation, it will readjust its weights to account for further information. These corrections are repeated until the model fits the data satisfactorily. By using supervised learning, businesses may increase their efforts to handle real-world issues, such as identifying and filtering spam emails [66],[67].

3.1.2. Unsupervised Learning

Unsupervised learning, often known as unsupervised machine learning, uses machine learning techniques. To analyze and classify unlabeled data, this method of learning is employed. Without the help of a human analyst, these algorithms can find previously unseen patterns or clusters in data. Exploratory data analysis, cross-selling methods, consumer segmentation, and image recognition are all improved by their ability to detect data similarities and differences. In addition, it has picture recognition capabilities [68], [69].

3.1.3. Semi-Supervised Learning

The main drawback of supervised learning is that it costs a lot to process and necessitates manual labeling by ML experts or data scientists. Furthermore, the range of applications for unsupervised learning is constrained [68]. The idea of semi-supervised learning is presented to address these issues with supervised learning and unsupervised learning methods. The training set for this algorithm consists of both labeled and unlabeled data. While there is a significant amount of unlabeled data, there is a relatively little amount of annotated data. An unsupervised learning technique is first used to cluster comparable data, and it also aids in labelling the unlabeled data into labelled data. This is why labelled data is more expensive to acquire than unlabeled data [70].

3.1.4. Reinforcement Learning

Reinforcement learning is instructing computer programs to behave a certain way through a series of decisions. The agent develops the abilities required to achieve a goal while functioning in an uncertain, potentially complex environment. In the reinforcement learning process, artificial intelligence is placed in an environment akin to a game. Before making a decision, the computer will work through the problem through trial and error. Artificial intelligence is given incentives or punishments for its actions depending on what the programmer wants the machine to do. Its goal is to obtain the greatest potential overall benefit [71].

3.2. SELECTION OF MACHINE LEARNING ALGORITHM

Choosing the best machine learning algorithm can be difficult because so many supervised and unsupervised machine learning algorithms take a different approach to the learning process. There isn't a single strategy that is best or appropriate for everyone. Even highly skilled data scientists cannot judge if an algorithm will work without first putting it to the test themselves, so finding the right algorithm frequently necessitates a process of trial and error.

High-flexibility models have the propensity to overfit the data by simulating extremely small changes that could be caused by noise. Even if simpler models are easier to understand, their accuracy could suffer. To choose the best algorithm, one must trade one benefit for another, which may entail the model's complexity, accuracy, or speed [72].

Trial and error is a key component of machine learning. The best course of action is to try a different strategy or algorithm if the first one does not yield the desired results. With the help of the tools offered by MATLAB, you will be able to experiment with various machine-learning models and choose the most suitable one. Figure (3.2) details all the algorithms under the machine learning technique.

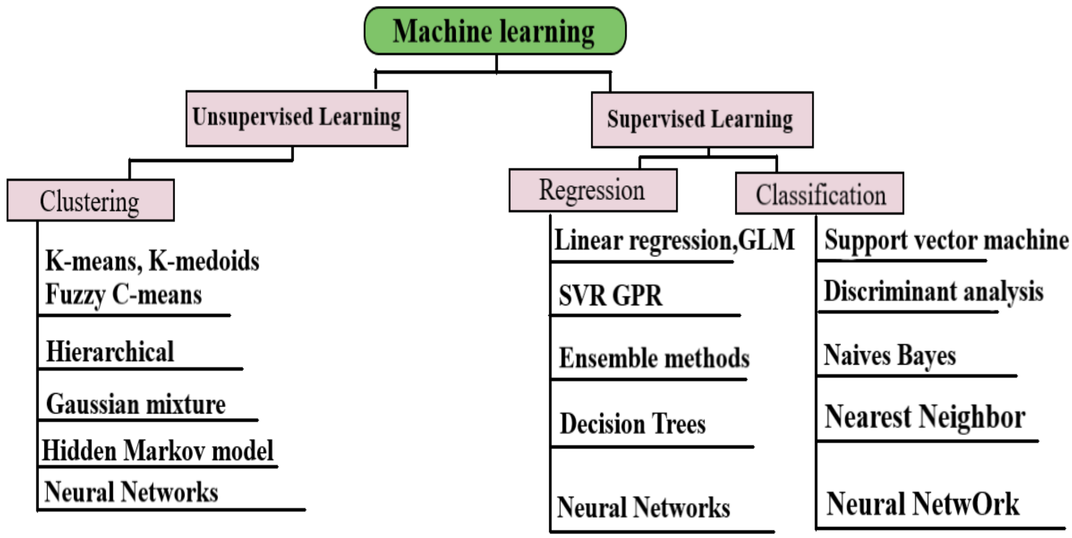

Figure 3.2. Algorithms of machine learning**.**

3.2.1. Support Vector Machine (SVM)

One of the most popular supervised learning algorithms is the Support Vector Machine, also known as SVM. It can be used to solve classification and regression-related problems. However, its main use is in machine learning, for example, when dealing with classification issues [73].

The Support Vector Machine (SVM) technique aims to produce the optimum decision boundary or line to categorize an n-dimensional space. This will simplify us to classify any new data points in the future. A hyperplane is a term used to describe this optimal decision boundary.

SVM is used to choose the extreme points and vectors that go into making the hyperplane. The technology, known as the Support Vector Machine, is named after these uncommon situations, referred to as support vectors [73].

The goal of the support vector machine technique, as shown in figure (3.3), is to find a hyperplane in an N-dimensional space (where N is equal to the number of features) that puts the data points into clear and distinct groups.

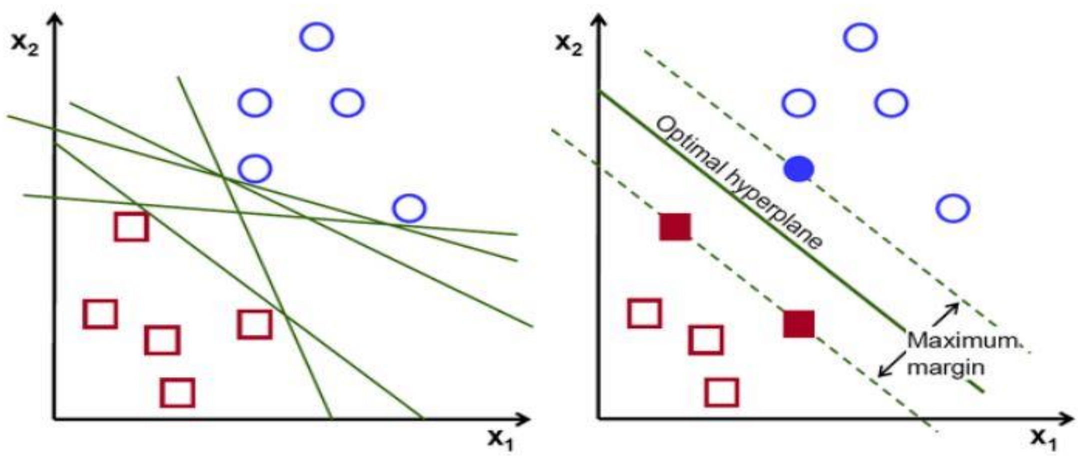

Figure 3.3. Algorithm of SVM.

3.2.2. K-Nearest Neighbors

The k-Nearest Neighbor (kNN) classification algorithm, one of the more simple techniques for machine learning, has been around long enough to reach a point of theoretical maturity. This method is based on the assumption that for the method to be relevant, a given sample must also fall into the same category as the k other samples that are most comparable to it (i.e., the samples closest to it in the feature space) [74].

As an example, think about developing an email spam filter. Consider for a second that we could recognize spam emails by simply looking at their headers. You can create a simple dataset by manually designating a sufficient number of emails as spam or not. The most similar emails should be identified by comparing each freshly received email to every email currently in the collection. The same label should be accepted for the message being received if the majority of k of the messages that are next to it or those that are most comparable to it are labelled as whatever (spam or non-spam). This technique makes it feasible to determine if each newly received email is spam or not[74],[75]. Figure (3.4) explain the steps of execution of the algorithm.

The steps execution of the algorithm:

- We begin by determining the value of the variable k, which represents the number of neighbours.
- Find the value that reflects the distance between the new example and the rest of the data set's examples.
- We take the appropriate steps to get the shoes close to the passengers.
- We create the class for the residents of the community.
- The dominant social class among the nearby neighbours.

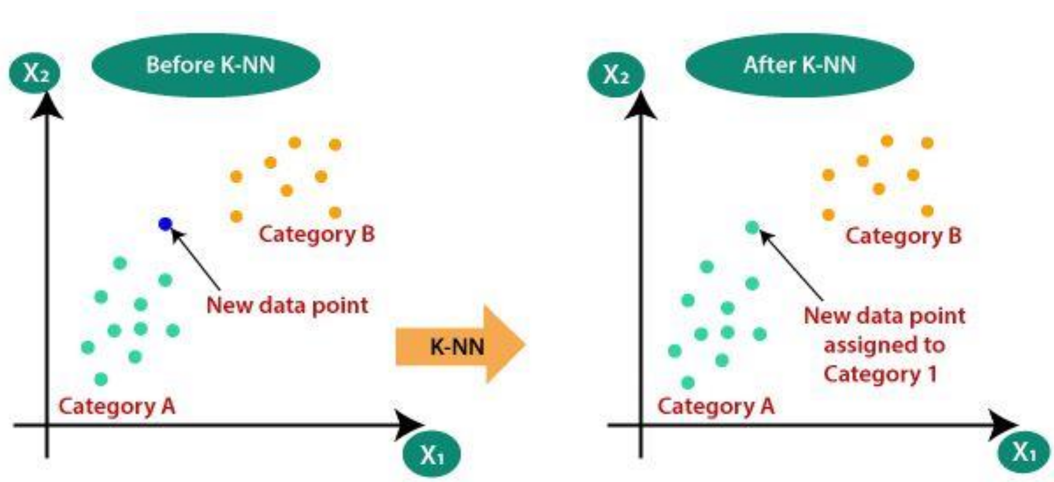

Figure 3.4. Algorithm of KNN.

3.3. CLUSTERING

This method is used in exploratory analysis to find unexpected links or clusters in big data. There are several real-world applications for cluster analysis, including in the study of perception, market research, and factor sequence analysis. For example, mobile network operators can use machine learning to optimize the locations at which they build cell phone towers by estimating the number of clusters of people using their masts. Given that a phone can only talk with one tower at a time, the team uses agglomeration algorithms to determine the best placement for cell towers to maximize signal reception for groups or clusters of consumers [76],[77], as shown in the example in Figure (3.5). The k-means and k-medoids clustering algorithms, graded agglomeration, Gaussian mixture models, hidden Mark models, self-organizing maps, fuzzy c-means agglomeration, and subtractive agglomeration are a few examples of broad agglomeration techniques.

The ability of a clustering method to expand with the data is a crucial aspect to take into account. Not all clustering algorithms scale effectively in machine-learning datasets containing millions of samples [78]. Many clustering algorithms base their work on the similarity between every pair of instances.

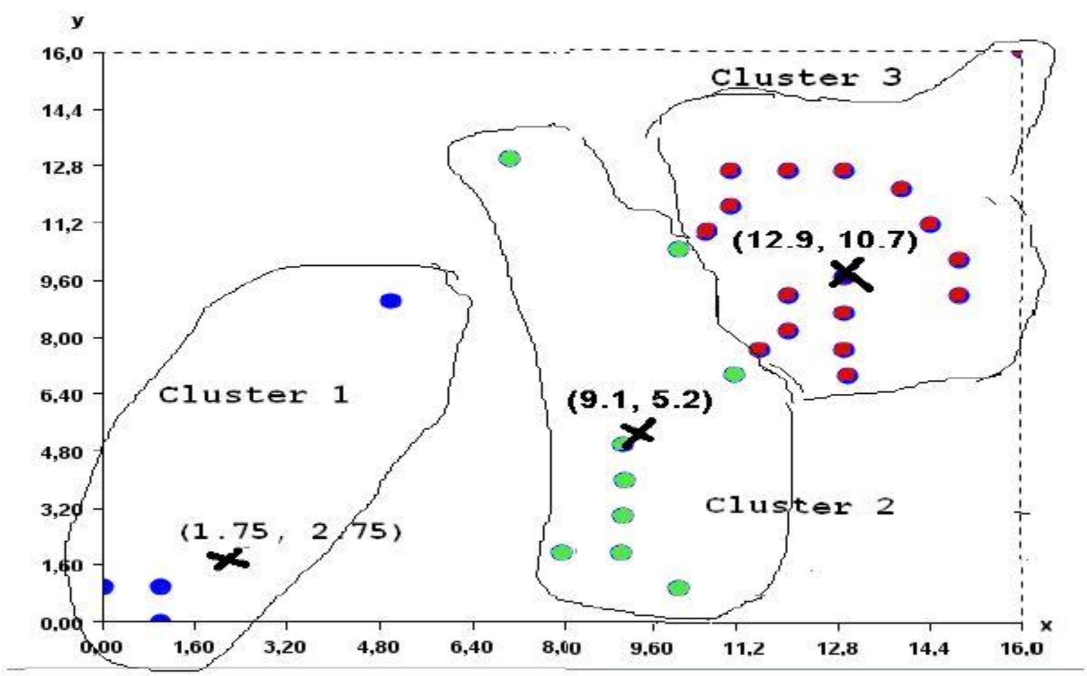

Figure 3.5. Example of clustering algorithms.

PART 4

MATERIALS AND METHODS

This chapter discusses the study methodology and the steps involved, starting with the dataset collection stage and ending with the testing stage. It's important to focus on the part of the recognition process where noise is applied to the images, and they are processed. MATLAB is used to process the images, which includes resizing the images, altering the type of images, and adding noise to the photos. All these steps are done to cut costs associated with processing the images. The pre-processing stage is the first step toward a successful recognition rule, which plays a crucial role. In addition to the time, it takes to execute, selecting an appropriate image size is essential because it directly impacts the program's output.

This research will identify leaf diseases particular to several distinct plant species instead of just one or two plants, as was done in earlier studies. The five subcategories covered by the suggested program are picture capture, image preprocessing, image segmentation, feature release, and feature classification.

Many different classification schemes can be used to diagnose plant diseases, and many methods have been used to achieve this goal. The classifiers used in this work for the detection process were the support vector machine (SVM) and the K-means clustering algorithm. Figure (4.1) illustrates the steps of diagnosis used in this study.

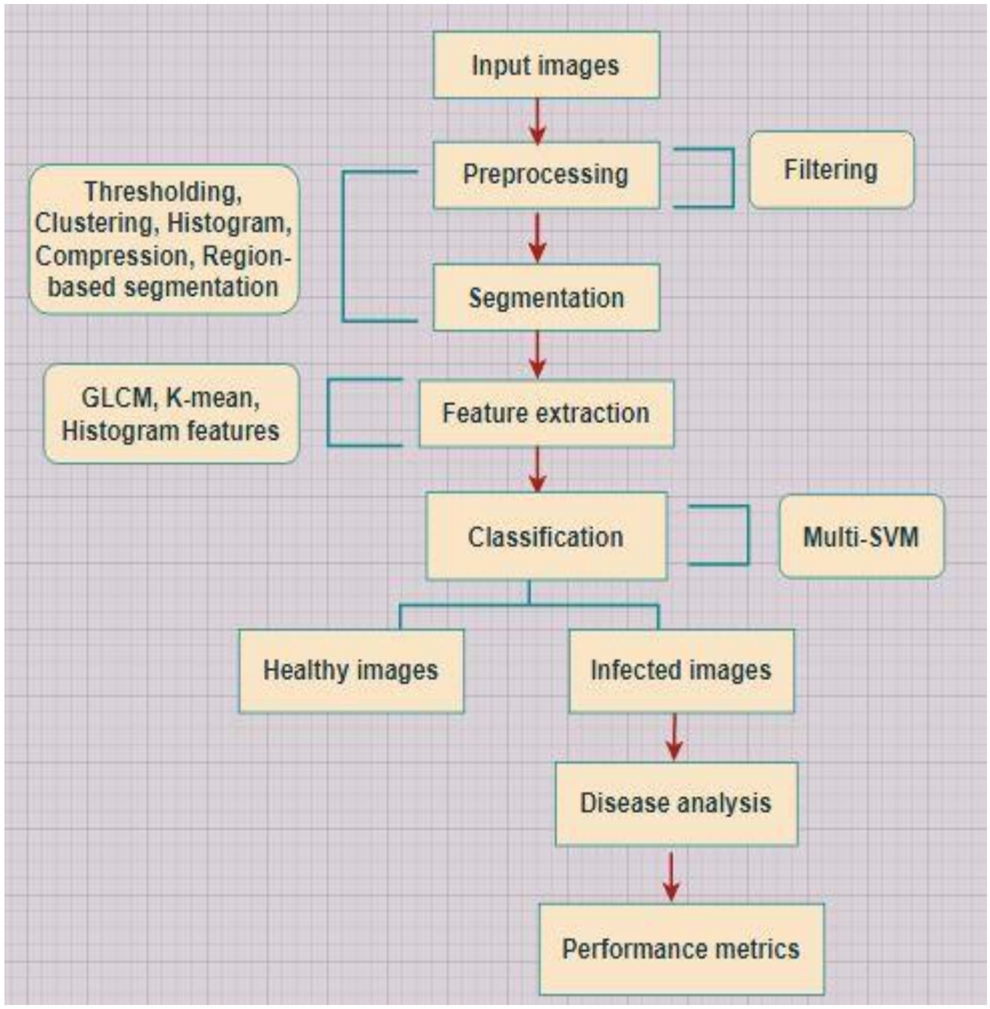

Figure 4.1. Flowchart of the proposed system.

Images of damaged plant leaves and various datasets related to living things are collected as part of the image capture stage and saved to a file. The presence of leaf diseases in a variety of plant species, including cassava, maize, tomato, pepper, cotton, and rice, is analysed in the database. Between these two phases, 80% of the images are used for training and 20% to assess the system's performance to guarantee accurate disease prediction. There will be a range of sizes for the image files that make up the database. Each shot is scaled down to the same dimensions before being swallowed by several artificial intelligence algorithms to lessen the possibility of any computer issues.

4.1. DATASET

The data collection used for this paper has 28 images, which will comprise images of healthy and damaged cassava leaves, maize plants, tomato plants, grape leaves, pepper plants, cotton plants, and rice plants, which will serve as the model's foundation. These pictures were taken from the data set (Starter: grape leaf disease 39aca76b-7). GitHub and Kaggle provided access to additional data sets, allowing the collection of additional photos. Grape leaf diseases are illustrated in the following figure (4.2), which draws its content from the plant dataset.

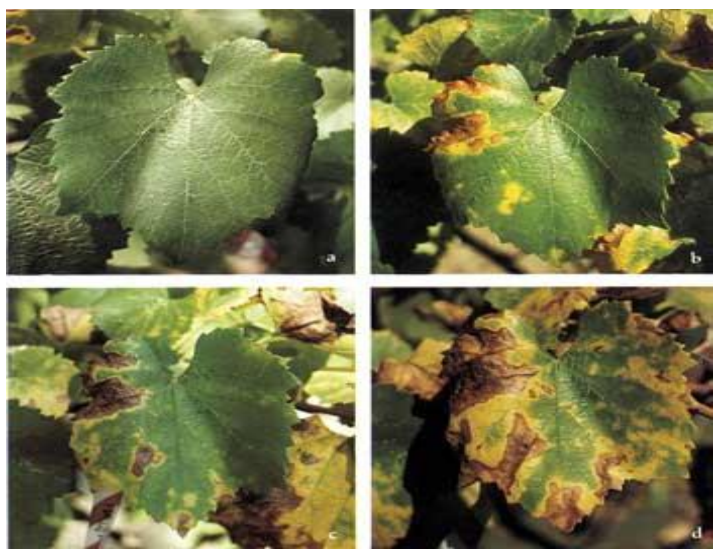

Figure 4.2. Different diseases of grape leaf plant.

4.2. PROPOSED SYSTEM

The AI engine will use the SVM separator phase illness solution previously selected and provided to the user to divide the image into its parts. The date and location of the images are being changed to reflect the illness that has been found. A cloud-hosted website has been created to offer details on various ailments. The user is provided with this information in the form of text and a map so they can follow it and take the appropriate action. Compared to conventional approaches, it can identify infectious illnesses early. Plant disease analysis comprises three main stages: feature categorization, feature background research, and feature editing. The Feature Classification category is identified by the letter K.

This procedure uses both the subtractive clustering strategy and the integration algorithm. When the textural elements are eliminated from the Quotation section, the GLCM is created. In Feature Divide, diagnostics based on the SVM Classifier are used. A matrix and accuracy will be used to evaluate one's level of success in this industry. The total value of the photos is multiplied by 100 times the total number of precisely separated images to determine the accuracy of the results.

The confusion matrix is a dimension 2 x 2 that displays the performance of the SVM separator while utilizing the correct of the true, false of the true, good of the false, and false, wrong parameters. In the earlier study, a K-means integration technique, a subtractive clustering algorithm, and an SVM classifier were used to identify the ailment more precisely. If any new needs are discovered, this research may be extended to address them. Additionally, it used AWS's cloud computing infrastructure and database to find every possible site.

The system may be expanded to accommodate new data sets spanning a larger range of diseases and plant varieties by simply altering parameters in AWS clouds. Currently, the suggested method considers many variables other than current and potential leaf diseases. It is feasible to think about and use various techniques since the future endeavor being imagined is as varied and potent as imaginable. Aside from the many opportunities that AI offers the agricultural sector, there remains a persistent knowledge gap regarding modern machine learning techniques. Data training equipment is necessary for artificial intelligence systems to produce accurate forecasts and predictions. This obstacle should be removed by making AI solutions accessible on open platforms, lowering their costs, and hastening their adoption rate.

4.2.1. General Experiment

Numerous tests were run on the image sets until the correct classifier parameters were identified in the training process. A set of sequences containing clean and noisy image data were used to train the multi-SVM model. The first phase featured typical images of numerous leaf varieties. The other sequences' order was muddled using the MATLAB program. The plants in training were exposed to various noises, including salt and pepper and Gaussian noise. The MATLAB code has been added to create a wide range of different noise examples.

Figure (4.2) shows an example of the original image and an image tainted by noise (salt and pepper noise with an intensity of 0.08). The graphic illustrates the comparatively low amount of noise added to the training images. These images undergo a preliminary processing stage before being uploaded to the network. The steps involved in photo preparation can be divided into several groups. The next step is to convert the image to a grayscale format after reading the image on the RGB scale.

Each pixel in the grayscale image is represented by an unsigned eight-bit integer (0-255). This figure shows what proportion of all pixels is white. When a pixel's value is zero, it has a black focus; as the value rises, the white focus rises as well, until it reaches the highest possible level of white, represented by the value 255. The RGB color space uses four distinct values to represent each pixel in an image. These three numbers represent various concentrations.

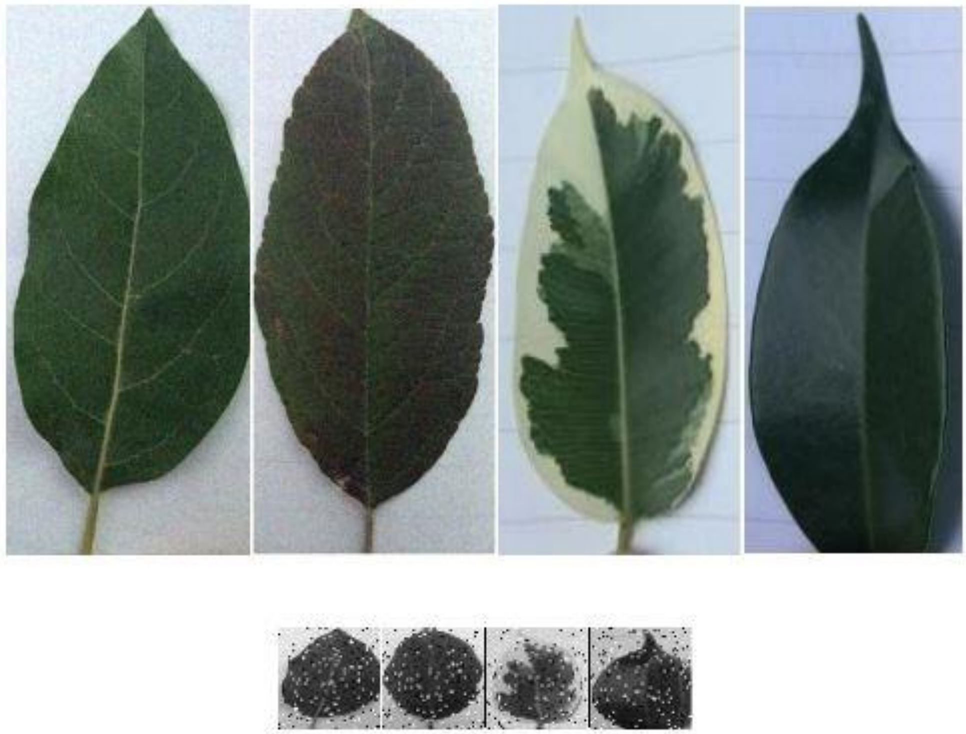

Figure 4.3. Training image examples.

4.2.2. Hue Saturation Intensity (HSI)

It is advised to use the Hue Saturation Intensity (HSI) model to create an RGB model. The (HSI) model is comparable to how colors are encoded by human vision according to hue. The HSI color space was created to "precision" regulate the color and better understand how different objects perceive and interpret color. HLS displayed with color (Hue, Lightness, and Saturation).

Traditional image processing methods like convolution, equalisation, histogram, and others perform best in the HSI color space. Because utilising equal proportions of the colors R, G, and B, these strategies are effective with light values [79],[80].

In contrast, it accepts any value between 0 and 360 degrees. When H is calculated using the cross-function, the output is always an integer between 0 and 180 degrees. The angle H must be greater than 180 degrees if B is bigger than G. Continue counting

H as you did previously, and if B exceeds G, use (360 degrees minus H) to calculate the appropriate hue value. The hue in an RGB-subspace distance triangle connects related white and the distance separating white from full color. The triangle's edges are lined entirely in a contrasting color. When a pure color is exposed to white light, the degree of purification is measured as fullness. The word "hue" describes the sober shades of yellow, orange, or red.

4.3. K-MEAN APPROACH

There is a chance that both positive and negative consequences could result from each of the several ways that can be used to learn more about a wide variety of plant diseases. K-Means, a recently verified approach for computer vision and machine learning, is now publicly accessible to anyone interested in using it. The commodities can be divided into K distinct groups using this strategy.

4.3.1. K-Mean Clustering

K-means is one of the simplest unconventional teaching techniques for solving an acknowledged integration challenge. The process is simple: apply the appropriate k groupings to divide the provided data set into the necessary number of groups. The basic idea is to create k centroids, one for each collection. It is crucial to place these centroids because various places have varied effects. The best course of action is to maintain the greatest feasible physical separation between one another [81],[82]. The K-means flowchart is displayed in figure (4.4).

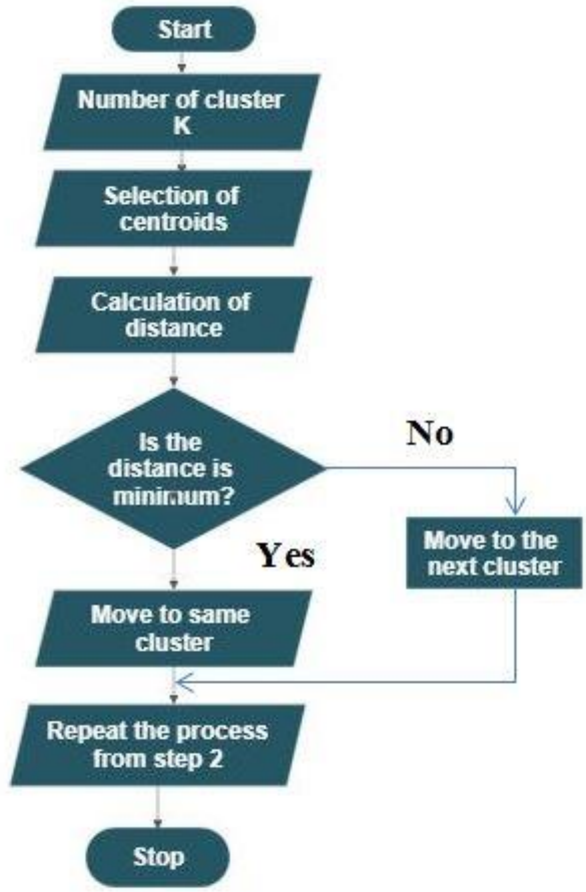

Figure 4.4. Diagram depicting the k-means clustering process [83].

Connecting each data point to the nearest centroid is the next step that needs to be taken. The first stage is finished right away, and the ages of the participants in the first group are established because there is no waiting period. The next step is to calculate the new k centroids necessary as the result of an earlier phase is a barycenter of collections. The points from the same data set will need to be bound together with a new centroid nearby once we have obtained these new k centroids and built a loop. As can be seen, this loop causes the positions of the k centroids to change until they eventually stop shifting altogether gradually. As a result, this does not affect centroids. The K-Means clustering algorithm produces numerous separate, unconnected groups [84]. Predictable, numerical, repeatable, and supervised, K-Means is a method. Hierarchical clustering is widely used to categorize photographs.

K collections will constantly be accessible for use. There will always be at least one item in every collection. There isn't a hierarchical or fragmented I collection to be discovered. Each collection member is physically closer to the "Center" of the collection than any other, despite proximity not always connecting one thing to another. Each observation is a collection member with a description in K-means, a sort of cluster analysis that splits n views into k groups [85],[86].

The K-means clustering method divides the data into groups by reducing the squares of the distances between the picture intensities and the cluster centroids. The method of iterative K-means clustering, Lloyd's algorithm, divides the data into k groups whose centroids specify how they should be grouped. Then, exactly one of those clusters is given to each n observation. The steps of the algorithm are listed below:

- Ascertain the k clusters' starting centers (centroid).
- Calculate the distances between each cluster's center and all observations, using points as the unit of measurement.
- Insert each observation into the cluster whose centroid is closest to the observer.
- Before going on to step 5 to find k new centroid sites, find the mean data for each cluster.
- Keep going back and forth between steps 2 and 4 until the cluster assignments stay the same, or the allotted number of repetitions has been reached, whichever comes first.

4.3.2. K –means Clustering for Color Image

The process of grouping pixels with similar properties is called clustering. The user-specified number of clusters and distance metric is necessary for k-means clustering. Segmentation is done when the input RGB picture is transformed to L*a*b color space format.

Since K merging disperses K inches randomly among the data, it is also known as Hierarchical Separation [87]. K ought to be produced randomly at centroids. Choosing random data points will allow you to create those first points. If not, each choice will require you to choose K numbers randomly. Which centroid is closest to each data point should be determined. Use a scale like the Euclidean or Cosine range. Give the centroid that is closest to it a data point. Calculate the centroid by deducting it from the average points provided for each cent. Replicate the past three phases until the centroid distribution does not change. The algorithm is said to be "integrated" when no modifications are made [88],[89].

The light a thing generates determines its hue as the sun's rays pass through the glass prism. Instead of being white, the light that manifests is a continuous spectrum of colors that ranges from violet on one side to red on the other. The visible light spectrum spans 900 to 700 nanometers. Gloss, light, and light are three concepts that describe color. The RGB color model, a fundamental color model, provides the foundation for color models [90].

4.4. IMAGE PREPROCESSING

Digital images are collected and integrated to produce the finished output of specialized photos, such as realistic or realistic interior item alterations. The phrase describes how images are printed, displayed, and stored in a congested environment. Thinking about how an image will seem after transforming from its analog to digital form after image capture can be depressing. Image enhancement, which involves adding extra effects to digital photos, can help display many images or perform image analysis. Identifying the most important traits is simple because you can do things like turning off the music, emphasizing particular elements of an image, and so forth. Different facets of exploitation include the camera's depth of field, the volume of noise, and the blurriness of movement. The two objectives of picture restoration techniques are lower noise levels while restoring any lost resolution [91].

Understanding both the physics of light and color is necessary for image processing. People can access color-coded information about meals, locations, objects, and times of the day. They are free to utilize this information however they see fit. It is possible to distinguish between the various picture-processing strategies using diverse tones.

When the Decorated image crosses the sky, clouds, trees, and flowers, you must use a brush with a level that varies according to the size of the text graphics. Waves and strokes serve as the contrast. Wavelet conversion is a practical method that can be used while creating visuals. The wavelet conversion makes it possible to investigate various visual possibilities. Because image compression reduces the storage and transfer space needed, digital cameras must function adequately [92].

Using optical character recognition (OCR), a technique, handwritten or scanned text can be transformed into digital text that computers can read. It frequently refers to the accessibility of records that start with introducing new data sources, such as paper, an invoice, a bank statement, income, business cards, the number of printed records, or email. Paper, an invoice, a bank statement, an income statement, and an email are examples of this data source. A technique frequently used when creating digital print manuscripts is to create documents that can be filtered, searched, and stored in computer processes such as text-to-speech, data retrieval, machine translation, and online display. A substantial body of work in computer intelligence, pattern recognition, and concept recognition is the Optical Character Recognition (OCR) project [93],[94].

4.5. IMAGE SEGMENTATION

It is a technique for condensing the representation of an image into intelligible shapes and separating objects of interest from the background of an image. The use of picture segmentation can reduce the size of a photograph. The images in this publication were categorized using the k-means clustering technique. Based on centroids, iterative K-means clustering allocates each n view to precisely one of the k groups [95].

The distance Euclidean metric method, used by default in the MATLAB k-means function to compute point-to-cluster centroid distances, yields square computations by design. Here, the collection's centroid mean points and the corresponding Euclidean distances between them are shown. The following numbers are used in the Euclidean approach of k-means compilation [96].

A wide range of collections is investigated to establish the ideal number. Subordination may result in segregation if numerous collections are used. On the other hand, as shown in table (4.1), which shows the quality analysis for k-means, having low cluster numbers can result in the separation being excessively precise.

Table 4.1. Quality analysis.

No. of clusters	Analysis
1	For k-means segmentation using clustering parameter of the majority area of leaves is generalized into one cluster. This results in cluster overfitting.
2	For k-means segmentation using clustering parameter of the majority area of leaves is generalized into one cluster. This results in cluster overfitting. While the background is clustered into two separate clusters, resulting in underfitting.
3	For k-means segmentation, the clusteringparameter of the area of the leaves is clustered well and not overfitting or underfitting. The background is also well clustered into one cluster.

4.6. FEATURE EXTRACTION

Data that may be utilized to determine a sample's value is extracted from feature releases. The most significant traits are shape and color. The word "texture" describes

how a surface appears in a picture. Identifying Powdery Color Marks and Downy Mildew takes a few mildew traits. Therefore, the system's color and texture components are retrieved for detection in this instance.
These methods are used to compute color characteristics:

- The first RGB to HSV color conversion was performed.
- The picture is split into equal 3X3 chunks.
- To determine the medium color (H/S/V) for each of the nine blocks, apply the following formula:
- Xc = 1 N xi, where xi is the pixel size and N is the number of pixels (2). In this instance, it denotes "considered one of the variables."
- The formula below is used to determine the difference in each block.
- The tilt of each (H, S, and V) block is determined.

The GLCM characteristics are retrieved from the picture after segmentation. The statistical technique for analyzing texture, known as the Gray-Level Cooccurrence Matrix (GLCM), considers the spatial connection between pixels. By calculating the spatial connection between the pixels in the pictures, the GLCM functions describe the texture of the images. From this matrix, statistical measurements are obtained. An array of offsets that indicate pixel associations of various directions and distances must be given when creating GLCMs. In the suggested technique, contrast, energy, homogeneity, and correlation are extracted characteristics. The (I, j) element in the normalized Gray-Level Cooccurrence Matrix will be represented by Pi. The quantized image's N stands for the number of unique grey levels [97],[98],[43]. Diagram for identifying and categorizing leaf diseases The programming is shown in Figure (4.5).

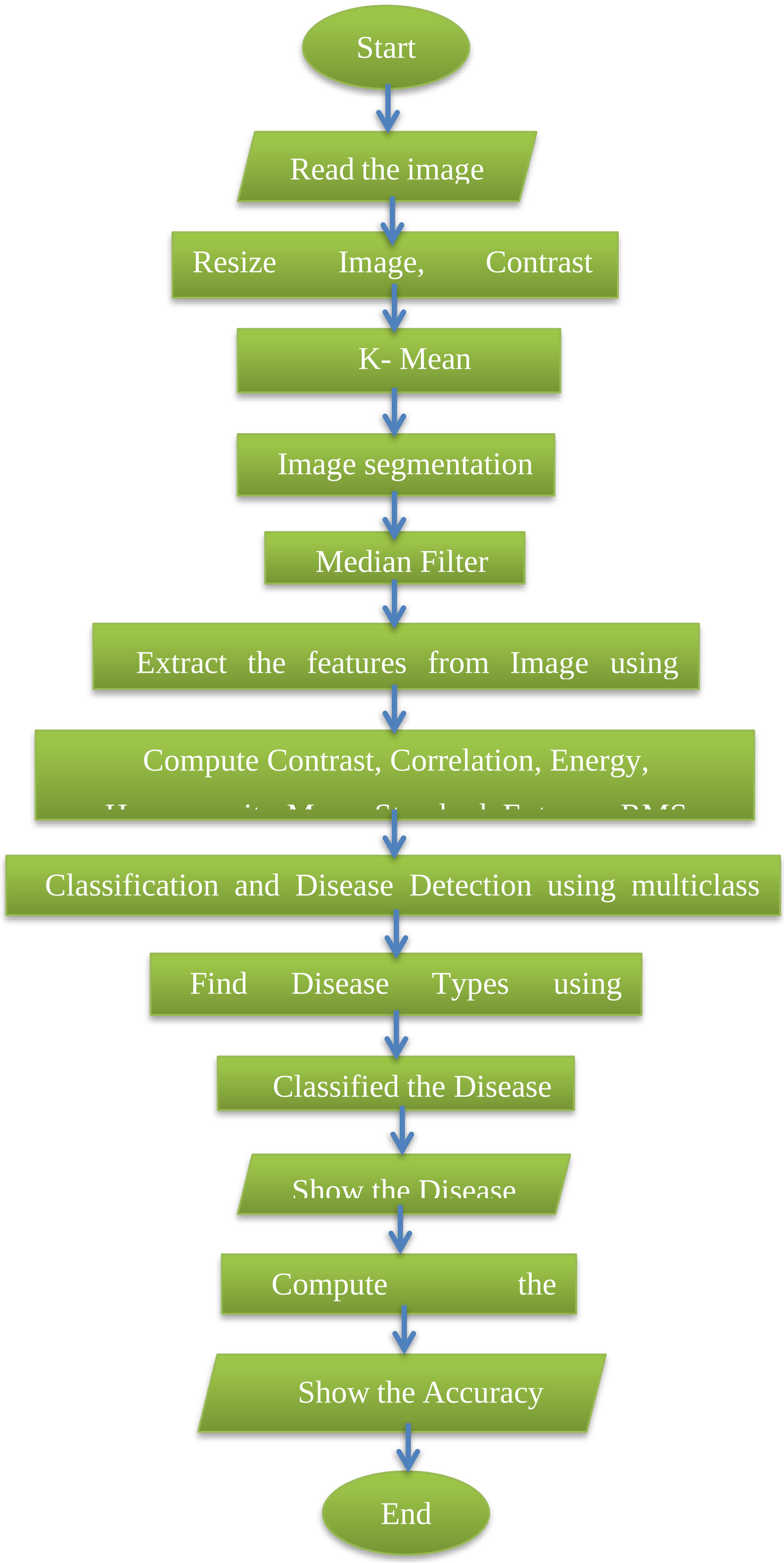

Figure 4.5. Leaf disease diagram coding.

The system's methodology for identifying and categorizing leaf diseases is as follows:

- Examine the inserted picture.
- Image resizing.
- Image enlargement versus image resizing.
- Switch the RGB to HSI color space.
- Employ the k-mean group function.
- The picture has three auxiliary components.
- Select a diseased area.
- To the picture, apply the center filter.
- Eliminate the feature from the picture using the Gray-Level Cooccurrence Matrix (GLCM).
- Make the following calculations: root means square, entropy, homogeneity, standard, relationship, brightness, and power.
- To distinguish between infections, use a multiclass support vector machine.
- Calculate accuracy.
- Show precision.

4.7. THE PROCEDURE FOR THE DIAGNOSIS

The main goal of this study is to identify and describe leaf diseases in various plants. The five stages that we took might be considered the diagnostic process for these illnesses:

- The dataset was used to initialize all images, which are subsequently stored in various formats. (.jpg,.png), (.bmp), (.gif), etc.
- Use the K-means strategy to divide the photos into specific groups.
- The texture properties accessible from the different groups were obtained using a grey-level co-occurrence matrix.
- The SVM classifier is used to stratify various plant leaf diseases after removing

all readily available compositional information.

- The completed data is uploaded to the cloud server after modification.

We got a lot of the data needed for this study thanks to Web Services (AWS), which offered a variety of sophisticated algorithmic strategic planning tools and a complete platform for resource measurement. We could access the Internet thanks to AWS, which made this possible.

4.7.1. K-Means Algorithm work Steps

A novel computer vision and machine learning algorithm has been tested and published. Using a partitioning algorithm, K-Means divides up the items into K groups.

The steps of the algorithm are as follows:

- To begin, we will generate K random centers for the set.
- Computing the range of each pixel for every kth aggregation center we have.
- Giving each pixel the address of its immediate neighbor.
- We first calculate the common pixel definition and the group value to modify the neighboring group.
- We won't stop until you've gotten to the very last pixel.

The algorithm is used for a powerful picture, such as I(x,y) with x = 0,1,2... y = 0,1,2 and M-1. The following are the processes for obtaining an N-1:

- We begin by choosing K centres at random where I = 0, 1, 2, K1, the centre is the initial aggregation centre.
- To get the pixel distance from each cluster position, we use the formula Distance = | Center - I (x, y) |, where I = 0, 1, 2,... 1 - K (1).

- Work out the shortest route. Set this pixel to the institutional reference of the appropriate group using the formula Tp = minimal distance p, the value of I for which the distance is small.
- Verifying that the centre of clustering centres equals centre plus I (x, y). Up until the final pixel, repeat these steps.

Following these steps, the image will be split into k separate images, each composed of a group of pixels with the same intensity.

MCPDA implementation (Modified Color Recovery Detection Algorithm) The digit count can be decreased with this approach. In this instance, incorrect image pixel values can be noticed in the impacted area of the image. Finally, our method finds the red pixels of the unhealthy region in the image. By using a red pixel image once to create adequate thresholds for color processing algorithm acquisition, this algorithm is available for a more accurate diagnosis. This application correctly detects yellow plant illnesses hidden between the leaves.

The following are the steps:

- Changing the RGB image that was rendered into a grayscale image value (GSV) GSV equals 0.298 R, 0.587 G, and 0.114 B. GSV equals 0.298 R, 0.587 G, and 0.114 B. GSV equals 0.298 R, 0.587 G, and 0.114 B. GSV = (3).
- Using RIV as the red image strength values and BIV and GIV as the blue and green image strength values, we define each fraction of R, G, and B.

The damage to a plant's leaves that results from a disease is frequently thought to be the first indication that the sickness will spread to other plants. The suggested system will conduct a study. Here are a few leaf ailments:

- Leaf spots: Leaf spots result from the fungus adhering to a leaf. It starts as a little brown patch on the leaf's skin. It then spreads throughout the entire plant, seriously harming it.

- Leaf Rust: The lower leaves of plants with this fungus, typically seen on mature plants, rust.
- Powdery mildew: A fungus infects plants, causing white, powdery leaf holes.
- Downy mildew: Affected plants (oomycete organisms) produce white or grey letters in the veins of their leaves, which facilitate crop harvesting.
- The mosaic virus infects plants and leaves them with white, yellow, or green (light or dark) dots.
- Yellow Leaves: Low or excessive watering causes moisture stress, which results in the yellowing of the leaves.
- Leaf curl: This illness causes leaves to twist, curl, and change color. Both mildew and a virus are the culprits.

4.7.2. Mechanism of Multi-SVM

The classification tool known as the Vector Support Machine employs the learning technique known as kernel-based learning. The distance between the data used for training and the class borders widens while using the SVM training method. The decision-making function that results from using only training data, or support vectors, is shown in Figure 2. The decision limit is close to where this decision-making process is located. It performs remarkably well when the size number in the elevation space is greater than the quantity of training data points. The support vector machine (SVM) can transform data from the input space into the feature area with the largest possible size with the aid of the kernel function [99].

At a location with a large magnitude, indirect data can be sorted using a hyperplane. The complex computer known as Kernel Hilbert Space downloads it (KHS). The feature vector is the input being given to the director. The site picture's vector portion divides the training and testing vectors. The moderator will initially get some experience by dealing with the training set before starting to split the test set. A separator's effectiveness can be assessed by contrasting expected and actual labels.

SVMs, or "vector support equipment," are in use today and have been extensively used in several settings. SVM uses the Structural Risk Minimization (SRM) principle, in contrast to most other techniques. Consider decreasing the upper limit of the standard error rather than focusing on training error [100]. There are several similarities, the most noticeable of which are the high overall effectiveness, the lack of space minima, and the compact solution description.

By projecting the input data onto a higher dimensional space that incorrectly includes space and necessitates the inclusion of a divisive hyperplane in the feature region, SVMs can solve segregation difficulties. Another potential use for the kernel technique is the modification of feature maps. Nonlinear support vector machines (SVMs) use kernel functions like the radial basis function (RBF), polynomial functions, and others to match data sets with more complex decision regions [101].

Like SVM models for many other indirect distinctions, SVM models for indirect kernels must deal with the encryption problem, which can arise when few training models are available due to the large VC size[102].

In this essay, we concentrate on a few issues with binary division. Xi E Rd is the input vector corresponding to Ith samples tagged Yi E I, I by its category. A separator with a specific label checks a collection of N samples (X1, YI),..., (Xi, Yi),..., (XN, YN) (1).

The indirect separation boundary can be created by splitting the local line boundary nJ X bj, where j = 1,..., M, into M pieces and then joining them, as illustrated in Figure (4.6). As many times as necessary, this procedure can be performed.

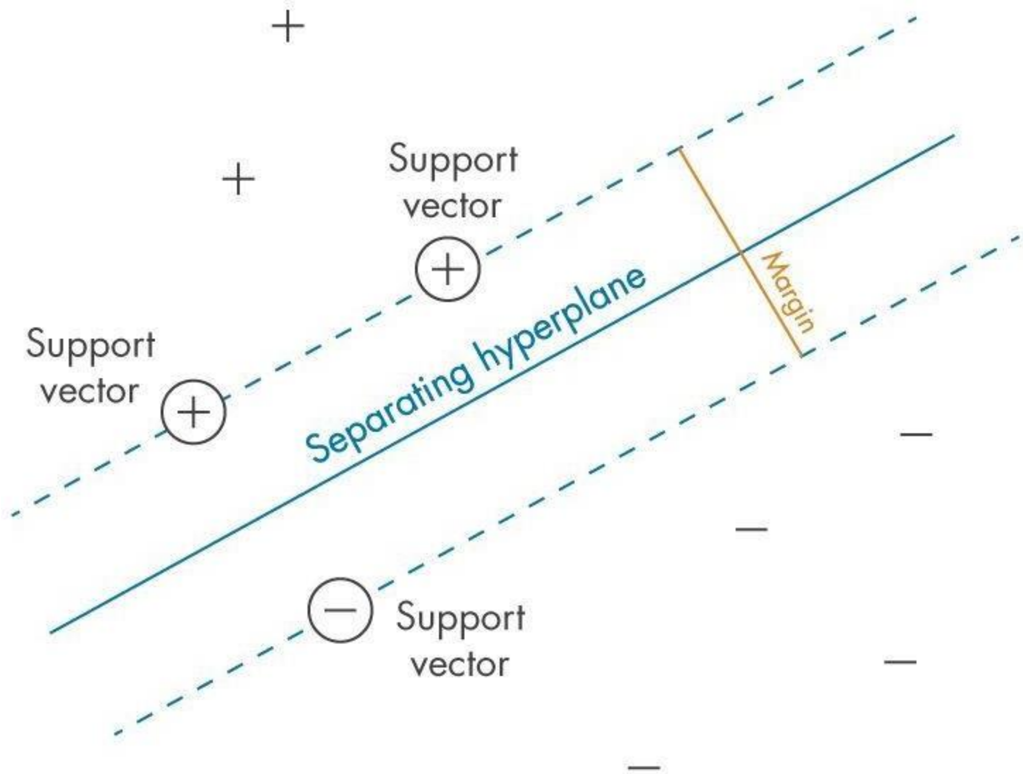

Figure 4.6. Classificaton using multisvm.

4.7.3. The Categories of Multi-Svm

MATLAB could be a helpful program for the SVM classifier training process. The various classes are divided based on the following feature categories using a multi-vector support machine:

- Line SVM, an indistinguishable data structure with different results in either a high- or low-magnitude region, is converted utilizing a line of kernel work with the default kernel scale.
- Quadratic SVM, the polynomial kernel function of polynomial order, q 2, is combined with an automatic kernel scaling while doing a quadratic SVM analysis.
- In contrast, the SVM Cubic algorithm uses a default kernel size and a polynomial kernel function in polynomial order.

- Assistance with the SVM using precise Gaussian coefficients. The Fine Gaussian Support Vector Machine uses circular Gaussian kernels with a kernel rating of 0.71. Data sets can be separated according to the most important characteristics or parameters thanks to the unbounded dimensionality that Gaussian characters allow.
- Central Gaussian SVM, the Medium Gaussian SVM, on the other hand, uses a different kernel size from the other kernels while still using the same Gaussian kernel as scale (5).

PART 5

RESULTS AND DISCUSSION

5.1. ACCURACY

The classification system was chosen based on the predicted effects of the employed plant diseases. Table (5.1) lists the 28 photographs in the dataset, which were distributed as follows: 5 images of anthracnose, 3 images of bacterial blight, 7 images of citrus canker, 9 images of grey mould, 2 images of powdery mildew, and 2 images of unharmed nature (5.1).

Table 5.1. Distribution of the disease based on categories.

ID	Name of disease	Number
ID 1	Anthracnose	5
ID 2	Bacterial blight	3
ID 3	Citrus canker	7
ID 4	Grey mould	9
ID 5	Powdery mildew	2
ID 6	Normal images	2

SVM recognizes ailments like plant sickness, and a graphical user interface (GUI) idea was developed for the diagnosis. The classifier works wonderfully, with an accuracy rate of 98.387%, when there are clear indicators of sickness. It can be shown that Multi-class SVM yields acceptable results when classifying more than two classes; the accuracy of the SVM classification system is shown in table (5.2)

Table 5.2. SVM classification accuracy.

Type	Affected Area	SVM Classification
ID 1	53.5633%	96.7742%
ID 2	15.1108%	96.7742%
ID 3	15.0093%	96.7742%
ID 4	18.2951%	98.3871%
ID 5	5.7651%	96.7742%

5.2. BASIC CALCULATION

5.2.1. RGB To HIS Math

The minimal value is between red, green, and blue. The formula for calculating the value of the HSI color filling space is as follows: $S = 1 - m/I$ if I is greater than 0, and $S = 0$ if I is equal to 0.

Using of the following formulae to convert the hue, H, of a whole color to an angle scale:

$$H = \cos{-1}[(R - ½G - ½B)/\sqrt{R^2 + G^2 + B^2 - RG - RB - GB}]$$

if $G \geq B$, then $H = 360 - \cos{-1}[(R - ½G - ½B)/\sqrt{R^2 + G^2 + B^2 - RG - RB - GB}]$ if B is greater than G, where the output of the inverse cosine is measured in degrees.

5.2.2. K-mean Algorithm Calculation

Assume we have a value for K and x1, x2, x3, and xn data points (number of sets required). This is how the method works:

Initial K, choose K, and random K designations for the first centroids in the database. Get the Euclidean range for each point in the database using the K-marked points (cluster centroids). Assign a data point to the closest centroid using the distance obtained in the previous step. To construct a new centroid, average the results for each group in the sample. Until the centroids stay the same or you have completed a certain number of repetitions, repeat steps 2-4.

Each data point x is allocated to a cluster based on its centroid if each cluster is designated by the letters ci.

Distance Arg min dist(C1,x)2 Equation 1

EUCLIDEAN DIMENSION

D(p,q)=((q1-p1)2/+(q2 - p2)2/1)

EUCLIDEAN DISTANCE EQUATION 2

Finding new the centroid from the clustered group at points

$$C_i=(1/|S_i|)* \Sigma_{xi£s} X_i \quad (5.1)$$

5.3. STATISTICAL CALCULATION

Table (5.3) displays the results obtained from the chosen datasets after the testing phase was completed; the characteristics and equations are shown in table (5.4).

Table 5.3. Provides values from tested datasets.

S.n	Contrast	Corr.	Energy	Homo.	Mean	S.D	Entropy	RMS	Var.	Smooth-ness	Skew	IDM
1	0.77	0.86	0.07	0.78	92.14	64.56	7.7281	105.79	2.79	1.0000	2.37	0.69
2	1.1909	0.77	0.05	0.72	103.38	57.23	7.7327	110.44	1.38	1.0000	1.93	0.05
3	0.7021	0.87	0.06	0.78	100.45	72.31	7.7569	120.45	4.17	1.0000	2.46	0.66
4	0.8246	0.85	0.06	0.76	130.01	61.46	7.87	139.85	2.58	1.0000	2.32	0.07
5	1.0456	0.79	0.05	0.72	128.84	60.92	7.8141	137.89	2.39	1.0000	2.22	0.27
6	0.7104	0.87	0.07	0.78	92.35	63.87	7.7686	105.74	2.72	1.0000	2.35	0.64
7	1.0724	0.77	0.06	0.72	115.56	46.16	7.5345	122.16	1.49	1.0000	2.34	0.02
8	0.5808	0.89	0.07	0.81	116.77	63.07	7.8314	127.44	2.30	1.0000	2.28	-0.08
9	0.3834	0.89	0.12	0.86	94.97	60.28	7.5249	104.79	1.53	1.0000	2.11	-0.04
10	0.3581	0.92	0.10	0.85	149.79	64.77	7.8096	161.0	3.32	1.0000	1.76	-0.13
11	0.4943	0.90	0.09	0.83	77.37	60.22	7.4934	91.00	2.18	1.0000	1.88	0.35
12	0.4342	0.91	0.10	0.84	77.03	62.32	7.5871	91.62	2.39	1.0000	2.25	0.63
13	0.4865	0.89	0.09	0.83	57.07	50.75	7.0925	66.78	1.17	1.0000	2.30	0.62
14	0.2871	0.93	0.12	0.89	126.71	75.97	7.6694	147.1	5.32	1.0000	1.43	-0.16
15	0.5580	0.87	0.09	0.83	68.58	54.87	7.2520	80.80	1.84	1.0000	2.04	0.39
16	0.4054	0.85	0.35	0.87	80.11	43.42	5.6806	85.00	8.63	1.0000	2.71	0.75

Table 5.4. Features and formula.

No.	Feature	Formula
I	Contrast	Maximum pixel intensity - minimum pixel intensity.
2	Infirmity (Energy)	Σ Σ(c(I,j)^2
3	Maximum probability	P(A or B) = P(A)+ P(B) –P(A and B)
4	Homogeneity	Σ Σ(c(I,j)/1+(I,j)^2)
5	Entropy	Σ Σ(p(I,j) log p(I,j)
6	Correlat1on	Σ Σ((i-μ)*(j-μ) c(i.j)/σ1σ2)

- **CONTRAST**

This function gives a pixel's neighbor the brightness ratio for every pixel in the picture (size (GLCM, 1) - 1) 2 In terms of width, [0 (size (GLCM, 1) -1) 2]. The brightness is 0 in the static picture. Property contrast is often referred to as variance and inertia.

- **CORRELATION**

Returns a measurement of the relationship between a pixel and its neighbors in the picture. The breadth of [-1 1]. One or one of an image that is favorably or negatively connected with another is referred to as a connection. The URLs are just static NaN pictures.

- **ENERGY**

The total number of squares is sent to GLCM. The static image's zoom level is [0 1] power. Location Power is referred to as similarity, power similarity, and the second angular moment, displaying both characteristics and values, is a table (5.3).

Table 5.5. Features and values.

Features	ID 1	ID 2	ID 3	ID 4	ID 5
Contrast	0.5136	1.3937	1.9776	1.1178	0.2878
Correlation	0.7101	0.7532	0.7700	0.8307	0.9573
Energy	0.8941	0.6954	0.5852	0.4824	0.3633
Homogeneity	0.9715	0.9257	0.8975	0.8960	0.9573
Mean	17.137	23.3472	33.7035	35.4631	45.2763
SD	35.541	59.2528	70.1363	63.6863	61.3455
Entropy	2.8453	2.5652	2.2144	3.1349	4.0536
RMSE	10.453	6.5303	6.6518	8.2058	10.3502
Variance	1.16e+033	3.38e+033	4.38e+033	3.32e+033	3.2le+03
Smoothness	1.0000	1.0000	1.0000	1.0000	1.0000
Skewness	4.6761	2.5459	1.7394	1.6646	0.9597
Kurtosis	27.5418	8.2026	4.3163	4.6652	2.3851
Inverse Difference Moment	255	255	255	255	255

5.4. RESULT FROM SCREENSHOT

Figure (5.1) depicts multiple plant diseases using a K-mean cluster; figure (5.2) depicts a positive result of a plant leaf disease, figure (5.3) depicts a display window, the figure (5.4) depicts a comparison of K-Means clusters 1 versus 2 versus 3, the figure (5.5) depicts the final detection and results.

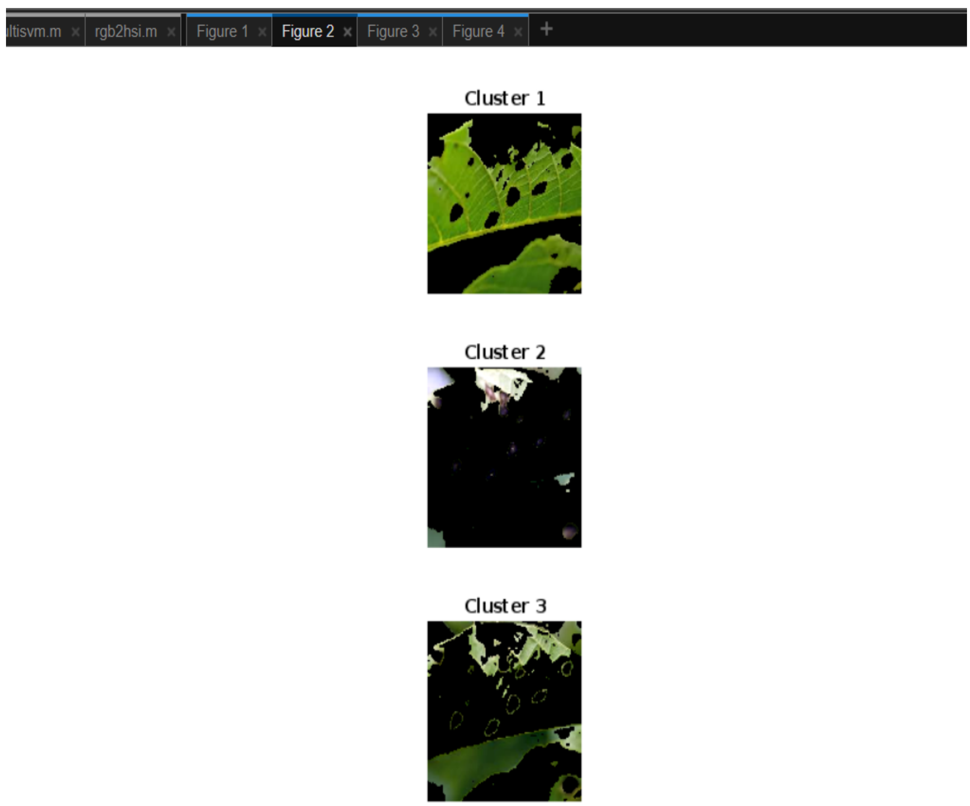

Figure 5.1. Plant disease detection using K-mean cluster.

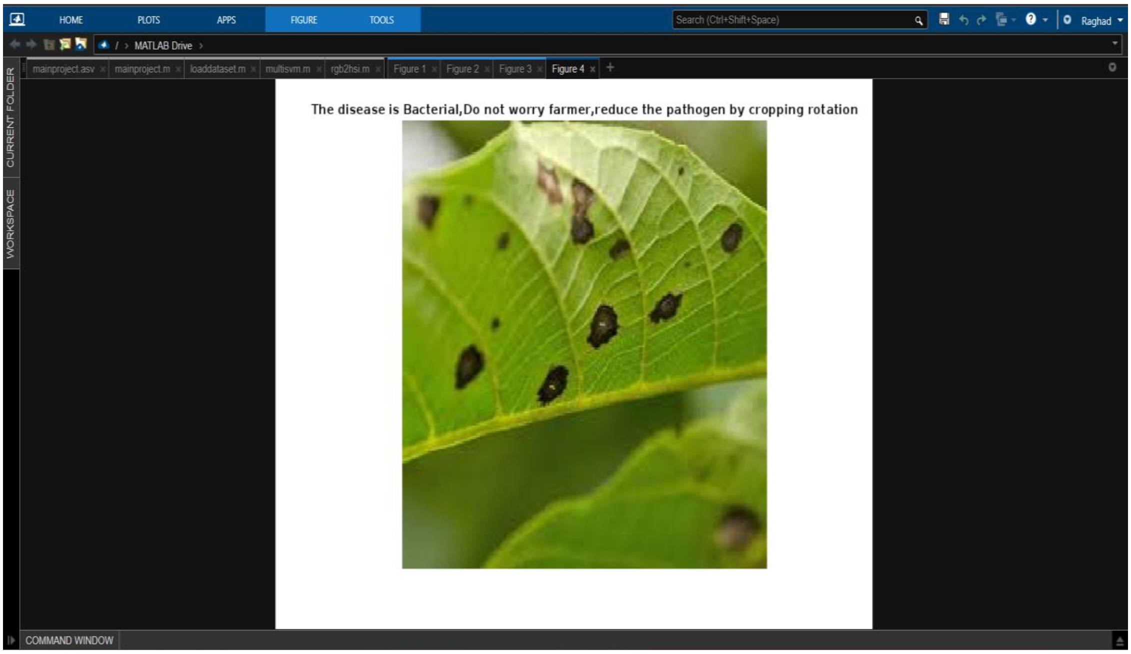

Figure 5.2. A positive result of a plant leaf disease.

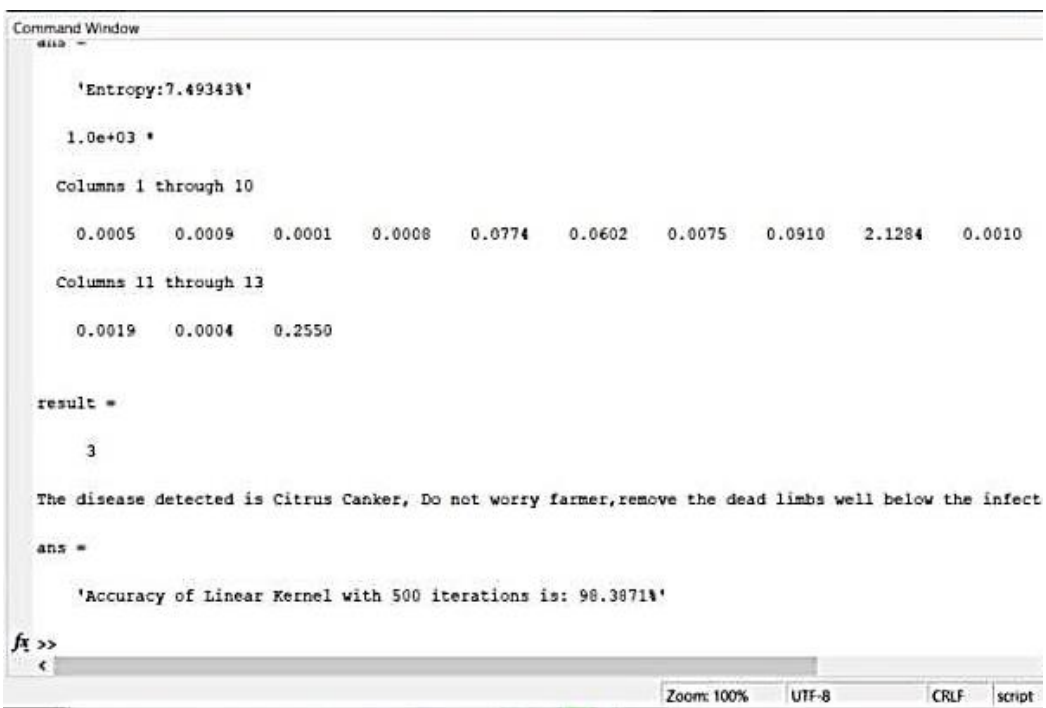

Figure 5.3. Display window of accuracy

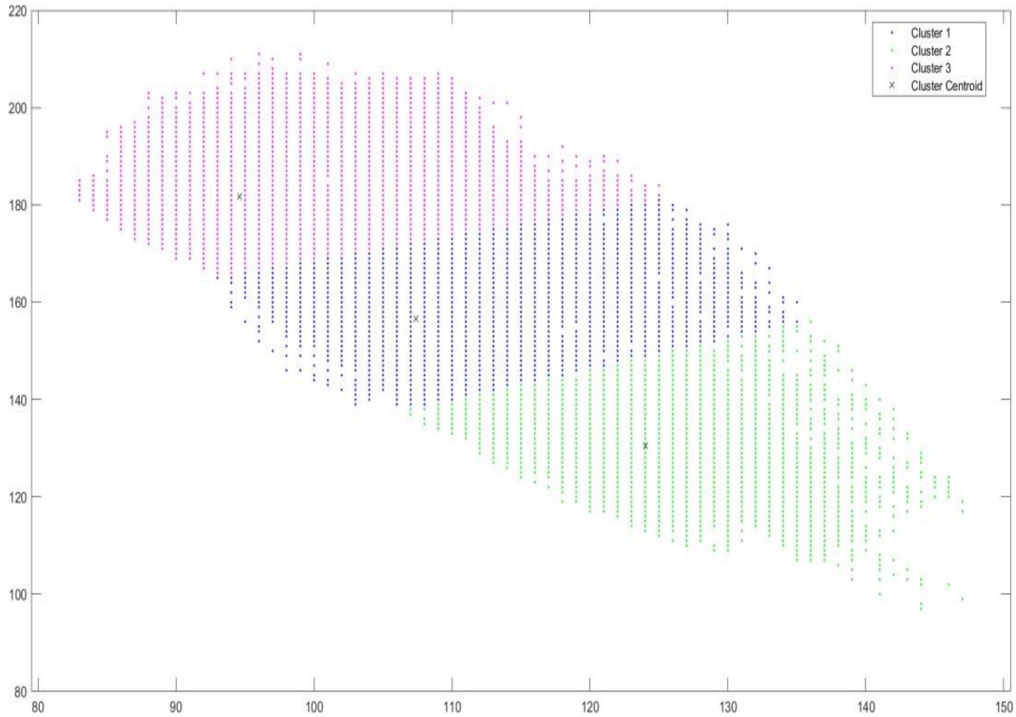

Figure 5.4. K-means clusters 1 VS 2 VS 3.

Figure 5.5. Final detection and RES.

5.4. DISCUSSION

Support vector machines (SVMs) were developed to categorize data into binary categories. The process of figuring out how to broaden it into categorization is still ongoing. Combining a few binary class dividers has been suggested to create multi-stage dividers. Some writers also provide strategies for simultaneously taking into account every class. Due to the high mathematical cost of solving multi-stage problems, a thorough comparison of approaches exploiting large problems has not been made. Testing has traditionally been restricted to small data sets, which poses a significant development challenge, particularly for multi-phase SVM solutions in one step. In this work, we discuss the declining usage of two such "whole" approaches:

The effectiveness of the three binary-based approaches—" one versus all," "one against one," and DAGSVM was then contrasted. Our results show that the "one against the other" and DAG tactics outperform other methods. The findings also imply that a few supporting vectors are typically needed in cases with significant difficulty when looking at all the data simultaneously.

SVMs (Support Vector Machines) were developed to separate binary data. The process of figuring out how to broaden it into categorization is still ongoing. Currently, there are two different kinds of multiclass SVM modes. Option one is to make and combine several binary categories. For the majority of these techniques, there are no comparable comparisons at this time. The multi-phase SVM problem resolution design in one step is more adaptable regarding the number of classes. The method for identifying and classifying leaf diseases makes use of multiclass SVM. A section of the sick leaf was separated using K-Means separation.

Consequently, multi-classroom SVM systems must have a few binary separators, or there will be a significant efficiency issue. Consequently, it is usually more costly theoretically to solve a multi-stage issue than a binary one with the same number of variables. Little-sized data sets have as far been evaluated. This work will discuss the fading of two such "whole" techniques. Then, we contrasted their findings with those of three binary classification techniques: DAGSVM, "one versus all," and "one against one."

Areas of cotton leaves unaffected are identified using the K and methodology for combining K and color conversion structure, where RGB is transformed into a Web color space. Because only a tiny set of k algorithms can yield good results, it stands out among the numerous collections of pictures from various places since it has many k. Different starting divisions may lead to different end collections.

The advantage of K suggests that the approach is straightforward and efficient. This technique works effectively when collections are not precisely segregated. The GLCM texture features are removed, and the partition is separated using Multiclass SVM. The method is now being tested to detect illness in plant leaves. Future research should aim to increase classification precision while detecting illnesses in diverse leaf and plant species.

PART 6

CONCLUSION

This paper's primary topic of discussion is an automated plant leaf diagnosis tool that the researchers developed. The primary purpose is determining which leaf portions of the various plant fruits impact the illness or condition. In the beginning, we utilized preprocessing to reduce the amount of noise within the database. The Haralic approach is applied to obtain the most pertinent characteristics of the tissue that is being investigated. Finally, several machine learning techniques, including retrieval, random forest, and support vector machine analysis (SVM), were utilized to differentiate features obtained from the plant leaf disease website. The results that the SVM generates are superior in all three models. Real-world applications, such as determining which papers are healthy and which are unhealthy, might use this functionality to differentiate between the two categories. A wide range of permutations and combinations are tested to calculate the number of combinations that get the best results. Dependency can easily lead to class when there are many possible choices. On the other hand, having a low cluster number could result in excessive space between each cluster. As a result, and by reviewing the values mentioned in Table 5, it is clear to us that we have obtained a homogenization ratio (97.15%, 92.57%, 89.75%, 89.60%, and 95.73%) for the five models used in the study (ID 1, ID 2, ID 3, ID 4, ID 5) respectively, and accuracy of the linear kernel with 500 iterations is 98.38%.

REFERENCES

1. Padol, P. B. and Yadav, A. A., "SVM classifier based grape leaf disease detection", ***Conference On Advances In Signal Processing, CASP 2016***, 175–179 (2016).

2. Chakraborty, S., Paul, S., and Rahat-Uz-Zaman, M., "Prediction of Apple Leaf Diseases Using Multiclass Support Vector Machine", ***International Conference On Robotics, Electrical And Signal Processing Techniques***, (February): 147–151 (2021).

3. Islam, M., Dinh, A., Wahid, K., and Bhowmik, P., "Detection of potato diseases using image segmentation and multiclass support vector machine", ***Canadian Conference On Electrical And Computer Engineering***, (April): (2017).

4. Potter, M. C., "Bacterial diseases of plants", ***The Journal Of Agricultural Science***, 4 (3): 323–337 (1912).

5. Lindenthal, M., Steiner, U., Dehne, H. W., and Oerke, E. C., "Effect of downy mildew development on transpiration of cucumber leaves visualized by digital infrared thermography", ***Phytopathology***, 95 (3): 233–240 (2005).

6. Singh, A. and Kaur, H., "Potato plant leaves disease detection and classification using machine learning methodologies", ***IOP Conference Series: Materials Science And Engineering***, 1022 (1): (2021).

7. Jaisakthi, S. M., Mirunalini, P., Thenmozhi, D., and Vatsala, "Grape leaf disease identification using machine learning techniques", ***ICCIDS 2019 - 2nd International Conference On Computational Intelligence In Data Science, Proceedings***, (February): 1–6 (2019).

8. Huang, Z., Qin, A., Lu, J., Menon, A., and Gao, J., "Grape Leaf Disease Detection and Classification Using Machine Learning", ***Proceedings - IEEE Congress On Cybermatics: 2020 IEEE International Conferences On Internet Of Things, IThings 2020, IEEE Green Computing And Communications, GreenCom 2020, IEEE Cyber, Physical And Social Computing, CPSCom 2020 And IEEE Smart Data, SmartD***, (January): 870–877 (2020).

9. Kulkarni, P., Karwande, A., Kolhe, T., and Kamble, S., "Plant Disease Detection Using Image Processing and Machine Learning", ***ArXiv Preprint ArXiv:2106.10698***, (2021).

10. Kothari, J. D., "Plant Disease Identification using Artificial Intelligence : Machine Learning Approach", ***International Journal of Innovative Research in Computer and Communication Engineering*** (November 2018): (2020).

11. Communication, S., "A comparative analysis of machine learning approaches for plant disease identification", ***Advancement in thre life science***, 4 (4): 120–126 (2017).

12. Yang, X. and Guo, T., "Machine learning in plant disease research caused by pathogens lead significant crop yield loss globally . Numerous researchers have been studying", ***VOL*** 34 (8): 1–5 (2019).

13. Zadeh Shirazi, A. and Mohammadi, Z., "A hybrid intelligent model combining ANN and imperialist competitive algorithm for prediction of corrosion rate in 3C steel under seawater environment", ***Neural Computing And Applications***, 28 (11): 3455–3464 (2017).

14. Abdu, A. M., Mokji, M. M., and Sheikh, U. U., "Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning", ***IAES International Journal Of Artificial Intelligence***, 9 (4): 670–683 (2020).

15. Ryu, M., Yun, J., Miao, T., Ahn, I. Y., Choi, S. C., and Kim, J., "Design and implementation of a connected farm for smart farming system", ***2015 IEEE SENSORS - Proceedings***, (November): (2015).

16. Shashwathi, N., Borkotoky, P., and K, S., "Smart Farming: A Step towards Techno-Savvy Agriculture", ***International Journal Of Computer Applications***, 57 (18): 45–48 (2012).

17. Tiwari, S., "A comparative study of deep learning models with handcraft features and non-handcraft features for automatic plant species identification", ***International Journal Of Agricultural And Environmental Information Systems***, 11 (2): 44–57 (2020).

18. Rani, R. U., "Pest Identification in Leaf Images using SVM Classifier", ***International Jornal of computational intelligence and informatics***, (June 2016): (2018).

19. Dhaware, C. G., "A Modern Approach for Plant Leaf Disease Classification which Depends on Leaf Image Processing A Modern Approach for Plant Leaf Disease Classification which Depends on Leaf Image Processing",International conference on computer communication and informatics (ICCCI), 2017 .

20. Mokhtar, U., Alit, M. A. S., Hassenian, A. E., and Hefny, H., "Tomato leaves diseases detection approach based on support vector machines", ***11th international computer engineering conference (ICENCO)*** ,246–250 (2015).

21. Zaw, K. K., "Support Vector Machine Based Classification of Leaf Diseases",***Int J. sci. eng. apl.***, 7 (08): 143–147 (2018).

22. Singh, M. K. and Chetia, S., "Detection and Classification of Plant Leaf Diseases in Image Processing using MATLAB", ***International of life science search***, (January): (2018).

23. Madiwalar, S. C., "Plant Disease Identification : A Comparative Study", ***Interrnational conference on data Management , Analytic and Innovation (ICDMAT)***, 13–18 (2017).

24. Dandawate, Y. and Kokare, R., "An automated approach for classification of plant diseases towards development of futuristic Decision Support System in Indian perspective", ***2015 International Conference On Advances In Computing, Communications And Informatics, ICACCI 2015***, 794–799 (2015).

25. Pydipati, R., Burks, T. F., and Lee, W. S., "Identification of citrus disease using color texture features and discriminant analysis", ***Computers And Electronics In Agriculture***, 52 (1–2): 49–59 (2006).

26. Hlaing, C. S. and Maung Zaw, S. M., "Tomato Plant Diseases Classification Using Statistical Texture Feature and Color Feature", ***Proceedings - 17th IEEE/ACIS International Conference On Computer And Information Science, ICIS 2018***, (1): 439–444 (2018).

27. Agrawal, N. and Singhai, J., "Grape leaf disease detection and classification using multi-class support vector machine", ***2017 International Conference On Recent Innovations In Signal Processing And Embedded Systems (RISE)***, 238–244 (2017).

28. Eaganathan, U., Sophia, J., Luckose, V., and Benjamin, F. J., "Identification of Sugarcane Leaf Scorch Diseases using K-means Clustering Segmentation and K-NN based Classification", ***International Jornal of Advances in computer scince and technology (IJACST)***, 3 (12): 11–16 (2014).

29. Kaur, S., Pandey, S., and Goel, S., "Semi-automatic leaf disease detection and classification system for soybean culture", (2018).

30. Parikh, Aditya, et al. "Disease detection and severity estimation in cotton plant from unconstrained images." ***2016 IEEE international conference on data science and advanced analytics (DSAA)*. IEEE**, (2016).

31. Suresha, M., K. N. Shreekanth, and B. V. Thirumalesh. "Recognition of diseases in paddy leaves using knn classifier." ***2017 2nd International Conference for Convergence in Technology (I2CT)*. IEEE**,(2017).

32. Kurniawati, Nunik Noviana, et al. "Investigation on image processing techniques for diagnosing paddy diseases." ***2009 international conference of soft computing and pattern recognition*. IEEE**, (2009).

33. Pandey, Rashmi, Sapan Naik, and Roma Marfatia. "Image processing and machine learning for automated fruit grading system: A technical review." ***International Journal of Computer Applications*** 81.16 (2013).

34. Tsaftaris, S. A., Minervini, M., and Scharr, H., "Machine Learning for Plant Phenotyping Needs Image Processing", ***Trends In Plant Science***, 21 (12): 989–991 (2016).

35. Maniyath, S. R., Vinod, P. V., Niveditha, M., Pooja, R., Prasad Bhat, N., Shashank, N., and Hebbar, R., "Plant disease detection using machine learning", ***Proceedings - 2018 International Conference On Design Innovations For 3Cs Compute Communicate Control, ICDI3C 2018***, (July 2019): 41–45 (2018).

36. Vaidya, M. V, "Classification of Leaf disease using Texture Feature and Neural Network", ***2018 International Conference On Inventive Research In Computing Applications (ICIRCA)***, 1–6 .(2018).

37. Moghadam, Peyman, et al. "Plant disease detection using hyperspectral imaging." ***2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*. IEEE**,.(2017).

38. Sharath, D. M., Kumar, S. A., Rohan, M. G., and Prathap, C., "Image based Plant Disease Detection in Pomegranate Plant for Bacterial Blight", ***2019 International Conference On Communication And Signal Processing (ICCSP)***, 645–649 (2019).

39. Shrestha, Garima, Majolica Das, and Naiwrita Dey. "**Plant disease detection using CNN.**" ***2020 IEEE Applied Signal Processing Conference (ASPCON)*. IEEE**., 6–10 (2020).

40. Al Bashish, Dheeb, Malik Braik, and Sulieman Bani-Ahmad. "A framework for detection and classification of plant leaf and stem diseases." ***2010 international conference on signal and image processing*. IEEE**, 113–118 (2010).

41. Arivazhagan, Sai, et al. "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features." ***Agricultural Engineering International: CIGR Journal*** 15 (1): 211–217 (2013).

42. Anjomshoae, Sule T., and Mohd Shafry Mohd Rahim. "Enhancement of template-based method for overlapping rubber tree leaf identification." ***Computers and Electronics in Agriculture*** 122: 176–184 (2016).

43. Kumari, C. U., Prasad, S. J., and Mounika, G., "Leaf Disease Detection : Feature Extraction with K-means clustering and Classification with ANN", ***2019 3rd International Conference On Computing Methodologies And Communication (ICCMC)***, (March): 1095–1098 (2019).

44. Thushara, Battala Lakshmi, and T. Mahaboob Rasool. "Analysis of plant diseases using expectation maximization detection with BP-ANN classification." ***XIII (VIII)*** (2020).

45. Li, Yang, and Xuewei Chao. "ANN-based continual classification in agriculture." ***Agriculture*** 10.5 (2020).

46. Rastogi, Aakanksha, Ritika Arora, and Shanu Sharma. "Leaf disease detection and grading using computer vision technology & fuzzy logic." ***2015 2nd international conference on signal processing and integrated networks (SPIN)*. IEEE,** (2015).

47. Husin, Nur A., et al. "Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning." ***Agronomy*** 10.11 (2020).

48. Sethy, Prabira Kumar, et al. "Image processing techniques for diagnosing rice plant disease: a survey." ***Procedia Computer Science*** 167 (2020): 516-530. 167 (2019): 516–530 (2020).

49. Nababan, Marlince, et al. "The diagnose of oil palm disease using Naive Bayes Method based on Expert System Technology." ***Journal of Physics: Conference Series***. Vol. 1007. No. 1. IOP Publishing,(2018).

50. Mondal, Dhiman, et al. "Detection and classification technique of Yellow Vein Mosaic Virus disease in okra leaf images using leaf vein extraction and Naive Bayesian classifier." ***2015 international conference on soft computing techniques and implementations (ICSCTI)*. IEEE**,166–171 (2015).

51. Sari, Wahyuni Eka, Yulia Ery Kurniawati, and Paulus Insap Santosa. "Papaya Disease Detection Using Fuzzy Naïve Bayes Classifier." ***2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE**, (2020).

52. Yue, Ding Min, and Feng Qin. "Plant Leaf Recognition Based on Naive Bayesian Classification and Linear Discriminant Analysis Model." ***2019 4th International Conference on Communication and Information Systems (ICCIS)*. IEEE,** 2019. 191–196 (2019).

53. Zhang, Y. C., Mao, H. P., Hu, B., and Li, M. X., "Features selection of cotton disease leaves image based on fuzzy feature selection techniques", ***Proceedings Of The 2007 International Conference On Wavelet Analysis And Pattern Recognition, ICWAPR ’07,*** 1: 124–129 (2007).

54. Gite, S., and H. Agrawal. "Proceedings of the Second International Conference on Computer and Communication Technologies." , ***Springer***, (2015).

55. Sabrol, Hiteshwari, and Satish Kumar. "Fuzzy and neural network based tomato plant disease classification using natural outdoor images." ***Indian J. Sci. Technol*** 9 (November): 1–8 (2016).

56. Majid, Kholis, Yeni Herdiyeni, and Aunu Rauf. "I-PEDIA: Mobile application for paddy disease identification using fuzzy entropy and probabilistic neural network." ***2013 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE**, (2013).

57. Srinidhi, V. V., Apoorva Sahay, and K. Deeba. "Plant pathology disease detection in apple leaves using deep convolutional neural networks: Apple leaves disease detection using efficientnet and densenet." ***2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE**,1119–1127 (2021).

58. Swaminathan, Akshay, C. Varun, and S. Kalaivani. "Multiple Plant Leaf Disease Classification using Densenet-121 Architecture." ***International Journal of Electrical Engineering and Technology (IJEET)*** 12.5 (2021): 38-57.

59. Hassan, Sk Mahmudul, et al. "Plant disease identification using shallow convolutional neural network." ***Agronomy*** 1–20 (2021).

60. Rezende, V. and Santos, A., "Image Processing with Convolutional Neural Networks for Classification of Plant Diseases", ***2019 8th Brazilian Conference On Intelligent Systems (BRACIS)***, (October): 705–710 (2019).

61. Nagasubramanian, Koushik, et al. "Explaining hyperspectral imaging based plant disease identification: 3D CNN and saliency maps." ***arXiv preprint arXiv****:1804.08831* (Nips 2017).

62. Shrestha, Garima, Majolica Das, and Naiwrita Dey. "Plant disease detection using CNN." ***2020 IEEE Applied Signal Processing Conference (ASPCON). IEEE***,(12): 2106–2112 (2021).

63. Islam, Md Tariqul. "Plant Disease Detection using CNN Model and Image Processing." ***INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*** *Volume* 9 (2022).

64. Wallelign, Serawork, Mihai Polceanu, and Cédric Buche. "Soybean plant disease identification using convolutional neural network." ***FLAIRS conference.*** (2018).

65. Nasteski, V., "An overview of the supervised machine learning methods", ***Horizons.B***, 4 (December 2017): 51–62 (2017).

66. Jiang, Tammy, Jaimie L. Gradus, and Anthony J. Rosellini. "Supervised machine learning: a brief primer." ***Behavior Therapy*** 51 (5): 675–687 (2021).

67. Cunningham, Pádraig, Matthieu Cord, and Sarah Jane Delany. "Supervised learning." ***Machine learning techniques for multimedia*. Springer, Berlin, Heidelberg**, 21-49.(2008).

68. Yau, K. A., Elkhatib, Y., Hussain, A., and Al-fuqaha, A. L. A., "Unsupervised Machine Learning for Networking : Techniques , Applications and Research Challenges", ***IEEE Access***, 7: 65579–65615 (2019).

69. Sinaga, Kristina P., and Miin-Shen Yang. "Unsupervised K-means clustering algorithm." ***IEEE access*** 8: (2020).

70. Reddy, Y. C. A. P., P. Viswanath, and B. Eswara Reddy. "Semi-supervised learning: A brief review." ***Int. J. Eng. Technol*** (2018).

71. Alharin, Alnour, Thanh-Nam Doan, and Mina Sartipi. "Reinforcement learning interpretation methods: A survey." ***IEEE Access*** 8: (2020).

72. Raschka, Sebastian. "Model evaluation, model selection, and algorithm selection in machine learning." ***arXiv preprint arXiv****:1811.12808* (2018).

73. Ma, Yunqian, and Guodong Guo, eds. *Support vector machines applications*. Vol. 649. New York: ***Springer,*** (2014).

74. Dang, Yijie, et al. "Image classification based on quantum K-Nearest-Neighbor algorithm." ***Quantum Information Processing*** 17.9 (2018).

75. Anchalia, Prajesh P., and Kaushik Roy. "The k-nearest neighbor algorithm using MapReduce paradigm." ***2014 5th International Conference on Intelligent Systems, Modelling and Simulation. IEEE,***513–518 (2014).

76. Gan, G., "Insurance : Mathematics and Economics Application of data clustering and machine learning in variable annuity valuation", ***Insurance: Mathematics***

And Economics, 53 (3): 795–801 (2013).

77. Raj, Jennifer S. "Machine learning based resourceful clustering with load optimization for wireless sensor networks." ***Journal of Ubiquitous Computing and Communication Technologies (UCCT)*** 02 (01): 29–38 (2020).

78. Chen, Xingyu, et al. "Proportionally fair clustering." ***International Conference on Machine Learning***. PMLR, (2019).

79. Chaudhary, Piyush, et al. "Color transform based approach for disease spot detection on plant leaf." ***International journal of computer science and telecommunications*** 3 (6): 4–9 (2012).

80. Arnal Barbedo, Jayme Garcia. "Digital image processing techniques for detecting, quantifying and classifying plant diseases." ***SpringerPlus*** 2.1 (2013).

81. Bansal, Arpit, Mayur Sharma, and Shalini Goel. "Improved k-mean clustering algorithm for prediction analysis using classification technique in data mining." ***International Journal of Computer Applications*** 157.6 (2017).

82. Saini, Gursharan, and Harpreet Kaur. "A novel approach towards K-mean clustering algorithm with PSO." ***Int. J. Comput. Sci. Technol,*** 5 (4): 5978–5986 (2014).

83. Bustamam, A., et al. "Application of K-means clustering algorithm in grouping the DNA sequences of hepatitis B virus (HBV)." ***AIP Conference Proceedings.*** Vol. 1862. No. 1. AIP Publishing LLC, 030134:(2017).

84. Warne, Pawan P., and S. R. Ganorkar. "Detection of diseases on cotton leaves using K-mean clustering method." ***International Research Journal of Engineering and Technology (IRJET)*** 2.4: 425-431, (2015):

85. Sandhu, Gurleen Kaur, and Rajbir Kaur. "Plant disease detection techniques: a review." ***2019 international conference on automation, computational and technology management (ICACTM). IEEE,*** (2019).

86. Tete, Trimi Neha, and Sushma Kamlu. "Detection of plant disease using threshold, k-mean cluster and ann algorithm." ***2017 2nd International Conference for Convergence in Technology (I2CT).*** **IEEE**, 523–526 (2017).

87. Osman, M. K., et al. "Color image segmentation of tuberculosis bacilli in Ziehl-Neelsen-stained tissue images using moving k-mean clustering procedure." ***2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation.*** IEEE, 0–5 (2010).

88. Patil, Rupali, et al. "Grape leaf disease detection using k-means clustering

algorithm." ***International Research Journal of Engineering and Technology (IRJET)*** 3.4 (2016).

89. Bashir, Sabah, and Navdeep Sharma. "Remote area plant disease detection using image processing." ***IOSR Journal of Electronics and Communication Engineering*** 2.6: 31-34. (2012).

90. Sridhathan, Senthilkumar, and M. Senthil Kumar. "Plant infection detection using image processing." ***International Journal of Modern Engineering Research (IJMER)*** 8.7 (2018).

91. Verma, Shradha, Anuradha Chug, and Amit Prakash Singh. "Application of convolutional neural networks for evaluation of disease severity in tomato plant." ***Journal of Discrete Mathematical Sciences and Cryptography*** 0529: (2020).

92. Singh, V. and Misra, A. K., "Detection of Plant Leaf Diseases Using Image Segmentation and Soft Computing Techniques", ***Information Processing In Agriculture***, (2016).

93. Hungilo, Gilbert Gutabaga, Gahizi Emmanuel, and Andi WR Emanuel. "Image processing techniques for detecting and classification of plant disease: a review." ***Proceedings of the 2019 international conference on intelligent medicine and image processing*****.** (2019).

94. Zhang, S. W., Y. J. Shang, and L. Wang. "Plant disease recognition based on plant leaf image." ***J. Anim. Plant Sci*** 25 (42): 25–28 (2015).

95. Devaraj, A., Rathan, K., Jaahnavi, S., and Indira, K., "Identification of Plant Disease using Image Processing Technique", ***2019 International Conference On Communication And Signal Processing (ICCSP)***, 749–753 (2019).
96. Kshirsagar, Gaurav, and A. N. Thakre. "Plant disease detection in image processing using MATLAB." ***International Journal on Recent and Innovation Trends in Computing and Communication*** 6.4: 113-116. (2018)

97. Sabrol, H., and S. Kumar. "Intensity based feature extraction for tomato plant disease recognition by classification using decision tree." ***International Journal of Computer Science and Information Security*** 14 (9): 622–626 (2016).

98. Azim, Muhammad Anwarul, et al. "An effective feature extraction method for rice leaf disease classification." ***TELKOMNIKA (Telecommunication Computing Electronics and Control)*** 19 (2): 463–470 (2021).

99. Mokhtar, Usama, et al. "SVM-based detection of tomato leaves diseases." ***Intelligent Systems'*** Springer, Cham, 641–652 (2014).

100. Deepa, R. N., and C. Shetty. "A machine learning technique for identification of plant diseases in leaves." ***2021 6th international conference on inventive computation technologies (ICICT)*.**(2021).

101. Padmanabhuni, ms sri silpa, and pradeepini gera. "an extensive study on classification based plant disease detection system." ***Journal of mechanics of continua and mathematical sciences*** 15.5 (2021).

102. Dharani, M. K., et al. "Review on crop prediction using deep learning techniques." ***Journal of Physics: Conference Series*.** Vol. 1767. No. 1. IOP Publishing, (2021).

RESUME

Raghad Mula ALYAS graduated first and elementary education in iraq . she completed high school education at (touz high school for girls) in tikrit Governorate , then, she obtained bachelor's degree from University of IMAM JAAFER AL-SADIQ of Software Engineering in 2015 . After graduation she worked with Tikrit University as teaching assistant in the computer lab . To complete M.Sc education, she moved to Karabuk/Turkey in 2019. She started her master education at the department of computer engineering in Karabuk University.