

**PLANT LEAF DISEASE DETECTION USING IMAGE
PROCESSING AND
SEGMENTATION TECHNIQUES**

A

PROJECT REPORT

Submitted in partial fulfillment of the requirement for the degree

of

BACHELOR OF TECHNOLOGY

IN

**COMPUTER SCIENCE AND ENGINEERING/INFORMATION
TECHNOLOGY**

Under the supervision

of

Dr. Pradeep Kumar Singh

by

PARV SHARMA (151290)

NISHANT SHARMA (151291)



Department of Computer Science & Engineering and Information
Technology

**JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY
WAKNAGHAT, SOLAN-**

173234, HIMACHAL PRADESH

CANDIDATE'S DECLARATION

I hereby declare that the work presented in this report entitled “ Plant leaf disease detection using image processing and segmentation techniques” in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering/Information Technology submitted in the department of Computer Science & Engineering and Information Technology, **Jaypee University of Information Technology Wagnaghat** is an authentic record of our own work carried out over a period from July 2018 to May 2019 under the supervision of Dr. Pradeep Kumar Singh ,Assistant Professor (Senior Grade) Computer Science & Engineering and Information Technology

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Parv Sharma (151290)

Nishant Sharma (151291)

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Signature of supervisors:

Dr. Pradeep Kumar Singh

Assistant Professor

Department of Computer Science

JUIT, Wagnaghat

Dated:

ACKNOWLEDGEMENT

On the submission of our thesis report on “Crop growth monitoring using image segmentation and machine learning techniques”, we would like to extend our gratitude and sincere thanks to our supervisor Dr. Pradeep Kumar Singh, Assistant Professor (Senior Grade), Department of Computer Science & Engineering and Information Technology for his constant motivation and support during the course of our work . We truly appreciate and value his esteemed guidance and encouragement from the beginning. We are indebted to him for having helped us shape the problem and providing insights towards the solution. And for providing a solid background for our studies and research thereafter. He has been a great source of inspiration to us and we thank him from the bottom of our heart. Above all, we would like to thank all our friends whose direct and indirect support helped us.

Parv Sharma (151290)

Nishant Sharma (151291)

Contents

CANDIDATE’S DECLARATION	2
ACKNOWLEDGEMENT	3
LIST OF FIGURES	6
LIST OF TABLES.....	7
ABSTRACT	9
CHAPTER-1	10
INTRODUCTION	10
1.1 INTRODUCTION	10
1.1.1 Image Definition	10
1.1.2 Semantic Network	10
1.1.3 Identifying Patterns	11
1.2 PROBLEM STATEMENT	12
1.3 PLANT DISEASE FUNDAMENTALS	12
1.3.1 Bacterial disease	12
1.3.2 Viral Disease	13
1.3.3 Fungal Disease.....	13
1.4 OBJECTIVES	13
1.4 ORGANIZATION.....	14
CHAPTER-2	15
LITERATURE REVIEW	15
2.1 LITERATURE SURVEY	15
2.1.1 Objective of First Paper:	15
2.1.2 Objective of Second Paper:.....	16
2.1.3 Objective of Third Paper	16
2.1.4 Objective of Fourth Paper:	17
2.1.6 Objective of sixth paper:	19
2.1.7 Objective of the seventh paper:	20
CHAPTER-3	21
SYSTEM DEVELOPMENT	21
3.1 TOOLS	21
3.1.1 Data attaching Tools.....	22
3.2 BASIC ARCHITECTURE	25
3.3 INFORMATION POSIT	26
CHAPTER-4	29
ALGORITHM.....	29

4.1 GENERAL	29
4.2 IMAGE ACQUISITION	29
4.3 IMAGE PREPROCESSING AND ENHANCEMENT	29
4.4 IMAGE SEGMENTATION	31
4.5 K MEANS COLOR BASED CLUSTERING	31
4.6 OTSU's CLASSIFIER	32
4.7 FEATURE EXTRACTION	34
CHAPTER 5	35
TEST PLAN	35
5.1 GENERAL	35
CHAPTER 6	41
RESULTS AND PERFORMANCE ANALYSIS	41
6.1 GENERAL	41
6.2 ALTERNARIA ALTERNATA	41
6.3 ANTHRACNOSE	42
6.4 BACTERIAL BLIGHT	43
6.5 CERCOSPORA LEAF SPOT	44
6.6 HEALTHY LEAF	44
CHAPTER 7	46
CONCLUSION AND FUTURE SCOPE	46
7.1 CONCLUSION	46
7.2 FUTURE SCOPE	46
REFERENCES	47

LIST OF FIGURES

Figure 1.1: Diagrammatical representation of neural network	11
Figure 3.1: Neural Fitting Tool	22
Figure 3.2: Neural Training Tool	24
Figure 3.3: Block Diagram of Basic Architecture	25
Figure 3.4: Histogram of healthy leaf	27
Figure 3.5: HIS representation of normal leaf	27
Figure 3.6: Histogram of diseased leaf	28
Figure 3.7: HIS representation of Diseased Leaf	28
Figure 4.1: Image Preprocessing Original Image.	30
Figure 4.2: Preprocessed Image	30
Figure 4.3: K Means Algorithm	31
Figure 4.4: Cluster One	33
Figure 4.4: Cluster Two	33
Figure 4.4: Cluster Three	33
Figure 4.5: Features of segmented ROI	34
Figure 5.1: GUI	35
Figure 5.2: Input Image	36
Figure 5.3: Enhanced Preprocessed Image	37
Figure 5.4: Segmented Image Box	38
Figure 5.5: Segmented Image Clusters	39
Figure 5.6: Input dialog Box	39
Figure 5.7: Segmented ROI	40
Figure 6.1: Detection of Disease	41
Figure 6.2: Disease Result 1	42
Figure 6.3: Disease Result 2	42
Figure 6.4: Disease Result 3	43
Figure 6.5: Disease Result 4	44
Figure 6.6: Disease Result 5	44

LIST OF TABLES

Table 6.1: Estimation of Time and Area affected

45

LIST OF ACRONYMS AND ABBREVIATION

ANN	Artificial Neural Network
CCV	Color Coherence Vector
GUI	Graphical User Interface
HIS	Hue Intensity and Saturation
ML	Machine Learning
MM	Markov Model
NN	Neural Network
ROI	Region of Interest
RGB	Red Green and Blue
SOM	Self Organizing Map
SVM	Support Vector Machine

ABSTRACT

There's a variety of development which has been made regarding digital image processing and machine learning algorithms which also include its various applications. Now we are living in an era where the problem regarding agriculture is a major issue nowadays. The major problem in crop growth is we have to take care of the health of the plants and crops .In this report we basically focused on classification of plants as different type of diseases. For this we use HSI color model and clustering algorithm .We also used MATLAB for our project.

CHAPTER-1

INTRODUCTION

1.1 INTRODUCTION

1.1.1 Image Definition

Image is a collection of pixels or dots which are stored in rectangular array. Each individual pixel is having certain kind of color. We can measure the size of the image by counting the no of pixels in that particular image. Different type of images are there such as Black and White and Grey scale images. Both types vary from each other .In black and white image each dot or pixel is either black or white, therefore only one bit is needed per pixel. Whereas Grey scale images uses 8 bits per pixel.

For color images things gets slightly difficult. In color images number of bits at every dot termed as the height of image. It is also referred as the bit plane. For bit plane consisting of x , $2x$ color are possible.

Different methods are available to store the color information of image. One of the method is RGB image also termed as true color image. For every pixel red, green and blue component is stored in three dimensional array.

1.1.2 Semantic Network

[1] Semantic network also termed as neural networks consist of layer of connected nodes like neurons in brain used for computing purposes. These networks has the ability to learn the data and get trained over that data, in order to identify patterns and also classification.

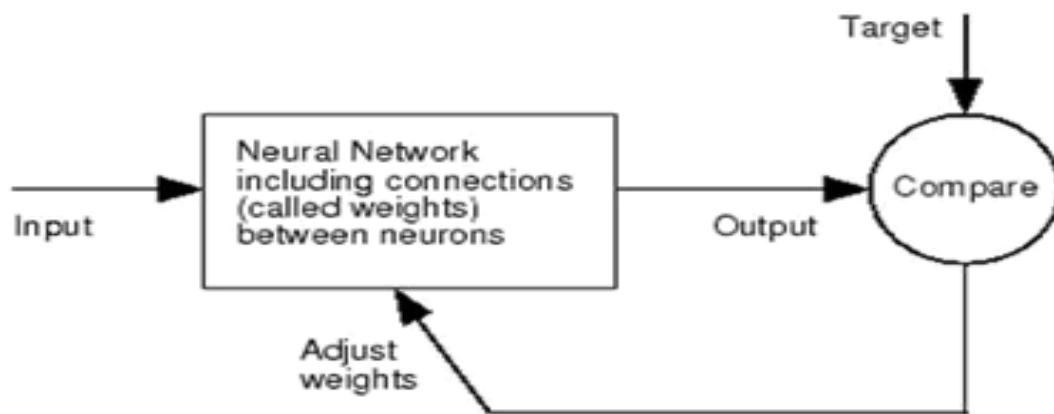


Figure 1.1 Diagrammatical representation of neural network

Neural network or semantic network has the advantage that they can solve the complicated problem very easily. To train the neural networks or to implement them in MATLAB we have NN toolbox.

Different commands are there to operate with NN toolbox. Different functions are there in NN toolbox for fitting, recognizing the pattern and clustering.

1.1.3 Identifying Patterns

In order to detect the same kind of pattern different pattern recognition techniques are used in MATLAB. Using these techniques we recognize the similar kind of the pattern in the problem. When same kind of pattern are detected then these can be used to generate outputs or solve the problems more efficiently. In order to recognize the pattern, we need to train the machine. For this first we need to classify the data .The data is classified using the key features .For classifying the data we have different type of learning modules is there such as supervised learning and unsupervised learning modules .Bothe of these modules are used to identify the patterns. In supervised learning module we train the machine by recognizing the patterns in the data set and then results which are generated are applied to the testing data set. We train the machine over the training dataset and test it over the testing data set. In unsupervised learning module, there are no visible pattern the dataset, so with the help of the some algorithm we try to catch the patterns. Clustering algorithm, classification algorithm such as Markov Model (MM) is there.

For recognizing the patterns we identify we have different techniques such as preprocessing, Extraction of features and classification. In preprocessing we try to filter out, smooth the data by normalizing in more ordered way. Filtering such as noise filtering is there. Feature extraction is usually done using the software which collect the information from the data. Sensors are also used for this purpose and the final phase is the classification.

1.2 PROBLEM STATEMENT

The main purpose is detect the diseased part of the plant. Using MATLAB convolutional neural networks are implemented in order to classify the diseased part. Aim is to detect the diseased part by finding the optimum way with minimum cost. In this problem we have considered fundamental five categories of the plant leaf disease which are Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora leaf spot and Healthy Leaves. All of these disease belongs to fungal, viral or bacterial type of the diseases. In our proposed solution we identify the percentage of the affected area and identify the disease. Our approach provide the result in minimum time span with maximum precision and accuracy in comparison to other existing approaches.

1.3 PLANT DISEASE FUNDAMENTALS

There are different types of plant disease exist, but majority of these disease can be categorize into the three different categories which are bacterial disease viral disease and the fungal disease. The most ideal way to detect the disease is the classification followed by detection. Classification is done on the basis of shape and texture features.

1.3.1 Bacterial disease

This is also known as bacterial leaf spot. Bacterial leaf spot is majorly detected in stone fruits such cherry, plum etc. In this disease black spots or dark spots occur on the different part of leaves. Yellow halos is also symptom of this disease. Spot size is of irregular nature. Bacterial spots occur on the different part on the top and bottom start occurring and if these spots cluster together in any section of the leaf then this results in killing of that section by this disease. Wet and cool formation also contribute to the formation of the bacterial disease in

the plant leaves. In these formations bacterial leaf spot can spread very quickly. Mostly bacterial leaf spot occur on the aged leaves but it can destroys the tissues of the new leaves too.

1.3.2 Viral Disease

Viral disease are caused by viruses and as virus are intercellular, so these diseases attacks inside out. Viral disease are sometime very difficult to identify. Virus can affect any region of the plants such leafs, roots, stem and others. Abnormal patterns are observed on the affected area green and yellow coloration is seen in leaves affected with the virus. The life span of the plant or its parts affected with the viral disease is very less. It directly affects the productivity and other factors. Wrinkles on the different part of the leaves is also primary symptom of these disease. Every virus life span is very high as compared to the other types of the disease, because each virus if not properly cured give rise to new type of the virus so it is important for timely prevention of these disease.

1.3.3 Fungal Disease

Fungal disease occur because of the fungi or fungal organism. One of the property of the fungi is that it spread with wind and the water. Gray green spots on the leaf of the plants are observed and if not properly cured they start getting spread toward the outer region of the leaf. Wilting, scabs are the primary symptoms fungal disease. Fungal disease attack on the plant leafs result in the yellowness of leaves at end.

1.4 OBJECTIVES

- To apply and implement the new algorithm to monitor the growth of the crop and detect the diseased part in the plant.
- To incorporate the global optimization strategy for enhancing the optimal solution.
- To modify the existing solution more efficiently to solve this problem is more genuine way.

1.4 ORGANIZATION

- Introduction: Chapter one gives the basic brief of the problem. It gives the idea that how images are stored, type of the images, RGB storage, semantic networks and
- Classification.
- Literature Survey
- Experimental Work
- Purposed Work

CHAPTER-2

LITERATURE REVIEW

2.1 LITERATURE SURVEY

2.1.1 Objective of First Paper: Rice Disease detection Pattern Recognition Techniques.

Published in 11th International Conference on Computer and Information Technology

Gist of Paper: The point of this paper is to depict a product model framework for the discovery of malady in rice plant based on different pictures of the rice plants. Pictures of the tainted piece of the rice plant are taken utilizing computerized camera. With the end goal to identify the abandoned piece of the plant different procedures like picture division, picture developing and so forth. By utilizing neural system the tainted piece of the leaf is grouped. Picture preparing and delicate processing procedures are joined on infected plant.

Procedures embraced in paper:

- Preparing & design examination strategies of images
- Binary cutoff methods
- Border layout calculation using eight-availability strategy
- Self-organizing map(SOM)

In this examination paper, the diseased part of the rice plant leaf is identified with the help of the self-organizing map. Testing is done using four different images of the crop. Infected region is extracted using neural networks pattern recognition techniques.

By utilizing effective example acknowledgment procedures, the framework will have the capacity to do the opportune finding of the field issue and the proposal will assist the ranchers with taking the suitable measure to build the nature of the harvest .It won't just decrease the improvement cost later on yet in addition spare the earth too.

2.1.2 Objective of Second Paper: Detection of plant leaf diseases using image segmentation and soft computing techniques. Published in Information Processing in agriculture.

Gist of Paper: This paper monitor the crop growth using the image segmentation techniques. Noise filtering is done and features are extracted and then image is further classified to detect the diseased part.

Strategies/Methodology embraced in paper:

- Support Vector Machine (SVM)
- Artificial Neural Network (ANN)
- Dispersion method
- Self-sorting out element

Using image segmentation techniques and machine learning algorithms the information for ripening stages of crop and infected part recognition is made.

There were a few issues for doing extraction of vague shading pixels from the foundation of the picture.

- Neural arrange don't permit better division of the grape leaf illness pixels.
- The framework will show programmed determination ability with extremely successful execution for the further agrarian item investigation/review framework improvement.

2.1.3 Objective of Third Paper: Remote Area Plant diseases detection using Image Processing. Published in IOSR Journal of Electronics and Communication Engineering.

Gist of Paper: Infected part in the plant can be detected with help of color, and other changing properties by using classification algorithm.

Strategy/Methodology embraced in paper:

- Segmentation
- RGB
- Color transformation
- Image acquisition
- classification

Different pixel information is extracted and Green leaves pixel and diseased leaf pixel are compared by finding the ratio of pixel corresponding to the healthy leaf to the pixel corresponding to the infected leaf. Background is remove and different region of the images are formatted after the image acquisition.

- Using image segmentation to extract the image feature is best. But important thing is the level of the results which derived using are how much reliable.
- Results exhibited intriguing enhancement in the forecast framework. It is possibly a promising option in contrast to existing expectation models.
- Further these approaches are scalable and can be modified as per the requirements

2.1.4 Objective of Fourth Paper: Image processing for smart farming: Detection of disease. Published in IEEE second International Conference.

Gist of Paper: The point of this paper is to separate the various diseases in the different parts and then apply the suitable algorithm and to design the approaches in order to detect the diseases using artificial neural network. Two database of the image is used one for training and other for the testing.

Strategy/Methodology embraced in paper

- Image Segmentations
- ML calculations

- Support vector machines
- Artificial Neural Networks

All the disease of the apple and grapes are correctly identified using digital image processing and machine learning neural networks. Results gets improved when contrasted with the discriminative models.

Machine Learning approaches are adaptable and furthermore can give secluded methodology to the information investigation particularly for the new area of 'plant pressure examination'.

It will likewise help in the quality revelation process and in addition the presentation of novel determinations conventions for the complex aggressive attributes like biotic and abiotic stress and yield.

2.1.5 Objective of the fifth paper: A digital image processing based algorithm which detect and recognize the plant diseases and various symptoms. Bio system Engineering Volume 102, Issue 1, January 2009, pp. 9-21.

Gist of the paper: The idea proposed here is to process and analyze the colored images to identify the affected area and various visible symptoms of the diseases

Strategy/Methodology embraced in paper

- Acquisition and Preprocessing
- Pixel contrast enhancement
- Segmentation
- Classification

The proposed methodology is applied to twenty different kind of images result is made on the basis of white and black colors in the resulting image. Black color is used for represent the symptom of disease and which for unaffected region. The results are also compared with the

binary images in which one represented the disease region and zero represented the not diseased part. Some variation were there in both the results.

2.1.6 Objective of sixth paper: Plant disease detection using image processing. 2015 international conference on the computing communication control and automation.

Gist of the paper: the paper provides different types image processing techniques which can directly be implemented in MATLAB for preprocessing of the image and also brief about the image segmentation and image classification. For the feature extraction the paper provide algorithmic methods which can easily calculate shape and color oriented features.

Strategies/Methodologies embraced in paper:

- Image preprocessing
- Image enhancement
- SVM classification
- Semantic networks
- K means clustering
- Neural network based classifier

The feature extraction and the image segmentation algorithm used in this paper are efficient with very high accuracy. Different disease are identified with very high precision rate and accuracy. Clustering algorithm approach is also very much efficient fast, the clustering algorithm segment the image in the different clusters in very short span of time . For small dataset of image the algorithm used in this paper are very effective. But if the dataset size gets larger then there may come some sort of distortion in the accuracy of the above approach.

2.1.7 Objective of the seventh paper: Smart Farming: Pomegranate Disease Detection Using Image Processing” Second International Symposium on Computer Vision and the Internet 2015.

Gist of the paper: The paper provides the image processing techniques and the algorithm which help the farmers to successfully identify the disease in pomegranate. Image acquisition and image processing of the input image is done using the filter commands in the MATLAB by pixel values of the input image gets more clarified and the disease which is present in the pomegranate is successfully detected. The approach also provide user two different options that is with intent search and without intent search.

Strategies/Methodologies adopted

- Image preprocessing
- Feature extraction
- Morphology
- Color Coherence Vector (CCV)
- Clustering
- Training and classification
- Intent search

Remote area farmer can also identify the disease in the pomegranate crop as this algorithm gives user option with and without intent. It let the user to upload the image in the system for further processing. This approach is very affective. Also experimental readings shows that the algorithm has eighty two percent accuracy considering average of both case. In almost every case pomegranate disease is identified. Web base approach is also very fast and there is no distortion occur in communication of images in transition of images in this approach. Pixel values remain maintained and cleared.

CHAPTER-3

SYSTEM DEVELOPMENT

3.1 TOOLS

MATLAB tool is used for the solving the disease detection problem. This provide strong support for the implementation of the advanced algorithms. Machine learning algorithms and classifier too are very easily implemented. MATLAB also allows user to testing and training of the data. User can set various cutoffs for all the attributes and train the data on a given dataset and also using testing tools we can verify the data. In our problem we measure various parameters such as mean, correlation, contrast, variance, smoothness .[2]All of these parameters can be easily calculated using existing methods in the MATLAB. By applying the SVM classifier in MATLAB and passing the above parameters we can easily detect the type of disease in any leaf. MATLAB allow users to plot hue, saturation and intensity values for all type of leafs i.e diseased or healthy. All requirements for the implementation of above methodologies can be done very easily in MATLAB. Below are the reason that MATLAB is used for the purpose

- Easy implementation of complex algorithm
- Previous history
- Accuracy and precision
- Advanced algorithm implementation
- Easy GUI implementation

Being a general programming language MATLAB is best fitted for the digital image processing. MATLAB allow users replication as it make sure that whatever steps user has implemented in the image processing they are properly documented by it . Moreover in MATLAB user can create interface very easily using guide command. There is no need to write code for the GUI as MATLAB automatically writes the coding modularities for the GUIs. All source codes are easily accessible for testing. Enhancement and segment process in image process is very easy in MATLAB.

3.1.1 Data attaching Tools

Neural networks are best used for fitting and attaching using various fitting functions. The fact is that the neural networks are designed in way that can arrange themselves along given set of ranges.

Data attaching is also the way toward separating precision fit. Experts and scientists use the data attaching sets which consist of number terminologies and factors in order to visualize the information which is collected.

One of important feature of the MATLAB is that it gives opportunity to load the data and adjust the data as per the recommendations. Fitting is achieved with the help of the fitting tool available in the MATLAB.

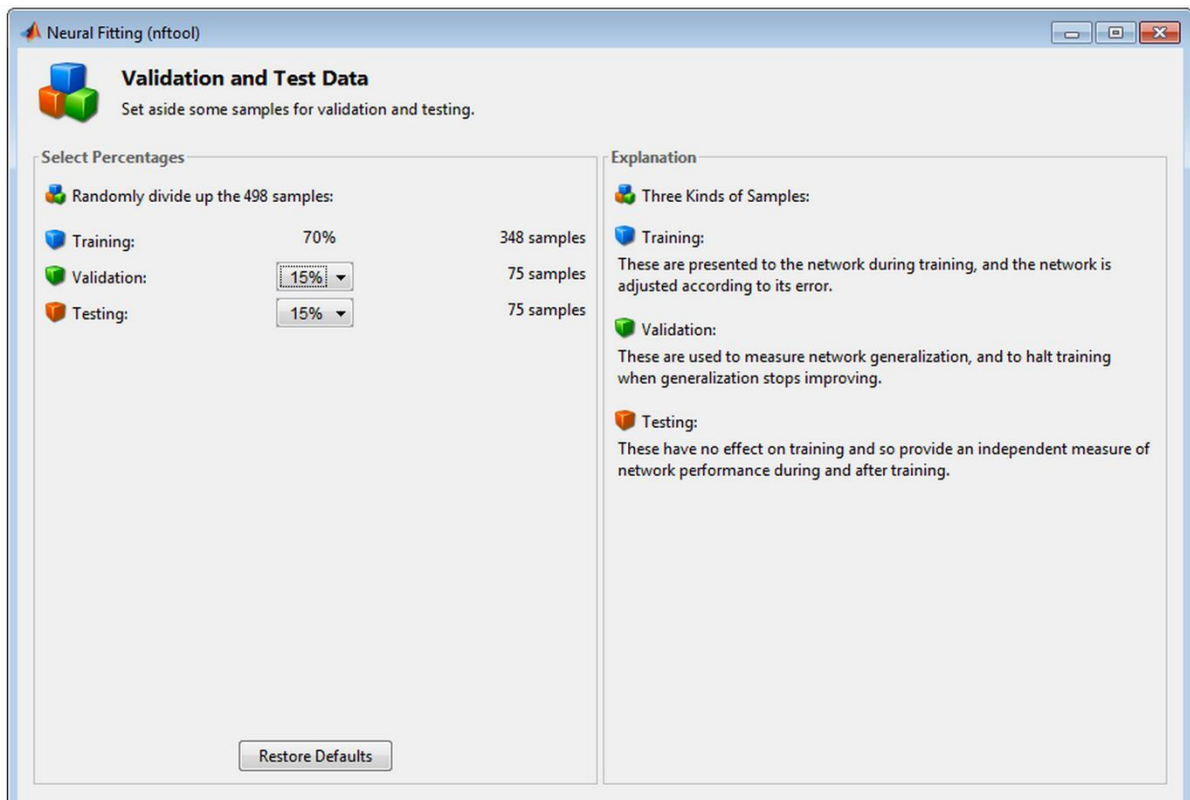


Figure 3.1: Neural Fitting Tool

The above figure is the basic structure of neural fitting tool available in the MATLAB.

Analysis: The above tool gives the idea of the how aimed vector are subjectively isolated in three types as such:

- Around seventy percentage of the data is to get started and getting ready.
- Fifteen percent of the data is used for the validation purpose means to check and validate that trained data is accurate.
- Rest of the fifteen percent of the data is used for an absolute self-sufficient preliminary of the framework theory.

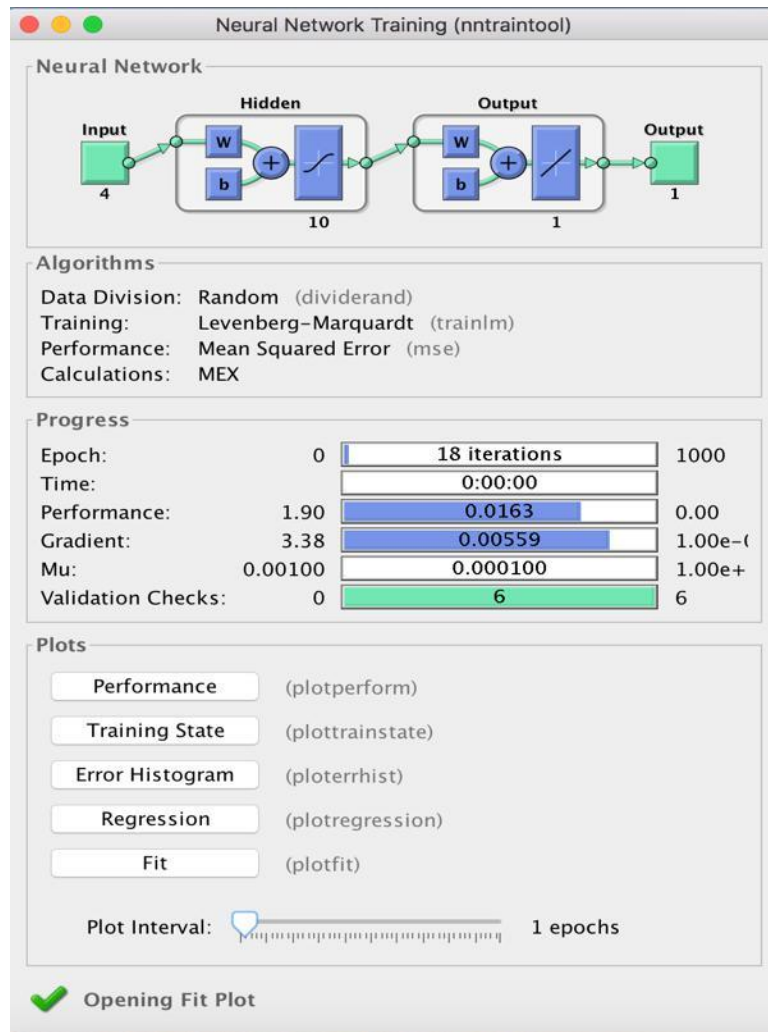


Figure 3.2: Neural Training Tool

Figure 3.2 is the neural network training tool which shows the block diagram of the neural network.

Analysis: The progress section consist of various parameters as epoch, time, performance, gradient, approval checks and each of whose value is specified there.

3.2 BASIC ARCHITECTURE

[10] Digital devices such as digital camera or smartphones are used to take photos of the plant leaves and these images are used to separate out and measure the diseased part area. In order to properly identify the affected region we need image without and impurities so contrast enhancement other image processing techniques are applied on every input image. By enhancing the image using the image processing techniques we get different features of the images which are not visible from the human eye. Below is the flowchart depicting the basic architectural flow.[10]

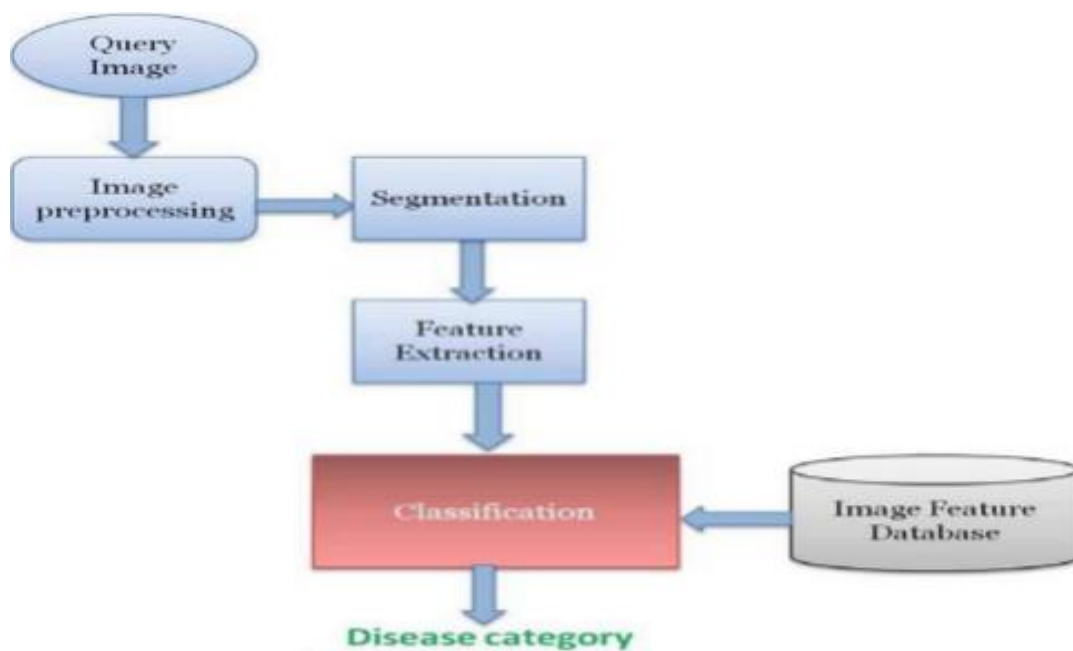


Figure 3.3: Block diagram of basic architecture

Image acquisition is the first step in the image processing. Any image is taken from the digital devices thereafter image preprocessing and segmentation are done and necessary feature are extracted from image. During classification, with the help classification algorithm different clusters are made. We used k means color based clustering in order to detect and identify the affected area. By selecting the one of the cluster our algorithm uses support vector machine for the prediction of disease.

3.3 INFORMATION POSIT

An information posit is an amassing of interconnected information of interconnected data which can be placed to solely, mix and administrated with everything taken into account.

An instructive accumulation is dealt with few sort of data structure. In a data collection, for example, an instructive accumulation can have a social affair of entrepreneurship data (information, bargains diagrams, and whatnot). The information posit is seen as an educational accumulation, as a collections of data it related to an explicit sort of information, for instance, bargains data for an explicit corporate office. Informational index utilized in the venture is put away in an exceed expectations document and further utilized for getting results in MATLAB programming.

Our informational index is an accumulation of four qualities which are as per the following:

- The main part of our information includes a diagram of the leaves which have the property of both diseased and diseases free leaves which again are investigated and various result are made appropriately.
- The second part of the image give description of the hue part which demonstrate the Overwhelming shading.
- The third stage is the saturation stage that demonstrate the immaculateness of image or the stretching on exposure to the light.
- The last stage gives the information of the intensity component of the image and also demonstrates how it varies.

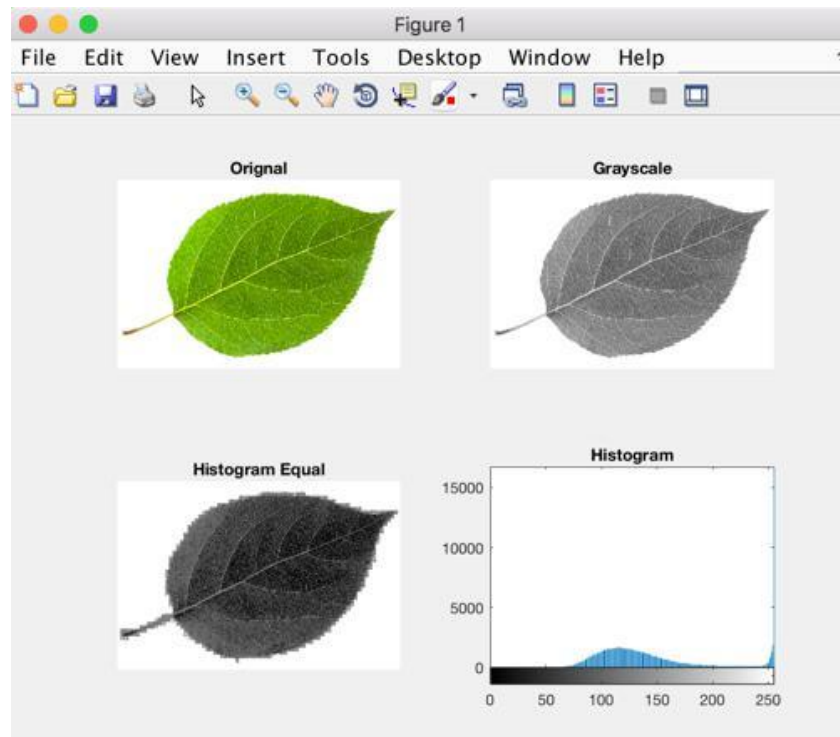


Figure 3.4: Histogram of healthy leaf

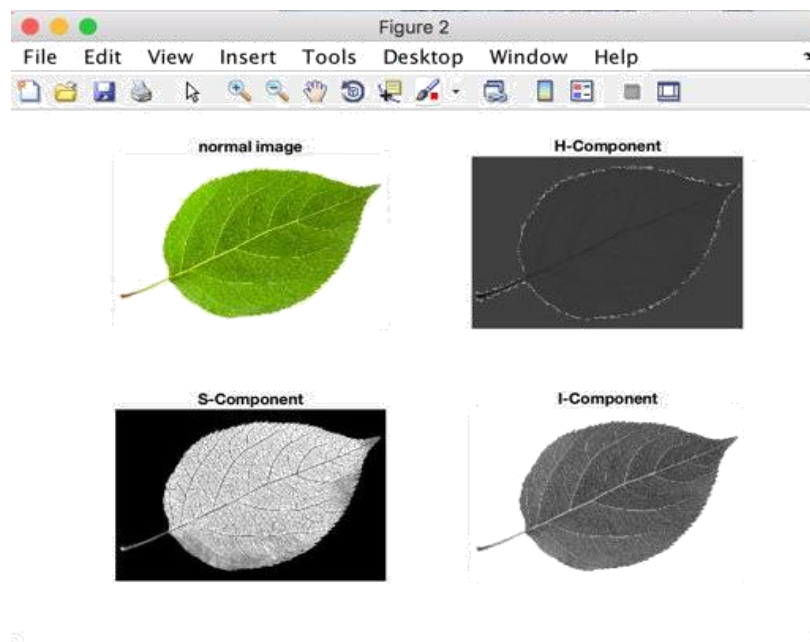


Figure 3.5: HIS representation of normal leaf

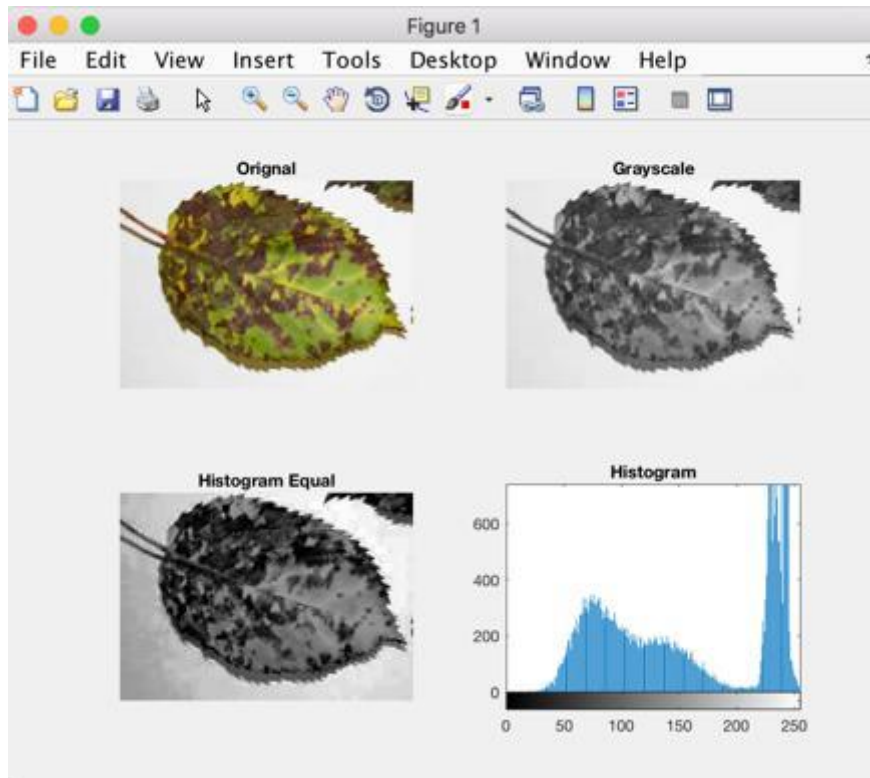


Figure 3.6: Histogram of diseased leaf

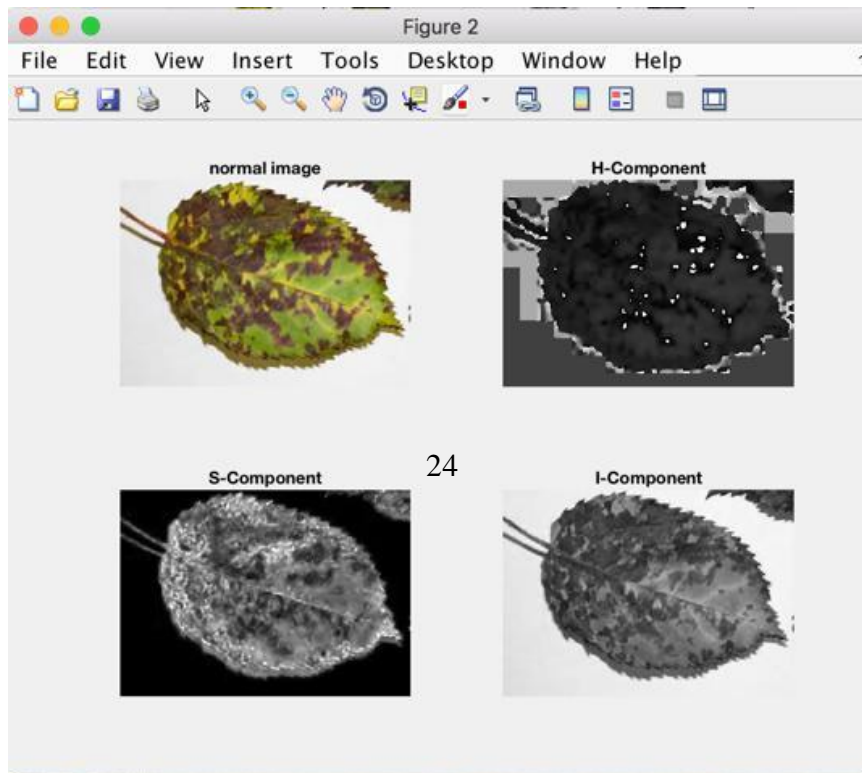


Figure 3.7: HIS representation of Diseased Leaf

CHAPTER-4

ALGORITHM

4.1 GENERAL

In this chapter we describe the implementation details of the algorithm and steps in which our algorithm works for the identifying the disease. K means color based clustering algorithm is used in this problem. Below are the steps in which an input image is processed.

- Image acquisition
- Image preprocessing and enhancement
- Image segmentation using k means and Otsu classifier
- Feature extraction

4.2 IMAGE ACQUISITION

Image acquisition means to collect different type of samples for the formation of the input dataset. Dataset images further go through the various steps. In order to provide best solution to any problem it is necessary that dataset cover majority of the different type of inputs. We have covered different plant leaves. Different image formats are taken in our dataset. Any other image excluding dataset can be used in our algorithm provided its size is matching and format is known, our algorithm and its classifier gives the prediction of the disease for that random image.

4.3 IMAGE PREPROCESSING AND ENHANCEMENT

It is the second phase in digital image processing. In this using MATLAB input image noise is reduced, pixel values gets more classified, spot reduction and contrast enhancement is there. The purpose of the image preprocessing and enhancement step is that after preprocessing the image its get easy to separate the infected area. Indirectly the classifier we use works better with preprocessed image without and impurities. Values of the pixels also get adjusted in the preprocessed image. MATLAB provides user number of different filters for the enhancement of the image.



Figure 4.1: Image Preprocessing Original Image.



Figure 4.2: Preprocessed Image

In figure 4.1 we can see the difference between the two of the images. The first image is the input image and the second one is the preprocessed image in which we enhance the contrast other features of the image using the MATLAB. Preprocessing let the users to reduce the noise in the image and overall quality of the image is improved. The leftover space in the first image is also gets separated in the second preprocessed image. For every different color the contrast gets more enhanced for every pixel value.

4.4 IMAGE SEGMENTATION

Image segmentation is the third step in our purposed methodology. After image acquisition and preprocessing we segment the image in to three different clusters. For clustering we use k means color based clustering algorithm and the classification is done through the Otsu classifier. After making three different clusters user is asked to choose one of the cluster and then disease is predicted using that particular cluster.

4.5 K MEANS COLOR BASED CLUSTERING

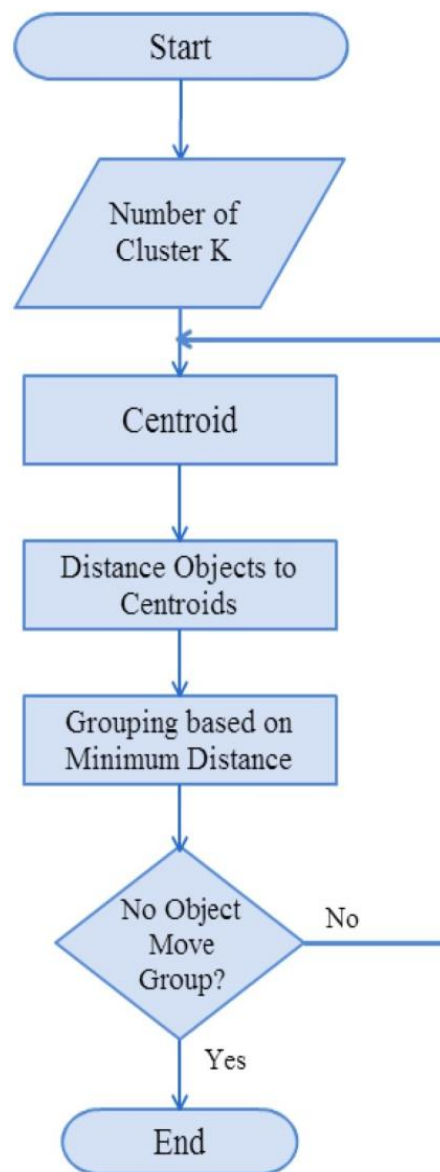


Figure 4.3: K Means Algorithm

[12]Above flowchart explains the normal working of the k means clustering algorithm. A centroid is chosen and distance of the other objects is measured from that centroid using Euclidean shortest distance method. In this way different clusters are made. Implementation of the K mean algorithm is as follows:

1. Input preprocessed image
2. Convert the given color space to lab color space
3. Normal color space of the image is RGB color space which consist of the primary colors red, green and blue.
4. It is easy to extract the different features from lab color space in comparison to the RGB color space.
5. Lab color space is also represented by the $L^*a^*b^*$, in which l stands for lightness, b stands for the blue/yellow and a stands for the red/green value. This color space is way to communicate with the colors and locate them. All the information of colors lies around a^* and b^* axis.
6. Different clusters are made on the basis of different colors using k mean clustering. Clusters are made using distance between the initial groups and the distance of the other from that group. The nearest one are chosen. For this we use Euclidean shortest distance formula.
7. Grouping of the clusters is based on the minimum distance from the centroid.

4.6 OTSU's CLASSIFIER

Otsu method is one of famous method for clustering based image thresholds. It is named after famous scientist Nobuyuki Otsu. This method automatically perform clustering based image Thresholding. This method automatically convert and reduce the image from the grey level to the binary one. The algorithm assumes that the image contains two classes of pixels following bi-modal histogram (foreground pixels and background pixels), it then calculates the optimum threshold separating the two classes so that their combined spread (intra-

class variance) is minimal, or equivalently (because the sum of pairwise squared distances is constant), so that their inter-class variance is maximal.



Figure 4.4: Cluster One



Figure 4.4: Cluster Two



Figure 4.4: Cluster Three

Above figures are the different clusters which are formed using the OTSU classifier.

4.7 FEATURE EXTRACTION

Various features color oriented, shape oriented are extracted using MATLAB commands. These features plays significant role in determining the affected region of the leaves. Various features like mean, standard deviation, variance and others are extracted in order to know the health of the leaf.

FEATURES	
Mean	48.6186
S.D	74.9293
Entropy	4.69867
RMS	9.88411
Variance	5360.22
Smoothness	1
Kurtosis	3.12776
Skewness	1.28105
IDM	255
Contrast	3.24236
Correlation	0.659386
Energy	0.34533
Homogeneity	0.816415

Figure 4.5: Features of segmented ROI

Above figure shows the various shape and color oriented features which are extracted for every image.

CHAPTER 5

TEST PLAN

5.1 GENERAL

In our methodology the initially seventy five different images are take which belongs to different leafs. After image preprocessing and image segmentation images are classified into three different clusters .Initially one of the cluster is chosen and disease is predicted for that particular cluster. There after users are also provide options to calculate the affected region area and also we can calculate the percentage of the area. The figure describes the GUI of our application.

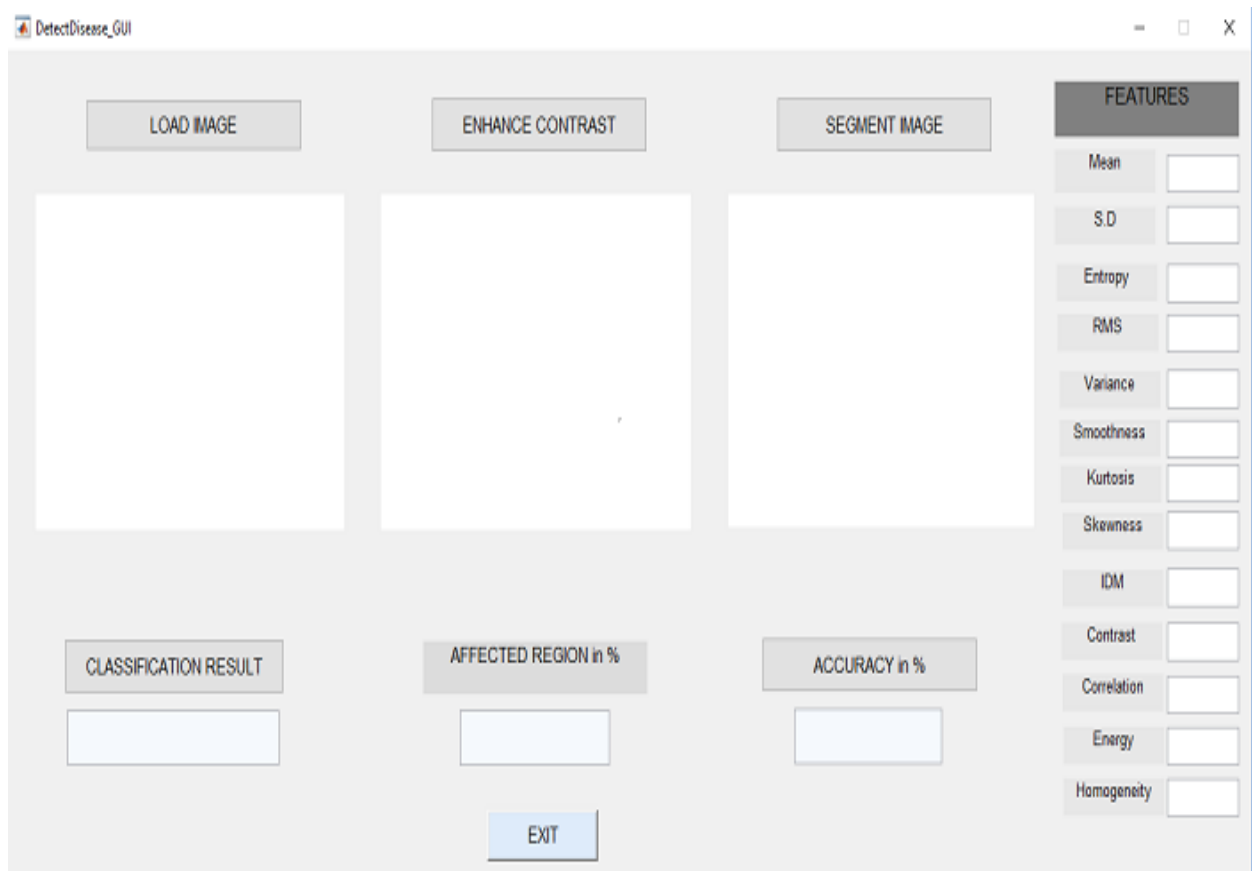


Figure 5.1: GUI

In user interface plot we have different buttons which are load image, enhance image and segment image. Above GUI is designed in MATLAB using guide command. Load image lets the user to choose image from the dataset. When user select one of the particular diseased or healthy leaf image then that particular image is loaded in the user interface.

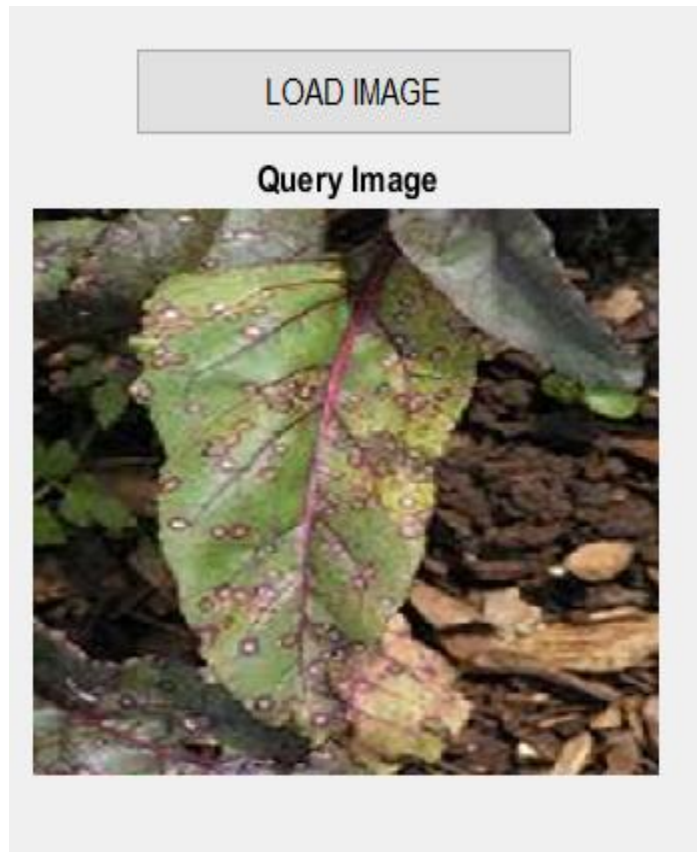


Figure 5.2: Input Image

Above figure shows user the image which is loaded from the input dataset. One of the image corresponding to particular type of the disease is chosen by the user and it gets loaded in the user interface for further process. Query image can be replaced with other images from the input dataset by clicking again on the load image button.

In figure we see the another button named enhance contrast. This tab is basically responsible for the preprocessing of the loaded image before sending it to the segmentation process. Noise reduction and pixel gets more clarified in this phase.

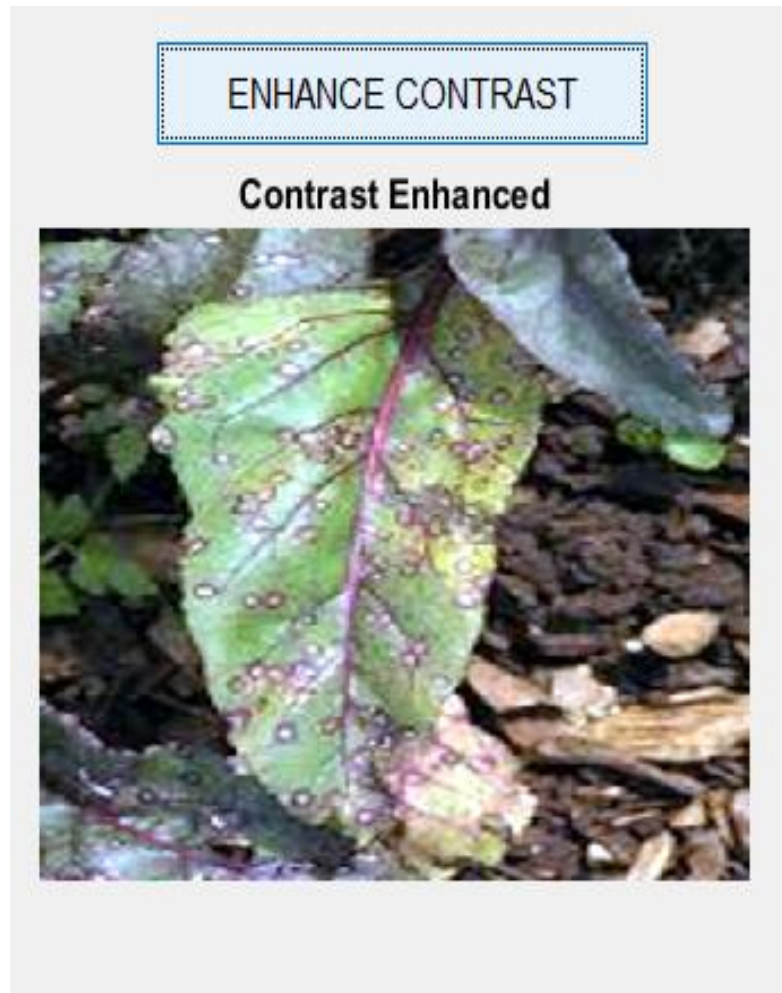


Figure 5.3: Enhanced Preprocessed Image

The above figure shows the contrast enhanced image of the input image. In this phase if in any diseased leaf of the plant dark level spot are there then contrast enhancement button highlight those buttons and it gets easy to segment the image into the different clusters for the classification and the prediction of the disease. There is improvement of the unwanted distortions and enhances the features of the image for further process.

The third button named image segmentation is third step in image processing of input image. In this phase main algorithm starts working. K means color based clustering algorithm and the classification using the Otsu method is done in this phase.

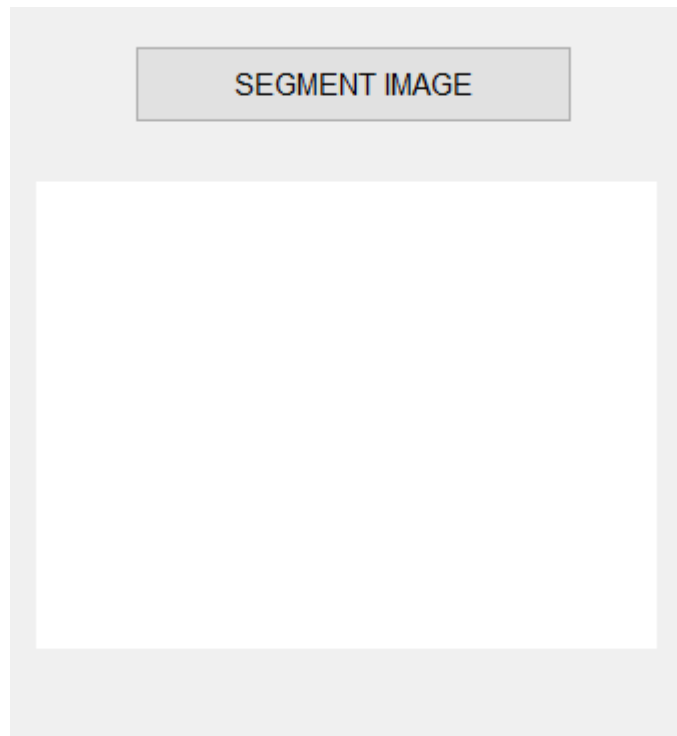


Figure 5.4: Segmented Image Box

The third button segment image button. When the user click this particular button then user is directed to new window show the original image and its three different clusters. Image segmentation is done using k mean clustering. Color based clustering is done in k means and best three clusters are displayed. For the classification purpose we use the Otsu classifier which classifies the type of the image in the chosen image.

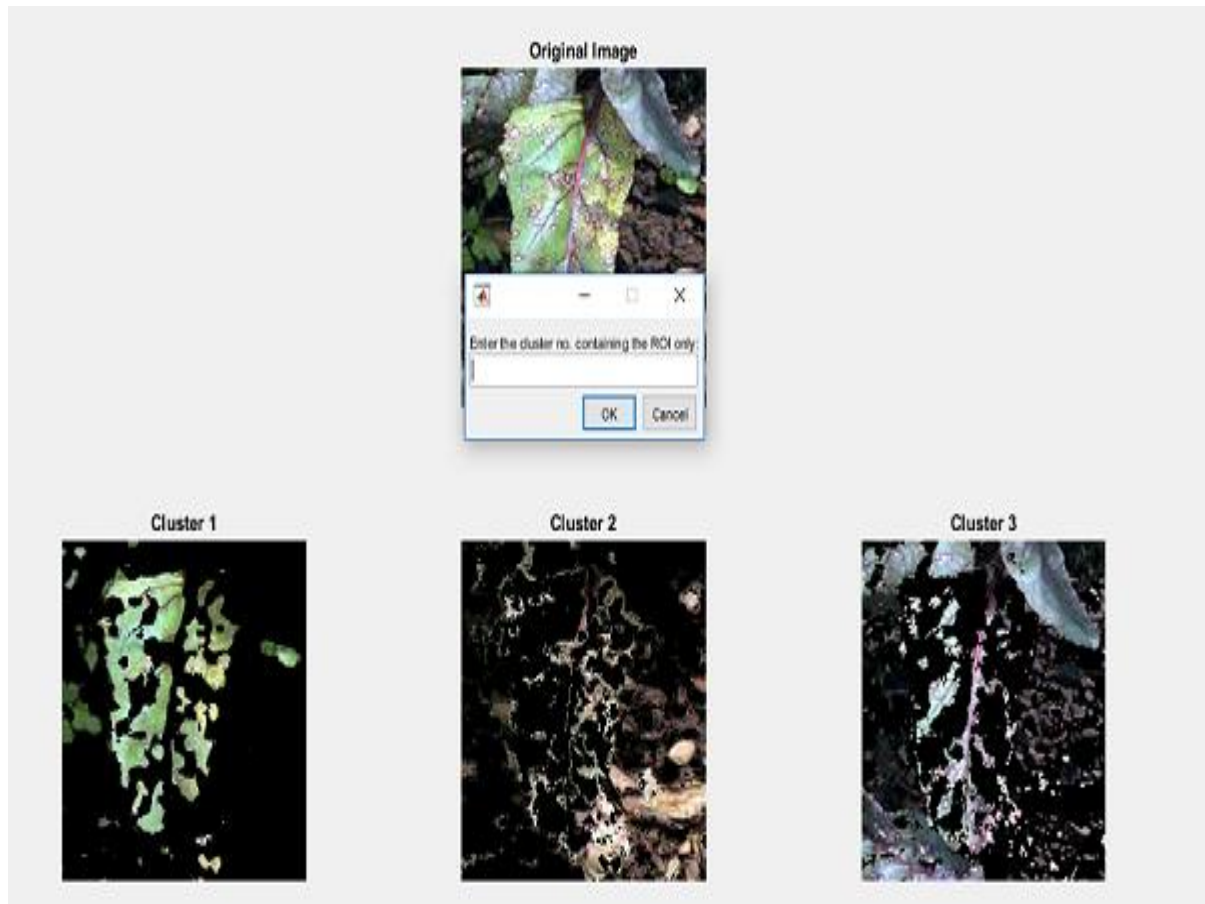


Figure 5.5: Segmented Image Clusters

Figure 5.5 shows the segmented image of the input image. Original image and three different clusters of that particular image are shown in the figure. Also input dialog box is shown which asks the user to enter the cluster number in which the user has the region of the interest. User is required to choose the cluster in maximum color separation is there.

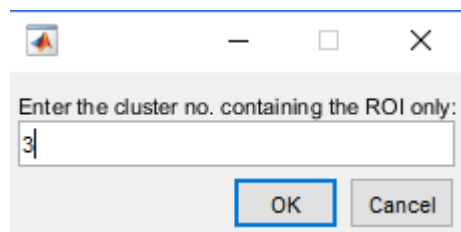


Figure 5.6: Input dialog box

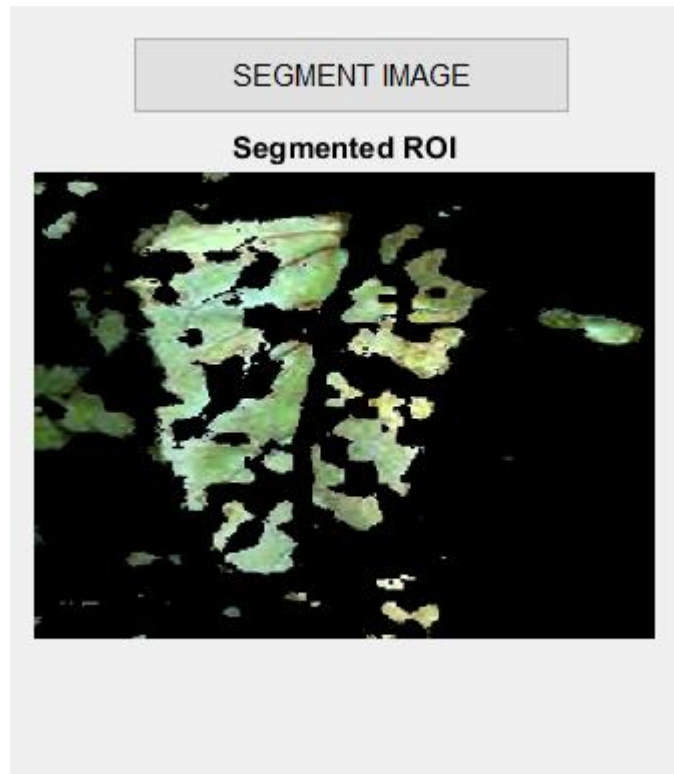


Figure 5.7: Segmented ROI

the above figure shows the cluster with the region of the interest gets loaded into the segment box. This image passes through OTSU classifier and different shape and color oriented features are calculated using those prediction of disease is made.

CHAPTER 6

RESULTS AND PERFORMANCE ANALYSIS

6.1 GENERAL

The proposed algorithm is made to run for each individual image. In our solution we have covered four different type of diseases which are Alternaria Alternata, Anthracnose, Bacterial blight, Cercospora leaf spot. Given below figures shows the detected disease for input image from a particular disease dataset.

6.2 ALTERNARIA ALTERNATA

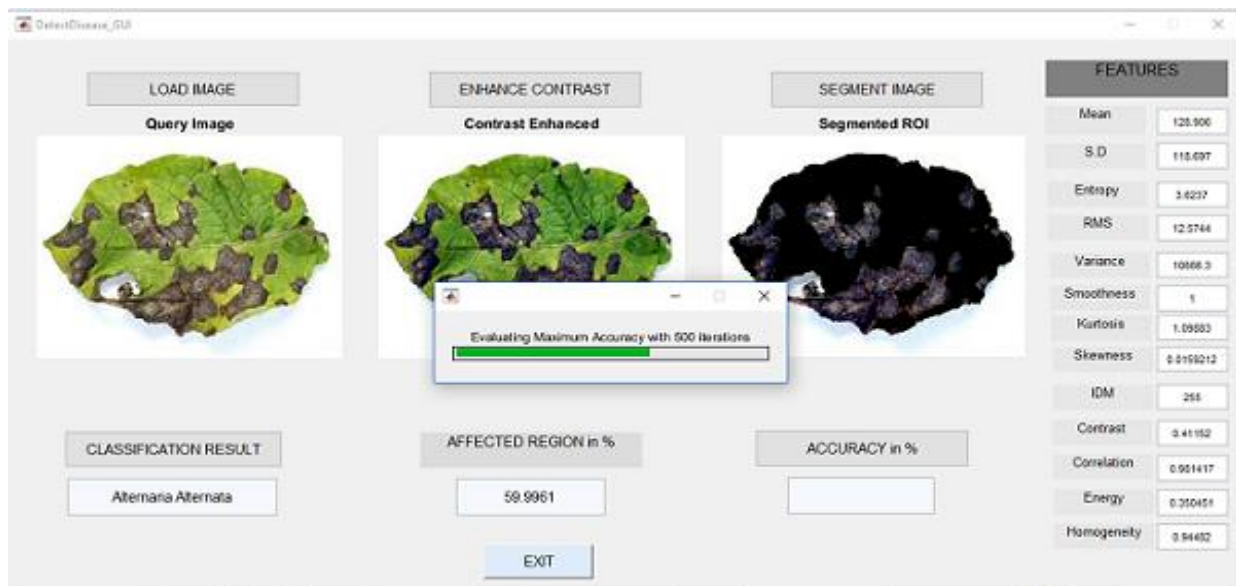


Figure 6.1: Detection of Disease

Analysis: Plant leaf infected with the alternaria alternate is loaded from database. Contrast enhancement and preprocessing of image is done in the second phase. In image segmentation column one of the cluster is loaded .As the above figure shows the disease classified as Alternaria Alternata .Also area of the affected region in percentage is also shown. To check the accuracy of the our proposed methodology the image is passed through five hundred iteration and every time different clusters is chosen by the algorithm and then accuracy is predicted. The figure shows the accuracy

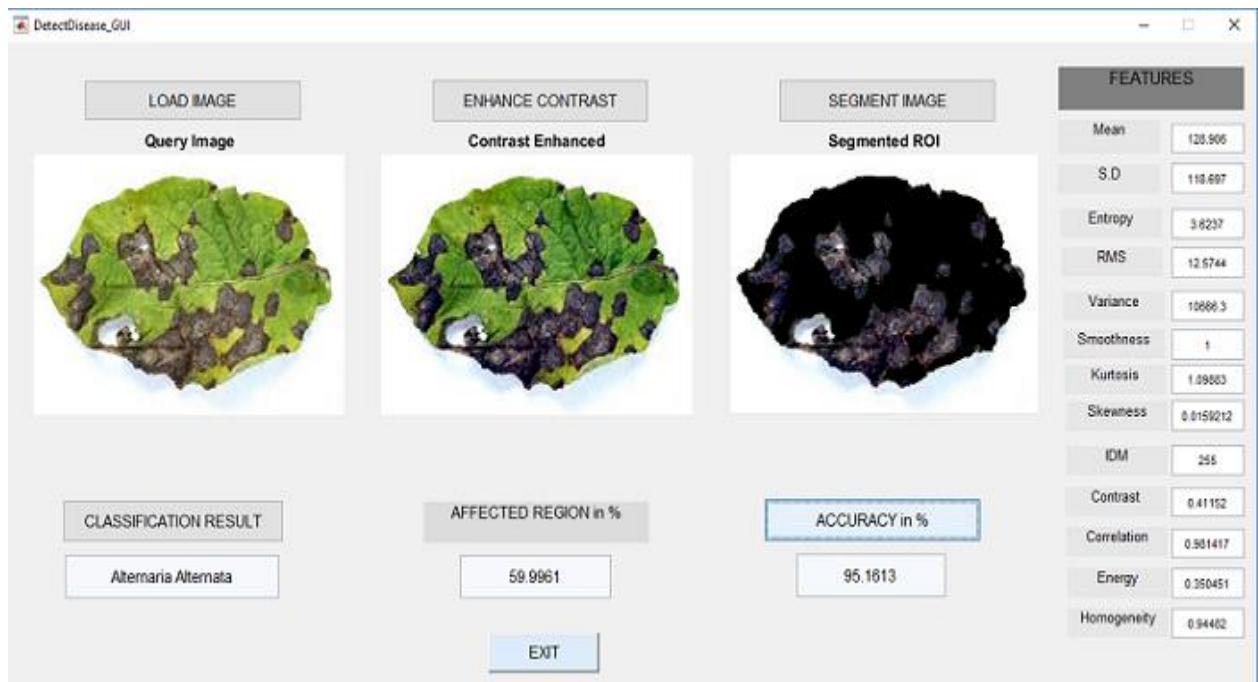


Figure 6.2: Disease Result 1

6.3 ANTHRACNOSE

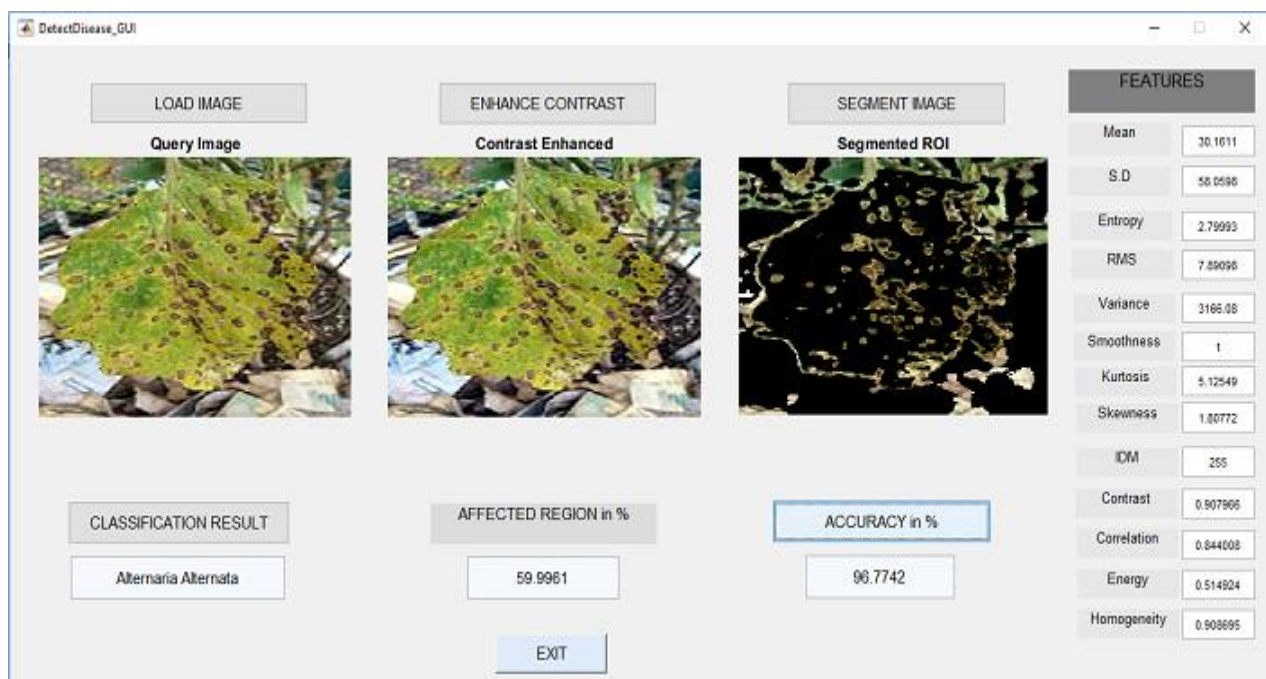


Figure 6.3: Disease Result 2

Analysis: the figure shows the disease classification and prediction of the leaf image infected with anthracnose. Same steps are performed. In this case we can see that disease is classified as the Anthracnose and nearly sixty percentage of area is affected by this disease and the accuracy comes out to be ninety six percentage. Various features which are extracted are also displayed on the right side of the image.

6.4 BACTERIAL BLIGHT

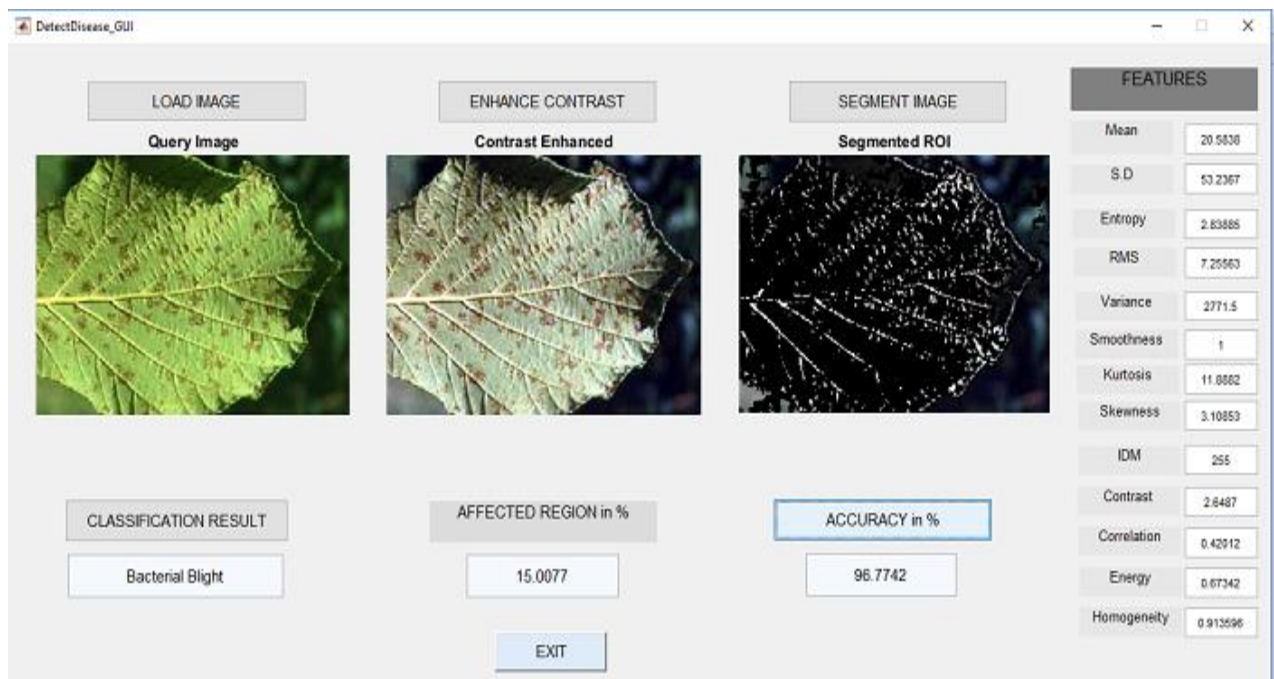


Figure 6.4: Disease Result 3

Analysis: In the figure we can see that the diseases identified for the query image comes out to be bacterial blight. Various features both shape color oriented are shown. Preprocessed image and segmented region of interest is shown. In this case disease affected area comes out to be fifteen percentage on the other hand accuracy comes out to be ninety six percentage.

6.5 CERCOSPORA LEAF SPOT

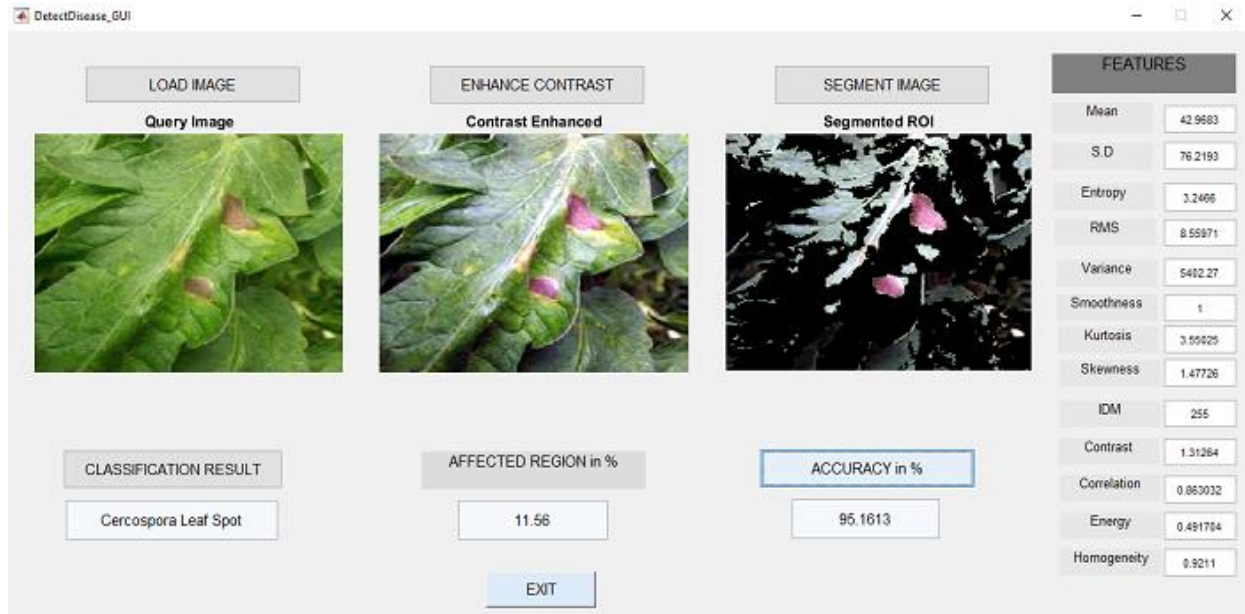


Figure 6.5: Disease Result 4

Analysis: disease classified is Cercospora leaf spot and percentage of the area affected is around sixteen percentage and the accuracy comes out to be approximately ninety five percentage.

6.6 HEALTHY LEAF

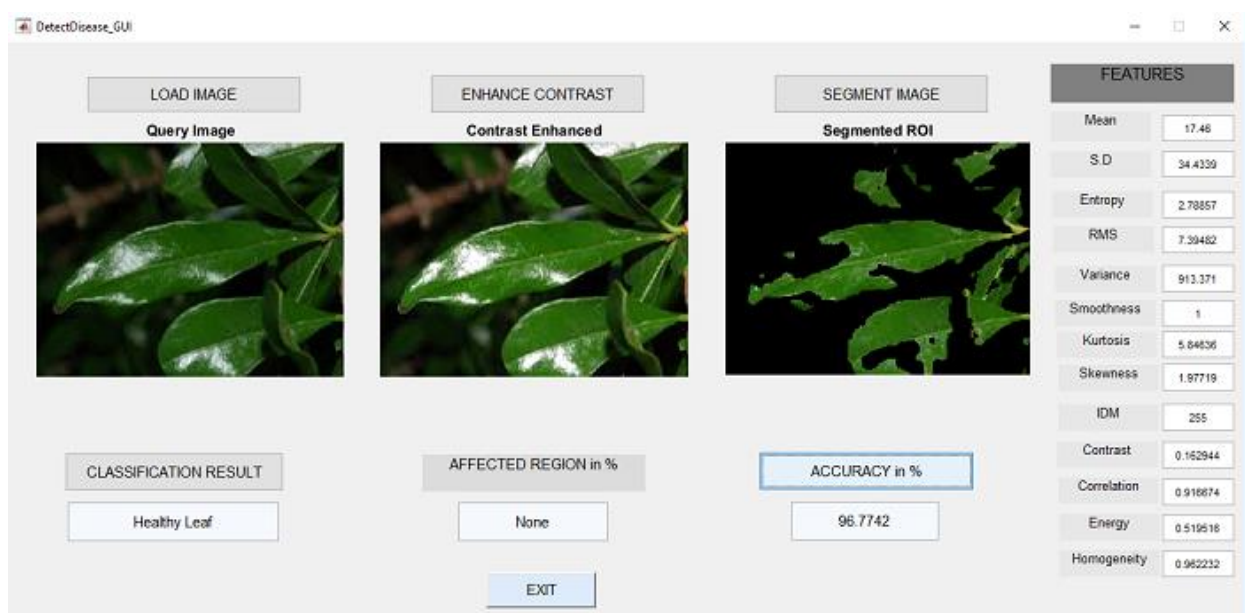


Figure 6.6: Disease Result 5

Analysis: when the query image of healthy leaf is taken then the result is classified as the healthy leaf with no affected region and accuracy of ninety six percentage.

Table 6.1: Estimation of Time and Area affected

Type of the disease	Number of the images	Clustering time (s)	Area of the affected region (%)
Alternaria alternata	22	Below 20s	15.0062
Anthracnose	23	Below 20s	15.0915
Bacterial blight	7	Below 20s	12.0093
Cercospora leaf spot	9	Below 20s	18.009

The tables shows the different number of samples taken and the average clustering time for each individual image. Affected area percentage is also shown.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

The data mining techniques are playing a significant role in the agriculture industry and other industries. Data mining algorithm are easy implement, we can solve complex problems using these techniques. Using image techniques and data mining algorithms we successfully identified the affected area in the plant leaf. Various features of the image are extracted with their numeric values. The algorithm used here is very much efficient and best case time space complexities are achieved. For ninety percentage of images average clustering and processing time is less then twenty seconds. Average accuracy of ninety plus percentage is achieved in every query image of the dataset.

7.2 FUTURE SCOPE

Web based image processing techniques can be implemented. In this user is provide with two modes with and without internet. In case of web base processing remote area users can upload image in system and whole image system techniques and classification algorithm will we implemented in the cloud itself. Real time monitoring of the data is there using the cloud platform.

REFERENCES

- [1] Evolutionary Artificial Neural Networks in Neutron Spectrology, Jose Mauel Ortiz-Rodriguez, Ma. Del Rosario Martinez-Blanco and Hector Rene Vega-Carrillo
- [2] S. Bashir, N. Sharma, Remote Area Plant disease detection using Image Processing, IOSR Journal of Electronics and Communication Engineering , Volume 2, Issue 6 ,pp.31-34,2012.
- [2] S. Arivazhagan, R. N. Shebiah, S. Ananthi, S. V. Varthini , Using texture detection features, recognizing unhealthy region of plant leaves and classification of plant leaf disease , AgricEngInt: CIGR Journal, Vol. 15, No.1, pp.211-217,2013.
- [3] J. Behmann, A. K. Mahlein, T. Rumpf, C. Romer, L. Plumer, A review of advanced machine learning methods for the detection of biotic stress in precision crop protection, Springer Media New York, Precision Agriculture, Volume 16, Issue 3, pp. 239–260, 2015.
- [4] J. G. A. Barbedo, Digital image processing techniques for detecting, quantifying and classifying plant diseases, Arnal Barbedo SpringerPlus , pp.1-12,2013.
- [5] F. Qin, D. Liu, B. Sun, L. Ruan, Z. Ma, H. Wang, Identification of Alfalfa Leaf Diseases Using Image Recognition Technology, Plos Journals, pp.1-15, 2016.
- [6] A. Meunkaewjinda, P. Kumsawat and K. Attakitmongkol, Grape leaf disease detection from color imagery using hybrid intelligent system, 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Volume: 1, pp. 513 - 516, 2008.
- [7] E. Omrani, B. Khoshnevisan, S. Shamshirband, H. Saboohi, N. B. Anuar, M. H. N. M. Nasir, 'Potential of radial basis function-based support vector regression for apple disease detection', Department of Biosystem Engineering, pp.2-19,2014.
- [8] A. Singh, B. G. Subramanian, A. K. Singh, S. Sarkar, Machine Learning for High-Throughput Stress Phenotyping in Plants, Trends in Plant Science, Volume 21, Issue 2, pp. 110-124, 2016.
- [9] V. Singh, segmentation 49,2017. [10] A. Camargo, J. S. Smith, An image-processing based algorithm to automatically identify plant disease visual symptoms, Biosystems Engineering , Volume 102, Issue 1, pp. 9-21, 2009.

- [10] Leaf Disease detection of the cotton plants using image processing techniques, Pranita P. Gulve, Sharayu S. Tambe, published in 2015
- [11] ,AnInvestigat -ion Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9, 2016.
- [12]Chen CH, Huang WT, Tan TH, Chang CC, Chang YJ, Using K nearest neighbor classification to detect abnormal Lung Sound, Sensors (base) .
- [13]M.Jhuria, A.Kumar, R.Borse, Image processing forSmart Farming : Detection of disease and food grading, Institute of Electrical and Electronics Engineering, 2013. [14] N. Bhardwaj, G. Kaur, P.K.Singh, A Systematic Review on Image Enhancement Techniques, Sensors and Image Processing, Advances in Intelligent Systems and Computing (AISC)
- [15] G. Kaur, N. Bhardwaj, P.K.Singh, An Analytic Review on Image Enhancement Techniques Based on Soft Computing Approach, Sensors and Image Processing, Advances in Intelligent Systems and Computing(AISC), Springer, pp. 255-265, 2018.
- [16] K. Vasudeva, P.K. Singh,Y. Singh, A Methodical Review on Issues of Medical Image Management System with Watermarking Approach, Indian Journal of Science and Technology, Vol 9(32), DOI: 10.17485/ijst/2016/v9i32/100188, ISSN: 0974-5645, 2016.
- [17] D. Agarwal, A. Gupta, P. K.Singh, A systematic review on Artificial Bee Colony Optimization Technique, International Journal of Control Theory and Application, Vol. 9(11), pp. 5487-5500, 2016,
- [18] A. Sharma, P.K. Singh, P. Khurana, Analytical Review on Object Segmentation and Recognition, published in proceedings of 6thInternational Conference - Cloud System and Big Data Engineering (Confluence), 14th - 15thJanuary, 2016, Noida, India, IEEE, pp. 524 – 530,2016.
- [19] R. Bhardwaj, P .K. Singh, Analytical Review on Human Activity Recognition in Video, published in proceedings of 6th International Conference - Cloud System and Big Data Engineering (Confluence), 14th-15thJanuary, 2016, Noida, India, IEEE, pp. 531 – 536,2016.

