Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach

Abebaytu Wodaje Mamie

A Thesis Submitted to

The Department of Electronics and Communication Engineering

School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Master of Science Degree in Electronics and Communication Engineering

Office of Graduate Studies

Adama Science and Technology University

June, 2023

Adama, Ethiopia

Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach

Abebaytu Wodaje Mamie

Advisor: Dr.Selewondim Eshetu

Co-advisor: Dr.Satyasis Mishra

A Thesis Submitted to
The Department of Electronics and Communication Engineering
School of Electrical Engineering and Computing

Presented in Partial Fulfillment of the Requirement for the Master of Science Degree in Electronics and Communication Engineering

Office of Graduate Studies
Adama Science and Technology University

June, 2023
Adama, Ethiopia

APPROVAL SHEET OF M.SC THESIS

We, the advisors of the thesis entitled "**Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach**" and developed by **Abebaytu Wodaje Mamie**, hereby certify that the recommendation and suggestions made by the board of examiners are appropriately incorporated into the final version of the thesis.

Major advisor/supervisor	Signature	Date
Co- advisor/co-supervisor	Signature	Date

We, the undersigned, Members of the Board of Examiners of the thesis by **Abebaytu Wodaje Mamie** have read and evaluated the thesis entitled "**Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach**" and examined the candidate during opened defense. This is, therefore, to certify that the thesis is accepted for the partial fulfillment of the requirement of the degree of Master of Science in Electronics and Communication Engineering.

Chair person	Signature	Date
Internal Examiner	Signature	Date
External Examiner	Signature	Date

Finally, approval and acceptance of the thesis proposal is contingent upon submission of its final copy to the Office of the Postgraduate Studies (OPGS) through the Department Graduate Council (DGC) and School Graduate Committee (SGC).

Department Dead	Signature	Date
School Dean	Signature	Date
Office of Postgraduate Studies Dean	Signature	Date

DECLARATION

I hereby declare that this Master Thesis entitled "**Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach**" is my own work and has not been submitted for the award of any academic degree, diploma, or certificate in any other university. All sources of materials that are used for this project have been duly acknowledged through citation.

Name of Student	Signature	Date

RECOMMENDATION

We, the advisor(s) of this thesis, hereby certify that we have read the revised version of the thesis entitled "**Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach**" prepared under my guidance by **Abebaytu Wodaje Mamie** submitted in partial fulfillment of the requirements for the degree of Master of Science in Electronics and Communication Engineering. Therefore, we recommend the submission of revised version of the thesis to the department following the applicable procedures.

Major advisor/supervisor	Signature	Date
Co- advisor/co-supervisor	Signature	Date

ACKNOWLEDGEMENT

First and foremost, I would like to express my profound gratitude to Almighty GOD and HOLY VIRGIN MARY for giving me the moral, psychological, and spiritual strength to accomplish this research work. And I would like to thank my advisor, Dr.Selewondim Eshetu, and Co-advisor, Dr.Satyasis Mishra, for their guidance, and full support starting from the beginning to the completion of this thesis work. Appreciate my advisor, Dr.Selewondim Eshetu his patience, suggestions, ideas and comments, especially during a time when I needed it most.

Finally, I would like to express my sincere gratitude to my mom for all the indescribable sacrifices she made so that I have become who I am today. I want to give great gratitude to, my friends, families, classmates, staff members, Department Head Mr.Eshetu Tessema and post graduate coordinator for their initiation for accomplishing the thesis in proper time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SVM	Support Vector Machine
RGB	Red, Green, and Blue Color
HIS	Hue, Saturation, and Intensity
GLCM	Gray Co-Occurrence Matrix
NN	Neural Network
AI	Artificial Intelligence
ANN	Artificial Neural Network
ML	Machine Learning
HOG	Histogram Oriented Gradient
SIFT	Scale Invariant Feature Transform
SURF	Speed Up Robust Feature
PCA	Principle Component Analysis
GUI	Graphical User Interface
IDM	Inverse Difference Moment
S.D	Standard Deviation
K-NN	K-nearest Neighbor

ABSTRACT

The pea is a grain crop that is consumed by both people and animals. The main protein source for under-privileged farmers. High altitude regions are where pea is produced, and for maximum grain production, a minimum average temperature of 30° C is needed. Crop diseases are one of the main variables affecting pea output and productivity enhancement. One of the key methods for reducing yield production loss is the early detection of pea illnesses, although this takes a significant amount of time, money, and effort. The study suggested a machine learning method for categorizing pea illnesses based on their leaves in order to solve these issues. The design science research technique was used to achieve this. To conduct this study, a total of 2000 images were collected from Ziquala Woreda Addislem villages, Waghimra Zone, and prepared. The researcher uses picture preparation techniques after gathering the required images, such as image scaling, normalising images, and noise removal. Techniques for data augmentation were also used. It is used to identify and pick crucial characteristics that contribute to a disease's symptom. This study focuses on Ascochyta blight, downy mildew, rust, and powdery mildew as the main diseases that reduce field pea productivity. However, in Ethiopia, weed, aphids, and storage pests are also a serious productivity barrier. The image was analysed in MATLAB, and support vector machine classification was used to determine the leaf's condition. With the help of machine learning, the k-Medoid clustering technique, and the multi-SVM algorithm, this suggested system provides an overview of the classification and detection of peat leaf illnesses using MATHLAB software and finally by inserting the pea leaf image on GUI detect, classify disease and get better accuracy result 98.3871% because instead of k-means in this work k-medoid clustering and other statistical feature performance.

***Keywords**:* pea crop, Image processing, Future extraction, Classification Machine learning, neural networks, K-means clustering, Multi SVM algorithm and MATHLAB.

CHAPTER ONE

1. INTRODUCTION

1.1 Background of the Study

There is a definite need to improve agriculture production in order to satisfy the needs of the growing population because the world's population is expected to reach 8.5 billion people by 2030, Weather factors like temperature, humidity, and precipitation are contributing to an increase in the proliferation of bacteria, viruses, fungus, nematodes, and other microorganisms. Plants are more prone to illness since there are so many germs nearby. Attacks by pests and diseases play a significant role in the fall in agricultural output. Application of appropriate prevention and protection measures is aided by accurate and timely prediction of plant diseases. Consequently, it aids in raising yield quality.

Various symptoms, such as lesions, colour changes, damaged leaves, damaged stems, aberrant growth of the stem, leaf, bud, flower, and/or root, etc., can be used to identify plant diseases. Additionally, as a disease indicator, leaves exhibit symptoms like spots, dryness, premature falling, etc. (Yimam, 2020). Plant diseases can be found using an analysis of these observable symptoms. A skilled or experienced person or people visually inspect a plant as part of the traditional disease detection process. However, the strategy calls for thorough understanding and proficiency in the field of disease detection. Additionally, it may lead to inaccurate predictions as a result of cognitive biases and visual illusions. Consequently, the method is not a workable solution for large agricultural.

Agriculture is a major source of income in Ethiopia. The majority of the people of the nation are involved in agriculture, either directly or indirectly. In order to maintain the nation's economic development, high-quality agricultural produce is required. Farmers make the appropriate decisions regarding the crops by observing and managing the necessary temperature, light, humidity and disease needs in order to produce crops of higher quality and productivity (A. Verma, 2019). Additionally, because of population expansion, climate change, and political unpredictability, the agriculture sector has begun to explore for novel ways to boost food production. This has drawn researchers to search out distinctive, inventive, and dependable technologies that will aid in improving agricultural productivity. There are still issues that farmers are having, such the need for

early detection of plant diseases. An automated expert system that can quickly identify the disease would be very helpful because it is not always feasible to see the type of disease on a plant's leaf with the human eye.

Ethiopia is a heavily dependent nation on agriculture, with agriculture providing the majority of the nation's GDP. In Ethiopia, agriculture is the main industry. The livelihood of the populace depends on agricultural productivity. Farmers grow a wide variety of crops here in Ethiopia. The production of the crops is influenced by a number of variables, including the climate, the soil, various diseases, etc. The standard method for identifying plant leaf diseases is simple observation. This is ineffective.

Crops that are not properly maintained and protected become more susceptible to illnesses, which, lowers their ability to produce. Farmers need automatic monitoring of plant disease in place of manual monitoring in order to increase the growth and productivity of crop fields. Manual disease monitoring cannot produce satisfactory results since it relies on the antiquated technique of naked eye inspection, which takes more time and requires a specialist (A. Cruz, 2019). For quick and precise plant disease identification, we used a digital image processing methodology to get over the limitations of the traditional eye watching method. The farmer can use this technology to determine what kind of diseases are affecting the plant. The signs of plant illnesses are visible on several plant components, although leaves are the most frequently seen area for spotting an infection. Therefore, using leaf photos as a starting point, researchers have attempted to automate the identification and classification of plant diseases (Y. Q. Xia, 2015).

Modern inventions have made it possible for human culture to provide sufficient nutrient to meet the need. However, the security of food supply continues to be jeopardized by a number of causes, such as environmental change, plant diseases, and other issues. Any nation's ability to sustain its agriculture depends on the quality and quantity of agricultural output, particularly plants. Ethiopia's economy is based on agriculture, and it was the initial human action to initiate the process of human development, evolution, and advancement. This was done in response to the country's growing population and rising food demand, which also served to create jobs in a variety of agricultural fields. Today, agriculture is a significant activity. Crop diseases continue to pose a serious threat to the global food supply. An increasingly verified dataset of photos of sick and healthy plants was needed in order to build perfect image classifiers for the reasons underlying

plant disease recognizable proof. The detection of infections is now done using a server- and portable-based method for disease differentiating evidence (M. K. Gayatri, 2015)

Ethiopian economy relies heavily on agricultural productivity. As a result, both the environment and humans benefit greatly from the contribution of food and cash crops. Several diseases strike crops every year. Many plants die because of poor disease diagnosis and a lack of knowledge about the symptoms of the disease and how to treat it. This production is drastically declining due to plant diseases, which occurs in various ways, which can be either natural or artificial. Plant diseases are one of the major reasons for crop failure this not only causes famine, but also results in economic crisis, which can lead to another great depression & starvation. The traditional and antiquated technique for identifying and diagnosing plant diseases relies solely on visual inspection, which is exceedingly slow and inaccurate. Due to the lack of knowledge in this nation, contacting experts to identify plant diseases is costly and time-consuming. Regular plant inspections prevent the development of numerous diseases that require more chemicals to treat and are dangerous to other animals, insects, and birds that are essential to agriculture.

Digital high quality photos are first captured before being processed using MATLAB. Images of the good and bad are taken and saved for testing. Images are then subjected to pre-processing for image improvement. K-means clustering is used to segment captured leaf pictures and create clusters.

1.2 Motivation

To conduct this study because of the agriculture sector is crucial for the daily needs of the Ethiopian population and for the country's economy, and pea is one of the most important cereal crops in this sector. However, the sector is facing challenges such as poor cultivation practices and frequent drought, which can lead to lower yields and economic losses. Additionally, farmers are facing difficulties due to plant pests and diseases, which can further reduce their yields and income. Therefore, the researcher likely wanted to investigate ways to improve pea production and reduce the impact of plant pests and diseases on farmers in order to support the agriculture sector and the livelihoods of farmers in Ethiopia. Many farmers in Ethiopia do not have access to information about pests and diseases and may not have the knowledge or resources to identify them. This can be due to several factors, including limited access to agricultural extension services, lack of education and training, and limited access to technology and information resources. As a result, farmers may not be able to effectively manage pests

and diseases, which can lead to reduced yields and economic losses. Improving access to information and resources on pest and disease identification and management is crucial for improving agricultural productivity and supporting the livelihoods of farmers in Ethiopia.

Machine learning-based plant disease identification and categorization is the finest method for raising crop quality and lowering a nation's economic crises. I have been observed the problem of pea leaf continuously for three years affected by disease due to that reason the production of pea decreased by decreasing rate totally kebele, woreda and zone. This leads to the problem of shortage of shiro and other by products and farmers are anger by this cause. They ask pesticide from zone and woredas of the society, but not get any solution for that problem. To solve this problem I am interested to do this area of study.

1.3 Statement of Problems

Agriculture is essential for a nation's development since it fosters growth, helps countries fight poverty, and transforms their economies. Because of this, Ethiopia has seen a remarkably rapid rise in this industry over the past ten years. Cereals of various varieties, including Teff, Wheat, Maize, Sorghum, Bean, Pea, and Barley, make up the bulk of Ethiopia's agriculture and food production. These pulse-covered areas have been covered by around 216,786.33 hectares of field pea, which represents for 12.73% of the 8.8 million hectares (83.23%) of field crops that are made up of cereals. This crop is planted in primarily high-altitude locations with a little amount of low-land area as well (Arnoldi A., 2015). When combined with cereal, peas are valuable and affordable sources of protein that have a key role in restoring soil fertility and in market export (Brummer Y., 2015). Improved crop production techniques in general and improved seeds in particular are not widely used in Ethiopia. One of the most significant challenges restricting seed output is a lack of technical expertise. Peas provide a remarkable variety of nutrients and are a significant source of proteins. But there are insects they feed pea plant leaf and as am observed there is change of color and no seeds are produced this might affect agricultural crop production of the pea. This leads to affect countries crop production of the pea and the investment. Researchers stated that these diseases are caused by pests, insects, and pathogens, there are several pea leaf diseases, among them, Ascochyta, powdery mildew, Leaf blight, and rust are the major ones and mostly occurred in a too hot area which is pea grown. To solve this problem image

processing is advantageous especially Image Based Pea Leaf Disease Detection with Classification Using Machine Learning Approach is very important.

1.4 Objectives

1.4.1 General Objective

The major goal of this research is to create and develop a machine learning model that can identify and categorize pea leaf diseases.

1.4.2 Specific Objectives

The ensuing particular objectives are established in order to accomplish the general ones.

- ✓ To review the literature of earlier investigations in order to comprehend the field.
- ✓ To identify and collect the required datasets for the pea disease.
- ✓ To develop the classifier model.
- ✓ To assess the model's effectiveness after it has been developed.
- ✓ To detect and classify healthy and affected part of the pea leaf for improving quality.

1.5 Significance of the Study

Agriculture is declining daily. There are several factors that contribute to decline, including the weather, a lack of ground water, pollution, and infections. Farmers should have a lot of money to deal with these issues. Costly pesticides and chemicals are applied to crops to treat illnesses. Since the sickness is widespread, it will be of greater use. However, if the illness can be identified, then the cost of production can go down. So, in this case, we are employing a soft computing strategy to detect disorders. The farmers will benefit from the lower-cost identification of the illness proportion. The principal pathogens that affect leaves are bacterial and fungal. The leaf is severely affected by this disease. It modifies colour, which is the sign of disease. Therefore, by analyzing an image of such kinds of leaves, we may calculate the disease's prevalence and treat it in its early stages. The production and maintenance of plants will improve with the use of this technology, which makes it simple to identify injured leaves. The Minimum Distance Criterion together with K-Mediod instead of K-Mean Clustering is first used to classify the data. There is no way to completely stop using artificial intelligence, yet it is a real fact that it has improved human lives. Therefore, it is up to the consumers to make wise use of it to raise their

standard of living. After the study is complete, pathologists, agricultural extension specialists, farmers, and the government will comprehend the relevance of the findings.

1.6 Main Contributions

Identification, detection, and classification of the various pea plant leaf diseases are the purpose of this thesis. It is necessary to increase the pea production and resists economic reduction by detecting the pea leaf diseases. I have observed the problem starting from the last three years up-to now the leaf of pea affected by different types of disease due to that reason the production of pea is decreasing year to year. Using the SVM classifier and K-Medoid clustering for image segmentation, this thesis solves the issues by raising the performance parameter values.

1.7 Scope of the Study

Making a boundary of task coverage is crucial for improved results because the detection and classification of leaf disease is a very important field of research to address. The study is conducted based on the data obtained from Zquala Woreda Addisalem village, Waghimira Zone. Therefore, pre-processing techniques were implemented to make the data smooth and cleaned. The development of a model for classifying pea leaves diseases is the major goal of this work.

To do so, a deep learning-based approach with training from scratch (i.e., setting different parameters and hyper-parameter from the beginning of building the model), and transfer learning (i.e., reusing previously trained model). Hence, These methods automatically extracting the representative features from the images and identifying the features of image elements that are given to a convolutional neural network classifier.

1.8 Organization of the Thesis

This thesis is divided into five chapters overall, without including the references and appendix. The introductory portion of the study, which was covered in Chapter One, included the study's history, problem statement, aims, importance, scope, and thesis organization. The second chapter focused on a review of the relevant literature, which is more significant than the first. It refers to past study that is still relevant today and has been previously investigated (researched). The materials and procedures used to conduct this research are discussed in Chapter 3. Chapter four deals the results and discussions. Chapter five presents' conclusions and (recommendation) future research work directions are offered at the end.

CHAPTER TWO

2. LITERATURE REVIEW

2.1 Related Works

Several research efforts have been conducted and proposed for plant disease identification and detection. However, each works is designed to address different plants or parts of a plant. Therefore, in this section, related research on employing machine vision systems to identify plant diseases by various researchers is examined.

Brown Spot, Downy Mildew, Sugarcane Mosaic, Downy Fungal, Red stripe, and Red Rot are the illnesses that the authors mentioned in their technique for sugarcane leaf disease identification. The pre-processing step involved converting the RGB image to grey scale and removing any unnecessary elements. Segmentation is used to identify areas that are healthy and those that may be sick. Disease detection uses linear, nonlinear, and multiclass SVM (Prajakta Mitkal, 2016).

The system described is designed to automatically detect diseases in corn plants using image processing, with a Raspberry Pi serving as the main interface. The user selects an input image, and the system identifies the disease and suggests remedies using Python or Java. The app displays this information, and the user can use it to turn on or off a sprinkler assembly that sprays pesticides or fertilizer mixed with water. The system includes various sensors, such as temperature, humidity, water, and moisture sensors, which are interfaced with the Raspberry Pi to monitor soil conditions and water levels. The motor driver and DC motor are used for system movement, which can help to monitor soil conditions at different locations. Finally, a relay driver and single pole double throw relay are used to control external devices like the sprinkler assembly (Budiarianto Suryo Kusumo, 2018).

Introduced a method for detecting mango fruit illness. In order to transform the original image into binary and compute the histogram, the authors created a movie of mango fruit disease throughout the image acquisition process. To find defective parts in a picture, the Watershed technique is used for segmentation, and features are retrieved using the Blob extraction with Template-Matching algorithm. The normalised correlation method is used to classify diseases and displays the affected area in fruit images (Pujitha N, 2016).

Apple rot and apple blotch are the diseases that are included in the detection method for apple fruit disorders. In the K-Means method, the fruit image is transformed from RGB to L*a*b* colour space and Euclidean distance is employed to determine the contaminated zone. For fusing more than two features, feature level fusion is used to extract colour, shape, and texture features. Global colour histogram and Colour coherence vector are examples of colour features. Gabor features, a complete local binary pattern, a local ternary pattern, and a local binary pattern are examples of texture features. The classification outcome is then applied using the Random Forest Classifier. The diseases mentioned by the authors in their technique for apple fruit disease detection are apple scab, rot infections, and blotch fungal disease. Following image collecting techniques, a technique is employed to identify the region of interest and choose only the infected portion. Following feature extraction and database storage, a support vector machine is set up to classify and identify diseases (BhaviniJ.Samajpati, 2016).

To find plant diseases, image processing methods were employed. Development of a software system for the automatic identification and classification of plant illnesses was the primary goal of this work (Mottaleb, 2021).The image processing approach included a number of processes, including image acquisition, image preprocessing, image segmentation, feature extraction, and image classification, to identify plant illnesses. The pictures of the leaves were used to find plant diseases. Thus, it was determined that image processing methods were extremely helpful in the identification and classification of infections in agricultural applications. In this work, advanced technology, such as image processing, was used to create an automatic detecting system. This approach was developed to provide aid to the farmers for the early detection of plant infections. This approach also supplied useful data for its handling. In future, this study would be extended further for detecting large number of infections.

The author reviewed various image processing methods for spotting plant diseases. This study's main goal was to detect the infections. Accuracy and speed were shown to be two key requirements for infection detection. Therefore, the development of modified algorithms for quick and accurate identification of infectious leaves would be the main focus of this study's expansion. This study involved a survey on several infection classification techniques. To find plant leaf infection, one could use these categorization strategies. An algorithm for picture segmentation was also presented in this study's mechanical discovery and classification of plant leaf infection of more advanced technology, like image processing. So, for the purpose of detection, correlated

infections for these plants were used. The best results were obtained with the least amount of computational work. These findings showed how well the suggested algorithm worked for identifying and categorizing plant diseases (Zare, 2021).

The unique infection recognition and classification method described by this author uses a variety of machine learning algorithms and image processing technologies. The sick component was initially located and captured. After that, the sick part received image preprocessing treatment. Additionally, the segments were completed, and the interest region was located. The segmented image was then used to extract necessary features, such as colour, shape, and contrast, in the following stage. Finally, SVM Classification models for obtaining results provided the achieved results. When it came to classifying plant infections, this classification model performed extremely well. The results of the tests demonstrated that the proposed strategy greatly outperformed previously used infection recognition methods (Li, 2022).

Various plant infection detection methods were thoroughly studied (al., 2021).It was decided that an illness must be detectable by both plants and people. Consideration of the right input was essential to have a significant impact on plant infection & systems in the farming region. In order to give a methodical technique for identifying and discriminating plant infections, various research difficulties were taken into account in this work. These methods could one day be very helpful to pathologists and farmers.

2.1.1 Types of Plant Diseases

Biotic plant diseases are those that are brought on by other living organisms. Numerous biotic illnesses are mostly caused by fungi, bacteria, and viruses. Abiotic conditions, on the other hand, are caused by non-living ecological phenomena like weather, chemical burns, spring frosts, etc. The majority of abiotic diseases are preventable, contagious, and non-transmissible. So, the document takes biotic problems into account. For numerous bacterial and fungal disorders, there are numerous works (Pujari JD, 2016). The main three categories used for identification and classification are spots (caused by either fungi or bacteria), mildew, and rust. The possibility of automatic detection of nutritional deficiency is also looked at.

2.1.2 Plant Disease Detection System

The capture, pre-processing, segmentation, feature extraction, and classification (or recognition) modules make up the architecture of a straightforward plant disease detection system. The training and testing are its two phases. Its two stages are testing and training. The first step in the

training phase is to take a picture of a particular component, such as a leaf, a stem, a root, or a branch. The pre-processing of images to improve blur, reduce noise, and correct various geometric misrepresentations is optional. A training leaf image that has been infected is segmented to identify one or more regions of interest and isolate them from background. The collected feature vectors from the regions of interest are then used to train the classifier. During the testing stage, a test leaf picture goes through pre-processing, segmentation, and feature extraction modules. The trained classifier determines whether the test image represents an infected or healthy sample. Later in this part, there is a brief discussion about the necessity of all the modules.

2.2 Overview of Pea

Pisum sativum L., also known as field pea, is a significant cool-season pulse crop and a crucial element of sustainable cropping systems (Dahl W. J., 2012). Of all the pulses grown in Ethiopia, the field pea (Pisium sativum L.) accounts for the biggest national production and area share. When combined with cereals, which are low in key amino acids, they are valuable and affordable sources of protein. Pulses are important for both the export market and restoring soil fertility. Despite their significance, productivity and production fall far short of their potential for a variety of reasons, including a lack of superior variety seeds in the market. Over the past 20 years, a variety of enhanced field pea cultivars have been made available. Farmers have become more aware of the value of enhanced seeds for diverse crops thanks to the current Extension Package Programe at the national level. Despite all of this, farmers in need still do not have access to enough seeds of improved field pea types. Because more farmers will discover the advantages of using quality seeds, it is anticipated that more kinds will be released from research and that the need for high quality seed will rise. One of the main reasons restricting seed production in Ethiopia is a lack of technical skill. This is because improved agricultural production technology in general and improved seeds in particular are not widely used in Ethiopia. Various crops have been described as providing important agro-ecological services related to their capacity to develop symbiotic nitrogen fixation as well as their function as break crops for the decrease of pest and disease pressure (MOA and NR, 2016). Field pea is thought to have originated in the highlands of Ethiopia, the Mediterranean, and central Asia. Field pea has been farmed in Ethiopia since ancient times, and the country's mountains have long been home to the species' wild and primitive variants. Due of this, Ethiopia is regarded as one of the centers of field pea variety. The

field pea is cultivated all throughout the world for its nutritious dried seeds, soft green pods, and fresh green seeds. After soybean, groundnut, and common bean in terms of output volume in 2014, field pea came in fourth with 17.4 and 11.2 million tons of dry and green peas, respectively. Fresh seedlings, immature pods, and seeds make a green vegetable, and whole or powdered dry seeds are prepared in a variety of cuisines. Peas are ingested by humans in a broad variety of ways. Dry pea seeds are being used to extract high-quality starch, protein, or oligoside isolates, and entire seed structural and functional features have been evaluated for food development. (Brummer Y., 2015).Pisum sativum is farmed at high altitudes (1800-3200) in Ethiopia (Tsegay., 2013). Field pea ranks third in importance among Ethiopia's highland pulse crops, after faba bean and common bean, as a major food legume crop. 'Shiro wet' is mostly made in Ethiopia out of field pea. (Yimam, 2020).

Figure 1: Sample image of pea crop (Yimam, 2020)

Table 1: Production of pea in ziquala woreda bureau of agriculture sector.

Years of production	Production of pea in quintal	Production pea in percent
2017	67500	100%
2018	45000	66.67%
2019	32000	47.41%
2020	25600	37.93%

The above tabular information is received from ziquala woreda agricultural bureau report and symptom of disease by interviewing them in addition to I have already seen.

2.2 Pea Diseases

A number of authors (Kemal, 2015) noted that a number of illnesses can affect field peas, some of which can result in major harm or loss. The amount of money lost to sickness each year varies, frequently dependent on the local weather. Field peas are susceptible to Ascochyta blight, downy mildew, rust and powdery mildew, among other serious ailments. Fungi and bacteria-based illnesses can spread via insects, contaminated seed, drainage water, waste and stable manure, farm animals and equipment, and wind. It makes grain yields more unpredictable and makes farmers less confident about raising field peas. It happens in important field pea growing regions. On mature plants, this disease results in stem, leaf, and pod spot; on seedlings, it results in foot rot. According to reports, this disease caused field pea yield losses of 50–75% in the United States, 45% in England, 33% in Canada, and 20–53% in Ethiopia.

2.2.1 Ascochyta Diseases

One of the most serious diseases affecting field peas, ascochyta blight, is also known as "black spot disease" and is present in practically all of the major pea-growing regions. In many areas where field peas (Pisum sativum L.) are grown, Ascochyta blight, an infection brought on by a complex of Ascochyta pinodes, Ascochyta pi-nodella, Ascochyta pisi, and/or Phoma koolunga, is a destructive disease that results in large losses in grain output. The main obstacles are diseases, specifically Ascochyta blight (Ascochytapisi), Powdery and downy mildew (Erysiphepolygoni), which result in significant yield loss and yield instability. According to reports, Ascochyta blight and powdery mildew are the two main field pea diseases in the mid-altitudes, and they can diminish yield by 20–30% when they're moderately severe (O, 2019).

2.2.2 Powdery Mildew

Erysiphe pisi, a fungus parasite that causes the disease, forms a white, powdery, dust-like covering on the surface of the leaves and, less frequently, on the petioles of the leaves, stems, and pods. The leaves are yellowed, small, and occasionally noticeably deformed. The topmost foliage shows the first signs. On the surface of leaves, powdery mildew typically appears as a white or greyish powdery material. The damaged areas may also exhibit browning, yellowing, and leaf distortion or stunting. They germinate and grow on the surface of plant tissues without the need for water. reportedly the main field pea disease in the mid-altitudes, causing yields to drop by

20–30%. (S. Koenig, 2022). Powdery mildew usually forms a white powdery coating on the surface of leaves.

Absence of rain and even alight dew favor the development of sickness. Rain can stop the spread of the illness by removing spores. Field pea germplasm spread across the globe suffers an 86% loss due to the powdery mildew disease, which reduces production potential. In the mid-altitudes with moderate severity, powdery mildew disease has been reported to reduce field pea yield by 20–30%. When the days are warm and dry and the nights are cool enough for dew to accumulate, powdery mildew is a bothersome disease. Due to the severity of powdery mildew on a native field pea cultivar from a plot in Sinana, South Eastern Ethiopia, yield losses of 21.09% have been documented (Teshome E, 2017).

Figure 2: powdery mildew disease

2.2.3 Leaf Blight (Downy Mildew) Pea Leaf Disease

On the upper surface of leaves, downey mildew appears as a yellow or light green patch. On the underside of leaves, it has a grayish-purple or brownish appearance. Additionally, the impacted leaves may twist, curl, or become yellow or brown. Fungi that cause downy mildew develop inside the tissues of plants and need moisture to spawn and develop. On the upper surface of leaves, downy mildew appears as a yellow or light green area, with a grayish-purple or brownish appearance on the underside (Teshome E, 2017).

All of the major pea-growing regions of the world have reported finding or observing the leaf blight disease, which is brought on by the fungus Exserohilum turcicum (Das, 2017). Leaf blight is regarded as a prevalent and serious disease of productive cultivars in our content. This illness

slows down or stops plant growth and development, which lowers the yield of both grain and feed. Before flowering, leaf root causes yield losses of up to 50%. (H. Goitoom, 2019).

Figure 3: leaf blight pea leaf disease

2.2.4 Rust Pea Leaf Disease

The obligate fungal pathogen Puccinia purpurea Cooke is responsible for pea rust. Almost all regions of the world where peas are grown are affected by this disease, especially East Africa and India (R. L. Christopher and P. Ramasamy, 2019).

Figure 4: pea rust leaf disease

The majority of current agricultural technologies are focused on machine learning (ML) algorithms, which give farmers more control over crop selection, crop yield forecasting, crop disease forecasting, weather forecasting, minimum support price, and smart irrigation systems.

2.3 Digital Image Processing

An image can be represented mathematically as a two-dimensional function, f(x, y), where x and y are spatial (plane) coordinates. The intensity or grey level of the image at any given place is determined by the amplitude of f at that location. We refer to the image as a digital image when

x, y, and the amplitude values of f are all finite, discrete quantities. Using a finite number of point cell elements, commonly referred to as pixels (picture elements), a digital image is a matrix representation of a two-dimensional image. Each pixel is represented by numerical values. (Tyagi, 2018).

In imaging science, image processing is any type of signal processing where the input is an image, sequence of pictures, or video frame, such as a photograph or video frame, and the output is the processed image (Abdullah, 2016). As a result, image processing is divided into two categories: analogue and digital image processing. For tangible copies like printouts and photos, analogue image processing can be utilized. When applying these visual tools, image analysts employ a variety of interpretational fundamentals. Processing digital photographs on a digital computer is referred to as "digital image processing." Keep in mind that a digital image is made up of a finite number of elements, each of which has a specific place and value.

In this investigation, several infection types were also taken into consideration for further research .According to the author, quick and accurate identification of plant diseases was essential for the healthy and successful development of crops (Shivani K. Tichkule, 2016). It was found that techniques for image processing were essential for identifying plant diseases. The diseases in various plants might be clearly identified and classified with the aid of image processing technologies. K-means Clustering method was utilized in this study to find sick sections. On the other hand, neural networks were especially employed to achieve precision in the identification and classification of illnesses. Consequently, these methods could be used.

There are different Image processing technologies. These are

2.3.1 Detection using Support Vector Machine

To address classification and regression problems, the Support Vector Machine (SVM) supervised machine learning method can be utilized. We also examine the drawbacks of regression, but classification gains more from its use (Rani, R. U., et al., 2018) presented a technique that can identify pests based on pictures of leaf damage. After contrast boosting, the photos were segmented using the k-means technique. The features of this image are extracted, and identification is then produced using an SVM classifier. It was assessed what percentage of regions were impacted.

There are several methods to ascertain whether plants are impacted, according to (Dhaware, C. G. et al., 2017). Preprocessing entails transforming each image's RGB to HSV. Background subtraction with clusters is used for segmentation. Other notions include energy, homogeneity, and correlation. Images can be categorized by the SVM technique. Despite the small sample size, the results were reliable. There was no classification of the infected leaves by sickness. (Mokhtar, U., Alit, M. A. S., Hassenian, A. E., and Hefny, H., 2015) Identified tomato leaf diseases using SVMs. They applied morphological techniques for image preprocessing, including leaf image separation, image scaling, and background subtraction. Featured are colour and texture. Texture is eliminated using the Gabor filter. Application of SVM to categorization.

By (Zaw, 2018) Using MATLAB R2017a, an SVM classifier system was created to categorise leaf diseases such Alternaria, Cercospora Leaf Spot, and Bacterial Blight. HIS (Hue Saturation Intensity) is used in this study to transform the RGB data. When segmenting, the Grey Level Co-occurrence Matrix (GLCM) uses k-means clustering to extract the features of the damaged areas. Prior to feature extraction, noise is removed by the median filter. Finally, leaf sickness categorization and accuracy are determined by a support vector machine (SVM). The best accuracy of the method is 83% (Agrawal, 2017).

2.3.2 Detection using K-Nearst Neighbour Approach

The K-Nearest Neighbours approach has been used in machine learning to find patterns, compute statistics, and classify data. Utilizing the KNN classifier, a survey on identifying plant diseases was carried out. A detection technique for sugarcane disease was proposed. Utilizing image processing methods, feature extraction is carried out. The diagnosis of leaf scorch disease in sugarcane leaves has a 95% success rate.

(Kaur, S., Pandey, S., and Goel, S., et al., 2018). Used to distinguish between the three separate illnesses Septoria leaf blight, frog eye, and downy mildew in soybean agriculture. Their analysis of a sizable dataset revealed an average categorization accuracy of around 90%.

(Parikh, Aditya, et al., 2016) created a method for diagnosing and gauging the severity of the cotton plant disease. Using 40 images, Grey Mildew Disease could be identified with an accuracy rating of 82.5%.

A method for diagnosing diseases in rice leaves was provided by (Suresha, M., K. N. Shreekanth, and B. V. Thirumalesh et al., 2017). This technique involved taking pictures of paddy leaf illnesses while the leaves were still on paddy fields. The region of interest in the image

was found using Otsu segmentation after the RGB images had been converted to the HSV colour space. With an accuracy of 68.1%, Geometric metrics like area, perimeter, minor axis length, and major axis length were extracted and categorized using the K-Nearest Neighbour algorithm.

Digital image processing with back propagation neural network (BPNN), (Maniyath, 2018 (july 2019))diagnosed plant illnesses. Researchers have discovered methods for identifying plant illnesses by looking at photographs of leaves. To separate portions of sick leaves, Otsu's thresholding, boundary detection, and spot detection methods were applied. Based on factors including colour, texture, morphology, edges, etc., they categorized plant diseases. BPNN is used to categorize and locate plant pathogens.

Plant leaf diseases may be identified and categorized by (Vaidya, 2018) using KNN and neural network techniques. The ill plant leaf dataset is taken into account. The leaves in this collection are both healthy plant leaves and leaves that have been impacted by early blight, late blight, and black rot. Based on their categorization, they have calculated the percentage of the leaf that is altered by their results.

2.3.3 Detection using ANN

Successfully completed a proposed study for plant disease recognition utilizing the feed forward back propagation algorithm . They used an ANN classifier to evaluate early scorch, late scorch, cottony mould, and microscopic whitening disorders for the diagnosis of plant illnesses.

(Kumari, 2019) Disease was identified using imaging, segmentation, feature extraction, and classification. Features of the affected population were obtained by K-mean aggregation. The variables that are collected during image segmentation include variance, correlation, power, homogeneity, mean, and standard deviation. The illness classifier made use of group traits. NN classification was employed in this investigation. Target localization accuracy for bacterial leaf spot and cotton leaf disease was 90%.

According to (Li, 2022), the many properties of the plant leaf, including intensity, colour, and size, are recovered in order to categorize the illness the plant is experiencing.

A memory-based classification technique was given by (Li, 2022)GAN (CNN + GAN) combined. In order to produce accurate classification results, CNN only needs a little amount of raw data because it learns commonalities between input paired data. For upcoming work, GAN abstracts images from finished projects. During ongoing work (pest and plant classification), the conventional CNN model forgets.

For identifying plants based on leaf characteristics, (Rastogi A, 2015) proposed a two-phase strategy in their study. In the first step, leaf image preprocessing, feature extraction, artificial neural network training, and classification are all included. In phase two, K-Means, feature extraction, and ANN are used to categorize leaf diseases.

2.3.4 Detection using Naive Bayes

In order to diagnose oil palm plant illness using Nave Bayes, (Nababan, Marlince, et al., 2018)designed an input-output approach for an application based on intelligence. The study's findings from the Bayes technique could be used to make a diagnosis of the disease affecting oil palm plants and recognized symptoms.

(Sethy, Prabira Kumar, et al., 2020).chosen to study the diseases of rice and apple leaves (speckled deciduous disease, yellow-leaf disease, round spot diseases, and mosaic diseases). Colour, texture, and shape were all extracted from the apple leaf spot image. Disease classification was done using the BP neural network model, which had an average recognition accuracy of 92.6%.

(Sari, 2020).The results of the disease classification were determined using the Naive Bayes Classifier. Additionally, the test was run using forward-chaining search techniques. Compared to the FNBC method's accuracy of 88%, the forward chaining strategy has a 90% accuracy rate.

2.3.5 Support Vector Machine

Support vector machine developed by (Vapnik., 2018) is a learning algorithm for pattern classification and non – linear regression which has been widely applied in several sector. It is able to address both linear and nonlinear issues. Finding the ideal linear hyper- plane that minimizes the anticipated classification error for test samples that have not yet been seen forms the theoretical underpinning of SVM. SVM identifies a hyper-plane that separates the data with the greatest margin for linearly separable data. This surface's functionality is determined by:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} a_i{}^* y_i\,(x_i * x) + b^*)(x_i, y_i) \in R^N * \{-1,1\} \quad (2.1)$$

B* is the bias where a_i* is a Lagrange multiplier. Kernel functions were utilized to transform the linear inseparability problem in low dimension space into a linear partition problem in high dimension space for linearly inseparable data (Ying-xue Su, 2017). The function of this surface is defined by:

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{n} a_i{}^* y_i k\,(x, y) + b^*) \quad (2.2)$$

Where $k(x, y)$ is the kernel function. Some commonly used kernel functions were as follows.

The linear kernel function $k(\mathrm{x},\mathrm{y}) = x * y$ (2.3)

The polynomial kernel function $k(\mathrm{x},\mathrm{y}) = (1 + x * y)^q \, , q = 1, 2, \ldots . N$ (2.4)

The radial basis kernel function $k(\mathrm{x},\mathrm{y}) = exp(-\|x - y\|)^2$ (2.5)

SVM was originally designed for two classification problems. However, later it was extended to the multiclass problem.

2.3.6 Artificial Neural Network

A model known as an artificial neural network (ANN) simulates and mathematically models the anatomy of the human brain to be able to learn patterns from data. An ANN design typically consists of three different sorts of layers: an input layer, an output layer, and at least one hidden layer. A group of neurons is represented by each layer. There are nodes in the input and output layers that correspond to the input and output variables, respectively. Based on the specific problem to be solved, the number of concealed nodes varies. Edges are weighted connectors that allow data to travel between layers. As learning progresses, the weights are changed. A node receives information from the layer before it and calculates a weighted total of all of its net inputs (Siti Khairunniza-Bejo, 2014):

$$t_i = \sum_{j=1}^{n} \left(w_{ij} x_j + b_i \right) \quad (2.6)$$

Where n denotes the total number of inputs, w denotes the strength of the link between nodes i and j, x denotes the input from node j, and b_i denotes the basis. A transfer function, f_i, is applied on the weighted value t_i inputs to determine the output node, o_i:

$$o_i = f_i(t_i) \quad (2.7)$$

Activation functions come in a variety of forms, including Identity Function, Binary Step Function, Sigmoid Function, Linear Function, Tanh Function, Relu Function, and Radial Baisis Function. The sigmoid function is a non-linear function that, regardless of the input value, translates all inputs between [0, 1]. This is how it is described:

$$f(x) = \frac{1}{1+e^{-x}} \quad (2.8)$$

The models were performed to predict a risk value expressed as low, medium and high. With the sigmoid function, the Support Vector Machine was run. The Artificial Neural Network, however, was trained using the back-propagation process.

Table 2: Details of some existing plant disease detection (Hareem Kibriya, 2021).

References	Crop name	Type of diseases	Classifier	Accuracy%	Gap
(X. E. Pantazi, 2019)	Vine leaves	Downey and powdery mildew	One-class SVM	95	Single class
(S. Zhang, 2019)	Tomato	Downey and powdery mildew, gray mold and early and late blight	SoftMax	89.2	Complexity, resolution, number of sample dataset
(S. Zhang X. W., 2017)	Cucumber	Downey and powdery mildew, gray mold and anthracnose	SVM,NN	85.7	resolution, number of sample dataset
(Shijie, 2017)	Tomato	Leaf blight, curl virus, bacterial, gray spot	softMax, SVM	89	Numberof sample dataset
(G. Hu, 2019)	Tea leaf	Tea red leaves and tea red scabs	SVM	90	Single,number of sample
(R, 2020)	Corn, paddy, turmeric, tomato, sugarcane	Leaf spot, bore hole, sheat blight, alternia solani, rust	SVM, NN	94.51	Many plants But small dataset

CHAPTER THREE

3. MATERIALS AND METHODS

3.1 Introduction

This chapter covers the study methodology and the processes that were taken, from the dataset collection stage to the testing stage. It's crucial to concentrate on the step of the recognition process where noise is given to the images and they are processed. The images are processed using MATLAB, which involves cropping, changing the type of image, and introducing noise to the pictures. These actions are taken to reduce the price of picture processing. The pre-processing phase is the initial step towards a successful recognition rule, which plays a significant role. In addition to the execution time, choosing a suitable image size is crucial because it directly affects the program's output.

This research will determine leaf diseases particular to pea plant leave instead of just several distinct species of plants, as was done in earlier studies. 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.

3.1.1 Machine Learning in Mathlab

MATLAB is a powerful programming language and software environment for numerical computing and data analysis. It provides a wide range of built-in functions and tools for mathematical and scientific computing, as well as the ability to create custom functions and scripts.

MATLAB greatly simplifies the process through which knowledge is acquired. Because it offers tackles and objectives for undertaking more extensive data and applications that make engine knowledge available, MATLAB is the right setting for integrating engine knowledge into your information analytics. Because of MATLAB, engineers and information scientists can quickly access prebuilt programs, Wade toolboxes, and particular applications for grouping, organizing, and reversing data. The following can be used with MATLAB:

- Provides a powerful set of tools for creating, manipulating, and performing calculations on matrices and arrays (Matrix manipulation).
- Provides tools for analyzing and visualizing data, including statistical analysis, curve fitting, and plotting (data analysis and visualization).

- Offers tools for solving optimization problems and building mathematical models of complex systems (optimization and modeling).
- Provides tools for processing and analyzing signals and images, including filtering, Fourier analysis, and image segmentation (Signal and image processing).
- Offers tools for designing and analyzing control systems, including feedback control, state-space modeling, and simulation (Control system design and analysis).
- Gives tools for building and training machine learning models, including deep learning, reinforcement learning, and classification models (machine learning).

3.1.1.1 Machine Learning

An algorithm is developed using machine learning to model the knowledge contained in the data. Additionally, it is a kind of data analytics that instructs computers to learn from experience the same way that people and other animals do. A subset of artificial intelligence that covers a wider variety of issues is machine learning. With the help of experience rather than explicit programming or human interaction, machine learning tries to build intelligent systems or computers that can learn and train themselves. In this sense, machine learning is a constantly changing process. In order to include the dataset's data structure into ML models that companies and organizations can utilize, machine learning strives to understand the dataset's data structure. Image processing is a technique used in machine learning to transform an image into digital form and

Machine learning algorithms learn to understand things by analyzing large amounts of data and identifying patterns and relationships within that data. These relationships are used to make predictions or decisions on new data. With the use of machine learning, artificial intelligence (AI) systems may learn from their experiences. Machine learning algorithms apply computer approaches to "learn" facts directly from the data without relying on a preconceived equation as a model. As the quantity of learning examples increases, the algorithms adaptively get better at what they do. Two categories of learning methods are employed in machine learning (Nasteski, 2017):

1. Supervised learning: which, trains a model to predict future outcomes using input and output data that are known.
    - Create a predictive model based on input and output,
    - Train it to map inputs to outputs, and then use it to forecast the outcome of fresh inputs.

2. Unsupervised learning: This method uncovers underlying patterns or structures in the input data.

- On the basis of simply the input data, group and analyses the data.
- It is employed to infer conclusions from datasets made up of input data without labeled answers.

Supervised machine learning creates a model that makes predictions based on data when there is uncertainty. Using a known set of input data and known reactions to the data (output), supervised learning trains a model to generate precise predictions for the response to incoming data. Use supervised learning if the outcome you are attempting to predict has data that is known.

Supervised machine learning aims to build a model that can produce predictions in the face of uncertainty by using data as a foundation. A known input data set and a known output data set that represent known responses to the input data can be given to an algorithm to teach it how to learn under supervision. The program might then be able to predict how the model will respond to fresh data. By integrating classification (models classify input data into categories) and regression (predict continuous responses) techniques, predictive models can be created using supervised learning.

Unsupervised learning finds hidden patterns or basic data structures. The technique is used to draw inferences from datasets with input data but unlabeled responses. Clustering is the most used unsupervised learning technique. To find hidden patterns or data groupings, it is used in exploratory data analysis. A few applications for cluster analysis include object recognition, market research, and gene sequence analysis.

Without human supervision, learning reveals inherent data structures and occult patterns. This technique allows conclusions to be formed from datasets that only include input data and no tagged responses.

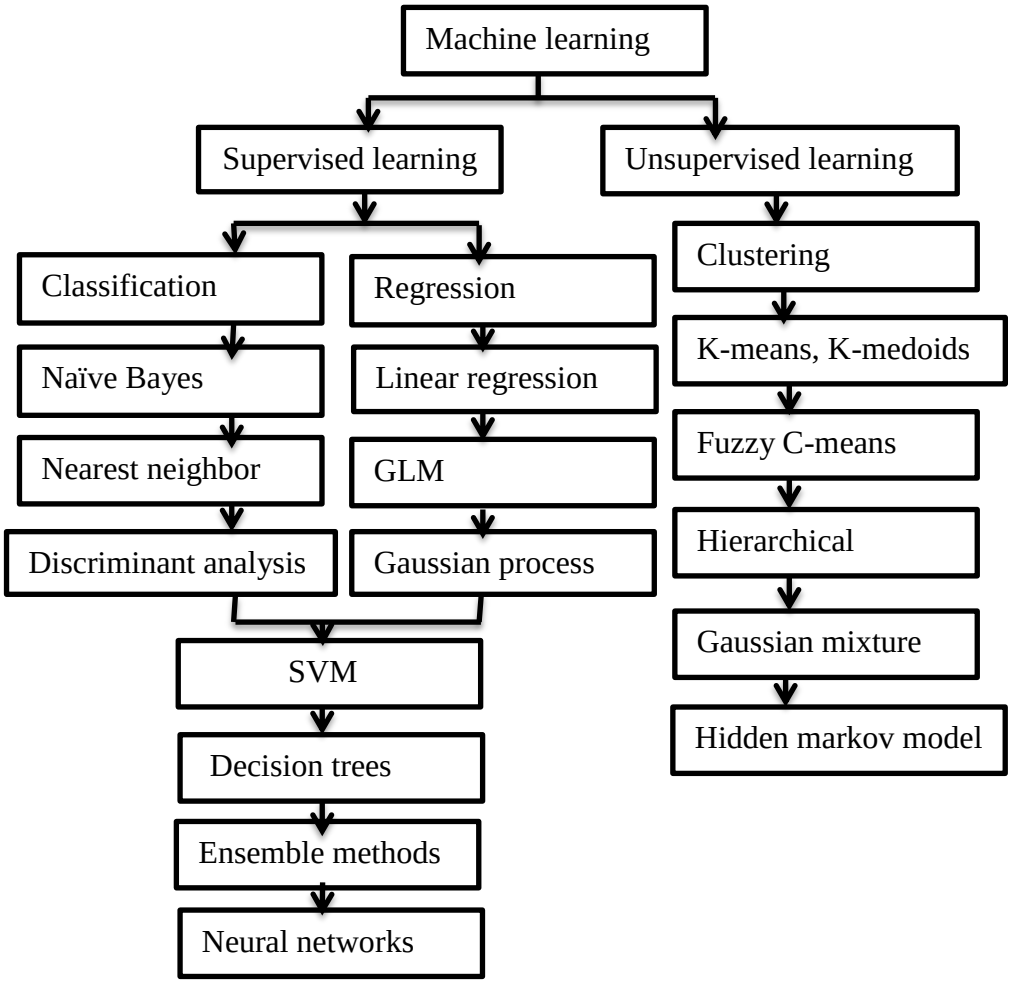


Figure 5: Machine learning techniques.

1. Unsupervised learning

Unsupervised machine learning, often known as unsupervised learning, makes use of machine learning techniques. This method of learning is employed to categorizes and examine unlabeled data. Without the assistance of a human analyst, these algorithms can identify patterns or clusters in data that were previously missed (Yau, K. A., Elkhatib, Y., Hussain, A., and Al-fuqaha, A. L. A., 2019). Their ability to perceive data contrasts and similarities aids in cross-selling tactics, customer segmentation, and image identification. It has picture recognition skills as well. (Sinaga, Kristina P., and Miin-Shen Yang., 2020).

Unsupervised machine learning is predicated on the notion that computers are capable of learning without human supervision. Un-labeled data, which is essentially raw data that can be discovered "in the wild" and is typically unstructured and unprocessed, is used for learning. Unsu-

pervised machine learning methods have a lot of drawbacks, as is only natural. There are only a few jobs that they can execute because they lack a beginning point for their instruction (Iryan, 2021).

b. Clustering

Means that the computer can classify objects into distinct groupings (clusters) depending on how distinctively they are naturally. It won't be able to tell you what kind of object is in each cluster, though, at the same time. Tasks like spam filtering, fraud detection, main personalization for marketing, and hierarchical clustering for document analysis can all be solved with clustering. etc (Iryan, 2021).

To discover unanticipated connections or clusters in large amounts of data, exploratory analysis employs this technique. The study of perception, market research, and factor sequence analysis are just a few examples of the real-world uses for cluster analysis. 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 (Chen, Xingyu, et al., 2019). Many clustering algorithms base their work on the similarity between every pair of instances.

2. Supervised learning

The use of labeled datasets to train algorithms with high accuracy in data classification and prediction is a significant difference between supervised and unsupervised learning.

A model will update its weights to take into account additional information when new data is added during cross-validation. Up until the model provides a satisfactory fit to the data, these corrections are repeated. Businesses may boost their efforts to handle real-world difficulties, such spotting and filtering spam emails, by utilizing supervised learning (Jiang, Tammy, Jaimie L. Gradus, and Anthony J. Rosellini., 2021).

Supervised learning is adaptable, thorough, and covers many of the prevalent ML jobs that are in great demand right now. Data with labels is necessary. As a result, the models are trained using processed (cleaned, randomized, and structured) and annotated data. The training process is overseen by a person through the processing and annotation (labeling) of the data, hence the term "supervised learning." It is a rather labor-intensive and time-consuming operation that is typically outsourced to free up time for the essential company duties (Iryan, 2021).

a. Regression

Finding a correlation between various variables with continuous values typically requires analysis. Finding correlations and making output predictions are helpful. Based on a set of an object's features, this kind of supervised method is frequently used to estimate the cost or value of the thing.

b. **Classification**

Enables- teaching the AI to classify various objects (values) into multiple groups. The computer is aware of which class each item belongs to. It is the process of categorizing or labeling an image according to its visual information. This entails using a dataset of labeled photos, where each image is connected to a certain category or class, to train a machine learning model. The model can be used to predict the class of brand-new, undiscovered photos when it learns to identify patterns in the image data that are indicative of each class.

3. **Semi- Supervised learning**

The primary drawback of supervised learning is that it necessitates expensive, time-consuming human tagging by ML experts or data scientists. There are also only a limited number of uses for unsupervised learning (Sinaga, Kristina P., and Miin-Shen Yang., 2020). The idea of semi-supervised learning is presented to address these issues using supervised learning and unsupervised learning methodologies. In its training set, this algorithm uses both labeled and unlabeled data. While there is a substantial amount of unlabeled data, there is only a little amount of annotated data. Utilizing an unsupervised learning technique, comparable data are initially clustered, aiding in the conversion of the unlabeled data into labeled data. As a result, purchasing tagged data is more expensive than purchasing untagged data (Reddy, Y. C. A. P., P. Viswanath, and B. Eswara Reddy., 2018).

4. Reinforcement learning

Through repeated trial-and-error encounters with a dynamic environment, a computer agent learns to execute a task using the machine learning process known as reinforcement learning. By choosing a set of activities that will maximize a reward metric, the agent can execute the task without the help of a human and without being particularly programmed to do so. Artificial intelligence is utilized for the process of reinforcement learning in a game-like environment. Before making a decision, the computer will attempt and fail its way through the problem. Artificial intelligence is rewarded or punished for its actions depending on what the programmer intended

the machine to do. Its goal is to produce the most total benefit feasible. (Alharin, Alnour, Thanh-Nam Doan, and Mina Sartipi., 2020).

3.1.1.2 Machine Learning Algorithm

The principles, processes, and statistical models that make up a machine learning algorithm allow computers to carry out particular tasks without having to be explicitly programmed. Algorithms that use machine learning discover patterns in data and forecast or decide based on those patterns. A dataset is used to train a machine learning system, and the input data is labeled with the desired outcome. On the basis of fresh, unforeseen data, forecasts may be made. Machine learning algorithms are able to foresee outcomes, find hidden patterns in data, and improve performance based on past results.

3.1.1.3 Sorting of Machine Learning Algorithm

Selecting the best machine learning algorithm can be difficult because so many supervised and u supervised machine learning algorithms take a different approach to the learning process. There is not 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 over -fit 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 (Raschka, 2018).

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.

3.2 Required Material

Both MATLAB R2015a and MATLAB R2019b are used in this research project.

3.3 Dataset

The data collection used for this work has 2000 images, which will comprise images of healthy and damaged pea plant leaves, which will serve as the model's foundation. These pictures were taken from the dataset (Starter: pea plant (Pisum sativum) leaf disease. and Kaggle provided access to additional data sets, allowing the collection of additional photos and by taking pictures

through camera and also by searching each pea leave diseases on google. pea leaf diseases are illustrated in the following figure.

Figure 6: Different diseases of pea leaf plant.

Table 3: disease dataset collected from Ziquala Woreda Addis Alem in 2022.

Name of leaf	Type of disease	Number of samples
Pea	Ascochyta disease	400
Pea	Powdery mildew	400
Pea	Downey mildew leaf disease	400
Pea	Rust leaf disease	400
Pea	Healthy pea leaf	400

3.4 Proposed Methodology

This part includes the implementation steps and methods consisting of creating and deploying identification, classification techniques. Illnesses of the pea plant have been found, and treatments for recovering from the leaf illnesses will be offered. Using image processing, the leaf's impacted area is visible. The database has been preprocessed, including image resizing, reshaping, and array form conversion. The method is a step-by-step collection, acquisition, preprocessing, segmentation using the k-means clustering method, feature extraction using the GLCM method, classification and detection, or system training using the SVM algorithm. Two folders, titled train and test, are seen here. Testing the system and determining its accuracy is the main component of the training and test used to create the system on plant leaves. Below is a block diagram of the suggested system.

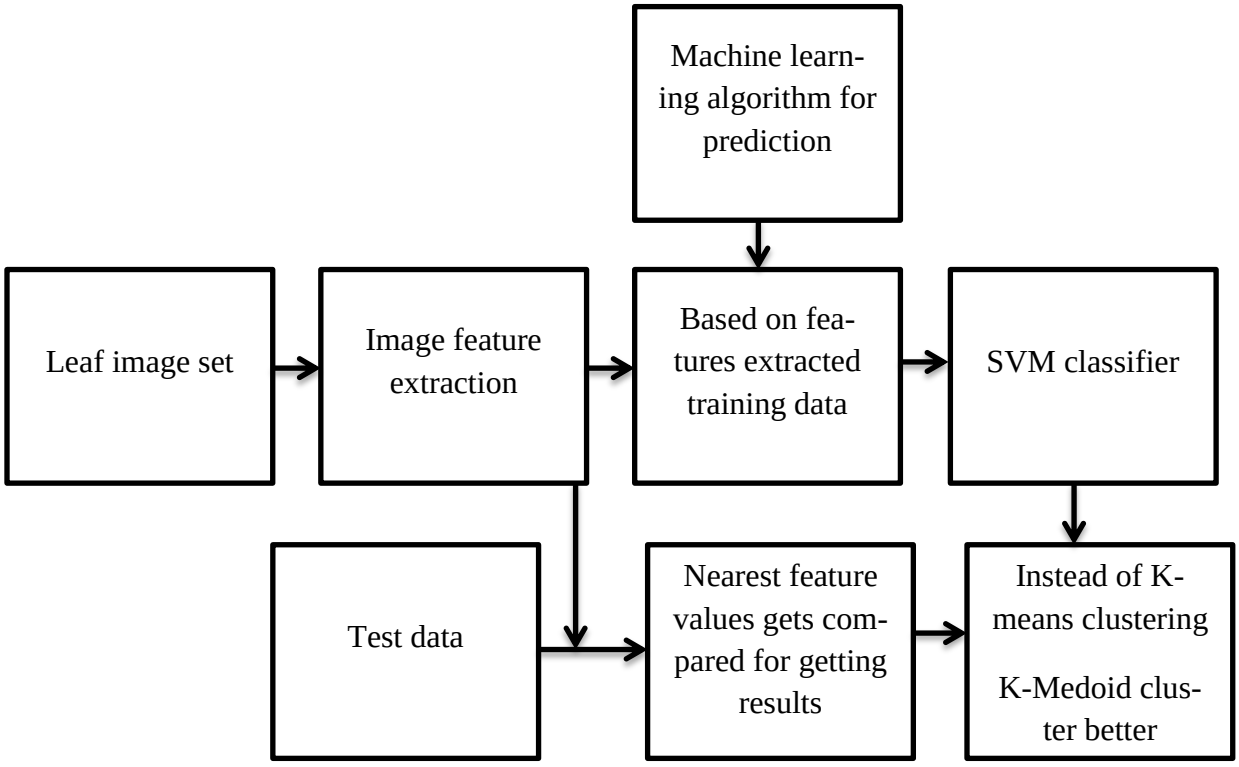


Figure 7: Proposed Work Block Diagram

For training and validation purposes, a data collection comprising photos of pea leaves with various illnesses and photographs of healthy leaves is employed. The algorithm predicts if the leaf is affected by a disease and the name of the disease based on the feature extraction parameter. Support vector machine (SVM) and K-means clustering are used by the machine learning algorithm

to classify disorders and forecast the health of the leaf. Kernel-based learning is a learning method used by the classification tool known as the vector support machine. With the help of the kernel function, the support vector machine (SVM) may convert data from the input space into the feature area with the maximum feasible size, improving the accuracy of the job. Below is a block diagram of the suggested system (Mokhtar, Usama, et al., 2014).

Support vector machine (SVM) is a collection of supervised learning techniques that evaluates data used for classification in the process of identifying leaf diseases. For labeling purposes, an SVM classifier is utilized. It is adaptable, effective in high-dimensional spaces, and memory-efficient. Results are improved and more precise as a result. An algorithm for unsupervised learning is K-means clustering. The objective is to divide N observations into K-clusters. A cluster is a collection of data that has been clustered together because of comparable characteristics. It serves as a labeling tool (Mokhtar, Usama, et al., 2014).

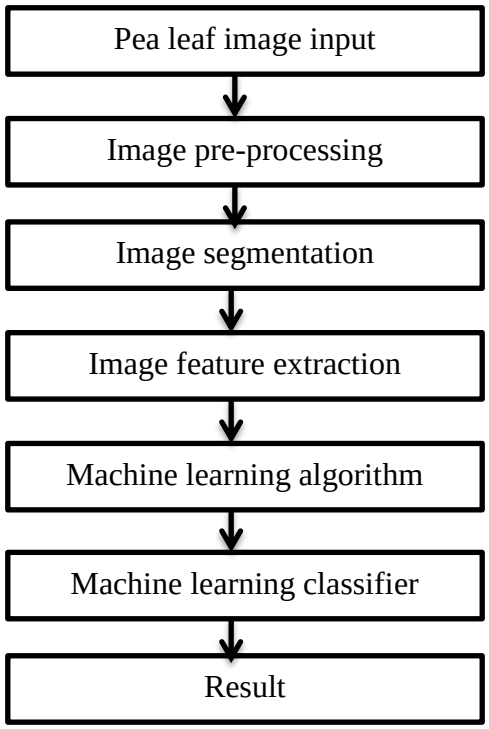


Figure 8: Basic Steps for Plant Disease Detection (Khandekar, 2019)

The steps of pea leaf parameter analysis, disease detection and classification are image acquisition, preprocessing, segmentation, feature extraction and apply machine learning algorithm.

There are different steps involved in image processing: The first step is known as Image Acquisition. Second -step preprocessing in which images are resized and various noises from images are

removed with the help of filter. The third step is segmentation of leave disease in which first of all disease spot is detected by hue histogram on HSI model then K-mean clustering í technique on L*a*b* í color í model is applied. After the K-means clustering all clusters are converted into HSV color model and a cluster that contains disease is segmented by analyzing maximum mean hue value among all clusters. The fourth step of image processing is feature extraction in which color and texture features are extracted from the segmented cluster. The fifth step training and classification and the last step recognition and detection.

The sequentially suggested method includes gathering a database of leaf images, pre-processing those images, segmenting those images using the k-means clustering method, extracting features using the GLCM method, and then training the system using the SVM algorithm.

3.4.1 Input Image (Image Acquisition)

The initial technique for processing digital images, referred to as image collection, entails taking an image using a digital camera and saving it on a digital medium for use in subsequent MATLAB operations. To gather the necessary dataset for the classifier's training, a method called image acquisition must be used. Retrieving an image from hardware so that it can go through additional processing is another activity that is being performed. We photographed healthy and sick leaves using a digital camera for my business. The first step in the procedure is gathering data from the public repository (field). For additional processing, it uses the image as input. The photos are captured in a way that allows any input format (.bmp,.jpg,.gif) to be used in the process.

3.4.2 Image Preprocessing

It is described as a procedure on images that enhance the key elements needed for processing and analyzing the images while also correcting distortion. This process reduces image duplication without introducing any new information to the image. The basic goal of image pre-processing is to enhance certain image attributes for subsequent processing, improve image data that contains undesired distortions, remove noise from the image, and modify pixel values. It improves the image's quality. Pre-processing techniques include altering the size and shape of images, removing noise, converting images, improving images, and performing morphological procedures. Here utilized a variety of MATLAB codes to resize images in this project (Bhadur, 2020).

It comprises techniques for converting images, removing noise, enhancing contrast, and reducing undesired distortion. Due to the fact that the photographs were taken in the actual field, noise

such as dust, spores, and water spots may be present. The goal of data preparation is to reduce image noise so that the pixel values can be adjusted. It improves the image's quality. Different facets of exploitation include the camera's depth of field, the volume of noise, and the blurriness of movement. The objectives of picture restoration techniques are lower noise levels while restoring any lost resolution (Verma, 2020).

1) Spot Identification on HSI model

The fundamental method for representing the colour information of images in image analysis was the RGB colour model. (A. K. Singh, 2021). An RGB model is transformed into the derived-color model known as HSI by applying formulae. The HSI model provides us with the best understanding of surface colour and leaf change after disease damage when compared to other models. In the HSI colour space, H (hue) stands for colour purity, S (saturation) denotes how much white makes up a particular colour, and I (intensity) denotes brightness (B. Zhang, 2021). The conversion from RGB to HIS format is as follows:

$$H = \theta, \text{ if } \mathrm{B} \le G \tag{3.1}$$

$$360 - \theta, \text{ if } \mathrm{B} \ge \mathrm{G} \tag{3.2}$$

$$\theta = \cos^{-1}\{\frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-B)^2+(R-B)(G-B)^2]}\} \tag{3.3}$$

$$S = 1 - \frac{3}{(R+G+B)}[MIN(R,G,B)] \tag{3.4}$$

$$I = \frac{1}{3}(R+G+B) \tag{3.5}$$

.

The conversion from HIS to RGB for ($0° \le H \le 120°$).

$$B = I(1-S) \tag{3.6}$$

$$R = I[1+\frac{S\cos(H)}{\cos(60^\circ - H)}] \tag{3.7}$$

$$G = 3I - (R+B) \tag{3.8}$$

$$120^\circ \le H \le 240^\circ \rightarrow H = H - 120^\circ \tag{3.9}$$

$$240^\circ \le H \le 360^\circ \rightarrow H = H - 240^\circ \tag{3.10}$$

The ranges for hue, saturation, and intensity are [0, 1], [0, 360], and [0, 1], respectively. The leaf disease's coloration is highlighted by the Hue adjustment. Selecting a cutoff value that separates diseased areas from plant leaves is crucial. Find the disease colour between the highest and lowest peak points (within the illness spot ranges) in the histogram for the hue component. These two points, designated as "1" and "2," serve as thresholds. In order to maximize the amount of information available for identifying and isolating the disease spot within the leaf image, the hue value is adjusted in the next phase.

$$H_{ij}^{'} = \begin{cases} H_{ij} & ; H_{ij} < \tau_1 \\ H_{ij} + \eta & ; \tau_1 \le H_{ij} < \tau_2 \\ H_{ij} & ; H_{ij} \ge \tau_2 \end{cases} \quad (3.11)$$

Where, Hij and Hij′ is the image's original and adjusted hue values, respectively. is a threshold that raises the colour value in the image in order to distinguish between the diseases. In order to achieve the enhanced image, combine the updated H with the unaltered S and I components to create the final enhanced HIS colour image, then change the final enhanced HSI colour map back into the RGB colour map (Prakash M. Mainkar, 2015).

3.4.3 Segmentation

It is the division of an image's pixel content into groups. The goal of segmentation is to make a visual representation more straightforward or practical. Image segmentation is a technique for breaking up a digital image into smaller chunks and converting an image into a form that makes it easier to analyze. Finding the image's objects and border lines requires the usage of image segmentation. With the aid of the Otsu classifier and the k-mean clustering algorithm, the segmented images are grouped into various sectors. The RGB colour model is converted into the Lab colour model before to grouping the photos (Y. Zhang, 2020).

The use of the Lab colour model allows for the easy clustering of segmented images. K-means clustering method for image clustering in which at least one portion of the cluster has an image with a large region of diseased part. The k-means clustering algorithm is used to divide the objects into K classes based on a set of features. Image Clustering captures similar attributes in the same group. Because the disease symptoms are grouped together, it is easier to find the disease. The classification is accomplished by minimizing the sum of square distances between data objects and their corresponding clusters. The image is converted from RGB Colour Space to L*a*b* Colour Space, which has a brightness layer 'L*', chromaticity layers 'a*' and 'b*'. The 'a*'

and 'b*' layers include all of the colour information, and colours are categorized in 'a*b*' space using K-Means clustering. K-means findings are used to label each pixel in the image, and segmented images including diseases are generated (BhaviniJ.Samajpati, 2016).

Image segmentation is used to determine the region of interest (ROI). It is the process of dividing one image into several pieces. Normally, it is used to recognize objects and boundaries in images. Using the Otsu classifier and the k-mean clustering approach, the segmented images are clustered into various sectors. The RGB colour model is converted into the lab colour model prior to grouping the photos (Devaraj, 2019). The use of the lab colour model allows for the easy clustering of segmented images. The distance Euclidean metric method, used by default in the MATHLAB 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 distance between them are shown (Kshirsagar, 2018).

In this study I, For good segmentation results, utilize a segmentation algorithm that partitions the input image into three groups.

1] Segmentation using the Boundary and Spot Detection Algorithm: The RGB image is transformed into the HIS model for segmentation. Boundary detection and spot detection aid in locating the affected portion of the leaf.

2] Clustering with K-means: K-means clustering is used to categorize objects into K classes based on a set of features. Objects are classified by minimizing the sum of the squares of the distance between the object and the corresponding cluster. The K-means Clustering algorithm:

a) Choose the center of the K cluster at random or using a heuristic.
b) Assign each pixel in the image to the cluster with the shortest distance between it and the center of the cluster.
c) Once again, compute the cluster centers by averaging all of the cluster's pixels. Steps 2 and 3 should be repeated until convergence is achieved.

3] Thresh-holding generates binary images from grey-level images by setting all pixels below a certain threshold to zero and all pixels above that threshold to one.

i) Separate pixels into two groups based on the threshold.
ii) Then compute the mean for each cluster.
iii) Calculate the square of the mean differences.

iv) Add the pixels from one cluster to the pixels from the other affected leaf. Therefore, the degree of greenness in the leaves can be utilized to identify the area of the leaf that is affected. The image's R, G, and B components are taken out. The Otsu's approach is used to determine the threshold. If the green pixel intensities are below the calculated threshold, the green pixels are then mask and eliminated.

3.4.4 Feature Extraction

The regions based on a chosen representation are described using it. The region's boundary serves as a representation of it, and the boundary's characteristics such as texture, colour, and shape describe it. The crucial step is to accurately anticipate the contaminated area. Extraction of form and textural features is done here. Calculations are made for shape-oriented feature extraction such as area, colour axis length, eccentricity, solidity, and perimeter. Similar to this, the health of each plant is assessed using the texture-oriented feature extraction of contrast, correlation energy, homogeneity, variance, skewness, and mean. (Prakash M. Mainkar, 2015). When selecting a feature extractor, there are various factors that should be taken into account, such as the kind of layers. A higher number of parameters increases the system's complexity and has a direct impact on the system's speed and output. Each network has been created with unique properties that are intended to improve accuracy while lowering computational complexity. After segmentation, the GLCM characteristics are extracted from the image. The grey level co-occurrence matrix (GLCM), a statistical method for examining texture, takes the spatial relationship between pixels into account. The GLCM functions explain the texture of the images by computing the spatial relationship between the pixels in the images. Statistical measurements are derived from this matrix. a collection of offsets indicating pixel correlations (Mrunmayee Dhakate, 2015). With the help of many criteria, including size, colour, and others, the image can be examined during the feature extraction process. Features are quantifiable descriptors of an object that identify basic traits like shape, colour, texture, or motion, among others. They are used to categorize different plant diseases.

Desired feature vectors, such as colour, texture, morphology, and structure, are extracted during feature extraction. It is a technique for utilizing the necessary resources to accurately describe a huge quantity of facts. The Grey level co-occurrence matrix (GLCM) formula for texture analysis yields statistical texture features, which are determined from the statistical distribution of observed intensity combinations at the given site in relation to other locations. Statistics are ranked

in order of first, second, and higher for the number of intensity points in each combination, which is also essential in GLCM. Many GLCM statistical texturing properties. For Mean, Kurtosis, Standard Deviation, Variance, and Skewness, use 1st order statistics. For correlation, energy, homogeneity, covariance, information measure of correlation, entropy, contrast, and inverse difference, use 2nd order statistics. (Mrunmayee Dhakate, 2015). The (i, j) element in the normalized gray-level co-occurrence matrix will be represented by Pi. The quantized image's N stands for the number of unique grey levels.

measure of correlation, entropy, contrast and inverse difference and difference entropy.

Colour Feature	**Formulas**	
Mean	$$\mathrm{E}_i = \frac{1}{N}\sum_{j=1}^{N}\rho_{ij}$$	(3.12)
Standard Deviation	$$\sigma_i = \sqrt{(\frac{1}{N}\sum_{j=1}^{N}(\rho_{ij} - E_i)^2)}$$	(3.13)
Skewness	$$\mathrm{S}_i = \sqrt[3]{\frac{1}{N}\left(\frac{1}{N}\sum_{j=1}^{N}\left(\rho_{ij} - E_i\right)^3\right)}$$	(3.14)

Here ρij resides at the jth image pixel and represents the colour value of the ith plane.

Energy, homogeneity, correlation, contrast, and other textures Using the grey level co-occurrence matrix (GLCM), features are retrieved.

Texture Feature	**formulas**	
Energy	$$f_1 = \sum_i 1 \sum_j \{p(i,j)\}^2$$	(3.15)
Contrast	$$f_2 = \sum_i 1 \sum_j \left|i-j\right|^2 p(i,j)$$	(3.16)
Correlation	$$f_3 = \frac{\sum_i 1 \sum_j (i,j)\, p(i,j) - \mu_x \mu_y}{\rho_x \rho_y}$$	(3.17)
Homogeneity	$$f_4 = \frac{p(i,j)}{1+\left\|i-j\right\|^2}$$	(3.18)
Entropy	$$f_5 = \sum_i 1 \sum_j p(i,j) \log p(i,j)$$	(3.19)
Sum average	$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i)$$	(3.20)
Sum variance	$$f_7 = \sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$$	(3.21)

Mean square error $$f_8 = \frac{\sum_i \sum_j |i - \mu|^2}{M * N} \quad (3.22)$$

3.4.5 Training & Classification

The principle behind a support vector machine is to maximize the shortest distance between the separating hyper plane and the nearest example. Basic SVM only supports binary classification, however extension SVM allows for multiclass classification cases. For addressing the separation of the various classes, extra restrictions and parameters are added to optimization problems in these modifications. Since SVM is a binary classifier, the class labels can only have two possible values: 1 and 0. A series of binary classifiers are built in this way: f 1, f 2,..., f M and each is trained to distinguish one class from the others in order to produce M-class classifiers.

$$argmaxg^j(x), whereg^j(x) = \sum_{i=1}^{m} y_i a_i^j k(x, x_i) + b^j,\ j = 1.....M \quad (3.23)$$

The $g^j(x)$ function returns the signed real value that can be interpreted as distance from Separation of hyper plane to point x. additionally, value can be thought of as a confidence value. One is more certain that the point x belongs to the positive class the higher the value. As a result, give point x to the class that has the highest confidence value for it. I substituted K-means clustering with K-The best multi-SVM technique for classifying and identifying leaf disease is medoid. In order to create a database, an image must first be taken and then go through preprocessing, segmentation, and feature extraction. After that, a disease name is chosen for a specific leaf, and the data is then recorded in the database.

3.4.6 Classification and Detection

The predefined patterns required to recognize and classify images into the appropriate categories are stored in a database that is the foundation of the image processing classification system. In order to interpret the diseased zone that has been taken from a picture and identify the type of disease infection in leaves, classification is used.

3.4.7 Hue Saturation Intensity (HSI)

It is advised to use the hue saturation intensity model (HSI) to create an RGB model. The (HSI) model is comparable to how colors are encoded by human vision according to hue. The HIS 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 convolutional, equalization, histogram, and others perform best in the HIS color space. Because utilizing equal proportions of the colors R, G and

B, these strategies are effective with light values (Reddy, Y. C. A. P., P. Viswanath, and B. Eswara Reddy., 2018).

In contrast, it accepts any values 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 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 degrees of purification is measured as fullness. The word hue describes the sober shapes of yellow, orange, or red.

3.5 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.

3.5.1 K-mean Clustering

K-mean 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 dataset into the necessary number of groups. The best course of action is to maintain the greatest feasible physical separation between one another (Bansal, Arpit, Mayur Sharma, and Shalini Goel., 2017).

By minimizing the squares of the distances between the picture intensities and the cluster centroids, the K-means clustering algorithm separates the data into groups. The method of iterative K-means clustering, divides the data into k groups whose centroids specify how they should be grouped (Sandhu, Gurleen Kaur, and Rajbir Kaur., 2019). Then, exactly one of those clusters is given to each N observation. The algorithm's steps are displayed below:

1. Ascertain the k clusters starting centers (centroid).
2. Calculate the distances between each clusters center and all observations, using points as the unit of measurement.
3. Insert each observation into the cluster whose centroid is closest to the observer.

4. Before going on to next step to find k new centroid sites, find the mean data for each cluster.
5. 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, which comes first.

3.5.2 K-Means Clustering for Color Image

Clustering is the method of assembling groups of pixels having comparable characteristics. 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. Several Algorithm Related to K-Means Clustering for Image Segmentation.

A type of centroid-based clustering called k-means uses a centroid to represent each cluster. The centroid is the mean of all the data points in the cluster. There are several other clustering algorithms that are similar to K-means clustering in that they also use centroids to represent clusters (Gope, 2021).

1. 1. K-Medoids: This k-means clustering variation represents clusters using medoids rather than centroids. A mediod is the data point in a cluster that is closest to the center of the clusters. Often used when the data contains noise or outliers, as it is more robust to such data.
2. Fuzzy C-Means: a soft clustering technique that, like K-means clustering, places each data point in a cluster according to how similar it is to the centroid. Fuzzy C-Means assigns each data point a membership value for each cluster, indicating the degree of the data point's membership to each cluster, as opposed to K-means clustering, which allocates each data point to a single cluster.
3. Hierarchical clustering: a kind of clustering technique where each data point begins as its own cluster and combines related data points together in a hierarchical framework. The clusters can be represented by a teach level of the hierarchy.
4. Gaussian mixture models: are a probabilistic clustering approach that makes the assumption that the data points were produced using a combination of Gaussian distributions. The programmed calculates the Gaussian distribution's parameters.
5. .Kernel K-Means: This K-means clustering variant maps the data points into a high-dimensional feature space so that the grouping can be done more accurately and quickly.

3.5.3 K-Medoids

A popular method for segmenting images is K-Medoid clustering, a variation of the K-Means clustering algorithm. K-Medoid clustering's fundamental principle is to divide a set of data points into K clusters, with each cluster represented by the data point closest to the cluster's center, or medoid. In image segmentation, the data points are usually the pixels in an image, and the goal is to group pixels that are similar to each other into clusters. K-Medoid clustering achieves this by iteratively assigning each pixel to the cluster whose medoid is closest to it, and then updating the medoids to be the pixel that minimizes the sum of distances to all the other pixels in the cluster (H. Wang, 2021). The starting medoids are chosen at random from K pixels in the image using the algorithm. Then, using a distance metric like Euclidean distance, it assigns each pixel, in each iteration, to the cluster whose medoid is closest to it. After all the pixels have been assigned to clusters, the medoids are updated by selecting the pixel that minimizes the sum of distances to all the other pixels in its cluster. Up to a certain number of iterations or until the medoids stop changing, this process is repeated. Once the algorithm has converged, each pixel is assigned to a cluster, and the image can be segmented by assigning each pixel the color or label of its corresponding cluster. The result is an image in which pixels that are similar to each other are grouped together. While pixels that are dissimilar are separated into different clusters. This can be useful for tasks such as object recognition and tracking, as well as for image (Y. Zhou, 2022).

3.6 The Comparison between K-Means Clustering and K-Medoids

Both K-means clustering and K-medoids clustering are unsupervised machine learning algorithms used for clustering or grouping similar data points together. However, the main difference between the two lies in the way they choose their cluster centers.

In K-means clustering, the cluster center is represented by the mean value of all data points in that cluster. The algorithm iteratively assigns each data point to the cluster whose mean is closest to it, and then recalculates the mean of each cluster. This process repeats until the means no longer change significantly or a maximum number of iterations is reached.

On the other hand, K-medoids clustering, also known as PAM (Partitioning Around Medoids), chooses medoids (or representative points) as cluster centers. Unlike K-means clustering, the medoid is not necessarily the mean of the data points in the cluster, but rather the data point that

minimizes the sum of distances to all other points in the cluster. The algorithm starts by randomly selecting k medoids and assigns each data point to the closest medoid. It then tries to swap each medoid with a non-medoid data point to see if this decreases the total distance between points and medoids. This process repeats until no further swaps can be made. In image processing, K-means clustering and K-medoids clustering can both be used for image segmentation, which is the process of dividing an image into multiple segments or regions with similar characteristics. K-medoids clustering is generally preferred in this context because it is less sensitive to outliers and noise in the data, which can be common in image processing.

K-Means and K-Medoids are both popular clustering algorithms used for grouping data into clusters. However, there are some key differences between the two algorithms:

Centroid vs Medoid : In K-Medoids, the center of each cluster is represented by its medoid, which is the data point that is closest to the center of the cluster (H. Wang, 2021).

Table 4: The comparison between K-Means and K-Medoid clustering (Gope, 2021).

K-Means cluster	K-Medoids cluster
Each cluster's center, which is the average of all the cluster's points, serves as a representation of the cluster's center.	The nearest data point to each cluster's center, or medoid, serves as a representation of the cluster's center.
Because it aims to reduce the sum of squared distances between each point and the designated cluster centroid, it is sensitive to outliers.	Less sensitive to outliers because it uses the medoids, which is less affected by outliers and noise.
Has a lower computational complexity compared to K-Medoids, especially for large datasets. Because only needs to compute the mean of the points in each cluster.	Needs to compute the distances between each point and every other point in the same cluster to find the medoid.
K-Means is guaranteed to converge to local minimum	K-Medoid is guaranteed to converge to a solution that is a local minimum of the clustering cost function.
Often faster and more suitable for datasets with Gaussian distributions well-separated clusters.	More robust to outliers and suitable for datasets with non-Gaussian distributions.

3.7 Leaf Parameter Calculation Method

For the purpose of calculating leaf parameters, images of the leaf and reference items that have known dimensions like length, width, and area are taken. Following image acquisition, the rgb gray (x) function is used to transform the RGB image to a grayscale version. The conversion equation is as follows:

$$f(x) = 0.299 * R + 0.587 * G + 0.114 * B \quad (3.24)$$

By using the thresh-holding approach, the grey image is transformed into a binary image. A morphological operation is used to remove any black spots or holes from the image (Sachin D. Khirade, A. B. Patil, 2015).

3.8 Algorithm Description

a. K-Nearest Neighbors

The classification algorithm k-Nearest Neighbour (KNN), one of the more simple techniques for machine learning, has been around long enough to reach a point of theoretical maturity. The k-nearest neighbours algorithm, often known as KNN or k-NN, is a supervised learning classifier that makes predictions or classifications about the clustering of a single data point using proximity. Based on the presumption that a sample must also fall into the same category as the k other samples that are most comparable to it (i.e., the samples that are closest to it in the feature space) in order for the method to be applicable, this approach was developed (Dang, Yijie, et al., 2018).

One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbour. It places the new case in the category that is most similar to the extant categories based on the assumption that the new case and the data are similar to the cases that are already available. A new data point is classified using the K-NN algorithm depending on how similar the existing data is and is stored. This means that utilizing the K-NN method, fresh data can be quickly and accurately sorted into a suitable category. Although it can be used for both classification and regression, classification issues receive the majority of its employment.

Since K-NN is a non-parametric technique, it makes no assumptions about the underlying data. It is also known as a lazy learner algorithm since it saves the training dataset rather than learning from it immediately. Instead, it uses the dataset to perform an action when classifying data. The KNN algorithm simply saves the information during the training phase, and when it receives new data, it categorizes it into a category that is quite similar to the new data (Anchalia, Prajesh P., and Kaushik Roy., 2014).

K-NN makes it simple to determine the category or class of a given dataset. Consider the diagram below:

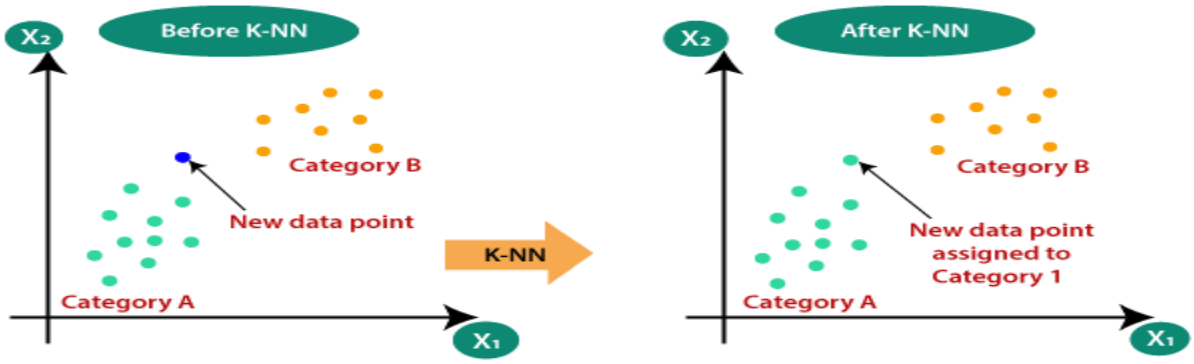


Figure 9: Algorithm of KNN (Anchalia, Prajesh P., and Kaushik Roy., 2014).

The following algorithm can be used to demonstrate how K-NN works:

Step-1: Choose the K th neighbor's number.

Step-2: Identify the Euclidean distance between K neighbours.

Step-3: Pick the K closest neighbours based on the Euclidean distance estimate.

Step-4: Count the number of data points in each category among these K neighbours..

Step-5: Assign the additional data points to the category where the neighbour count is at its highest.

Step-6: Model is prepared.

Suppose we have a new data point and we need to put it in the required category.

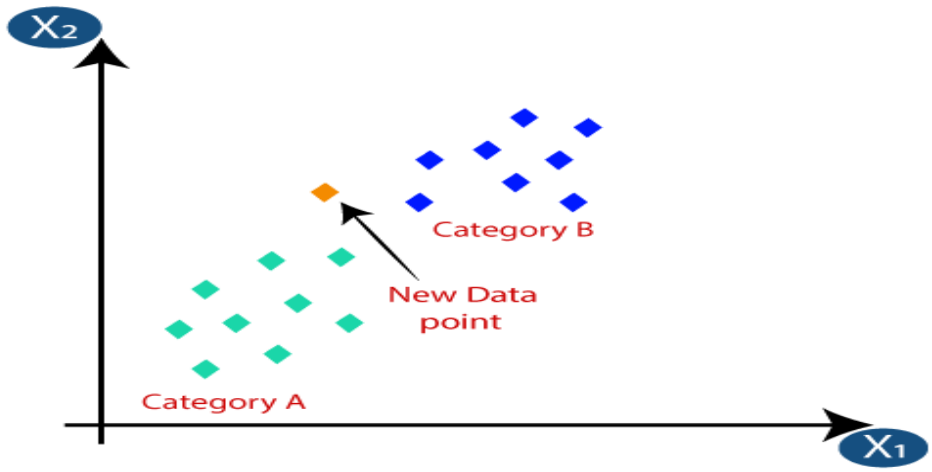
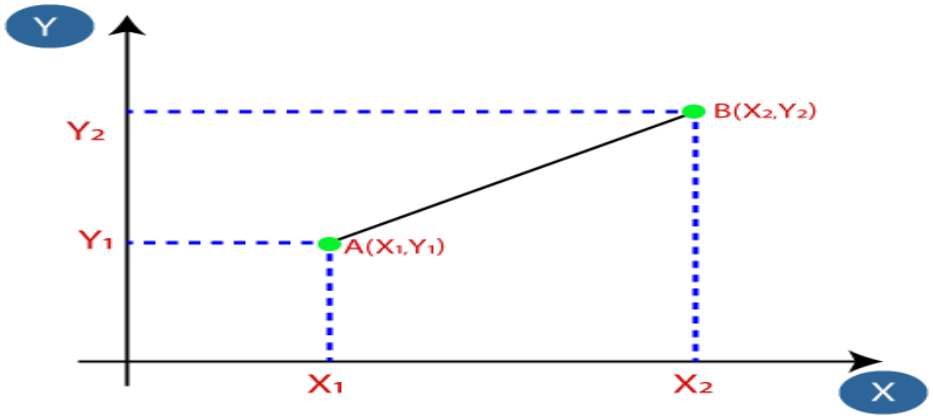


Figure 10: K-Nearest Neighbors (Dang, Yijie, et al., 2018).

The distance between two points, which we have already examined in geometry, is known as the Euclidean distance. It is calculable as:



Eculidean between A_1 and B_2 =$\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$

The last step in our processing of the leaf phase is the testing of various images and identifying the diseases. The algorithm used in the classification program is SVM. We use the variable for the model to represent healthy and unhealthy. Finally, give the data for the model and detect if it is healthy or diseased. SVM algorithm is more efficient for dividing a huge amount of data and it can be described as an efficient machine learning algorithm. As it is building on finding solutions to classification and identification tasks. It can be learning characters automatically on the dataset. This algorithm analyzes visual leaves more efficiently. The quality and type of training data collectively impact the capabilities of the model. Classifier accuracy depended on the data. The fundamental cause of each condition is categorized as either infectious or healthy.

3.8.1 Training and Testing Algorithm

Input: Providing an image of leaves localization.

Output: Classification of a review into healthy or diseased, it is diseased provides the remedies for overcoming the deficiency.

Step1: Start

Step2: Prepare a database (healthy or diseased)

Step3: Preprocessing normalization

Step4: Train SVM

Step5: Real images from Google or dataset

Step6: If the probability of healthy>probability of unhealthy display a healthy leaf, otherwise display a diseased leaf.

Step7: go to the fourth step.

Step8: stop

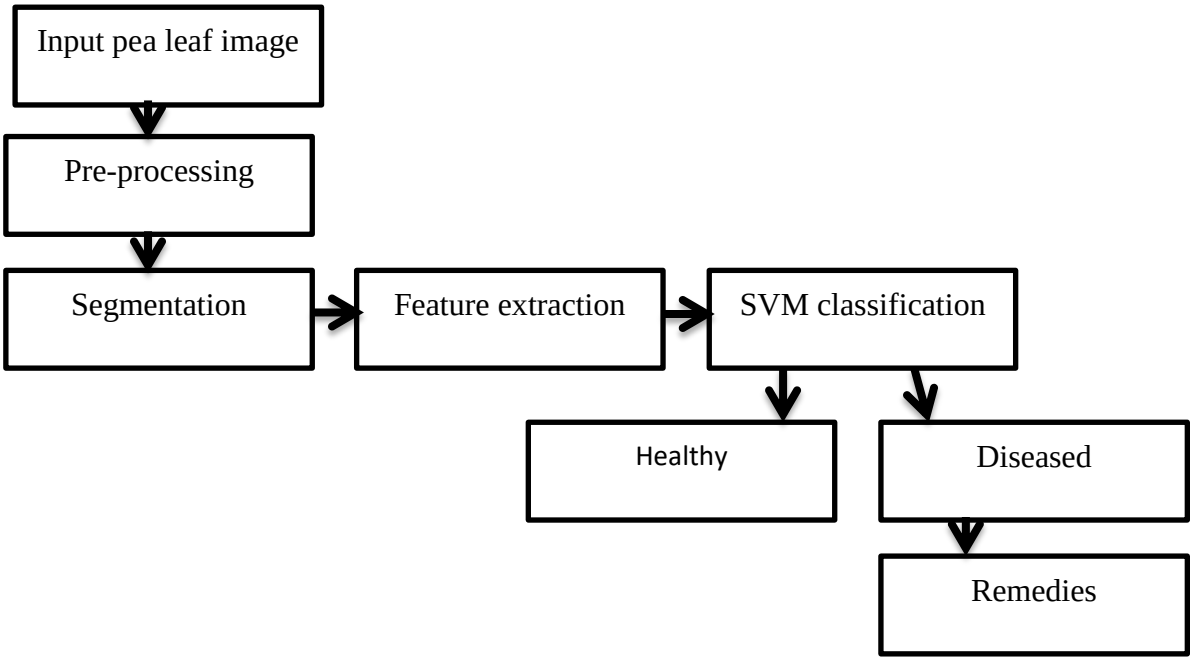


Figure 11: Block diagram of training and testing algorithm

3.9 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. The support vector machine can transform data from the input space into the feature area with the largest possible size with the aid of the kernel function (Mokhtar, Usama, et al., 2014).

In addition to SVM on the input image, the SVM approach for image diagnostics also requires image preparation. Support vector machines' categorization technique divides the image into classes. By removing the green pixels and applying a mask to the image, the presence of the sickness on the leaf may be determined and classified. The captured images of the leaf are divided into groups or segments using the K-means clustering method in the image segmentation block after being preprocessed for image enhancement. Before the clustering phase, the features are removed. (BhaviniJ.Samajpati, 2016). An algorithm that is employed for training, categorization, and ultimately, disease identification in leaves. The primary goal of the support vector machine algorithm is to do output classification.

At a location with a large magnitude, indirect data can be stored using hyper-plane. SVM's 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. Instead than concentrating on training error, think about lowering the upper limit of the standard error. (Deepa, 2021). SVMs can solve segregation difficulties. Another potential use for the kernel technique is the modification of feature maps. To match datasets with increasingly complex decision regions, nonlinear support vector machines (SVMs) use kernel functions such the radial basis function (RBF), polynomial functions, and others (Padmanabhuni, 2021). The SVM algorithm is used to predict the type of disease that will affect the leaf. The radial basis function is used as the kernel here (RBF) (Swati dewaliya, 2018). One dimensional (1D), two dimensional (2D), three dimensional (3D), or indefinitely dimensional (1D) data are all possible. RBF is used to select the location of the threshold that facilitates data splitting in infinite dimensions.

Radial basis function: The radial basis function, which serves as the SVM classification algorithm's default kernel, can be defined by the following formula:

$$k(x, x') = e^{-\gamma} \left\| x - x' \right\|^2 \tag{3.25}$$

$\|x - x'\|^2$ is the Euclidean distance between two feature vectors, where gamma can be set manually and has to be greater than zero.

A supervised learning approach known as a support vector machine, or SVM, can be applied to classification and regression issues. However, categorization issues are its main application. The objective of SVM is to construct a hyper-plane or decision boundary that can categorize datasets. The support vector machine (SVM), often known as a supervised learning method, is one of the most widely used. It can be used to address issues with classification and regression. Medical pictures, including x-rays, CT scans, and MRI scans, can be analyzed using SVM to identify them as either healthy or unhealthy and to forecast the risk of disease. SVM operates by identifying the ideal border or hyper-plane for dividing the various classes in the data. They are crucial in establishing the decision boundary because SVM looks for the decision border that maximizes the margin between positive and negative examples.

The Support Vector Machine (SVM) method attempts to produce the best decision boundary or line for categorizing an n-dimensional space. As a result, categorizing new data points in the future will be simpler for us. A hyperplane is a mathematical term for this optimal decision boundary.

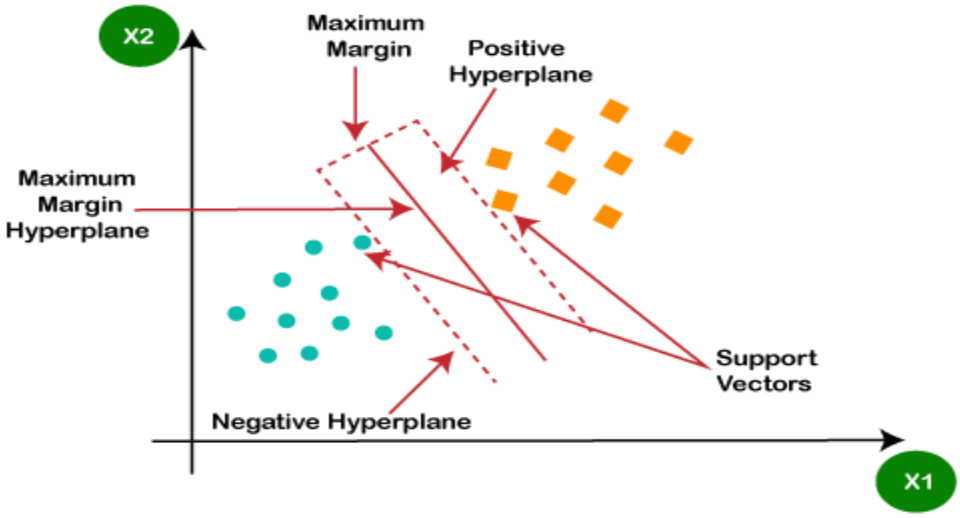


Figure 12: Algorithm of support vector machine.

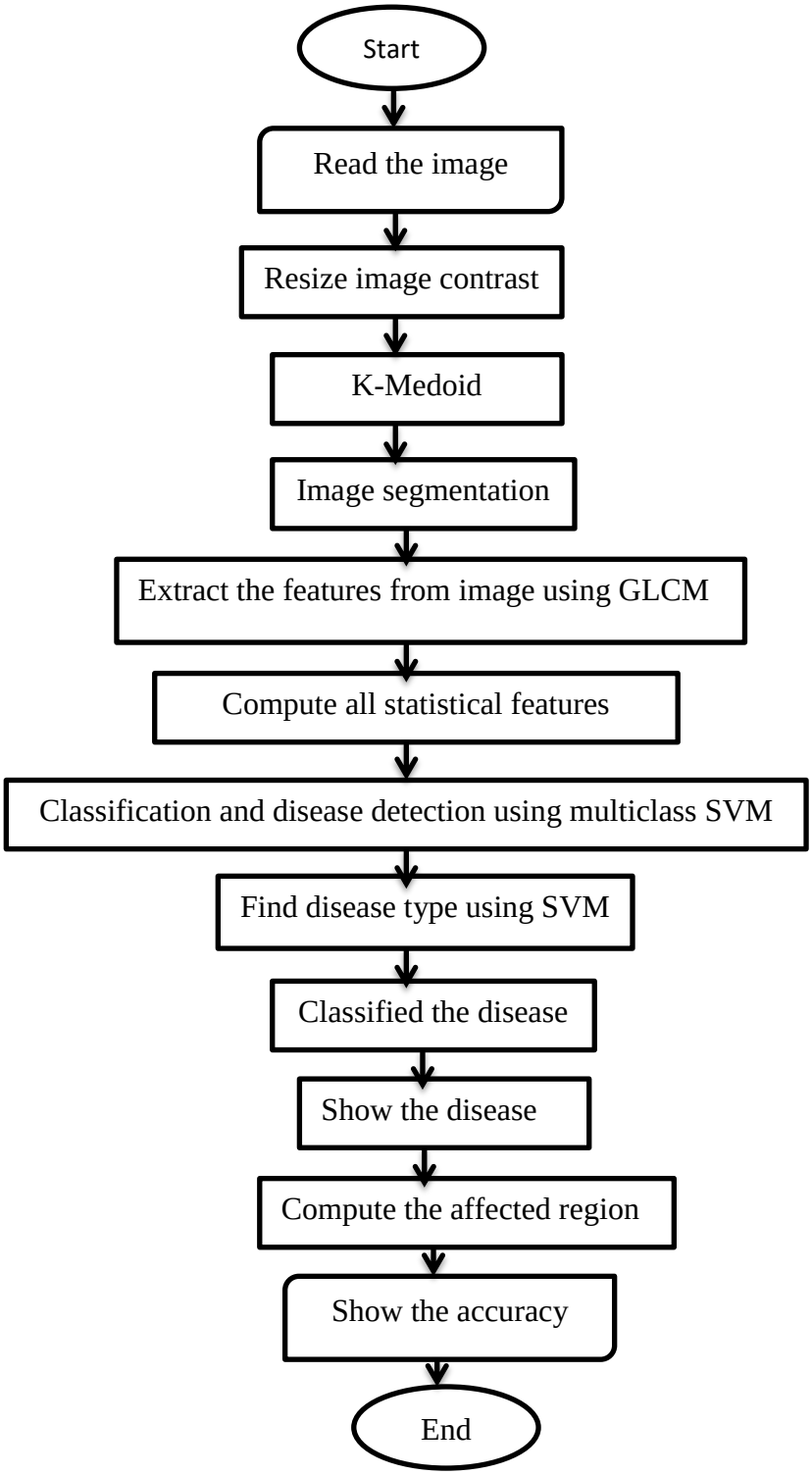


Figure 13: pea leaf disease detection and classification coding diagram.

CHAPTER FOUR

4. RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter gives the results of the Mathlab code simulation of the current and planned systems in order to meet the specific goal of the study. A classification structure was chosen based on what the employed plant diseases were expected to produce.

The current research uses K-Means clustering to identify and categorize diseases for segmentation and they takes 28 samples by k means as follow:

1. *Alternaria alternata* disease.

   Segmented by K-Means

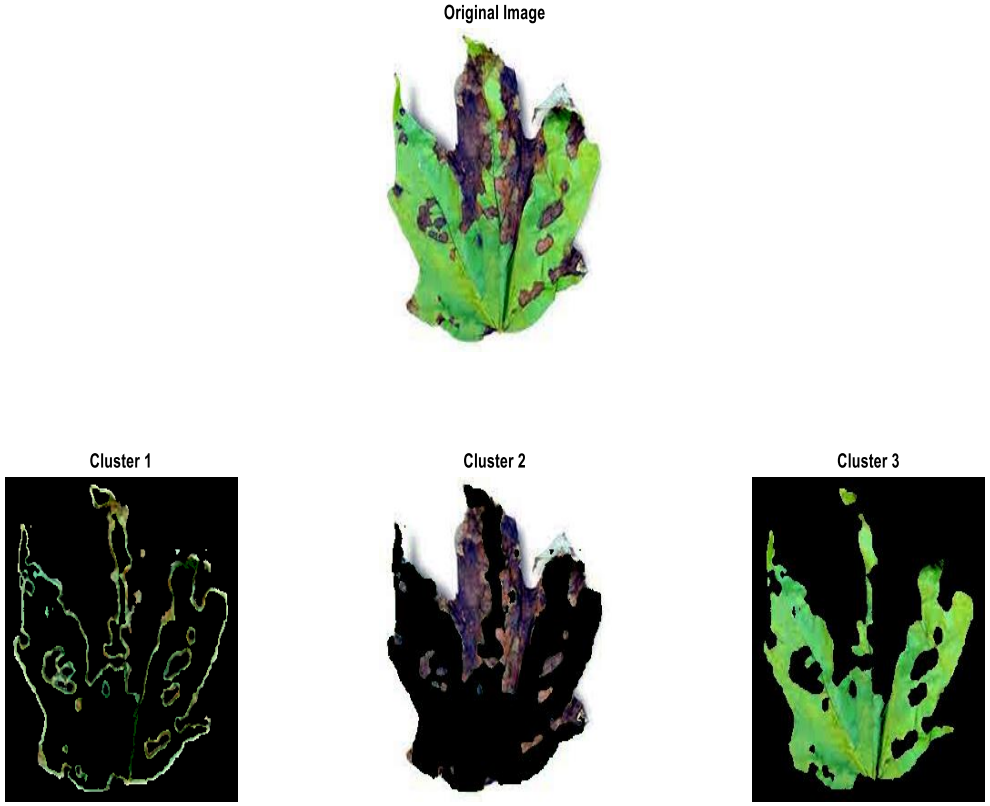
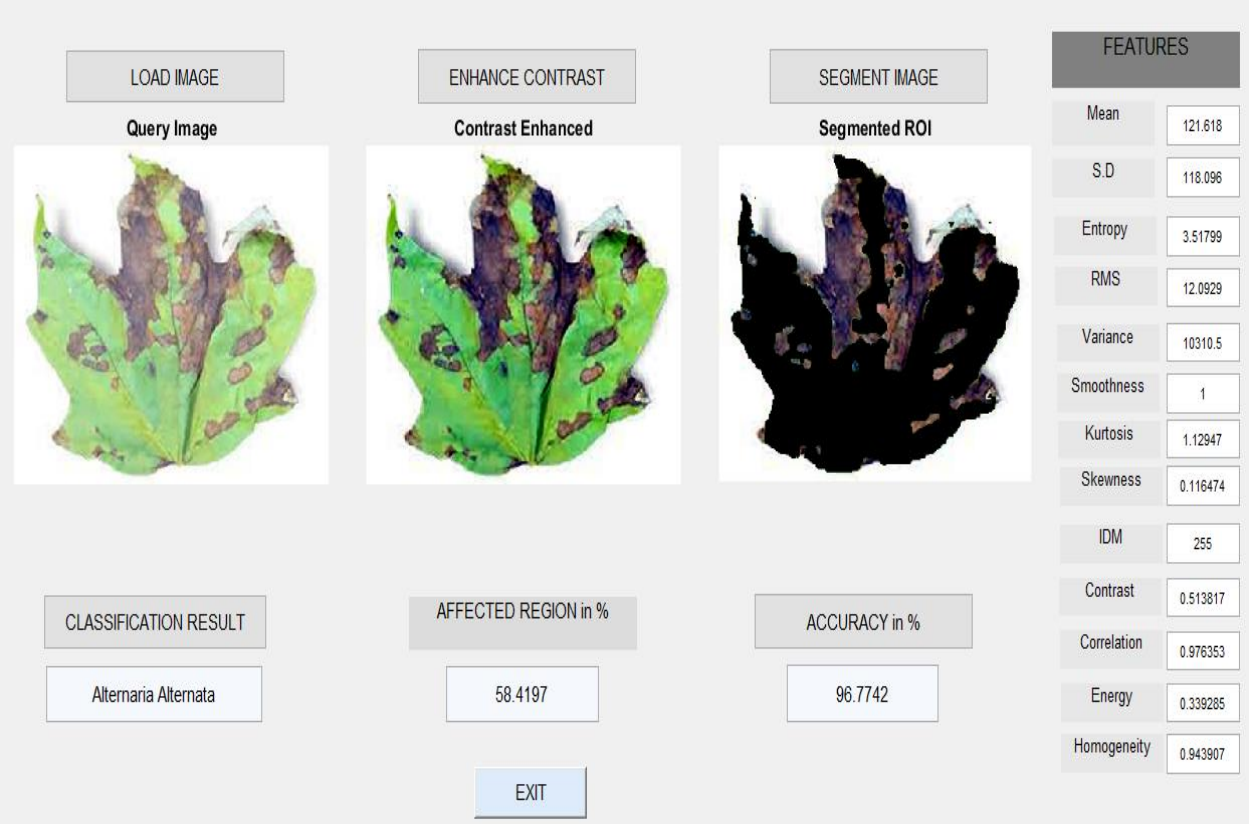


Figure 14: Segmented by K-Means cluster.

Figure 14 shows the segmented part of Alternaria alternata plant leaf diseases using k-means clustering technique studied by other related studies.



. Figure 15: The system shows alternaria alternata disease detection of plant leaf (Singh, 2017)

Figure 15 shows Alternaria alternata plant leaf diseases using k-means clustering technique by loading images from the dataset shows contrast enhancement, segmented region, feature values, classification results, affected region in percent, accuracy in percent, using GUI(graphic user interface) studied by other related studies.

2. *Cercospora leaf* spot disease

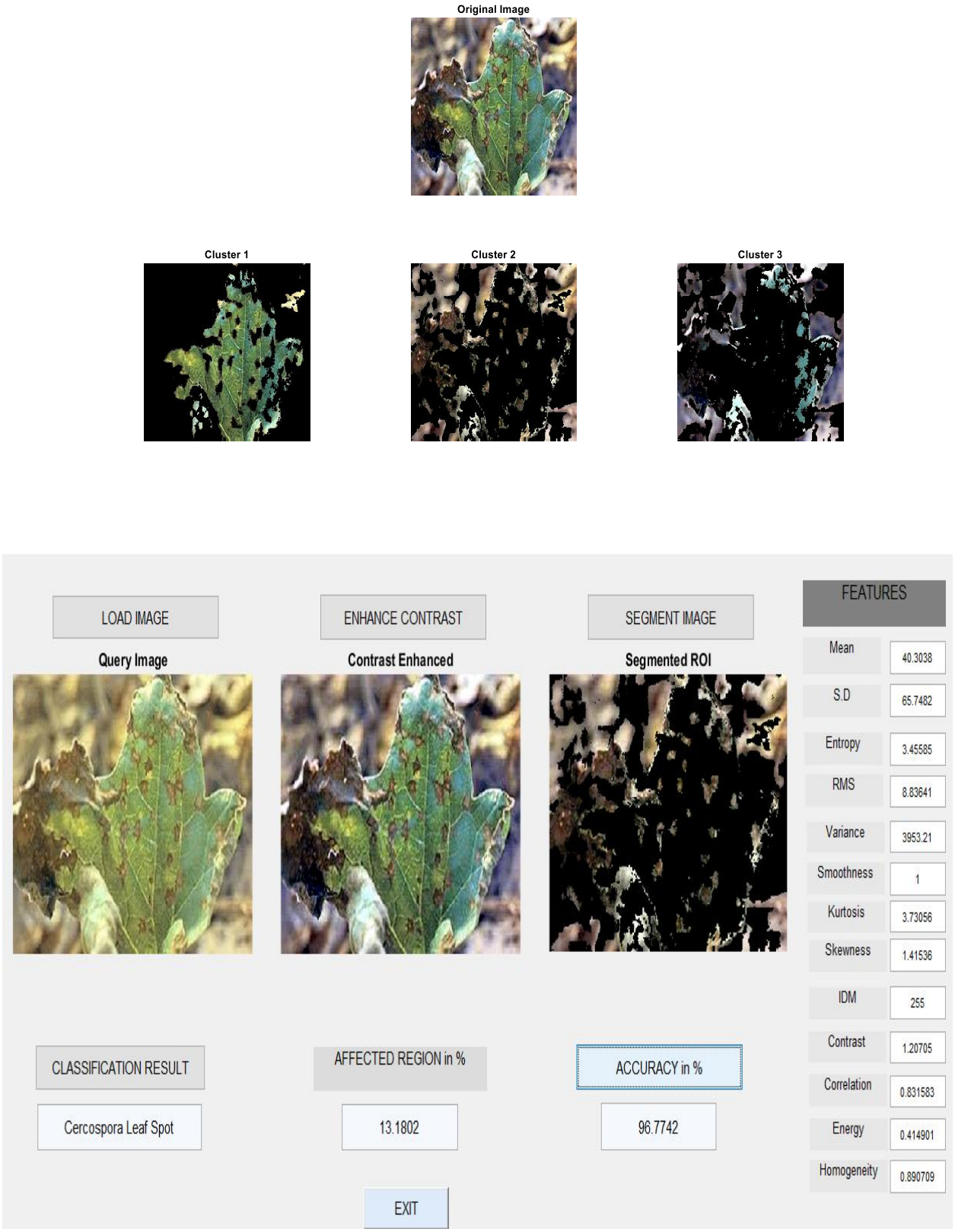


Figure 16: The system shows cercospora leaf spot disease detection (R, 2020).

Figure 16 shows the segmented part of *Cercospora leaf* spot disease plant leaf diseases using k-means clustering technique studied by other related studies and by loading the image dataset into the GUI it can possible shows contrast enhancement, segmented region, feature values, classification result, affected region in percent, accuracy in percent.

3. Bacterial blight disease

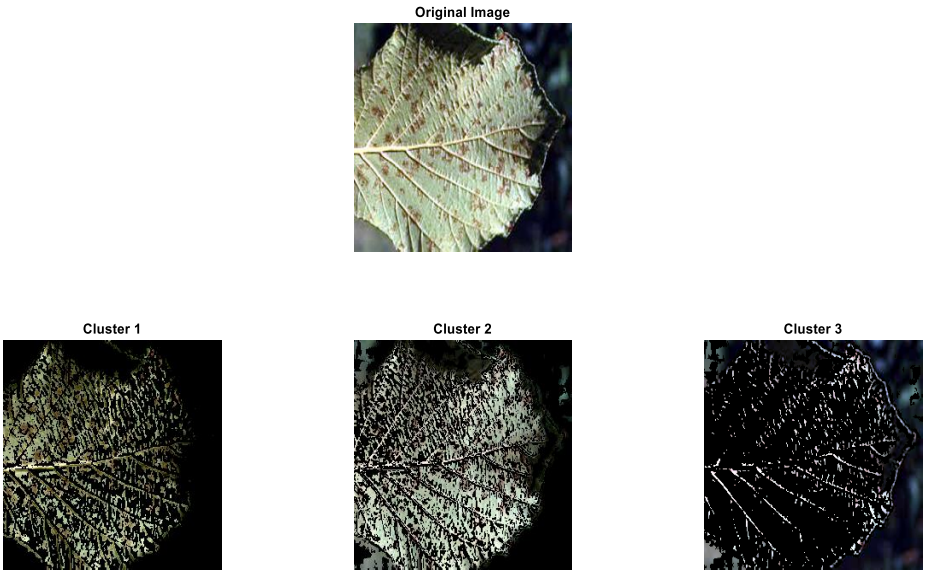


Figure 17: The system shows bacterial blight disease detection of a leaf (R, 2020).

Figure 17 shows the segmented part of bacterial blight disease plant leaf diseases using k-means clustering technique studied by other related studies and by loading the image dataset into the GUI it can possible shows contrast enhancement, segmented region, feature values, classification result, affected region in percent, accuracy in percent.

4. Healthy leaf

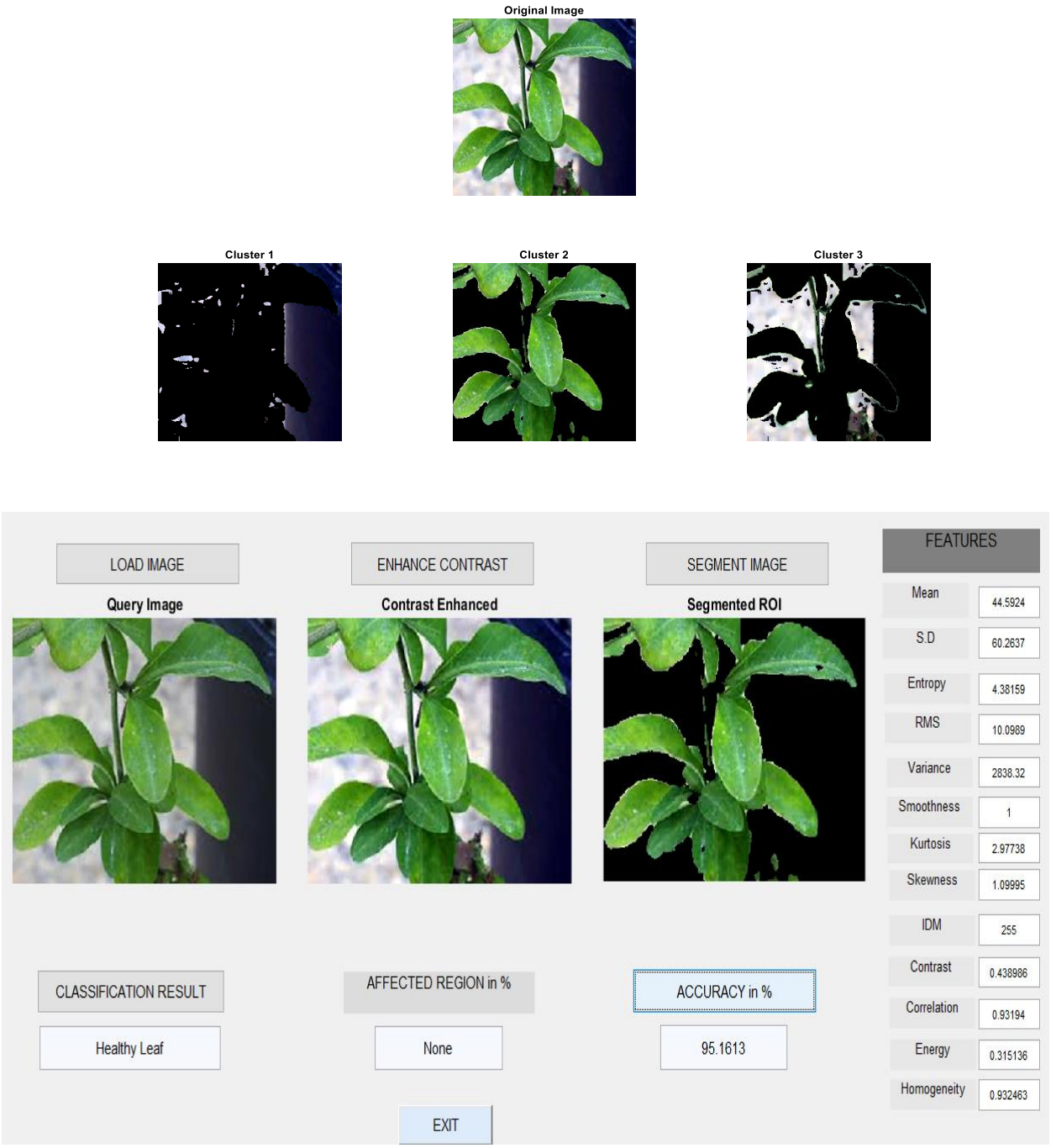


Figure 18: The system shows healthy leaf detection (R, 2020).

Figure 18 displays the portion of a healthy leaf that has been segmented using the K-means clustering technique, as investigated by previous studies in this field. By importing the image dataset into the graphical user interface (GUI), it becomes feasible to observe various features, including contrast enhancement, the segmented region, feature values, classification outcome, affected region in percentage, and accuracy in percentage.

The result of this thesis proposed work using K-Medoid segmentation technique with 2000 sample is shown as below: for this result improvement of accuracy using K-medoid instead of K-means.

I. Ascochyta pea leaf disease

Figure 19: Segmented by K-Medoid clustering

Figure 19 shows the segmented part of Ascochyta pea leaf disease using k-medoid clustering technique studied by this thesis. The main important of using k-medoid clustering is that it saves (reduces) the execution time on the same iteration and the best accurate.

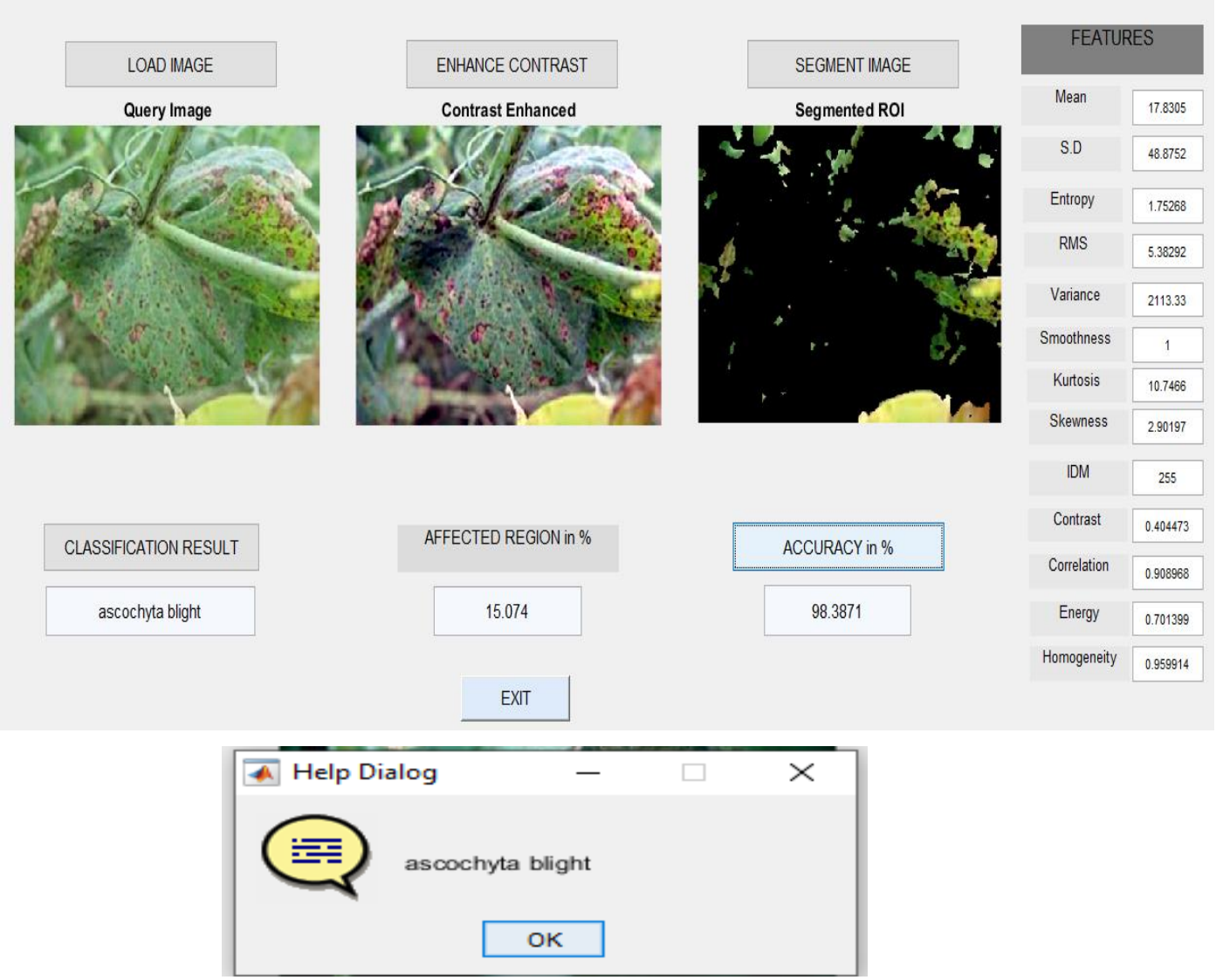


Figure 20: Ascochyta pea leaf disease detection segmented by K-Medoid.

Figure 20 displays the portion of Ascochyta pea leaf disease detection segmented by K-Medoid clustering technique, investigated by this thesis. By importing the image dataset into the graphical user interface (GUI), it becomes feasible to observe various features, including contrast enhancement, the segmented region, feature values, classification outcome, affected region in percentage, and accuracy in percentage.

II. Downy mildew pea leaf disease

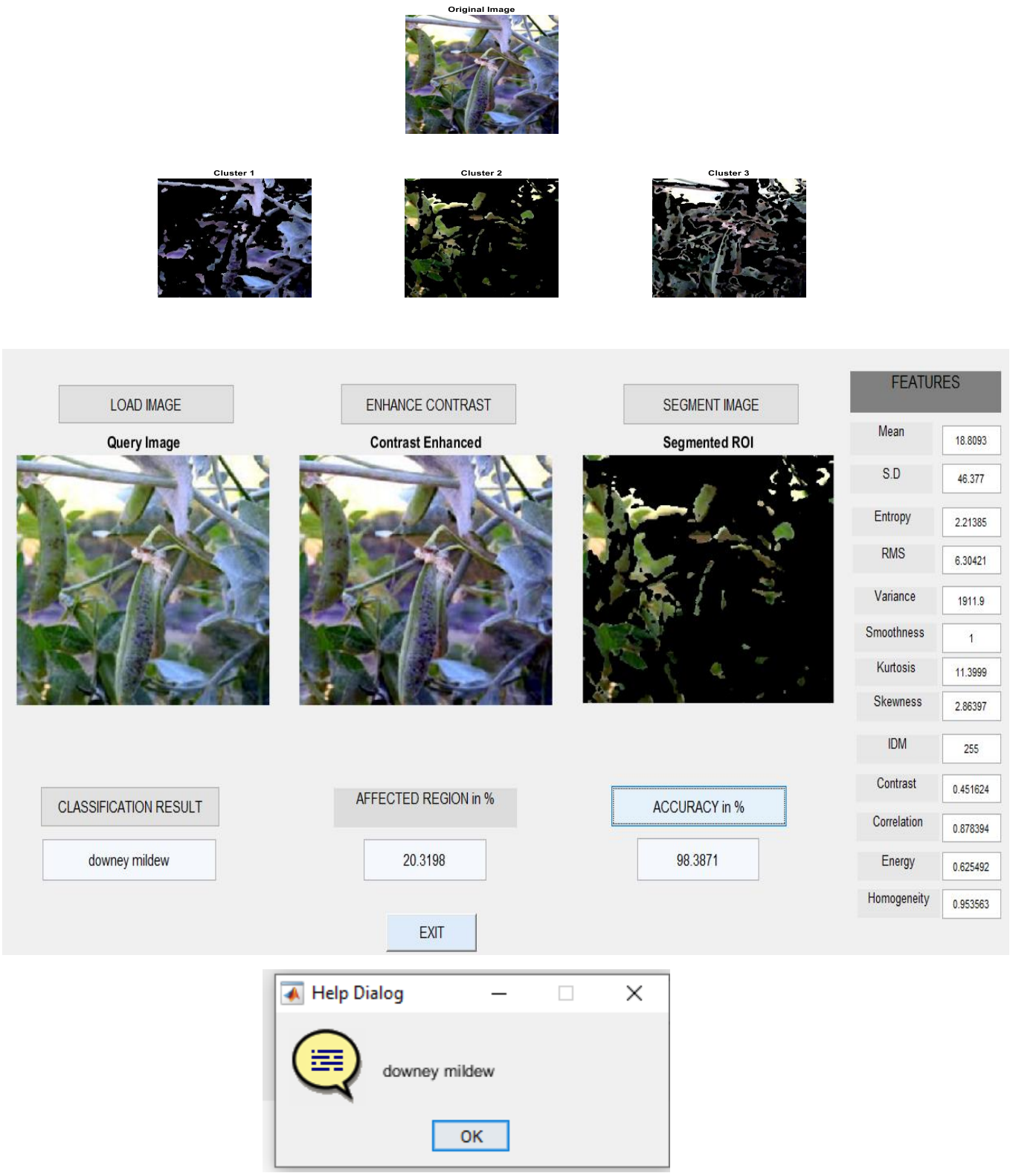
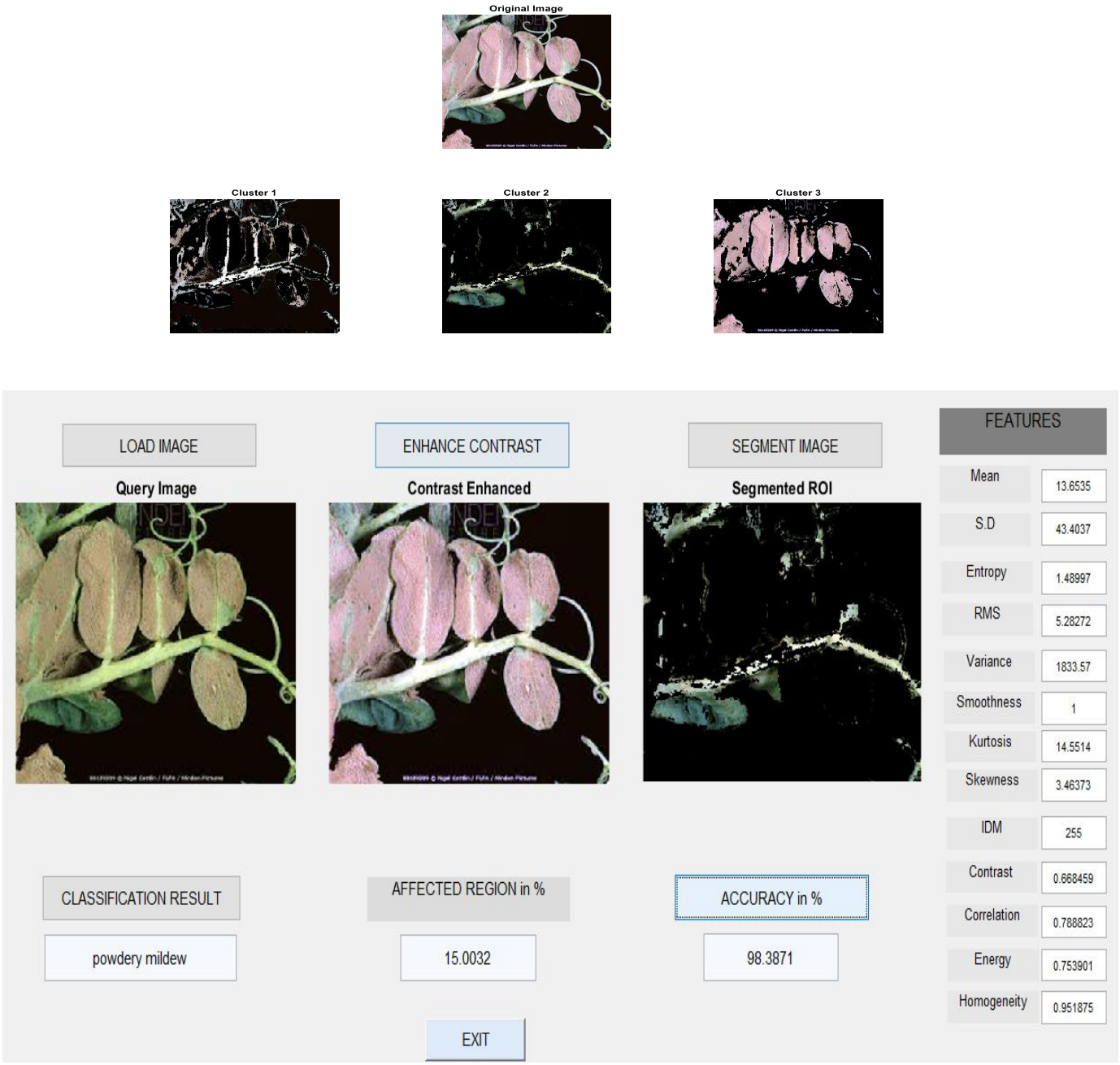


Figure 21: The system shows downey mildew pea leaf disease detection.

Figure 21 shows downey mildew pea leaf disease detection result by importing images from the dataset into GUI. There are different parameters displayed on the surface of GUI the contrast enhancement, the segmented part, the feature extraction parameters, the classification result, the affected portion of leaf in percentage and finally the accuracy in percentage with iteration 500.

III. Powdery mildew leaf disease



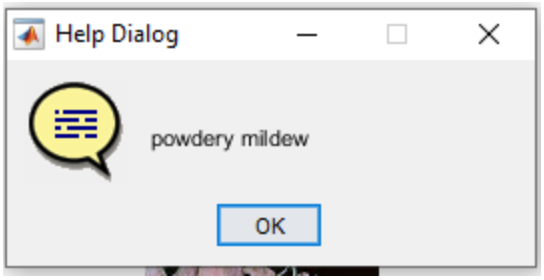
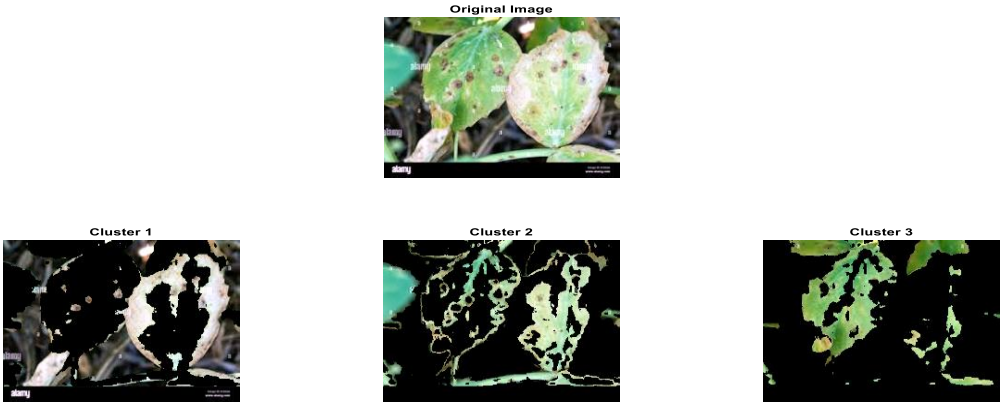


Figure 22: The system shows powdery mildew pea leaf disease detection.

The detection of powdery mildew pea leaf disease was demonstrated in Figure 22 using a graphical user interface (GUI). The GUI displayed various parameters including contrast enhancement, the segmented portion of the leaf, feature extraction parameters, classification results, the percentage of the leaf affected by the disease, and the accuracy achieved after 500 iterations. The GUI imported images from a dataset for analysis and provided a visual representation of the disease detection process. This allowed users to easily interpret and assess the results obtained from the analysis.

IV. Rust pea leaf spot



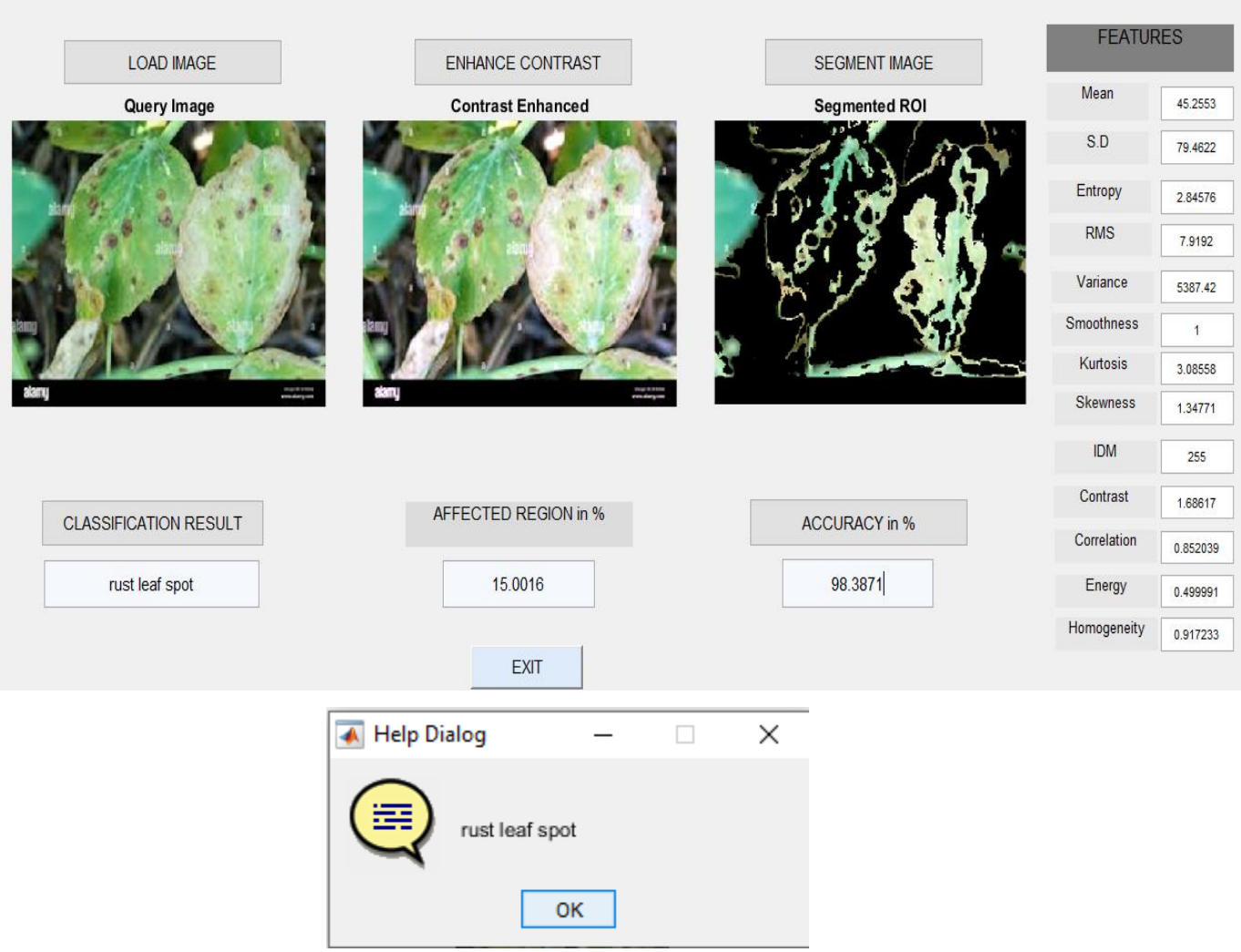


Figure 23: The system shows rust pea leaf spot disease detection.

Figure 23 demonstrates the detection of rust pea leaf spot disease using the K-Medoid clustering technique, as investigated in the thesis. The figure shows the segmented portion of the detected disease and was generated using a graphical user interface (GUI) that imported an image dataset. The GUI allowed for the visualization of various features, such as contrast enhancement, the segmented region, feature values, classification outcome, affected region in percentage, and accuracy in percentage. The use of the GUI made it possible to easily observe and analyze the results of the disease detection process.

V. Healthy pea leaf

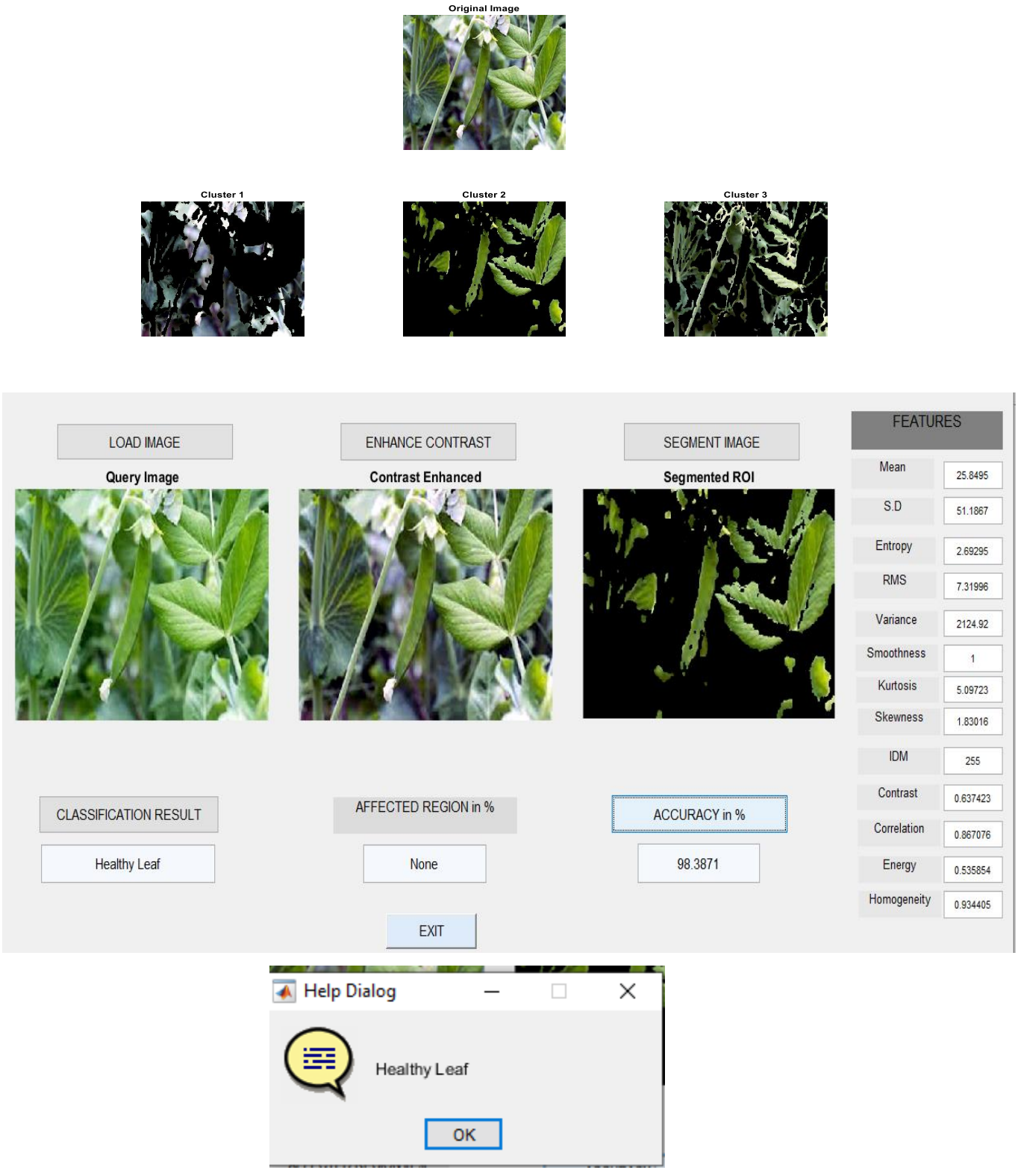


Figure 24: The system shows healthy pea leaf detection.

Figure 24 depicts the detection of healthy pea leaf using the K-Medoid clustering technique investigated in the thesis. The figure displays the segmented portion of the leaf that was identified as free from any disease, and it was generated using a graphical user interface (GUI) that imported an image dataset. The GUI provided various features for visualization, such as contrast enhancement, the segmented region, feature values, classification outcome, affected region in percentage, and accuracy in percentage. The use of the GUI made it easy to observe and analyze the results of the disease detection process and demonstrated the effectiveness of the K-Medoid clustering technique in identifying healthy pea leaf.

The following table illustrates how accurate the SVM classification method is.

Table 5: The results of the existing work's SVM classification accuracy.

Types of plant disease (classification result)	Affected region in %	SVM classification accuracy %
Alternaria alternate	58.4197	96.7742
Cercospora leaf spot	13.1802	96.7742
Bacterial blight	15.1067	96.7742
Healthy leaf	None	96.7742

Table 6: The proposed work's SVM classification accuracy using K-Medoid clustering.

Types of pea leaf (classification result)	Affected region in %	SVM classification accuracy %
Ascochyta blight	15.074	98.3871
Downey mildew	20.3198	98.3871
Powdery mildew	15.0032	98.3871
Rust leaf spot	15.0016	98.3871
Healthy leaf	None	98.3871

The above two tabular representation shows comparison between existing work accuracy of SVM classification using K-Means clustering and K-Medoid clustering. To get the above pro-

posed result there are so many testing trials done. The accuracy of linear kernel with 500 iteration is 98.3871%.

Table 7: The confusion matrix of testing.

Types of disease	Correctly classified or predicted	Incorrectly classified (predicted)
Ascochyta blight	80	320
Downey mildew	100	300
Powdery mildew	50	350
Rust leaf spot	100	300
Healthy leaf	100	100

This table shows during testing the categories of each disease types are incorrectly predicted and correctly predicted as we want.

4.2Statistical Feature Parametric Values

There are statistical feature parameter values displayed after testing the dataset or after execution of the code.

Table 8: Statistical feature value from tested dataset.

Features	**Ascochyta blight**	**Downey mildew**	**Powdery mildew**	**Rust leaf spot**	**Healhy leaf**
Mean	17.8305	18.8093	13.6535	45.2553	25.8495
S.D	48.8752	46.377	43.4037	79.4822	51.1867
Entropy	1.75268	2.21385	1.48997	2.84576	2.69295
RMS	5.38292	6.30421	5.28272	7.9192	7.31996
Variance	2113.33	1911.9	1833.57	5387.42	2124.92
Smoothness	1	1	1	1	1
Kurtosis	10.7466	11.3999	14.5514	3.06558	5.09723
Skewness	2.90197	2.86397	3.48373	1.34771	1.83016
IDM	255	255	255	255	255
Contrast	0.404473	0.451624	0.668459	1.68617	0.637423
Correlation	0.908968	0.878394	0.788823	0.852039	0.867076

Energy	0.701399	0.625492	0.753901	0.49991	0.535654
Homogeneity	0.959914	0.953563	0.951875	0.917233	0.934405

Contrast

Refers to the difference in brightness or color between the lightest and darkest part of an image. An image with high contrast has a large difference between the lightest and darkest parts implies that to have more visual interest, more vivid and can appear sharper , while an image lowest contrast has a small difference can appear flat and lacking in detail. To measure contrast in an image using statistical measure of such as the standard deviation of pixel values to identify edges in the image.

Correlation

Refers to a mathematical operation that involves comparing two images or a part of an image with a predefined template or kernel to determine their similarity. The correlation operation involves sliding the template or kernel over the image and computing the sum of the products of the corresponding pixel values in the template and the region of the image that it overlaps. This process is repeated for every location in the image resulting in a correlation map that shows how well the template matches each location in the image. The correlation operation is often used in image processing tasks such as object detection, pattern recognition, and image alignment.

Entropy

Is a measure of the randomness or uncertainty of the pixel values in an image. It is a statistical measure that quantifies the amount of information contained in an image or a signal. It is calculated based on the probability distribution of pixel values in an image. The entropy value is higher when the pixel values are more randomly distributed and lower when the pixel values are concentrated around a few values. Entropy is useful in image processing for tasks such as image segmentation and feature extraction. High entropy regions in an image may correspond to areas of interest, such as edges, textures, or objects, while low entropy regions may correspond to smooth or uniform areas. Entropy is often calculated using the Shannon entropy formula.

IDM (Inverse difference moment)

Is a texture feature used in image processing to quantify the spatial distribution of gray-level co-occurrence matrix (GLCM) values in an image. GLCM is a matrix that records the frequency of pairs of pixel values occurring together at a given offset or direction in an image. The inverse of

the sum of squared differences between the GLCM values and their mean value. It is a measure of the local homogeneity or uniformity of the texture in an image.

Energy

is a texture feature used to quantify the uniformity or smoothness of an image or a region of interest with in an image. It is a measure of the magnitude of the GLCM values and represents the sum of squared GLCM values.

Homogeneity

Is a texture feature used to quantify the uniformity or similarity of the gray-level values in an image or a region of interest with in an image. It is also known as inverse difference moment (IDM).

4.3 Discussion

Support vector machines are used to diagnose various pea leaf diseases using a graphical user interface (GUI) that can identify pea plant illness. The SVM is a popular machine learning method used for classification and regression analysis that works by finding a hyper-plane that separates different classes of data. The system has been complemented by a graphical user interface (GUI) to make the diagnosis of different types of pea leaf diseases more accessible to non-expert. When there are obvious signs of leaf disease, the classifier performs flawlessly, with the best accuracy rate of 98.3871%. The accuracy of the classification system depends on the quality and the size of the dataset used to train the SVM. However, It can be demonstrated that the Multiclass Support Vector Machine produces acceptable results when classifying more than two classes. From the above result table 5 is the existing work and table 6 is the proposed work. It is important for comparison of the two works. Table 7 shows during categorization of pea leaf in testing the dataset I have observe the incorrectly and correctly classification exist due to this reason confusion matrix is constructed. They are different statistical parametric values displayed during dataset testing and the disease classification result, affected region and accuracy also viewed on the excision body. To increase the accuracy of this proposed work by using K-Medoid clustering instead of K-Means and SVM is the best method used for classification purpose.

CHAPTER FIVE

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

The primary purpose is determining which leaf portions of pea fruits impact the illness or condition. To solve the problem of pea leaf I have been used machine learning from artificial intelligent. From the machine learning technique specially support vector machine used in this thesis work. The result that the SVM generates is superior in all of the other machine learning models. The SVM is used for classification purpose for this thesis work. There are many types of clustering techniques exist for segmentation and feature extraction. In the previous related work I have been observed they use K-Means clustering techniques by using this accuracy is 96.7742%. by reading the gaps I have improved the accuracy of this proposed work using another techniques of clustering. That is instead of K-Means in this thesis work K-Medoid clustering to improve the accuracy and saves the time of execution and also fast response to detect with best accuracy result to identify. The accuracy the linear kernel for 500 iteration of the proposed one is 98.3871%. Generally, no one done on the pea leaf detection this is one of the solution on the above starting mentioned problem. This thesis provides one of the solutions for our problems that are the reduction of pea crop starting from our society and over the country also.

5.2 Recommendation

In this work have been used machine learning SVM for future I have recommended to do by deep learning or combination of SVM with CNN is the better performance shift occurred. Future studies should focus on improving classification accuracy when identifying diseases in various pea leaf and other related plant species. In this work I want to show some automated based work related to the phone of farmer, but not done. For further studies the addition of hardware part is the best one. Pixel and resolution of the image also affects the accuracy or performance of the model for the future I recommended that best camera with fixed resolution of image during capturing the image.

Research Fund Acknowledgement

This thesis is funded by Adama Science and Technology University under the grant number ASTU/SM-R/805/23, Adama, Ethiopia.

6 REFERENCES

A. Cruz, Y. A. (2019). "Detection of grapevine yellows symptoms in Vitis vinifera L. with artificial intelligence,". *Computers and electronics in agriculture*, vol. 157, pp. 63-76.

A. K. Singh, S. K. (2021). "An Efficient Approach for Color-Based Image Retrieval Using RGB and HSI Color Spaces". *an efficient approach for color-based image retrieval using RGB and HSI color spaces which improves retrieval accuracy and computational efficiency*.

Abdullah, A. W. (2016). "Digital image processing analysis using Matlab.". *American Journal of Engineering Research (AJER)*, Vol. 5, No. 12, Pp. 143-147.

al., S. R. (2021). "A review of machine learning approaches for disease detection and diagnosis in plants". *published in Computers and Electronics in Agriculture.*

Arnoldi A., Z. C. (2015). *The role of grain legumes in the prevention of hypercholesterolemia and-hypertension-CRC*-Crit-Rev-Plant Sci.34, 144–168.-10.1080/07352689.2014.89790.

B. Zhang, J. C. (2021). "A Novel RGB-HSI Color Image Fusion Method Based on a Fusion Framework". *article introduces a novel RGB-HSI color image fusion method based on a fusion framework.*

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

Bhavini J. Samajpati, S. D. (2016). “Hybrid Approach for Apple Fruit Diseases Detection and Classification Using Random Forest Classifier”. *International Conference on Communication and Signal Processing* (pp. pp. 978-5090-0396). IEEE.

Brummer Y., K. M. (2015). Structural and functional characteristics of dietary fibre in beans, lentils, peas and chickpeas. *Food Res. Int.*, 67, 117–125.

Budiarianto Suryo Kusumo, A. H. (2018). Machine Learning based for Automatic detection of Corn-Plant Diseases Using Image . *International conference on computer, Control, Informatics and its Applications.* https://www.ijert.org/.

Das, A. A. (2017). "Leaf disease detection, quantification and classification using digital image processing. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation, and Control Engineering*, vol. 5, no. 11, pp. 45-50.

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

Devaraj, A. R. (2019). "Identification of Plant Disease using Image Processing Technique". *International Conference on Communication and Signal Processing (ICCSP).*

G. Hu, H. W. (2019). "A low shot learning method for tea leaf's disease identification,". *Computers and Electronics in Agriculture*, vol. 163, p.48-52.

Gope, A. T. (2021). "A Comparative Study of Clustering Techniques for Big Data Analysis in IoT Applications" . *comparative study of clustering techniques for big data analysis in IoT applications, including K-means, fuzzy C-means, and hierarchical clustering.*

H. Goitoom. (2019). Biological Control of Anthracnose of Sorghum using Trichoderma isolates under in vitro conditions. *Master's thesis in Social Science, Addis Abeba University, Addis Abeba Ethiopia, Unpublished,*.

H. Wang, P. Z. (2021). "A New K-Medoids Clustering Algorithm with Adaptive Distance Metric for Image Segmentation" . *Journal of Visual Communication and Image Representation*.

Hareem Kibriya, R. R. (2021). Tomato Leaf Disease Detection Using Convolution Neural Network. *18th International Bhurban Conference on Applied Sciences & Technology*. Islamabad, Pakistan: IEEE.

Husin, N. A. (2020). "Classification of basal stem rot disease in oil palm plantations using terrestrial laser scanning data and machine learning.". *Agronomy*, 10-11.

Kemal, S. A. (2015). Phenomics in crop plants: phenotypic methods of fungal.

Khandekar, M. S. (2019). A Review: Custard Apple Leaf Parameter Analysis and Leaf Disease Detection using Digital Image Processing. *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC)* (pp. ISBN: 978-1-5386-7808-4). IEEE Xplore Part Number: CFP19K25-ART; .

Kshirsagar, G. a. (2018). "Plant disease detection in image processing using MATLAB.". *International Journal on Recent and Innovation Trends in Computing and Communication*, 113-116.

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

Li, X. C. (2022). Deep learning for plant disease diagnosis. *A comprehensive review. Plant Disease*, (pp. Pp 106(1), 1-16).

M. K. Gayatri, J. J. (2015). “Providing Smart Agriculture Solutions to Farmers for better yielding using IOT,”. *International Conference on Technological Innovations in ICT for Agriculture and Rural Development.* IEEE.

MOA and NR. (2016). Plant Variety Release, Protection and Seed Quality Control Directorate, Crop Variety Register Issue No. 19. *Addis Ababa, Ethiopia*.

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

Mottaleb, M. A. (2021). Automatic detection of rice diseases using machine learning techniques. *A review. Plant Methods*, (pp. Pp 17, 1-23).

Mrunmayee Dhakate, I. A. (2015). “Diagnosis of Pomegranate Plant Diseases using Neural Network”. *IEEE* , (pp. pp. 978-1-4673-8564).

Nababan, Marlince, et al. (2018). "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*.

O, B. (2019). Comparing Yield Performance and Morpho-agronomic Characters of Landraces and Released Varieties of Field Pea (Pisum sativum L.) at Agarfa and Goro Woredas, Bale Zone, Oromia Region, Ethiopia. *International Journal of Genetics and Genomics*, D Vol. 7, No. 3 pp. 34-49. doi: 10.11648/j.ijgg.20190703.11.

Padmanabhuni, a. p. (2021). "An extensive study on classification based plant disease detection system.". *Journal of mechanics of continua and mathematical sciences*, Pp.5-15.

Prajakta Mitkal, P. P. (2016). “Leaf Disease Detection and Prevention Using Image processing using Matlab”. *International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 02,* .

Prakash M. Mainkar, S. G. (2015). “Plant Leaf Disease Detection and Classification Using Image Processing Techniques”. *International Journal of Innovative and Emerging Research in Engineering Volume 2, Issue 4, e-ISSN: 2394 – 3343,* p-ISSN:2394 – 5494.

Pujitha N, S. C. (2016). “Detection Of External Defects On Mango”. *International Journal of Applied Engineering Research ISSN*, 0973-4562 Volume 11, Number 7, pp. 4763-4769,.

R, R. (2020). Plant Monitoring and Leaf Disease Detection with Classification using Machine Learning-MATLAB. *International Journal of Engineering Research & Technology (IJERT).* Tamil Nadu, India: https://www.ijert.org/.

R. L. Christopher and P. Ramasamy. (2019). the Biology and Control of Sorghum Diseases, Kansas:. *American Society of Agronomy Crop Science Society of America Soil Science Society of America*.

Rastogi A, A. R. (2015). Leaf disease detection and grading using computer vision technology and fuzzy logic. *In:IEEE 2nd international conference on signal processing and integrated networks SPIN*, pp 500–505.

Reddy, Y. C. A. P., P. Viswanath, and B. Eswara Reddy. (2018). "Semi-supervised learning: A brief review". *International journal of engineering technology*, 50-60.

S. Koenig, P. N. (2022). International Conference on Robotics and Automation (ICRA). *International conferences.* IEEE.

S. Zhang, X. W. (2017). "Leaf image based cucumber disease recognition using sparse representation classification,". *Computers and electronics in agriculture*, vol. 134, pp. 135-141.

Sachin D. Khirade, A. B. Patil. (2015). "Plant Disease Detection Using Image Processing". *International Conference on Computing Communication Control and Automation,* (pp. pp. 978-1-4799-6892-3/15,). IEEE.

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

Sari, W. E. (2020). "Papaya Disease Detection Using Fuzzy Naïve Bayes Classifier.". *3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI).* IEEE.

Sethy, Prabira Kumar, et al. (2020). "Image processing techniques for diagnosing rice plant disease: a survey.". *Procedia Computer Science*, 167-170.

Shijie, P. a. (2017). "Automatic detection of tomato diseases and pests based on leaf images,". *Chinese Automation Congress (CAC)*, pp.2537-2510.

Shivani K. Tichkule, D. H. (2016). "Plant diseases detection using image processing techniques", . *Online International Conference on Green Engineering and Technologies (IC-GET).*

Singh, M. K. (2017). Detection and Classification of Plant Leaf Diseases in Image Processing using MATLAB. *International Journal of Life Sciences Research* .

Swati dewaliya, p. (2018). “Detection and Classification for apple fruit diseases using support vector machine and chain Code”. *International Research Journal of Engineering andTechnology (IRJET)e-ISSN*.

Teshome E, a. T. (2017). Comparative Study of Powdery Mildew (Erysiphe polygoni) Disease Severity and Its Effect on Yield and Yield Components of Field Pea (Pisum sativum L.) in the Southeastern Oromia, Ethiopia. *Journal of Plant Pathology and Microbiology, 8: 410. doi: 10.4172/2157-7471.1000410*.

Tsegay., H. Y. (2013). Characterization of dekoko (Pisum sativum var. abyssinicum) accessions by qualitative traits in the highlands of. *Southern Tigray, Ethiopia. African Journal of Plant Science, ,* Pp. 482-487.

Tyagi, V. (2018). “Understanding digital image processing”. *India CRC Press,*.

Verma, S. A. (2020). "Application of convolutional neural networks for evaluation of disease severity in tomato lant.". *Journal of Discrete Mathematical Sciences and Cryptography*.

X. E. Pantazi, D. M. (2019). "Automated leaf disease detection in different crop species through image features analysis and One Class Classifiers,". *Computers and electronics in agriculture,* vol. 156, pp. 96-104.

Y. Q. Xia, Y. L. (2015). “Intelligent Diagnose System of Wheat Diseases Based on Android Phone,”. *J. of Infor. & Compu. Sci,* vol.pp. 45-52.

Y. Zhang, L. L. (2020). "An Improved Mean Shift Clustering Algorithm for Image Segmentation". *an improved mean shift clustering algorithm for image segmentation*.

Y. Zhou, Y. Z. (2022). "A Hybrid K-Medoids Clustering Approach for Image Segmentation". *Pattern Recognition*.

Yimam, K. (2020). Diversity Analysis and Identification of Promising Powdery Mildew Resistance Genotypes in Field Pea (Pisum sativumL.). *American Journal of Biological and Environmental Statistics* , Vol. 6, No.1, pp.7-16.

Zare, M. &. (2021). A comprehensive review on machine learning-based plant disease detection using leaf images. Plant Disease., (pp. PP, 53-72.).