AUTOMATIC LEAF DISEASE DETECTION

USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

SUGANDAN S M	**921319106231**
VIGNESH KUMAR S	**921319106256**
VIJAYAKUMAR B	**921319106259**
VINUPRASATH S	**921319106264**

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

in

ELECTRONICS AND COMMUNICATION ENGINEERING

PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,

(An Autonomous Institution Affiliated to Anna University, Chennai)

DINDIGUL – 624622

APRIL 2023

PSNA COLLEGE OF ENGINEERING AND TECHNOLOGY,

(An Autonomous Institution Affiliated to Anna University, Chennai)

DINDIGUL-624622

BONAFIDE CERTIFICATE

Certified that this project report **"Automatic Leaf Disease Detection Using Machine Learning"** is the bonafide work of **SUGANDAN.S.M (921319106231), VIGNESHKUMAR.S(921319106256), VIJAYAKUMAR.B (921319106259),VINUPRASATH.S (921319106264)** who carried out the project work under my supervision.

SIGNATURE
Dr. G. ATHISHA, M.E., Ph.D.,
HEAD OF THE DEPARTMENT,
PROFESSOR
Department of ECE,
PSNA College of Engineering and Technology,
(An Autonomous Institution)
Dindigul – 624 622.

SIGNATURE
Mr . K. PALANIVEL RAJAN , M.E.,
SUPERVISOR
ASSISTANT PROFESSOR,
Department of ECE,
PSNA College of Engineering and Technology,
(An Autonomous Institution)
Dindigul – 624 622.

Submitted for the VIVA-VOCE Examination held on ______________

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

With warm hearts and immense pleasure, we thank the almighty for showering his blessings on us through this work. I would like to express our sincere thanks to (Late**) Thiru. R. S. KOTHANDARAMAN**, founder of our institution. We take this opportunity to express our sincere thanks to the respected Chairperson **Tmt. K. DHANALAKSHMI AMMAL**, who is the guiding light for all activities in our college.

We thank our Principal **Dr. D. VASUDEVAN, M.E., Ph.D.,** for being the beacon light in guiding every one of us to identify our strength successfully.

We heartly thank our Head of the Department of Electronics and Communication Engineering, **Dr. G. ATHISHA, M.E., Ph.D.**, for her constant support to us during the course of our project by being one of us. We heartily thank our Project Coordinator **Dr. M. REVATHY, M.E., Ph.D., ASSOCIATE PROFESSOR** Department of Electronics and Communication Engineering for her guidance offered during our project by being one of us.

We are extremely grateful to our Guide, **Mr. K. Palanivel Rajan, M.E, ASSISTANT PROFESSOR** Department of Electronics and Communication Engineering for his thoughtful comments and valuable suggestions which helped in moulding our project.

Last, we are extremely grateful to our parents and friends for their constant encouragement and moral support throughout our venture.

ABSTRACT

Automatic systems for disease detection in Farm land are expected to improve disease control, increase yield, and reduce pesticide application. Leaf Diseases are mainly caused by bacteria , fungi , virus etc . Our approach involves the segmentation of leaf images to isolate the leaf region from the background, followed by the extraction of features from the leaf region. Our results show that our approach is effective in detecting leaf diseases with high accuracy, making it a promising tool for disease management in agriculture. Our approach has the potential to be used in real-time applications, enabling farmers to take immediate action to prevent the spread of disease and minimize crop loss. Image Processing is one of the widely used technique is adopted for the plant leaf diseases detection and classification. This work is intended to aid in the detection and classification leaf diseases Alternaria alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot using Support Vector Machine (SVM) classification technique . First the diseased region is found using segmentation by K-means clustering , then both color and texture features are extracted. Finally classification technique is used to detect the type of leaf disease. The simulation results of this proposed work displays the Accuracy and Affected region of the leaf .

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

KNN	K-Nearest Neighbors
SVM	Support Vector Machine
GUI	Graphical User Interface
MATLAB	Matrix Laboratory
ROI	Region of Interest
RGB	Red Green Blue
CNN	Convolutional Neural Networks
HIS	Hue , Saturation , Intensity
OBM	Optimum Background Modeling
MRF	Markov Random Fields
PCA	Principal Component Analysis
GMM	Gaussian mixture models
MMA	Modified Moving Average

CHAPTER 1

INTRODUCTION

A product quality control is fundamentally required in order to gain more value added products. Many studies show that quality of agricultural products can be reduced from many causes. One of the most important factors of such quality is plant diseases. Consequently, minimizing plant diseases allows substantially improving quality of the products. Leaf diseases caused by various pathogens can severely affect the growth and yield of agricultural crops. This report focuses on the detection of leaf diseases caused by Alternaria alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot using an SVM Classifier .The study involves the collection of leaf images of infected and healthy plants from different crops. The leaf images are preprocessed, and relevant features such as texture, shape, and color are extracted using image processing techniques. The dataset is then split into training and testing sets, and an SVM Classifier is trained on the training set using the extracted features. The performance of the trained SVM Classifier is evaluated on the testing set, and hyperparameters are fine-tuned to optimize its performance. The results demonstrate that the SVM Classifier can accurately detect and classify the different types of leaf diseases .The use of an SVM Classifier can help in developing effective management strategies for these diseases in crops. The classifier's performance can be further improved by using more advanced image processing techniques and feature extraction methods. The findings of this study can be useful in developing automated systems for the detection and management of leaf diseases in crops. An abnormal condition that injures the plant or leads it to function improperly is called as a disease. This study aims to develop a prototype system to automatically detect and classify the Leaf diseases by using image processing technique as an alternative or supplemental to the traditional manual method.

1.1 Leaf Diseases

Disease damage to crops can greatly reduce yield. They are mainly caused by bacteria, viruses, or fungi . In most of the cases the diseases create visual symptoms, primarily creating spots or changing color on the leaf body, tip or stem of crops. The most common diseases of cultivated crops are Alternaria Alternata ,Bacterial Leaf Blight, Brown Spot, Narrow Brown Spot, Bacterial Leaf Streak, False Smut, Sheath Blight, Red Stripe, Stem Rot etc. For applying machine vision based disease recognition based on visual symptoms, this paper focuses on four diseases named Alternaria Alternata , Anthracnose, Bacterial Leaf Blight and Brown Spot.

1.1.1 Alternaria Alternata

Alternaria alternata is a fungal species that belongs to the genus Alternaria. It is a common saprophyte that can be found in soil, plants, and in the air. It is also known to be an opportunistic pathogen, causing infections in humans and animals. In humans, Alternaria alternata is associated with allergies, asthma, and respiratory tract infections. It produces spores that can cause allergic reactions in sensitive individuals. It is also known to cause cutaneous and subcutaneous infections, particularly in immunocompromised individuals.

Alternaria alternata is also an important plant pathogen, causing diseases in a wide range of crops, including tomatoes, potatoes, citrus fruits, and cruciferous vegetables. It can cause leaf spots, fruit rot, and blight, leading to significant economic losses in agriculture.

Alternaria alternata is a versatile and complex organism that has important implications for human health and agriculture.

Figure 1.1 Alternaria Alternata

1.1.2 Anthracnose

Anthracnose is a fungal disease that affects a wide range of plants, including trees, shrubs, and vegetables. The disease is caused by several species of fungi in the genus Colletotrichum and can result in significant economic losses in agriculture and forestry.

Anthracnose typically manifests as dark, sunken lesions on leaves, stems, and fruits. These lesions may be circular or elongated and may be accompanied by yellowing, wilting, and defoliation. In severe cases, the disease can kill the plant. The fungi that cause anthracnose can survive in plant debris and soil, and can be spread by wind, rain, and insects.

The disease is more common in humid, warm climates and can be exacerbated by poor plant nutrition and stress. Prevention and early detection are key to managing anthracnose and minimizing its impact on plants and crops.

Figure 1.2 Anthracnose

1.1.3 Bacterial Leaf Blight

Bacterial Leaf Blight is caused by Xanthomonas Oryza . It causes wilting of seedlings and yellowing and drying of leaves. The disease is most likely to develop in areas that have weeds and stubbles of infected plants. It can occur in both tropical and temperate environments, particularly in irrigated and rainfed lowland areas.

Figure 1.3 Bacterial Leaf Blight

1.1.4 Cercospora Leaf Spot

Cercospora leaf spot is a fungal disease that affects a wide range of crops, including soybeans, sugar beets, and tobacco. The disease is caused by the fungus Cercospora beticola, which infects the leaves of the host plant, causing circular to irregularly shaped lesions with a grayish center and reddish-brown margins. The disease can lead to significant yield losses if not properly managed. The fungus overwinters on infected plant debris, and spores are spread by wind, water, or equipment. The disease is favored by warm, humid conditions and can be more severe in areas with frequent rainfall or irrigation. Management of Cercospora leaf spot typically involves a combination of cultural practices, such as sanitation and crop rotation, and chemical controls, including fungicides.

Figure 1.4 Cercospora leaf spot

Table 1.1 Diseases and their symptoms

Disease Name	Pathogen Type	Host Plants	Symptoms
Alternaria alternata	Fungal	Tomatoes, potatoes, citrus fruits, cruciferous vegetables	Brown to black spots on leaves, stems, and fruits
Anthracnose	Fungal	Beans,peppers, cucurbits, tomatoes	Dark, water-soaked spots on leaves and fruits
Bacterial Blight	Bacterial	Beans, peppers, tomatoes, crucifers	Brown, water-soaked spots on leaves and fruits
Cercospora Leaf Spot	Fungal	Beans, soybeans, sugar beets, spinach	Yellow to brown spots with purple borders on leaves

1.2 Motivation

In today's digital age, it is important that the farmers get to use the latest technology for efficient management of their crops. The use of information access through mobile phones among the farmers has increased in recent years, which has made a positive impact on the output of the production. However, there is still a lacking of knowledge sharing between the farmers and the agriculture experts while it comes on a topic of proper crop management due to the challenges in training the farmers on a mass level on topics like disease identification and their management. As a result, in most of the cases the farmers rely on their experience and intuition for decision on identifying crop diseases and their treatments. The production might turn out not as expected if the symptoms are not treated in a proper manner, using appropriate amount of fertilizers guided by agriculture specialist. The motivation of this paper is to provide the farmers the access to a service which will connect them directly to the specialist for serving their needs on effective management of disease.

As a first step, Image processing techniques have been applied in this study for identification of diseases named Alternaria alternata, Anthracnose, Bacterial Leaf Blight, Brown Spot. It is possible to make this an autonomous system for disease identification and providing suggestions based on image analysis reports, but it is essential to govern the data by an expert so that the farmers get the best possible solution for their problems regarding cultivation.

1.3 Objective

There are three objectives to achieve in this project:

i. to develop the prototype of disease detection system

ii. to detect the disease by using image processing

iii. to apply image processing technique to analyze the pattern of disease.

1.4 SOCIAL RELEVANCE

Leaf diseases in crops can have significant social relevance due to their potential impact on food security and nutrition. Agricultural crops are a crucial source of food for humans and animals, and leaf diseases can lead to significant yield losses, affecting the availability and affordability of food. In developing countries, where agriculture is a major source of livelihood, crop losses can lead to poverty, food insecurity, and malnutrition, particularly among vulnerable populations, such as smallholder farmers and rural communities.

The economic impact of leaf diseases can be significant, as crops that are affected by these diseases can lose a substantial portion of their market value. For example, in tomato crops, early blight and late blight can cause yield losses of up to 50%, leading to significant revenue losses for farmers and traders. In citrus crops, greasy spot disease can cause yield losses of up to 60%, resulting in reduced income and employment opportunities for farmers and workers.

It is important that the farmers get to identify the condition of their crops well ahead of time before it is too late, in order to avoid any kind disaster that can be caused by the diseases. Accurate diagnosis and timely solving of leaf disease is thus a vital component of crop production management aiming for enhanced productivity leading to increased profits.

CHAPTER 2

LITERATURE SURVEY

1) Panuwat Mekha and Nutnicha Teeyasuksaet "Image Classification of Rice Leaf Diseases Using Random Forest Algorithm", May. 2021

The problem of rice diseases around the world make to damage and fall into a large number of rice. Caused by many of types, such as; fungi, Bacteria and Viruses. which are the main causes of rice disease affected to farmers. The classification of rice can be classified into several methods. In this research, image classification is used to classify the data set of rice leaf diseases, such as; Brown Spot Rice disease (BSR), Brown Spot Rice disease (BSR), Bacterial Leaf Blight disease (BLB), which is the rice leaf disease with severe outbreaks around Thailand. Moreover, image processing technology in the classification types of rice leaf disease, such as; Random forest classification algorithm, Decision tree classification algorithm, Gradient Boosting classification algorithm and Naïve-Baye classification algorithm, which is measured by the accuracy, precision and recall of each algorithms. The best result of performance in the image classification of rice leaf diseases is random forest algorithm equal to 69.44 percent.

2) Vandana Chaudhari and Manoj Patil "Banana leaf disease detection using K-means clustering and Feature extraction techniques", November. 2020.

In India, most of the people survive their life on farming. Farmers face many difficulties due to climate changes. One of them is loss in the yield and one of the reasons behind that is the diseases appear on the plant. Getting the expert eye necked opinion is not possible for all the farmers. There is need to recognize the disease in early stage using easiest way. Automated disease recognition can be done using image processing techniques and machine learning techniques. This work explains the automated system to identify the banana plant diseases by extracting color, shape and texture features. Support Vector Machine (SVM) Classification Techniques is used for classification of data. Proposed work showed the average accuracy of 85 % to identify four kinds of diseases as sigatoka, cmv, bacterial wilt and panama.

3) N. Krithika and A. Grace Selvarani "An individual grape leaf disease identification using leaf skeletons and KNN classification", February. 2018.

The most challenging process in agricultural applications is identification of leaf individually. In this paper, the classification of grape leaf diseases is proposed along with the leaf identification. Initially, the leaf skeletons are identified based on grape images. Since, the leaf skeletons are used for estimating the positions and directions of the leaves. The Tangential Direction (TD) based segmentation algorithm is proposed for retrieval of skeletons. If the grape leaf images are classified, then the histograms of H and a color channels are generated and the pixels values are observed to distinguish the healthy and diseased tissues. Then, extract the features and classify by using the KNN classification algorithm in order to find the leaf diseases.

4) G. Cerutti, L. Tougne, J. Mille, A. Vacavant, and D. Coquin, "Understanding leaves in natural images—A model-based approach for tree species identification,", Oct. 2013.

With the aim of elaborating a mobile application, accessible to anyone and with educational purposes, they present a method for tree species identification that relies on dedicated algorithms and explicit botany-inspired descriptors. Focusing on the analysis of leaves, they developed a working process to help recognize species, starting from a picture of a leaf in a complex natural background.

A two-step active contour segmentation algorithm based on a polygonal leaf model processes the image to retrieve the contour of the leaf. Features they use afterwards are high-level geometrical descriptors that make a semantic interpretation possible, and prove to achieve better performance than more generic and statistical shape descriptors alone. The results, both in terms of segmentation and classification, considering a database of 50 European broad-leaved tree species, and an implementation of the system is available in the iPhone application Folia.

5) R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods", Nov. 2012.

In this paper, they propose an effective superpixel-based saliency model. First, the original image is simplified by performing superpixel segmentation and adaptive color quantization. On the basis of superpixel representation, inter-superpixel similarity measures are then calculated based on difference of histograms and spatial distance between each pair of superpixels. For each superpixel, its global contrast measure and spatial sparsity measure are evaluated, and refined with the integration of inter-superpixel similarity measures to finally generate the superpixel-level saliency map. Experimental results on a dataset containing 1,000 test images with ground truths demonstrate that the proposed saliency model outperforms state-of-the-art saliency models.

Computer vision applications have come to rely increasingly on superpixels in recent years, but it is not always clear what constitutes a good superpixel algorithm. In an effort to understand the benefits and drawbacks of existing methods, they empirically compare five state-of-the-art superpixel algorithms for their ability to adhere to image boundaries, speed, memory efficiency, and their impact on segmentation performance. They then introduce a new superpixel algorithm, simple linear iterative clustering (SLIC), which adapts a k-means clustering approach to efficiently generate superpixels. Despite its simplicity, SLIC adheres to boundaries as well as or better than previous methods. At the same time, it is faster and more memory efficient, improves segmentation performance, and is straightforward to extend to superpixel generation.

6) A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," Dec. 2009.

A geometric-flow-based algorithm for computing a dense over segmentation of an image, often referred to as superpixels. It produces segments that, on one hand, respect local image boundaries, while, on the other hand, limiting under segmentation through a compactness constraint. The Berkeley database is used to quantitatively compare its performance to a number of over segmentation algorithms, showing that it yields less under segmentation than algorithms that lack a compactness constraint while offering a significant speedup over N-cuts, which does enforce compactness.

7) Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity", Apr 2004.

Objective methods for assessing perceptual image quality have traditionally attempted to quantify the visibility of errors between a distorted image and a reference image using a variety of known properties of the human visual system. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene, an alternative framework for quality assessment based on the degradation of structural information.

8) T. F. Chan and L. A. Vese, "Active contours without edges", Feb. 2001.

They propose a new model for active contours to detect objects in a given image, based on techniques of curve evolution, Mumford-Shah (1989) functional for segmentation and level sets. Our model can detect objects whose boundaries are not necessarily defined by the gradient. They minimize an energy which can be seen as a particular case of the minimal partition problem. In the level set formulation, the problem becomes a "mean-curvature flow"-like evolving the active contour, which will stop on the desired boundary. However, the stopping term does not depend on the gradient of the image, as in the classical active contour models, but is instead related to a particular segmentation of the image. They give a numerical algorithm using finite differences. Finally, They present various experimental results and in particular some examples for which the classical snakes methods based on the gradient are not applicable. Also, the initial curve can be anywhere in the image, and interior contours are automatically detected.

CHAPTER 3

METHODOLOGY

Methods are first based on analysis on white background and on the use of pairs of images in order to apply a background extraction process. Thereafter some method of segmentation of tree leaves are oriented towards the analysis on natural background based on a single image. There are in general two ways to describe object contours: explicit representations which are characterized by parameterized curves such as snakes and implicit representations, such as level sets, which represent a contour using a signed distance map. The level set representation is more popular than the explicit representation because it has a stable numerical solution and it is capable of handling topological changes. Active contour evolution methods are classified into three categories: edge-based, region-based, and shape prior-based.

3.1 EDGE-BASED METHODS:

Edge-based methods mainly consider the local information around contours, such as the grey level gradient. Kass et al. propose the snake model which is the best known edge-based active contour method. Caselles et al. propose a geodesic model which reflects more intrinsic geometric image measures than the snake, using the prior knowledge that the larger the gradient at a pixel, the higher the probability that the pixel belongs to an object's edge. Paragios and Deriche improve the geodesic model in using level sets to describe contours and using a gradient descent algorithm to optimize contours. The merits of edge-based methods are their simplicity, intuitiveness, and effectiveness for determining contours with salient gradient.

They have the following limitations:

a) They only consider the local information near to a contour, and thus the initial contour must be near to the object.

b) Contour sections lying in homogeneous regions of an image cannot be optimized.

c) They are of course sensitive to image noise.

3.2 REGION-BASED METHODS:

Region-based methods usually divide an image into object and background regions using statistical quantities, such as mean, variance, or histograms of the pixel values in each region. Chan and Vese approximate an image by a mean image with regions whose boundaries are treated as object edges. Zhu and Yuille present a statistical and variational framework for image segmentation using a region competition algorithm. Yilmaz et al. adopt the features of both object and background regions in the level-set evolution model. Mansouri proposes an algorithm for formulating contour tracking as a Bayesian estimation problem. For the region-based methods, prior knowledge of object color and texture may be incorporated into the contour evolution process. Color prior knowledge is usually represented using object appearance models such as color histograms, kernel density estimation, or Gaussian mixture models (GMMs). For example, Yilmaz et al. use kernel density estimation to model color features for estimating contours. Bibby and Reid use color histograms to model appearances and perform contour-based tracking at frame rates. Region texture features are usually modeled using the Gabor filer, the co-occurrence matrix, or Markov random fields (MRF), etc. For example, Sagiv et al. use Gabor features to perform textured image segmentation. Pons et al. fit an active contour using texture features which are extracted using unsupervised learning. Yilmaz et al. use the Gabor filter to model region texture features for determining contours. The merit of the region-based methods is that regions statistical information, together with the prior knowledge of object color and texture, can increase the robustness and accuracy of contour evolution. The limitation of the current region based methods is that the pixel values are treated as if they were independent for the posterior probability estimation. This independence assumption makes the obtained contour sensitive to disturbances caused by similarities of color or texture between the object and the background.

3.3 SHAPE PRIOR-BASED METHODS:

Shape prior-based methods statistically model object shape priors which are used to recover disturbed, occluded, or blurred contour sections. Leventon et al. project orthogonally a set of aligned training shape samples represented by the signed distance maps into a subspace using Principal Component Analysis (PCA). Paragios and Rousson construct a pixel-wise shape model in which local shape variability can be accounted for. Cootes et al. propose an active shape model for the different aspects of rigid objects in a shape prior formulation. Fussenegger et al. propose an online active shape model to perform region segmentation. The incremental PCA algorithm in is used to update the active shape model. Cremers proposes a linear dynamical shape model based on an autoregressive model for tracking a person with periodic motions using level sets. Yilmaz et al. propose a statistical method to learn object shape models which are used to recover occluded sections of a contour. Rathi et al. combine the particle filter with level set evolution. Occlusion is dealt with by incorporating shape information into the lights of the particles. Raviv et al. utilize the symmetry of rigid object shapes to deal with partial occlusions. The merit of the shape prior-based methods is that the disturbed, occluded, or blurred edges can be recovered. However, the current adaptive shape-based methods may distort undisturbed contour sections which can be found accurately using color features alone, while they globally recover the disturbed contour sections. In real world applications it is necessary to update the active shape model continuously in order to adapt to shape changes. However, the current method for updating the shape model does not simultaneously handle the multiple new shape samples, and fails to compute the sample eigen basis with sample mean updating. The previous dynamical shape model in for periodic motions of non-rigid objects is a simple data fitting process with no high-level understanding of shape changes. The model assumes that the underlying motion is closely approximated by a periodic motion, however human motion is rarely exactly periodic. Current contour-based tracking algorithms are subject to additional limitations as follows.

1) Tracking initialization often relies on a manually drawn closed contour around the object. Those methods, in which the boundaries of motion regions detected by background subtraction are the initial contours of moving objects such as in, are only effective for stationary cameras. In, the initial contour can be placed anywhere in the image, but it may take a long time to converge to the correct boundary. Quick and automatic initialization of contour tracking is still underdeveloped and demanding.

2) The existing level set-based tracking methods fail to track the contour of an object when the object moves abruptly. Related work in deals with the discontinuities induced by abrupt motion. However, the robust and effective handling of abrupt motion for contour-based tracking is still a difficult open problem.

Gradient based background subtraction

In this technique based on the gradient associated to each pixel the foreground is detected based on the Gaussian principle. Boundary of the image is obtained based on the neighbored ratio.

Advantages:

Real time applications can take advantage of this algorithm.

Disadvantages:

Location is used to detect the shadow which may not always be perfect method to achieve the needed functionality.

Intensity Information Based Approach

Standard deviation calculation is used to identify the intensity values at the point. Conditions are set for the shadowed pixel.

Advantages:

Mathematical calculation method of implementation will be mostly accurate to get the needed information.

Disadvantages:

Pixel intensity value is susceptible to illumination changes.

3.4 BACKGROUND MODELING:

1) Initial Background Model: The modified moving average (MMA) is used to compute the average of frames 1 through K for the initial background model generation. For each pixel (x, y), the corresponding value of the current background model Bt (x, y) is calculated using the formula as follows This is accomplished by making appropriate use of MMA which holds only the last background model Bt (x, y) and the current incoming video frame It (x, y) during the calculation procedure.

2) Optimum Background Modeling: In order to expeditiously determine background candidates, emphasis is placed on the first rapid matching phase in optimum background modeling (OBM). Unstable signals result only occasionally and indicate

the appearance of moving objects. For the optimum background pixel estimation in the video sequence, the main objective of OBM is to extract the stable signal of the incoming frame in the video sequence as shown.

It consists of the following steps:

1) immediate determination of background candidates via the rapid matching procedure;

2) use of the stable signal trainer in order to provide a measure of temporal activity of the pixels within the set of background candidates;

3) determination of the optimum background pixels via the accurate matching procedure.

3.5 RAPID MATCHING:

This procedure is used to quickly find a great quantity of background candidates by determining whether or not their respective pixel values for the incoming video frame $I_t(x, y)$ are equal to the corresponding pixel values of the previous video frame $I_{t-1}(x, y)$. If the values correspond, it indicates good candidate selection for the following stable signal trainer.

CHAPTER 4

PROPOSED SYSTEM

This project is intended to aid in the detection and classification of leaf diseases by SVM classification technique. First the diseased region is found using segmentation by Genetic K-means clustering, then both color and texture features are extracted. Finally classification technique of SVM is used to detect the type of leaf disease.

4.1 Image Acquisition

The images of the leaf is captured through the camera. This image is in RGB (Red, Green And Blue) form. Color transformation structure for the RGB leaf image is created , and then, a device-independent color space transformation for the color transformation structure is applied.

4.2 Image Pre-processing

To remove noise in image or other object removal, different pre-processing techniques is considered. Image clipping i.e cropping of the leaf image to get the interested image region. Image smoothing is done using the smoothing filter. Image enhancement is carried out for increasing the contrast. The RGB images into the grey images using color conversion using equation (1).

$$f(x)=0.2989*R + 0.5870*G + 0.114.*B \text{ - - - - - - - - - - - - - } (1)$$

Then the histogram equalization which distributes the intensities of the images is applied on the image to enhance the plant disease images. The cumulative distribution function is used to distribute intensity values .

4.3 Image Segmentation

Segmentation means partitioning of image into various part of same features or having some similarity. The segmentation can be done using various methods like otsu' method, k-means clustering, converting RGB image into HIS model etc.

1) **Segmentation using Boundary and spot detection algorithm**:

The RGB image is converted into the HIS model for segmenting. For boundary detection the 8 connectivity of pixels is consider and boundary detection algorithm is applied.

2)**K-means clustering**:

The K-means clustering is used for classification of object based on a set of features into K number of classes. The classification of object is done by minimizing the sum of the squares of the distance between the object and the corresponding cluster.

The algorithm for K –means Clustering:

1. Pick center of K cluster, either randomly or based on some heuristic.
2. Assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center.
3. Again compute the cluster centers by averaging all of the pixels in the cluster.

Repeat steps 2 and 3 until convergence is attained.

3) **Otsu Threshold Algorithm**:

Thresholding creates binary images from grey-level images by setting all pixels below some threshold to zero and all pixels above that threshold to one. The Otsu algorithm defined in is as follows:

i) According to the threshold, Separate pixels into two clusters

ii) Then find the mean of each cluster.

iii) Square the difference between the means.

iv) Multiply the number of pixels in one cluster times the number in the other The infected leaf shows the symptoms of the disease by changing the color of the leaf.

Hence the greenness of the leaves can be used for the detection of the infected portion of the leaf. The R, G and B component are extracted from the image. The threshold is calculated using the Otsu's method. Then the green pixels is masked and removed if the green pixel intensities are less than the computed threshold.

4.4 Feature Extraction

Feature extraction plays an important role for identification of an object. In many application of image processing feature extraction is used. Color, texture, morphology, edges etc. are the features which can be used in plant disease detection. Color co-occurrence Method : In this method both color and texture are taken into account to get an unique features for that image. For that RGB image is converted into the HSI translation. The proposed system architecture is shown in Figure.

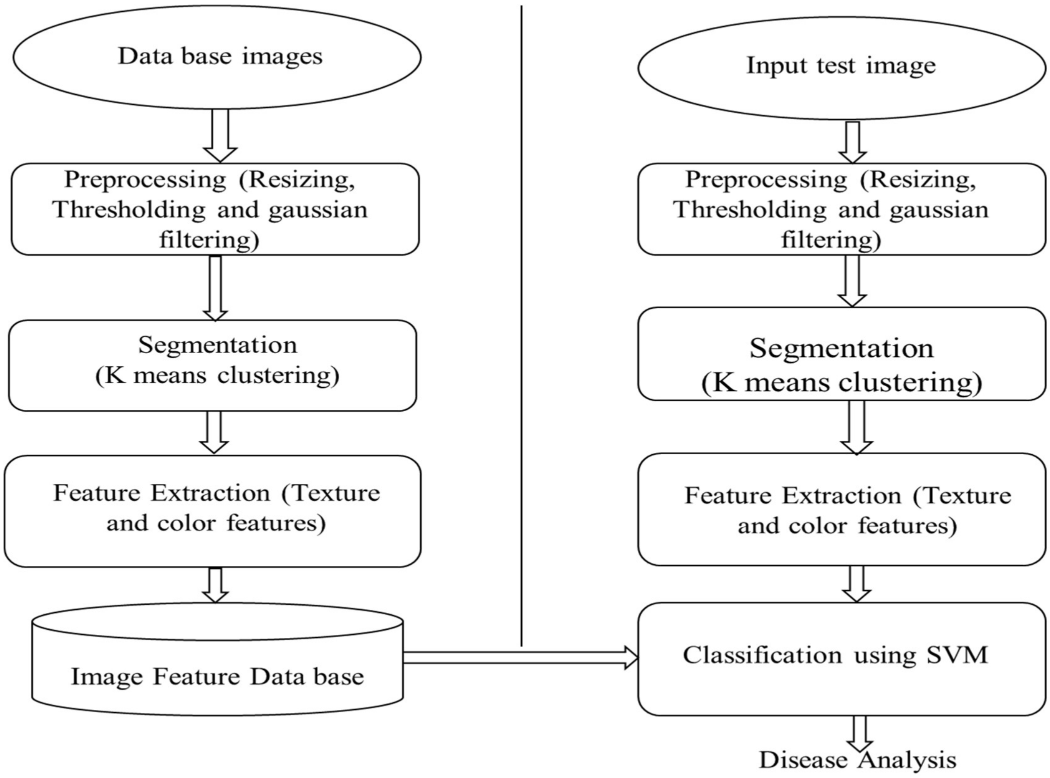


Figure 4.1 System Architecture

4.6 SVM

The information focuses or vectors that are the nearest to the hyperplane and which influence the situation of the hyperplane are named as Support Vector. Since these vectors uphold the hyperplane, thus called a Support vector.

SVM goes under the sort of managed AI calculation that gives an investigation of information for characterization and relapse examination. During a similar time, they can be utilized for relapse, SVM is generally utilized for order. We do plotting in the n-dimensional space. Estimation of each component is same as the estimation of the particular arrange. At that point, we discovering ideal hyperplane has the effect between the two classes.
These help vectors are the facilitate portrayals of explicit perception. It is an outskirts technique for isolating the two classes.

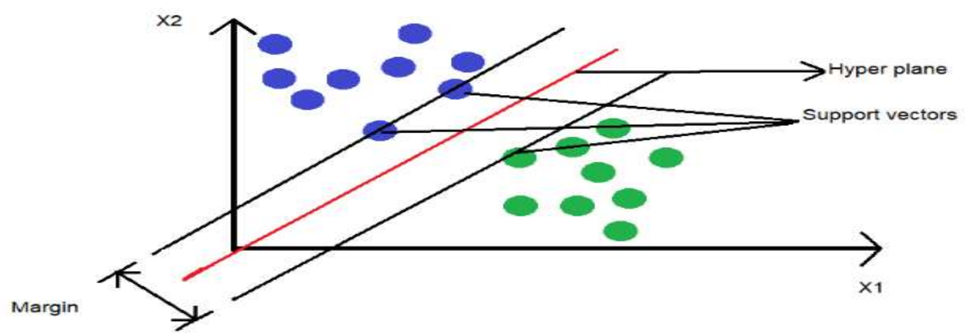


Figure 4.2 Linearly Non-Separable Samples Indicated in a Hyperplane

The essential rule behind the working of Support vector machines is direct – Create a hyperplane that separates the dataset into classes. Let us start with an example issue. Assume, for a given dataset, you need to classify red triangles from blue circles. Your point is to make a line that characterizes the information into two classes, making a contrast between red triangles and blue circles.

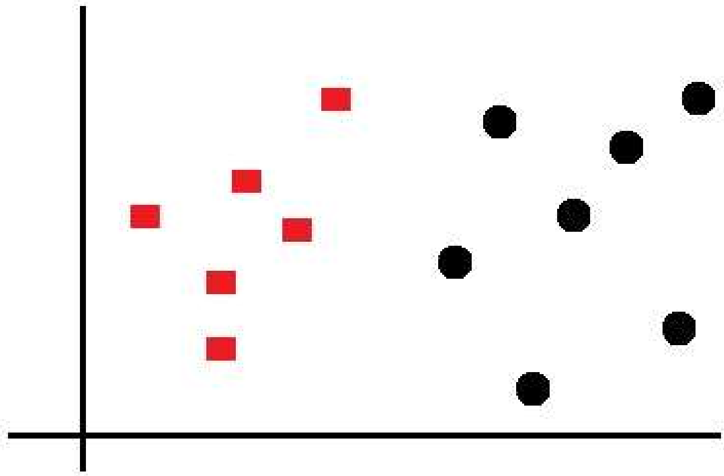


Figure 4.3 Dataset into classes

At some time one can envision a reasonable line that isolates the two classes, there can be numerous lines can capable carry out this responsibility. Accordingly, there is certifiably not a solitary line that you can concur on which we can play out this errand. Let us consider a portion of the lines that can have any kind of effect between the two classes as follows –

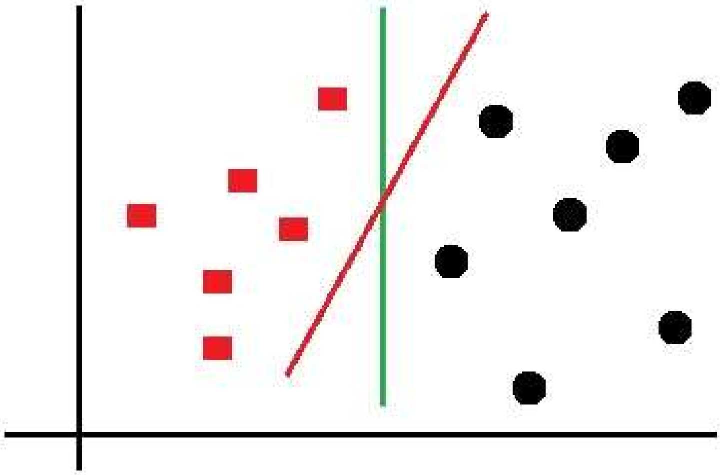


Figure 4.4 Clear partition between two classes

In the above picture, we have a green line and a red line. Which one improve separate the information into two classes? In the event that you pick the red line, at that point it is the ideal line that segments the two classes appropriately. Nonetheless, we actually have not represented the way that it is the widespread line that would group our information most effectively.

Multiclass Classification using Support Vector Machine :

In its most simple type SVM are applied on binary classification, dividing data points either in 1 or 0. For multiclass classification, the same principle is utilized. The multiclass problem is broken down to multiple binary classification cases, which is also called *one-vs-one*. In scikit-learn one-vs-one is not default and needs to be selected explicitly (as can be seen further down in the code). *One-vs-rest* is set as default. It basically divides the data points in class x and rest. Consecutively a certain class is distinguished from all other classes.

The number of classifiers necessary for *one-vs-one multiclass classification* can be retrieved with the following formula (with n being the number of classes):

$$\boxed{\frac{n*(n-1)}{2}}$$

In the one-vs-one approach, each classifier separates points of two different classes and comprising all one-vs-one classifiers leads to a multiclass classifier.

The SVM classifier has several advantages over other classification algorithms. It is effective in handling high-dimensional data, as well as noisy and incomplete data. It is also efficient in both training and testing phases, making it ideal for large datasets. SVMs are a type of supervised learning algorithm that can be used for both classification and regression analysis. They work by finding the hyperplane in a high-dimensional space that maximally separates the classes. SVMs are particularly useful when the data is linearly separable or can be transformed into a linearly separable space. SVMs are also known to be computationally efficient and have a low risk of overfitting. On the other hand, CNNs are a type of neural network that is especially useful for image classification and recognition. They work by using a convolution operation to extract features from the input image and then passing those features through a series of layers to arrive at a final classification. CNNs are particularly powerful because they can learn complex spatial hierarchies of features, which makes them well-suited for image recognition tasks. In general, CNNs tend to perform better than SVMs on image classification tasks, especially when dealing with large datasets with complex features. However, SVMs can be more effective than CNNs when the dataset is small or when the features are relatively simple. SVMs can also be more interpretable than CNNs, as they can identify the most important features in the data and assign weights to them.

In conclusion, the SVM classifier is a powerful and flexible algorithm for classification tasks. It provides an accurate and efficient way to classify data and has been successfully applied in a wide range of applications.

Table 4.1 Comparison for K-means , KNN and SVM algorithms

Algorithm	Strengths	Weaknesses	Suitable for
SVM	Effective for high-dimensional and complex data	Minimizes overfitting through regularization	Computationally expensive for large datasets Classification tasks with high-dimensional and complex data
KNN	Simple and easy to implement	Computationally expensive for large datasets	The choice of k can have a significant impact on performance. Classification tasks involving finding the similarity between data points
K-means	Simple and computationally efficient	Can handle large datasets. Requires a pre-determined value for k	Suitable only for unsupervised clustering tasks . Clustering tasks to group data points based on their similarity.

CHAPTER 5

SYSTEM ANALYSIS

5.1 Hardware and Software Specifications :

Hardware :

Processor : Intel® Core™ i3

Processor : 3.00GHz

RAM : 8 GB

Hard Disk : 916 GB

Software :

Operating System : Windows 11

Technology used : MATLAB

Version : Matlab 2014a

5.2 MATLAB

MATLAB is a widely used high-level programming language and interactive environment for numerical computation, visualization, and programming. It is an essential tool in various fields such as engineering, physics, mathematics, finance, and many others.

MATLAB offers a rich set of built-in functions and tools for data analysis, signal processing, optimization, and modeling. It also has a vast user community that develops and shares add-ons, toolboxes, and scripts that extend its capabilities.

MATLAB is a powerful tool for image processing due to its rich set of functions and toolboxes. The basic image processing operations, including image filtering, enhancement, restoration, segmentation, and feature extraction. MATLAB is an excellent tool for image processing that enables researchers and practitioners to tackle complex image processing tasks efficiently and effectively.

5.3 Implementation

The system consists of a GUI, which will enable the farmers to take images of crop leaf using their mobile phones and send it where the central system will analyze the pictures based on visual symptoms using image processing algorithms in order to measure the disease type. An expert group will be available to check the status of the image analysis data and provide suggestions based on the report and their knowledge, which will be sent to the farmer.

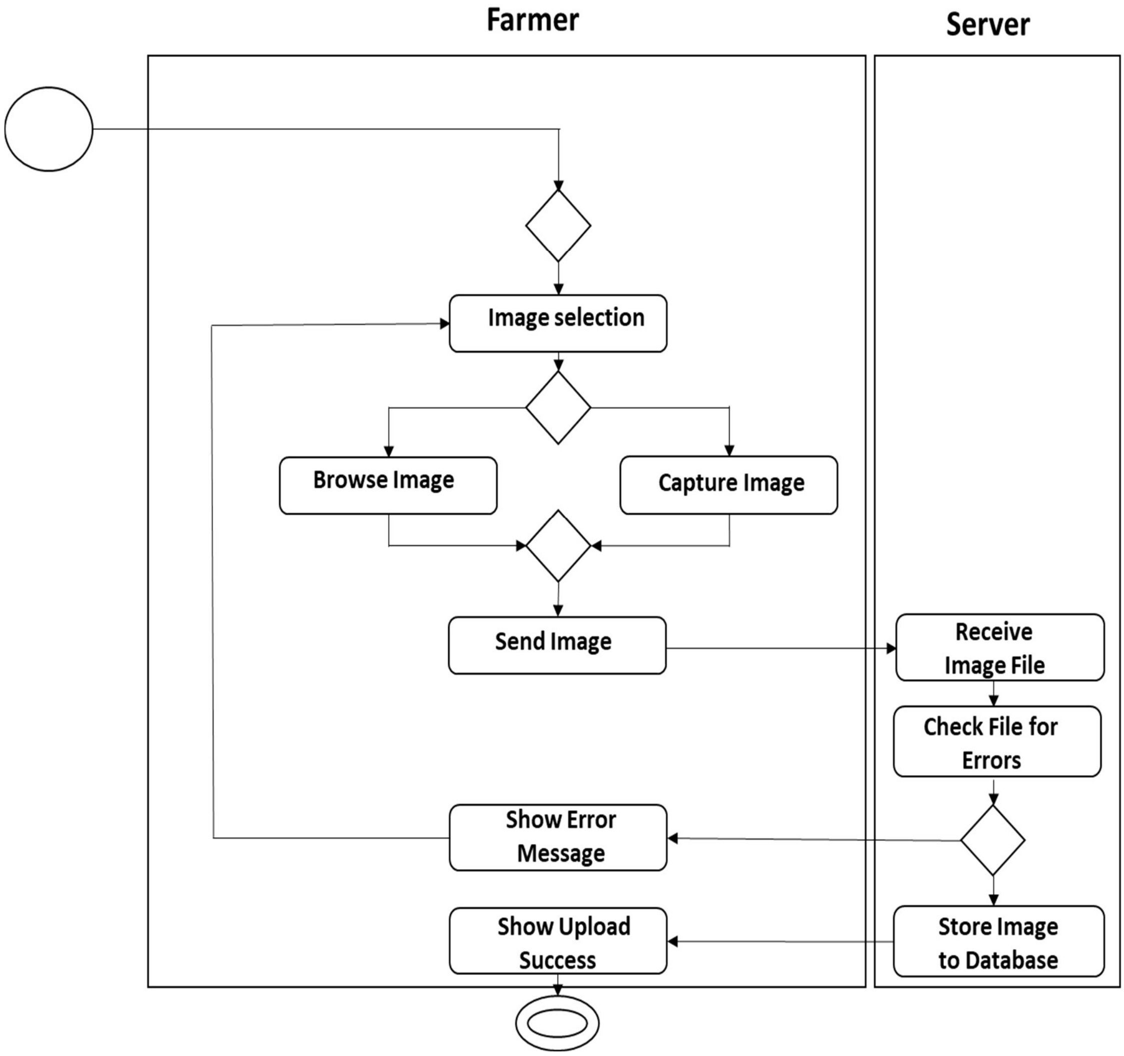


Figure 5.1 Functional diagram

The proposed application consists of 5 basic functionalities. They are 1) Image capture, 2) Image selection, 3) Image zoom and crop, 4) Share image with expert group 5) Receive the accuracy and affected region . At the very first page the application bar shows the icon for load image.

1) **Image capture :** At the very first page of the application, the application bar shows the icon for capturing image using the application. On navigation of the menu, the user gets to take image on shutter click event using the phone.

2) **Image selection :** In case of previously taken pictures of leaf , the application navigation menu also contains the option of selecting an image from the existing photo library of the phone.

3) **Image zoom and crop :** The leaf of crop is a very thin one, and it is important that the targeted area of the leaf gets focus in the image.

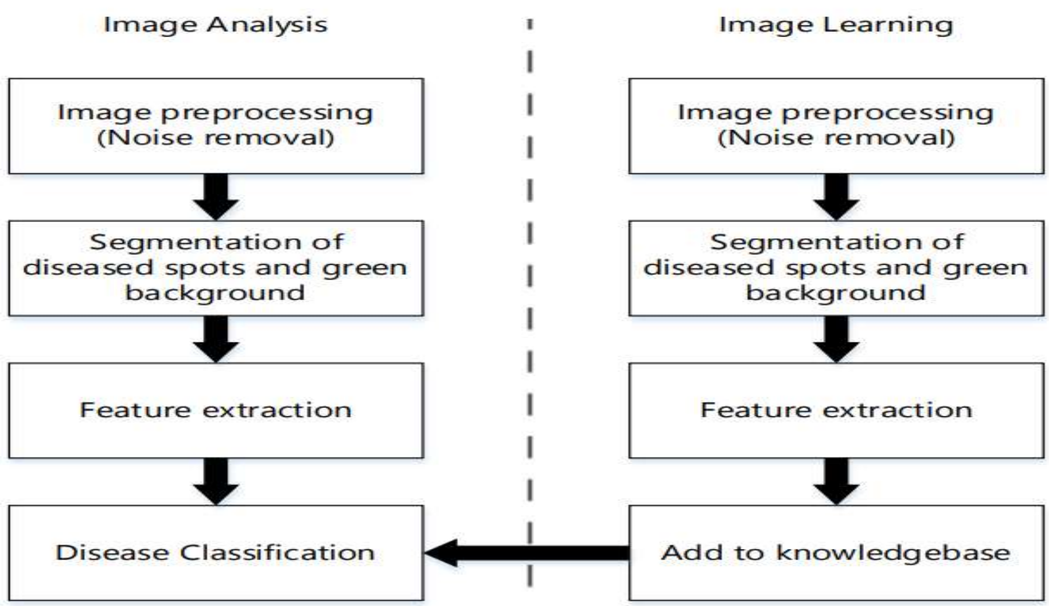


Figure 5.2 The basic procedures of image-processing for disease detection

Noise removal :

The uploaded images may contain noise, for which a bilateral smoothing filter has been for noise cancellation. The bilateral filter is a technique to smooth images while preserving edges where the intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels.

Segmentation :

A simple color feature-based approach has been followed for segmentation of the disease affected images of the leaf . Color is an important feature in color image processing, especially in crop images.

It has been observed for the leaf diseases that the RGB value of the affected region in a leaf can be key component for separating the target region from the non-affected one . The non-affected leaf is usually is green in color, leading the value of Green color to be higher than either of the Red or Blue color for each pixel.

However, in case of the affected spots, the value of Green pixel s less than either of the Red or Blue values. So after initial recognition of the green region within the image, we find mean RGB color value of the green part and convert it lab color space for finding delta E. Delta E is the color difference - between a pixel and its neighbors .

Once the mean LAB color value of the green area has been calculated, the distance delta - E ($\Delta EEaaaa$) is calculated for every pixel of the image using the following equation,

$$\boxed{\Delta E_{ab} = \surd\,((\,L_m - L_i)^2 + (A_m - A_i)^2) + (B_m - B_i)^2)}$$

$\Delta E a$b represents the distance of the pixel value from the mean green region pixel value and the non-green pixel value will be at far more distance, whereas the green pixels will tend to be close to the mean value and the ΔE will be minimal (close to 0).

The maximum color difference within the green region and the mean pixel value is set the threshold value for segmentation . In this process we segment the diseased leaf image and remove the pixels for which the value of $\Delta EEaaaa$ is greater than the calculated threshold value.

Feature Extraction :

The goal of the feature extraction is to characterize an object to be recognized by measurements whose values are very similar for objects in the same category, and different for objects in other categories . For the selected diseases that the designed system analyzes, the nature of visual symptoms provides the basis for which the features can be generated.

The distinctions in the pattern of the affected area in the diseased images are visible. It is noticeable that the lesions/ damaged part of the leaf are the key component of creating variety among different diseases. There are also other observations like color difference among the non-affected regions of the leaves.

CHAPTER 6

RESULTS

6.1 Execution :

1)Input Image

In the process of execution run the code . First, Load the input image from the system.

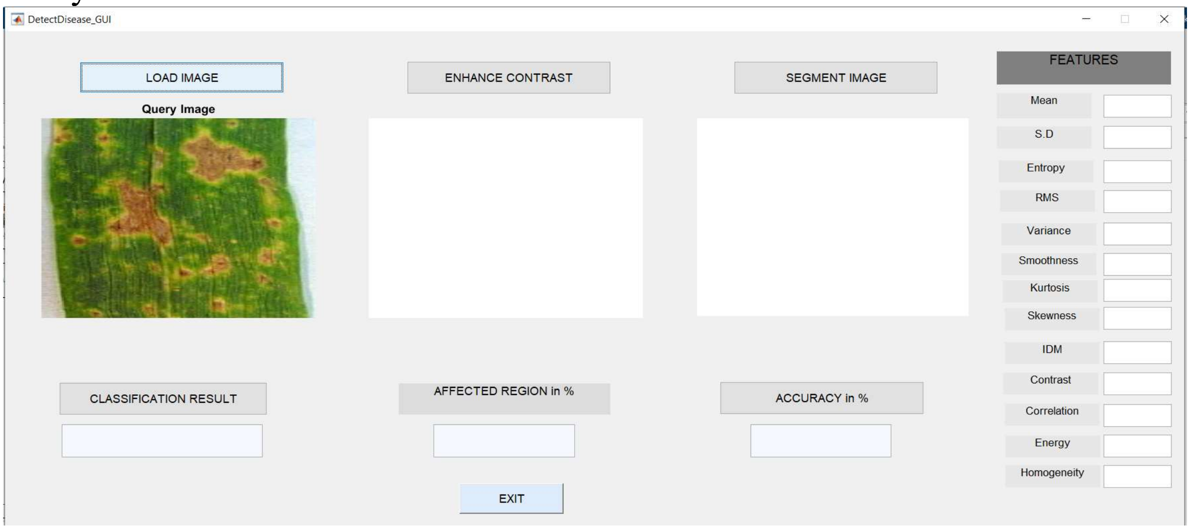


Figure 6.1 Input image

2)Enhance Contrast for the Input image

To improve the visual and appearance of an image and highlighting specific region and structure in the image .It helps in enhancing the performance of subsequent image analysis algorithms.

Figure 6.2 Contrast Enhanced image from the input image

3) Clustering the Input image

To detect the affected region clustering is done with the help of GA-K-means Clustering.

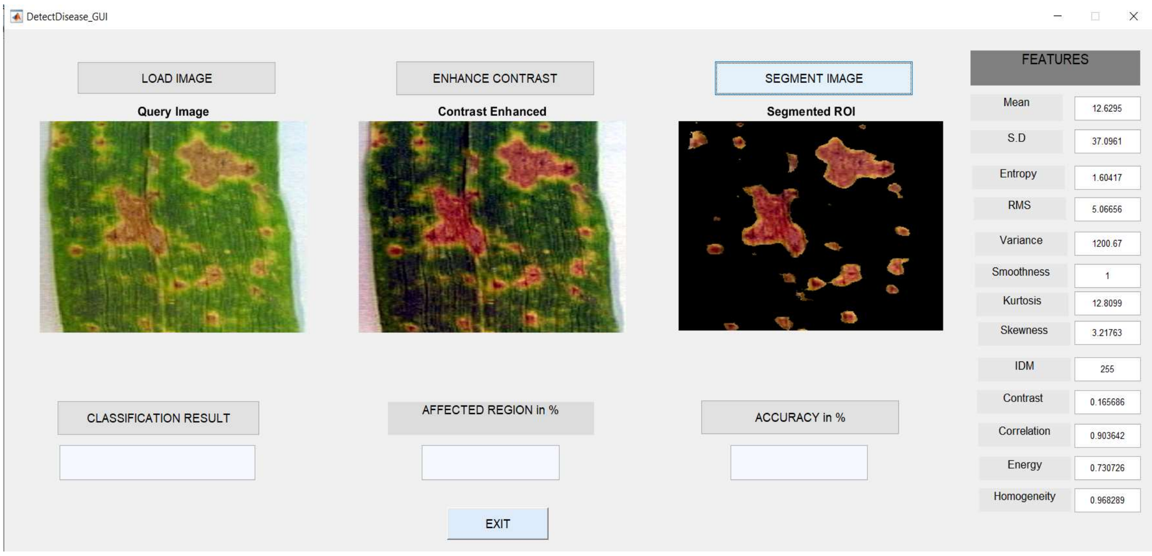


Figure 6.3 Clustering the input image based on k-means clustering algorithm

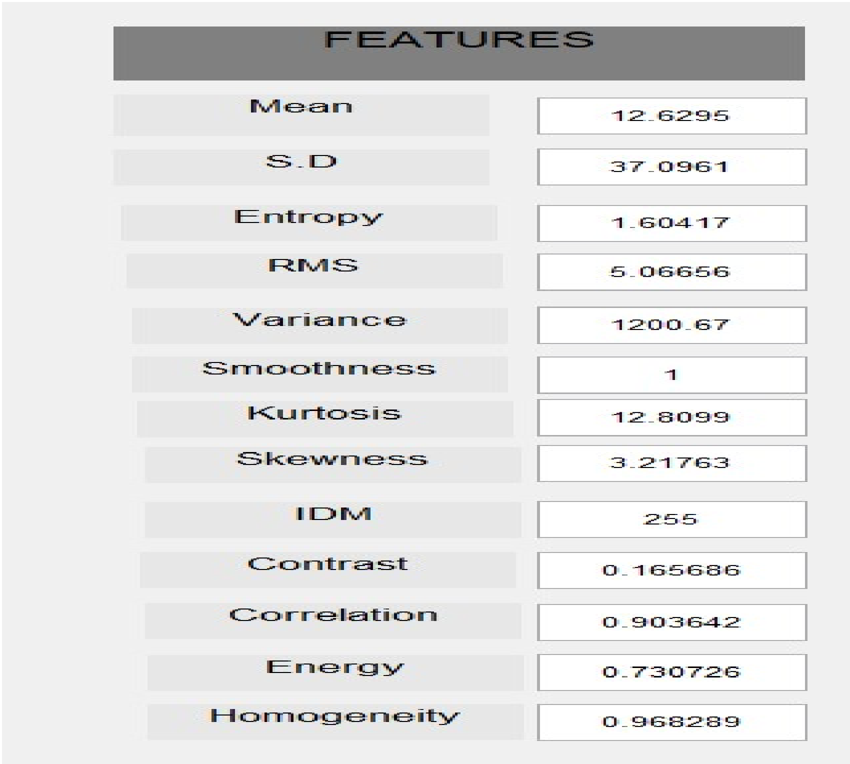


Figure 6.4 Feature extraction

To measure the performance of a crop disease test, the concepts sensitivity and specificity are often used. Say we test some people for the presence of a disease. Say some of the test leaves have disease and we call it true recognition (TR) if the system recognizes the disease properly.

In addition, if the system provides misleading results then it is called false recognition (FR). Thus, the number of true recognition and false 40 recognition add up to 100% of the set.

Finally we have calculated accuracy of the system using the following equation :

$$\text{Accuracy} = \frac{TR}{TR + FR}$$

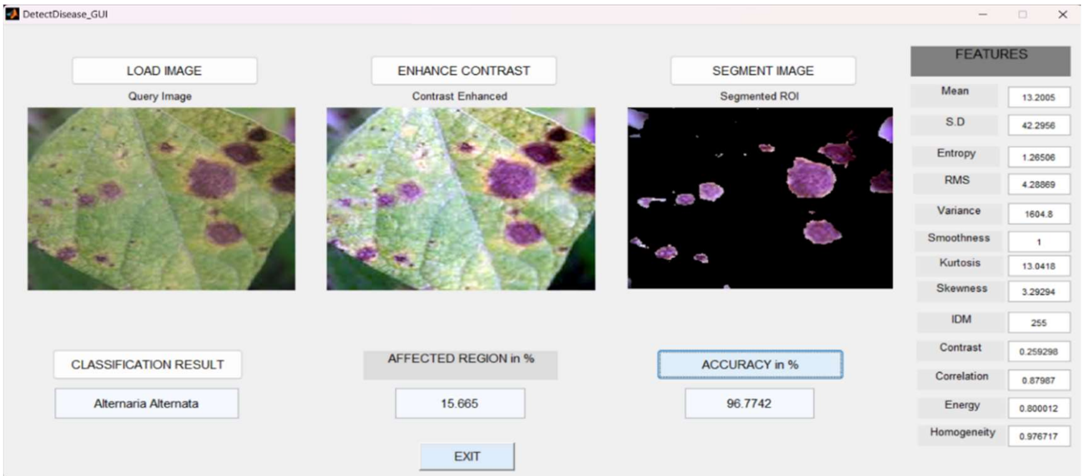
Command Window Description:

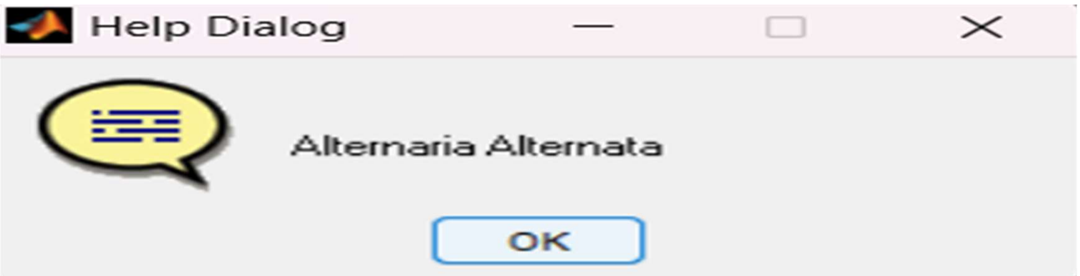
Ans = Affected Area is: 15.0015%

Ans = Accuracy of Linear Kernel with 500 iterations is: 98.3871%

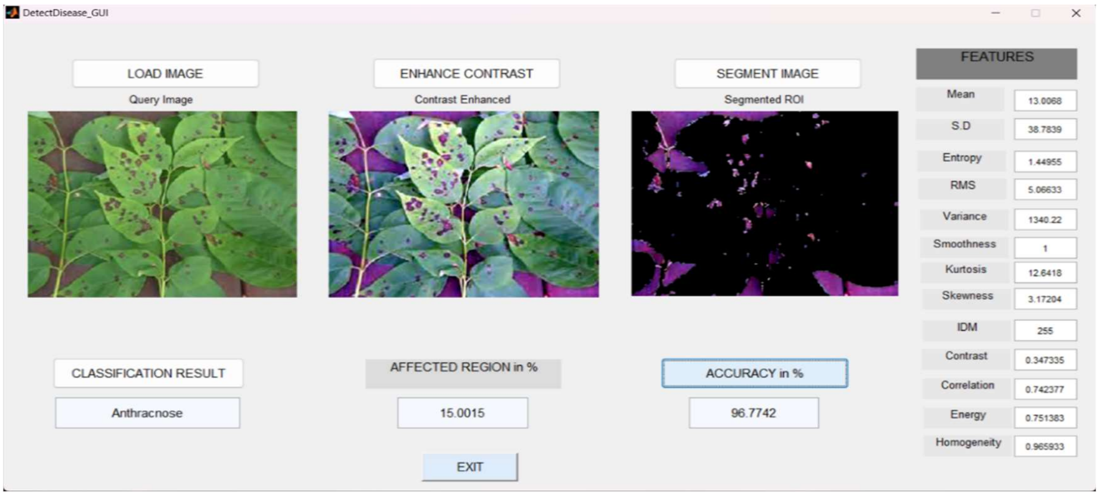
6.2 Disease Outputs

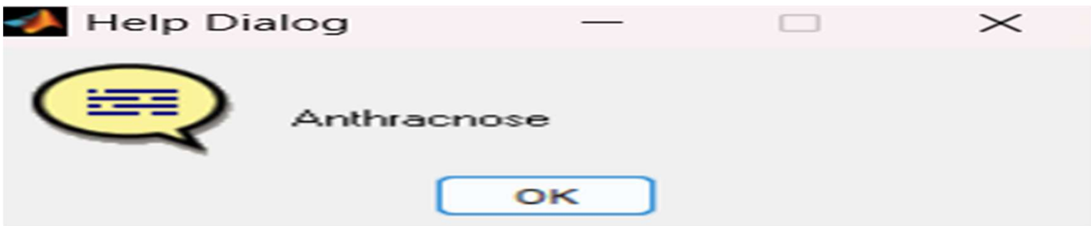
1. Alternaria Alternata



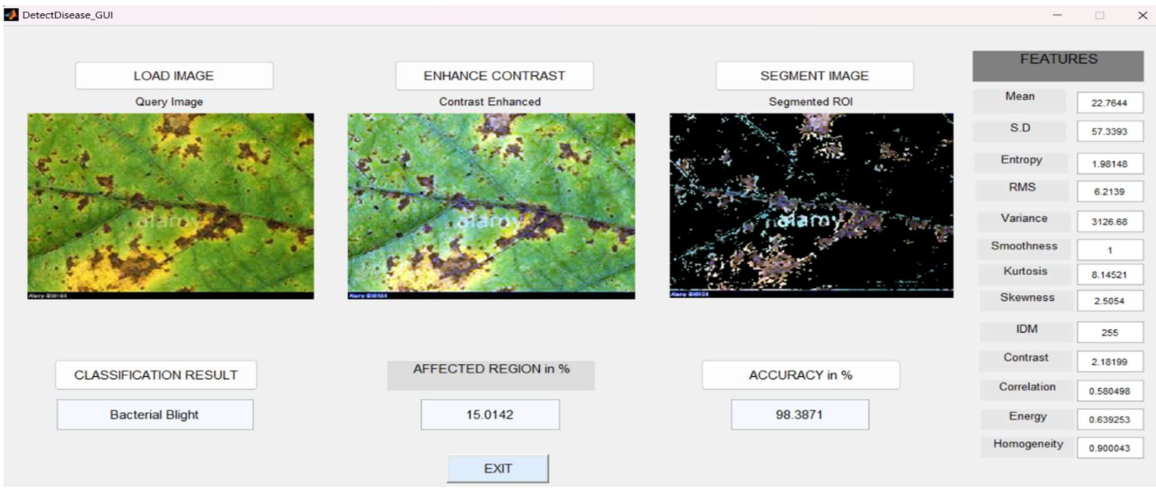


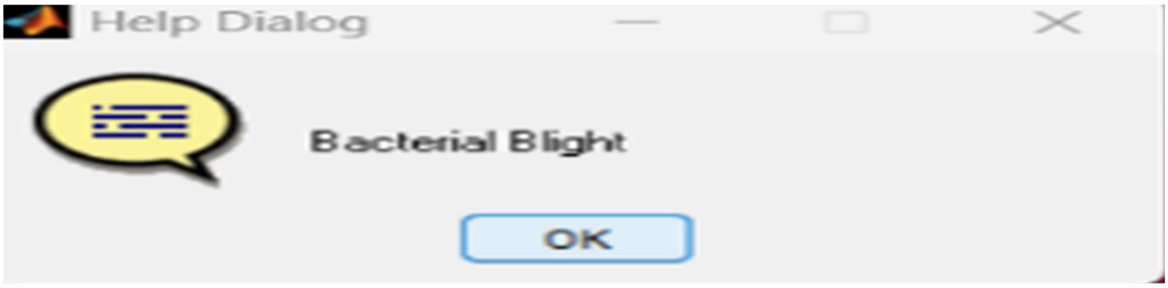
2)Anthracnose



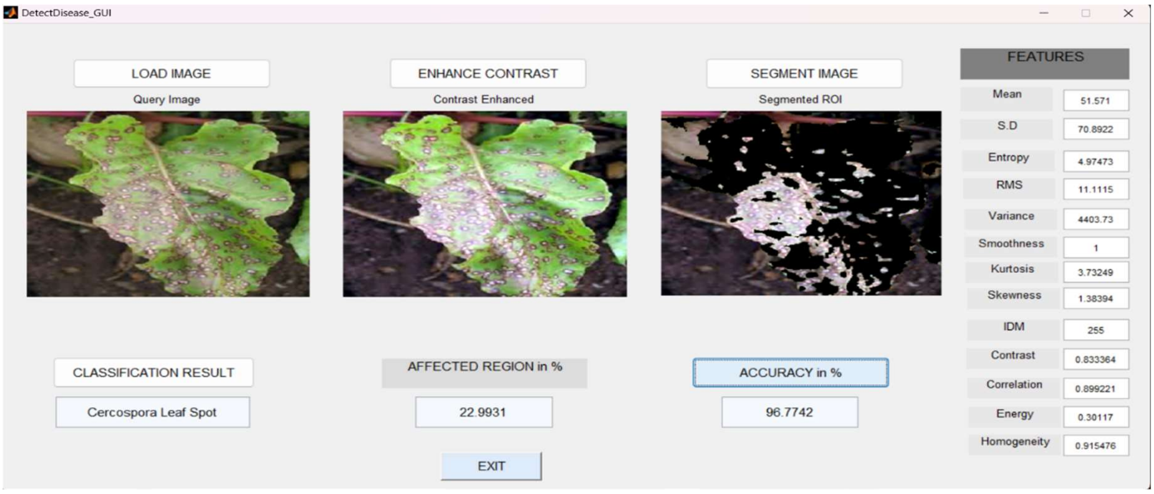


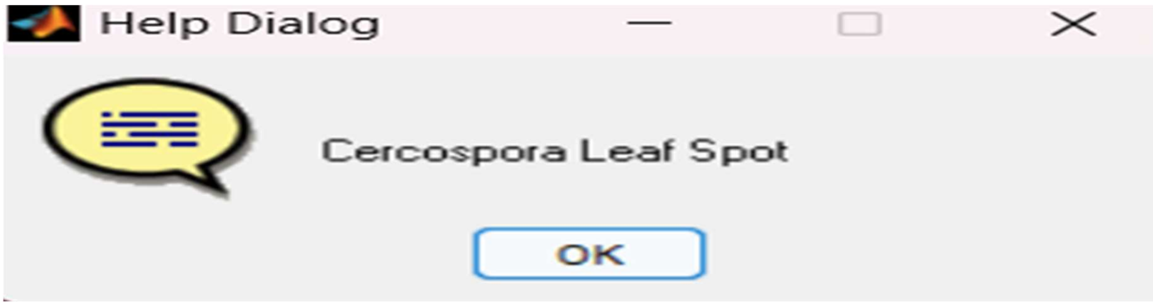
3. BacterialBlight



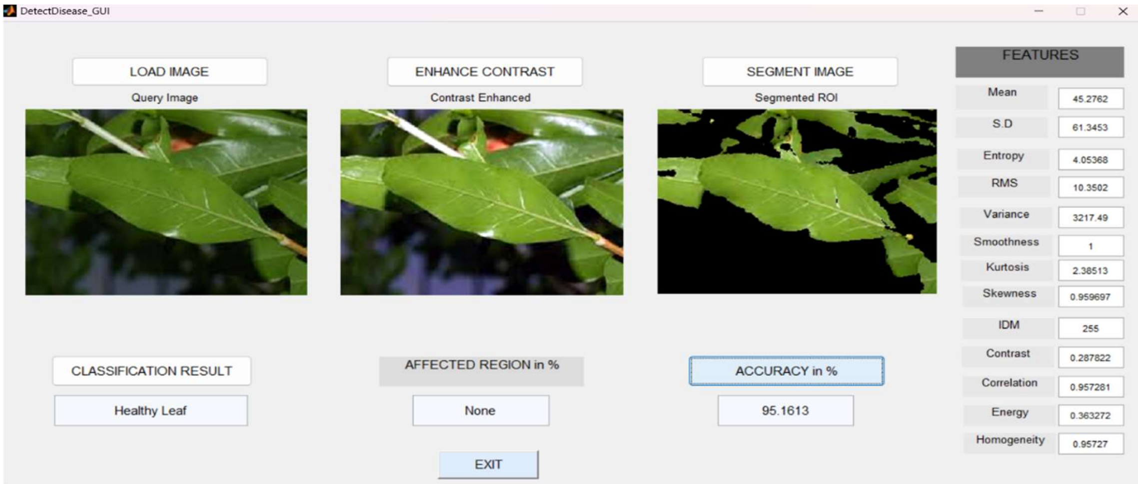


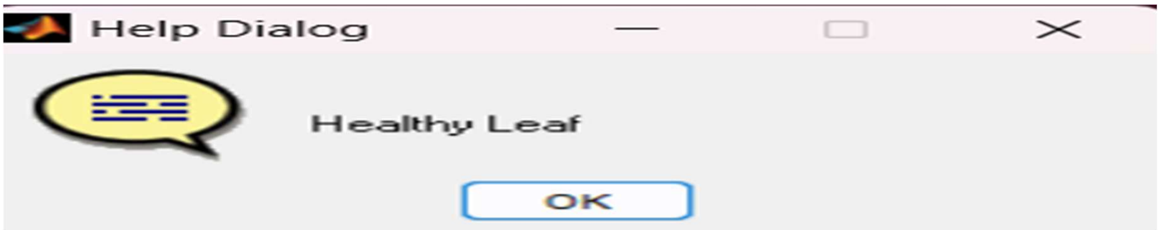
4. Cercospora Leaf Spot



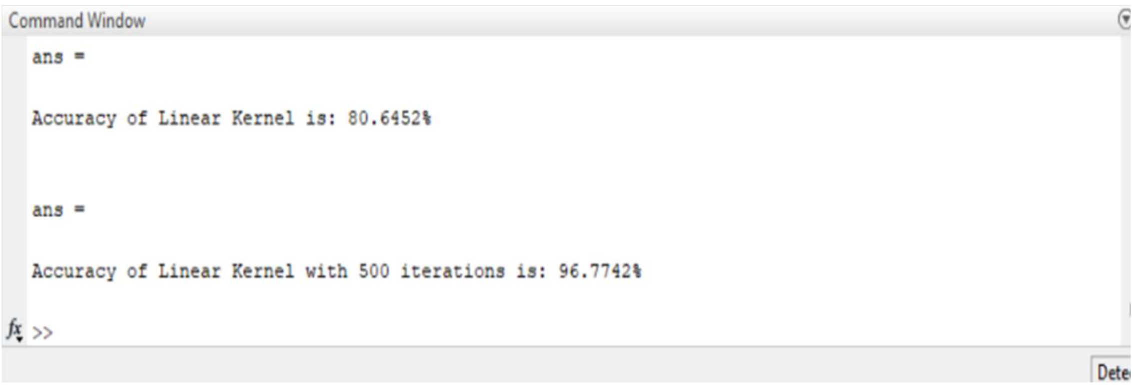


5. Healthy Leaf





6.3 Command line execution result :



CHAPTER 7

CONCLUSION

The system described above uses a combination of image preprocessing, segmentation, feature extraction, and classification techniques to accurately detect the type of leaf disease present in an image. The system employs resizing, thresholding, and Gaussian filtering for preprocessing, followed by K-means clustering for segmentation, and texture and color features for feature extraction. Finally, SVM classification is used for disease detection. The system has achieved an impressive average accuracy of 98.38% for testing images, indicating its effectiveness in detecting various types of leaf diseases. The high accuracy can be attributed to the efficient combination of various techniques used in the system. The image preprocessing steps help to enhance the image quality, while the segmentation techniques help to separate the leaf area from the background. The use of both texture and color features in feature extraction ensures that the system is capable of detecting different types of leaf diseases with varying visual characteristics. SVM classification is a robust and accurate method that performs well in the classification of complex data such as images. Overall, the described system provides an efficient and accurate method for the detection of leaf diseases, which can be useful in agricultural applications.

CHAPTER 8

FUTURE ENHANCEMENT

Future expansion of this work will be focused on following points:

1) To develop combinations of more algorithms by using fusion classification technique, so as to improve the detection rate of the classification process. Using fusion classification techniques involves combining the outputs of multiple classifiers to improve the overall classification performance. By taking advantage of the strengths of different algorithms, fusion classification can improve the detection rate of the classification process. This approach can also help to reduce the risk of overfitting and improve the generalization performance of the system. Different types of fusion classification techniques include decision-level fusion, feature-level fusion, and classifier-level fusion.

2) On the basis of detection of disease the proper mixture of fungicides will be provided to the grape farmer for further use in their farms. The detection of disease in grape plants can be used to provide the grape farmer with a proper mixture of fungicides for use in their farms. This approach involves identifying the specific disease affecting the grape plants and providing a customized fungicide mixture that targets that particular disease. By using a targeted fungicide mixture, the grape farmer can improve the effectiveness of their disease management strategy and reduce the risk of crop loss. This approach requires accurate and timely disease detection, which can be achieved through various techniques such as visual inspection, laboratory testing, and machine learning-based image analysis.

REFERENCES

[1] S. M. Hassan, A. K. Maji, M. Jasiński, Z. Leonowicz and E. Jasińska, "Identification of plant-leaf diseases using CNN and transfer-learning approach", Electronics, vol. 10, no. 12, pp. 1388, Jun. 2021.

[2] S. Khan and M. Narvekar, " Disorder detection of tomato plant (Solanum lycopersicum) using IoT and machine learning ", J. Phys. Conf. Ser., vol. 1432, no. 1, Jan. 2020.

[3] J. K. Patil1 and R. Kumar, “Advances in image processing for detection of plant diseases”, Journal of Advanced Bioinformatics Applications and Research, ISSN 0976-2604Vol 2, Issue 2, pp 135-141, June-2011.

[4] Weizheng, S., Yachun W., Zhanliang C., and Hongda W.,” Grading Method of Leaf Spot Disease Based on Image Processing” International Conference on Computer Science and Software Engineering - Volume 06 ,PP. 491-494,December 2008.

[5] S. S. Sannakki, V. S. Rajpurohit, V. B. Nargund,and P. Kulkarni, “Diagnosis and Classification of Grape Leaf Diseases using Neural Networks”, IEEE 4th ICCCNT, 2013.

[6] P. Chaudhary, A. K. Chaudhari, Dr. A. N. Cheeran and S.Godara, “Color Transform Based Approach for Disease Spot” International Journal of Computer Science and Telecommunications Volume 3, Issue 6, pp.65-70,June 2012

[7] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, “Fast and Accurate Detection and Classification of Plant Diseases”, IJCA, Vol-17,No.-1,pp. 31-38, March 2011.

[8] P. Babu, M. S. and Srinivasa Rao, “Leaves recognition using backpropagation neural network - advice for pest and disease control on crops”. Technical report, Department of Computer Science & Systems Engineering, Andhra University, India, www.indiakisan.net on May 2010.

/[9] Suman T. and Dhruvakumar T., “Classification of paddy leaf diseases using shape and color features”, IJEEE, Volume 07, Issue 01, PP.239- 250, Jan- June 2015.

[10] P. M. Mainkar, S. Ghorpade and M. Adawadkar, "Plant Leaf Disease Detection and Classification Using Image Processing Techniques", International Journal of Innovative and Emerging Research in Engineering, vol. 2, no. 4, 2015.

[11] S. B. Dhaygude and N. P. Kumbhar, "Agricultural plant leaf disease detection using image processing", International Journal of Advanced Research in Electrical Electronics and Instrumentation Engineering, vol. 2, no. 1, pp. 599-602, 2013.

[12] M L wang, X J niu, H M Zhang et al., "Research on feature extraction and recognition of common diseases of wheat leaf [J]", Computer Engineering and Applications, vol. 50, no. 7, pp. 154-157, 2014.

[13] B. Kim, Y. K. Han, J. H. Park and J. Lee, "Improved Vision-Based Detection of Strawberry Diseases Using a Deep Neural Network", Frontiers in Plant Science, vol. 11, pp. 1-14, 2021.

[14] T. Akiyama, Y. Kobayashi, Y. Sasaki, K. Sasaki, T. Kawaguchi and J. Kishigami, "Mobile leaf identification system using CNN applied to plants in Hokkaido", 2019 IEEE 8th Global Conference on Consumer Electronics GCCE, pp. 324-325, 2019.

[15] S. J. G. A. Barbedo, "A new automatic method for disease symptom segmentation in digital photographs of plant leaves", European Journal of Plant Pathology, vol. 147, no. 2, pp. 349-364, 2016.

[16] M. G. Du and S. W. Zhang, "Crop Disease Leaf Image Segmentation Based on Genetic Algorithm and Maximum Entropy", Applied Mechanics and Materials, vol. 713, pp. 1670-1674, 2015.

[17] J. Y. Bai and H. E. Ren, "An Algorithm of Leaf Image Segmentation Based on Color Features", Key Engineering Materials, vol. 474, pp. 846-851, 2011.

[18] N. Valliammal and S. S. N. geethalakshmi, "A Novel Approach for Plant Leaf Image Segmentation using Fuzzy Clustering", International Journal of Computer Applications, vol. 44, no. 13, pp. 10-20, 2012.

[19]L. Kaiyan, W. Junhui, C. Jie and S. Huiping, "A Real Time Image Segmentation Approach for Crop Leaf", 2013 Fifth International Conference on Measuring Technology and Mechatronics Automation, 2013.

[20] G. Prem Rishi Kranth, Hema Lalitha, Laharika Basava and Anjali Mathurh, "Plant disease prediction using machine learning algorithms", International Journal of Computer Applications, vol. 18, no. 2, 2018.

[21] H. Sabrol and K. Satish, "Tomato plant disease classification in digital images using classification tree", 2016 International Conference on Communication and Signal Processing (ICCSP), pp. 1242-1246, 2016.

[22] U. Mokhtar, M.A.S. Ali, A.E. Hassenian and H. Hefny, "Tomato leaves diseases detection approach based on support vector machines", 2015 11th International Computer Engineering Conference (ICENCO), pp. 246-250, 2015.

[23] V. Singh and A. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques", Information Processing in Agriculture, vol. 4, no. 1, pp. 41-49, 2017.

[24]www.inf.unibz.it/~mkacimi/svm.pdf

[25]http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/KMeans_Clustering_Overview.html