T.C.

ALTINBAS UNIVERSITY

Electrical and Computer Engineering

PLANT DISEASE DETECTION USING SVM CLASSIFICATION

Ghassan Faek Shakir Al Daban

Master Thesis

Supervisor:
Asst. Prof Dr.Sefer Kurnaz.

Istanbul, (2019)

PLANT DISEASE DETECTION USING SVM CLASSIFICATION

Ghassan Faek Shakir Al Daban

Electrical and Computer Engineering

Submitted to the Graduate School of Science and Engineering

In partial fulfillment of the requirements for the degree of

Master of Science

ALTINBAŞ UNIVERSITY

2019

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Engineering.

Ass Prof. Dr. Sefer Kurnaz

Supervisor

Examining Committee Members (first name belongs to the chairperson of the jury and the second name belongs to supervisor)

Asst. Prof. Dr. Sefer Kurnaz	School of Engineering and Natural sciences Altinbaş University	____________
Assoc. Prof. Dr.Yasa Eksioglu Ozok	School of Engineering and Natural sciences Altinbaş University	____________
Asst. Prof. Dr. Adil Deniz DURU	School of physical education and sports teaching Marmara University	____________

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Engineering.

Asst. Prof. Dr. Çağatay Aydın

Head of Department

Approval Date of Graduate School of Science and Engineering: ____/____/____

Prof. Dr. Oğuz BAYAT

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Ghassan Faek Shakir Al Daban

Director

DEDICATION

I dedicate this thesis to my family especially my mother who helped me a lot, classmates and all my friends for the support and encouragement throughout my education and life. Special dedication goes to my supervisor Asst,Prof Dr.Sefer.Kurnaz, my mother, my wife, daughter and my sisters and for their support and prayers during my research work.

ACKNOWLEDGEMENTS

I might want to offer my thanks to every one of the individuals who have upheld me all through the regularly extended periods of this voyage. I might want to thank my advisor, Asst.Prof Dr.Sefer .Kurnaz for being my compass notwithstanding when I believed I was lost and being in extraordinary part in charge of the zenith of this work. I might likewise want to thank my supervisors for their supportive exhortation, which incredibly enhanced the nature of this work. Finally, I thank this institution for hosting me during these years, securely earning its place as my home. I am very thankful Especially to my mother who helped me a lot in completing my studies and dad, whose values and education motivate me to keep asking questions; to my siblings and family for their infinitely appreciated love and to my country which, although inanimate, keeps me anchored and offers me an example of strength and challenging the difficulties.

ABSTRACT

PLANT DİSEASE DETECTION USING SVM CLASSIFICATION

Daban, Ghassan Faiq Shakir

M.S., Electrical and Computer Engineering, Altınbaş University,

Supervisor: Asst.Prof. Dr. Sefer Kurnaz

Date: May, 2019

Pages: 62

This paper introduces the research of techniques that we can use to classification different plant leaves diseases, Image Processing Using MATLAB and Support Vector Machine (SVM) to Detect Diseases that Occur on Plant Leaf. The importance of this research in particular in the detection of disease (Tuta absoluta) spread in Iraq, which caused the death of a large number of plants, which increases the desertification of the earth and lack of vegetation, it will help the Agriculture Officer or farmers to check the quality of crop without any manual supervision. Data obtained from farmers will help to identify the disease produced in a particular region and thus take proper action to prevent or control it. This program use algorithm to calculate variance and density in the infected area, segmentation the image by applying the techniques of image pre-processing in the leaf of the plant. is an important stage for detection and knowing the disease that infected in the plants automatically, the Disease spots are different in texture and color, in comparison with the healthy leaf color. Then the color transform of the color image (RGB) can be used for segmentation of infected area. In this paper, we will compare the effect of HSI components in the process of detection of the disease area in the plant leaves.

Keywords: Image process, feature extraction dataset, classification, support vector machine (SVM).

TABLE OF CONTENTS

Pages

ABSTRACT……………………………………………………………………………….…vii

LIST OF TABLES x

LIST OF FIGURES xi

LIST OF ABBREVIATIONS 13

1. INTRODUCTION.. 13

1.1 METHODS OF SPOT DISEASE DETECTION 14

1.2 A SIMPLE DESCRIPTION OF WORK STEPS 15

1.2.1 Input Image ………………………………..……………………..…...……15

1.2.2 Apply Contrast ……………………..……………………..……...……..……15

1.2.3 Segmentation Image ……………..………………………………….………..…16

1.2.4 Choose the Region of Interested (ROI) Area…………………….………….……16

1.2.5 Image Classification …………………………………………………………16

1.2.6 Leaf Image Database ………………………………………………………..…16

2. BACKGROUND... 18

2.1 METHODS OF SPOT DISEASE DETECTION. 18

2.2 USEFUL METHODS …………………...…………….…………………………18

2.3 METHODS TO ENHANCE THE PLANT LEAVES IMAGE ……….……..…..……20

2.3.1 Green Band Processing …….…………………..……………….………..….....21

2.3.2 Gray Scale Conversion ………………………….……………………...……….....21

2.3.3 Adaptive Histogram Equalization ……………….…………….…....…………….…22

2.3.4 Apply Smoothing Filter ……………………...……………….…....….………..22

3. PROPOSED WORK METHODOLOGY.. ……………………………..………………23

3.1 USING COLOR TRANSFORM ON LEAF IMAGE……….……………..….….……24

3.1.1 Use HIS Color Model (Hue Saturation Intensity)…………………………..….…25

3.2 IMAGE SMOOTHING ……………………………………………………...…..…….27
3.3 SEGMENTATION OF DISEASE AREA…………………………………..…..…..…….28
3.4 CLUSTERING USING K-MEANS TECHNIQUE……………………..……...……..29
3.5 FEATURE EXTRACTION……………………………………………………………...30
3.5.1 Training Dataset is Contain Training_Features and Training_ Label ……...…...…31
3.5.1.1 The Training_features …….……………………………………..……..……31
3.5.1.2 Training_label……..………………………………………………...…….31
3.5.2 Accuracy Dataset is also Contain Training_Features And Training_Label………..32
3.5.2.1 The training_features …………………………………….……..………32
3.5.2.2 Training_label………………………...………………………...…….……32
3.6 CLASSIFICATION USING SUPPORT VECTOR MACHINE(SVM)……….……42
3.6.1 Some Advantage of the Support Vector Machine ……………………..…..…..……43
3.6.2 The Shortcomings and Disadvantages of SVM…………………….……………..43
4. THE WORK PART (IMAGE PROCESS STEP ………………...…………………..…44
4.1 IMAGE RESIZE …………………………......…………………………………..….…44
4.2 USE IMAGE ADJUST………………………….…………………………………..…45
4.3 USE GLOBAL IMAGE THRESULD USING OTSU's METHOD………..…….………46
4.4 USE HSI FUNCTION (Hue Saturated Intensity)………………………….…………47
4.5 USE 'MAKECFORM FUNCTION ……………………………………………………51
4.6 USE K MEAN CLUSTER …………………………………………….…………52
4.7 USE BWCONNCOMP……………………………………………………………………55
4.8 USE REGIONPROPS (MEASURE PROPERTIES OF IMAGE REGIONS)…………..55
4.9 USE GLCM (CREATE GRAY-LEVEL CO-OCCURRENCE MATRIX FROM IMAGE)56
5. DISCUSSION……………………..………………………………………...………57
6. CONCLUSION………………………………...……………………………………59
6.1 FUTURE WORK SCOP………………………………………….……………..…60
REFERENCES... 61

LIST OF TABLES

Pages

Table 3.1: Feature extracted from 25 specimen of (Tuta Absulota disease)…… …………...…40

Table 3.2: Feature extracted from 25 specimen (Anthracnose disease)…………….. … … …..41

LIST OF FIGURES

Pages

Figure 1.1 : Images of plant leaves infected with bacterial diseases ………………..………..…...14

Figure 1.2: The disease occur in the plant leaves …………………………..……...……….………..14

Figure 1.3: Image Resize …………………………..……...............................…………….…..16

Figure 1.4: Contrast Enhanced for Leaf Image …………..……...............................………………..16

Figure 1.5: The disease occur in the plant leaves …………………………..……..…..…..………..17

Figure 1.6: Chosen Image from three Cluster images ………………………………...………..…..17

Figure 1.7: Types of a plant diseases that we will study [11][16]………………………..........…….18

Figure 2.1: Types and the different between Dicot and Monocot [6]……………………..……….19

Figure 2.2: Flow diagram of image enhancement...22

Figure 3.1: Flowchart for Disease area Detection [14]..24

Figure 3.2: leaf infected with (tuta absoluta) Plant Disease..26

Figure 3.3: Description of the Filtering process [21].…………………………..……………..……28

Figure 3.4: Plant Leaf Segmentation to find Disease spots …………………..……….…...…..……..29

Figure 3.5: Explains how the support vectors work with the separated dataset points[19]…....….43

Figure 3.6: SMV Recognition Area ………………………………………………………..……...44

Figure 4.1: Image resize process ………………………………………………………..…....……45

Figure 4.2: Image Enhancement Process ……………………………………………….…...……47

Figure 4.3: Apply Threshold Using Otsu's Method ……………………….……………………….48

Figure 4.4: Apply HIS function to the RGB image ………………………….....……….…………50

Figure 4.5: The Hue part image from the HIS …………………………………....………,...........…51

Figure 4.6: The Intensity part image from the HIS ……………………………..,….….……52

Figure 4.7: Image labeled by cluster index using K mean …………………………….…………….54

Figure 4.8: Result image after apply image processing steps …………………………..……..….55

LIST OF ABBREVIATIONS

HIS : Hue, Intensity and Saturation color space Model.

ROI : Region Of Interesting of Disease spot.

SVM : Support Vector Machine.

AHE : Adaptive Histogram Equalization

RMS : Root Mean Square

IDM : Data Message use to Automatic Meter Reading to read utility meters

GLCM : Gray-Level Co-Occurrence Matrix

1. INTRODUCTION

The plant diseases detection and treatment, is an important factor in human life, the plant diseases detection contributes to the preservation of the environment, provision of food, treatment of global warming and desertification. The farmers usually can identify diseases by observing and seeing diseased trees and plants with the naked eye. this method for detecting the disease visually is used just by the expert's farmers, and the agriculture staff who knowledge and experience in changes that occur in the color and tissue form of plant leave when plants are infected with plant diseases. This method is very hard, take a long time and impossible for large farms . can detect the same part of leaf disease in the leaf spot by different experts as a different disease. Then a quick and accurate approach to identify the plant leaves diseases is needed To increase the accuracy of the method is used. Some people interested in agriculture have used the techniques of image processing to get easy and accurate plant leaves diseases detection of [1]- [2]. The steps of work used by those researchers in the in classification and detection of plant leaf spot diseases: Take the pictures, use image processing, segmentation the disease area, extract the features from the disease area and classification the disease, The resulting accuracy of the method depends on the use of disease area detection. The noise in the images is the main obstacle in disease area detection, which is often due to light, change in illumination, ,camera flash, Preferably the background is white so it does not resemble the color background the color degree of green leaf because it is difficult to determine the diseased spot required ROI (Region Of Interest) in the image segmentation if it similar and presence of the vein in the plant leaf. Hence, we need a method to wipes out the noise (like the light or color points similar to the plant leaf) and provide the best disease area segmentation.

A new disease has been added, the plant leaf digger (Tuta absoluta), to be discovered in this research, the disease has spread in Iraq recently and caused of a lot of losses in agricultural crops and in the Vegetation of , where the Tuta Absoluta insect put eggs on the plant leaf and when the cocoons out from the eggs these cocoons digging tunnels in the tissue of the leaves and may reach the fruit and leg of the plant also, when the spread of these larvae on Most of the plant leaves will stop supplying the plant with food from the leaves by (photosynthesis), causing damage to crops and may also lead to the death of plants.

1.1 THE TYPE OF PLANT DISEASES.

The plant diseases occur on horticulture/agriculture, Plant diseases are divided according to the cause of the disease, namely, bacterial, fungal, insects (include the worms), viral, element lack. According to a study by the University of Agricultural Sciences, Dharwad, INDIA. The symptoms the signs of disease are observed on plant leaves, the methodology for identification of plant diseases affecting leaves of agriculture crops have been developed. In order to know and determine the type of disease that affects the plant for the purpose of prevention and treatment, figure 1.1 shows a sample of images infected by plant diseases in horticulture plants and the symptoms of plant disease exhibit different properties in the leaf as the texture, color and shape.

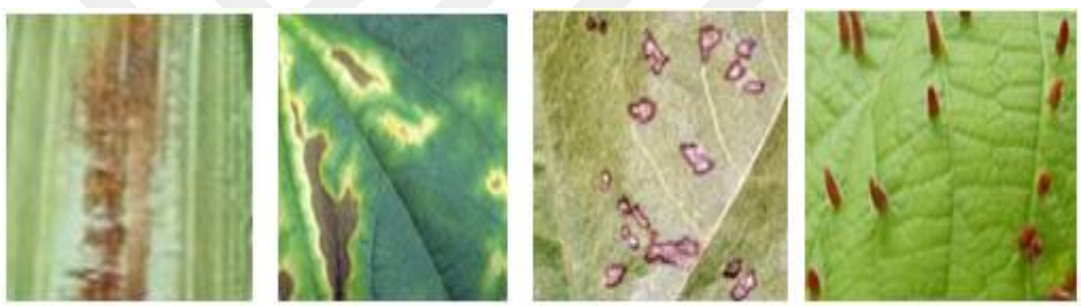

Figure 1.1: Images of plant leaves infected with bacterial diseases from the left 1) Stalk rot disease 2) bacterial wilt disease 3) angular leaf spots disease 4) galls of lime disease.

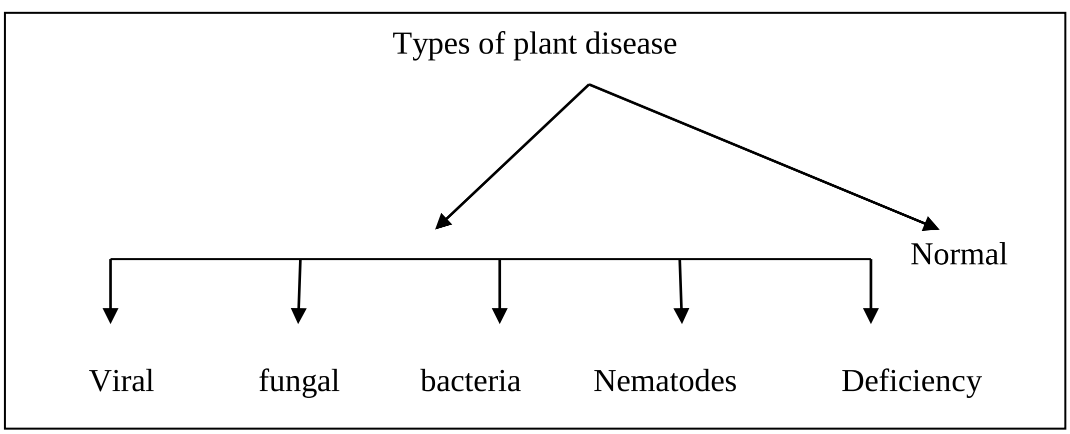

Figure 1.2: The disease occur in the plant leaves [14].

1.2 A SIMPLE DESCRIPTION OF WORK STEPS

Use of Matlab 2017 for image processing

1.2.1 Input Leaf Image.

First input image and resize the leaf images to (255,255)

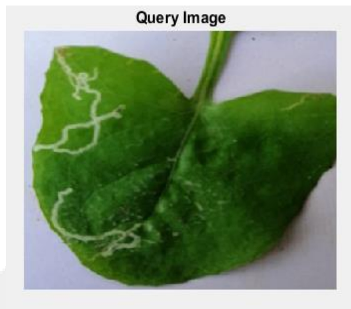

Figure 1.3: Image Resize

1.2.2 Apply Contrast-Enhanced

To clarify the area of the disease in the leaf plant, So that we can diagnose the disease better and more accurately

Figure 1.4: Contrast Enhanced for Leaf Image

1.2.3 Segmentation the Image.

Segmentation the image to three clusters by using K mean cluster.

Figure 1.5: Leaf Image Segmentation

1.2.4 Image Selection

Choose the Region of interested (ROI) The area displays the disease very clearly, and we choose cluster number three as an (ROI)

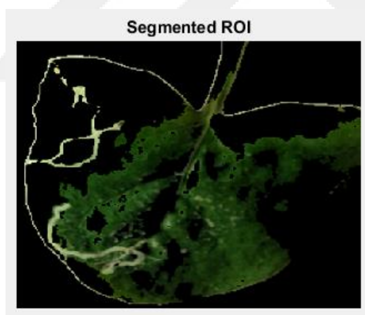

Figure 1.6: Chosen Image from three Cluster images

1.2.5 Image Classification.

Classification of the image by using an SVM support vector machine to detect the type of disease that infected the plant through the plant leaf.

1.2.6 Leaf Image Database

The digital image dataset consists of 100 digital plant leaf contain four types of diseases and the (Tuta Absoluta disease) images were taken using the imaging camera with white background. All the images are collected from an agricultural garden in Iraq in Tikrit city.

The images are stored as color jpeg images (Joint Photographic Experts Group) format with white background and resized to 256×256 pixels for all the objects. generally, The color image is made up of three channel images R.G.B (green, blue and red). The color intensity value of every channel images ranges from 0 to 255. The leaf images are collected from four plant disease categories such as Tuta Absoluta, Blight of Bacterial, Anthracnose and Cercospora Leaf Spot. Some of the sample images are shown in Figure 1.7.

the visual observation on these images of plant leaf disease is not sufficient to categorize these images into a different disease of the plant. therefore, some image process techniques are developed in this study for detecting the disease automatically.

Tuta Absoluta [11]

Bacterial Blight [16]

Anthracnose[16]

Cercospora leaf spot[16]

Figure 1.7: Types of a plant diseases that we will study this research [11],[16].

2. BACKGROWNED

In these sections, We focus on previous research that has yielded good results in the field of detection of plant diseases that serves humanity by providing food and green vegetation cover that reduces the severity of global warming, by cloud-based research, previously published, and in particular in the analysis the dataset that is extracted from the plant leaf image and classification using "Artificial Neural network" and the "support vector machine" SVM.

2.1 METHODS OF SPOT DISEASE DETECTION

In this paper, we study some techniques to segment the disease spots of the plant leaves images. The general concept for the algorithm of plant disease classification is taken digital images from an agriculture environment using a suitable resolution camera. and apply the techniques of image processing to the selected digital images to extract the features that are required for more texture and color analysis. then use the techniques of discriminating analytical to classify the images.
Selected the area of disease effects in the region of interest (ROI). and also discussed extraction Features from the infected leaves images and plant disease classification. we can use the methods of Artificial Neural Network (ANN) for classification the disease in plants leaves such as backpropagation algorithm, self- organizing feature map, but the SVMs From these methods is easier and faster in performance. can be used to more efficiently. to classify and identify various plant diseases accurately using an image pre-processing technique.

2.2 USEFUL METHODS

There are many Methods used to detect the disease from the leaf of the plant, the method introduced by Nunnik Noveana Korniawatee et al [3] for classification and detection of paddy leaves infected. this method used Otsu threshold to detecting the disease in leaves spot , removing the unnecessary spots (the spots that do not contain disease) and using a filter for image smoothing. GengYing et al [4] use the procedures of image processing To find out the disease in plants, to smoothing the image using a median filter. use the technique of threshold to convert filtered image into black and white binary(BW) image and then using the technique of edge detection to detect the infected area Using those techniques of disease detection in the plant leaf spot in leaves of plants that contain parallel veins and less visible. The problem occurs when the same techniques are applied in the discovery of diseases on plant leaves that contain large veins. The prominent veins will cause noise in the image texture , which requires more smoothing process to reach the desired disease area (ROI).

Parallel Veins Plant Leaf [5] Prominent Veins Leaf [6]

Figure 2.1: Types and the different between Dicot and Monocot leaf in the plant family

The color of disease area is different in density from leaf original color, so when we transform the image is from the device and applied some enhancement to the image like threshold on color component, that make the image get better. detection of infected area in leaf. Some studies convert the RGB image to the model of Hue, Saturation and Intensity (HSI) after that apply a threshold level on Hue components for segmenting the infected spot from the plant leaf [1], [7]. Di Cuii make a fast method to set the threshold depend on Hue Saturation Intensity color model to divide the area of disease [8]. Songi Kai [9] convert the RGB color image to (Y Cb Cr) color to detect the infected area.

Prof. Senjay, B. Dhaigude& et al… [10]. The application of color and texture statistics for detecting the disease in the plant leaf explained the colors transformation texture RGB image color is converted into HSI color space because the HSI descript the color very well. Filtering the green pixels to standard threshold level. Then segmentation using 32×32 patch the size and obtained suitable image segments. use these segments for texture analysis by color co-occurrence matrix. Then, if the texture features are compared to texture features of a healthy plant leaf the differences in parameters will be obvious.

In this research, we will study the classification and detection for four types of diseases often occur in the plant, which shows its effect on the leaves of plants clearly and these diseases are Tuta Absoluta, Blight of Bacterial, Anthracnose and Cercospora Leaf Spot in addition to distinction the healthy plants leaves of diseases, , shown in figure 1.7.

2.3 METHODS TO ENHANCE THE PLANT LEAVES IMAGE

The image enhancement task is an important step in the system of classification the image automatically. suitable plant leaves images are acquired to provide diagnostic evidence for many important plant diseases because of the light flash in taking image process, often these images have a noise show poor contrast, lighting variations. This problem may have a strong effect on the diagnostic process and its result. Pre-processing algorithms are applied to enhance the original image to increase the success rate in subsequent steps.[12]

Plant leaf images usually have a disease noise and various green texture backgrounds, which may lead to difficulties in extract the features. The plant leaf images commonly have very low contrast which is indicative on the green texture color, the disease spot area may also consist of green texture on the histogram.

to decrease these imperfections and generate a new image more convenient for extracting the features required in the classification, a pre-processing procedure consists of four stages to do this enhancement. The four stages are (1) Green band processing, (2) Grey conversion, (3) Adaptive histogram Equalisation and (4) filtering the image. The Figure 5 illustrates the flow chart for the image processing technology.

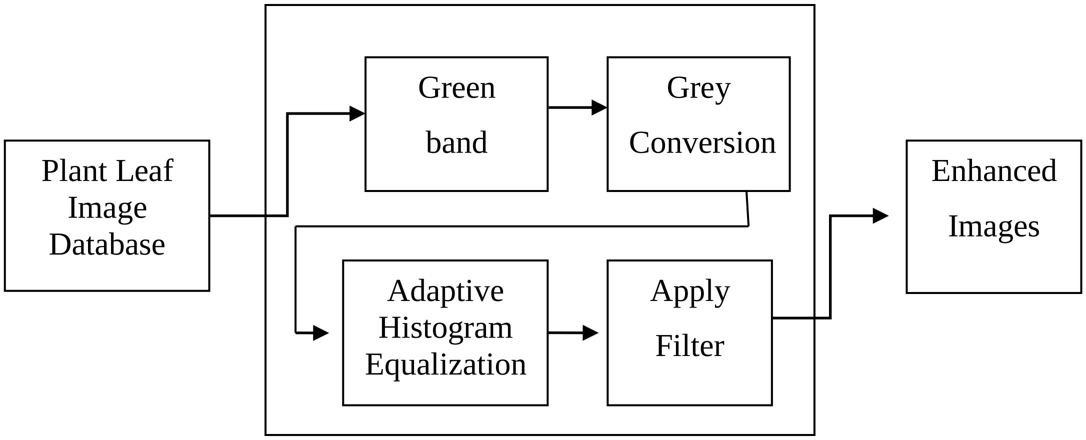

Figure 2.2: Flow diagram of image enhancement processing technique.

2.3.1 Green Band Processing

In order to improve the contrast of plant leaf images, some data is usually ignored before processing such as the blue and red constituents of the leaf image. Then, the green stripe is only widely used in treatment, showing the best variation of the veins and the largest variation between the disease spot and the green tissue of the leaf .and also the process of features extraction will easier and improve if one channel image is used. On the contrary, the blue and the red bands are hardly used by Modern automated applications because much information is not displayed accurately and clearly in these color channel. Then, Extraction the green pixel values from the input plant leaf image and stored this pixel value as elements in the matrix form.

2.3.2 Gray Scale Conversion:

the color image may have more than 16 million colors. the grayscale image has 256 shades of a gray color. the gray scale images are consist of shades of gray color only, Ranging between white at the high intensity (strongest) to the black lower intensity (weakest). These images are called monochromatic, Indicates no chromatic difference.They are frequently the measuring result of the power of light in each pixel of the color matrix in one domain of the spectrum electromagnetic, Then, to make the process easier, the green color channel of the image will translate to a grayscale image, Although, the loss of information is unavoidable, this loss is not very large to reduce the classification efficiency .

2.3.3 Adaptive Histogram Equalization (AHE).

The method of AHE is more suitable and effective than from the normal histogram equalization when it uses to detect the small area of disease Will provide low levels of contrast. and significantly decrease the intensities with the reduction of the Veins width. The AHE is applied after the removal of Veins reflex central line from the images.

2.3.4 Apply Smoothing Filter.

The use of the Mean filter to smoothing the image is a simpler and faster implementation in the process of removing noise from the plant leaf image in order to be clearer in the diagnosis and the Gaussian filter has some features where the main idea of Gaussian smoothing filter is to use the two-dimensional distribution as a "point-spread" which can be implemented by convolution because the image is stored as a discrete set of pixels, so a separate approximation of the Gaussian function must be obtained before convolution. theoretically, the Gaussian distribution does not have a value of zero everywhere and that requires that the kernel of the convolution be unlimited (infinitely)But in practices, it is virtually zero.

3. PROPOSED WORK METHODOLOGY.

The flowchart in Figure 3.1 shown The important steps to detect the disease spot, the format used in all images are JPEG. all these images enhanced the contrast color and transformed to the HIS Model of color space from RGB.

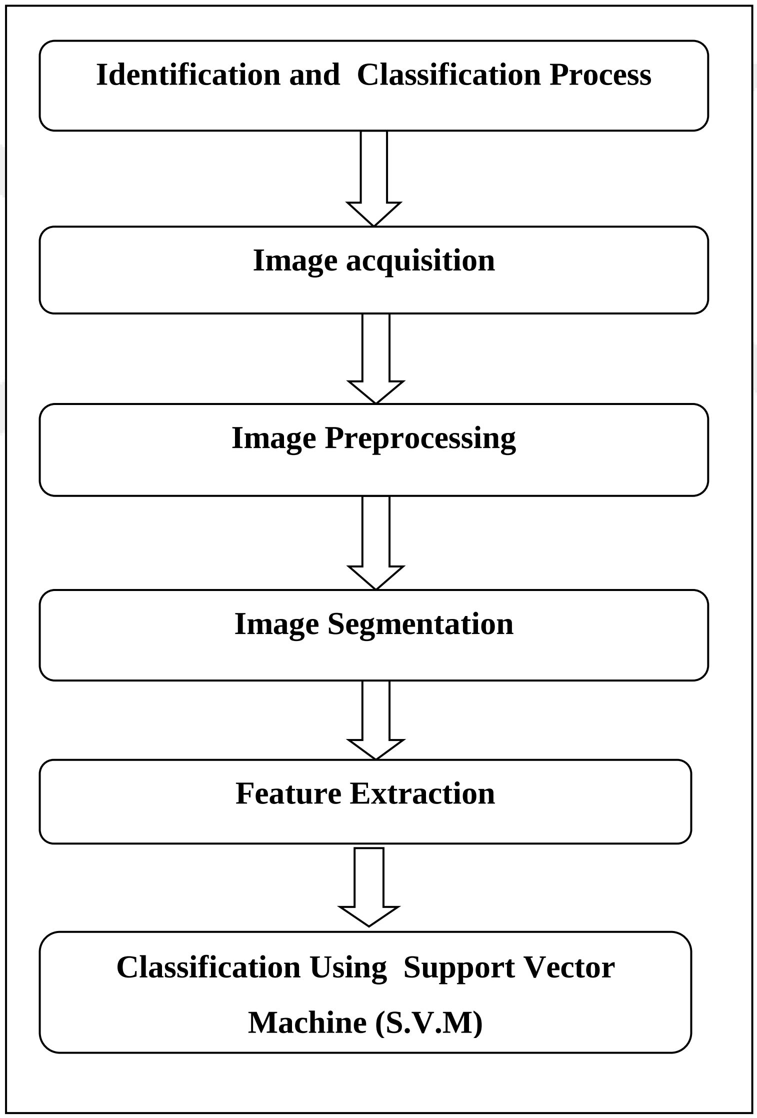

Figure 3.1: Flowchart for Disease area Detection[14]

The transformed images after resizing are passed on a filter to remove Some noise and improve images in next step after that, the Otsu threshold is used on color(RGB) image (im) the (H) Hue one of the component of the HSI color space is used to detect the disease area after classification by (SVM).

The features extracted from disease area of images segmented with other features of images obtained from healthy and infected leaves in the dataset is compared this method to get the easier and faster disease spot detection.

The main steps describing the proposed algorithm.

Step 1: acquisition images from the environment using a digital camera.

Step 2: Separate the RGB image and Create the color transformation structure from RGB to HIS Translation.

Step 3: Transform the value of color from RGB to specified space in color transformation structure

Step 4: Segmentation the infected leaf image by using Otsu method.

Step 5: Use the K-means algorithm for clustering.

Step 6: Features Extract by use GLCM function.

Step 7 : classify the disease by the result of features using Support Vector Machine(SVM).

This steps use to recognize the healthy (Normal) leaf with the infect leaf and diagnose the type of disease found in this plant leaf.

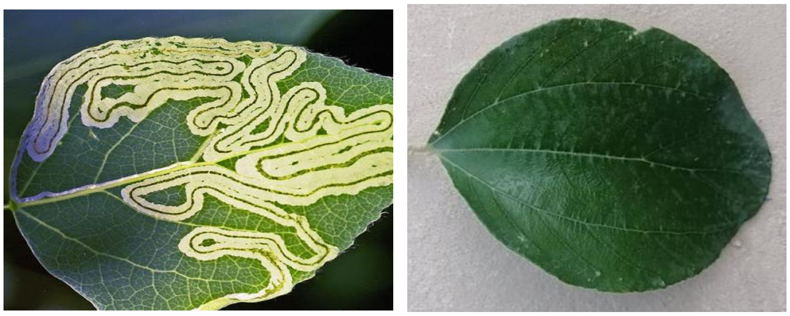

Figure 3.2: leaf infected with Tuta absoluta Plant Disease, Diseased leaf image in the left and the normal image in the right.

3.1 USING THE COLOR TRANSFORM FOR THE IMAGE

The vein in the leaf of plants is different in color intensity and the disease area is different in color, For the color of leaves plants in general. So if used Otsu threshold on grayscale image, the vein will be clearly visible in a binary(black and white) image with the infected area. But the region of interest (ROI) is only disease area in leaves, not vein.

To reduce the color intensity effect of the vein, we prefer transform the color of the RGB image before segmentation. After that we can apply Otsu threshold on color component to detect infected area accurately.

The particular specified and differing threshold value that is calculated for these pixels by using the Otsu's method, this matrix of green pixels usually are masked as this: if is the component of the green (G) pixels intensities is little than the Calculated in advance threshold value, the green, red and blue (RGB)components of these pixels are assigned to the zero value. That is mean these pixels have no weight or value for the disease classification steps, and maybe those pixels are represent healthy spots in the plant leaves. In addition, the time of the images processing will be less.

3.1.1 Use "HIS" Color Model ("Hue Saturation Intensity").

In this paper we used the HIS Color Model ("Hue Saturation Intensity"). the HSI based on human eyes on the Color Perception and dependent on the color model, the HSI consists of three parts the 'H' is (Hue) color model which is related to the wavelength of light represent a pure color, the Saturation 'S' is represented to the measures of "colorfulness" in HSI color Model, the Intensity 'I' represents, the amplitude of the light Convert the color images (RGB) of plant leave into a representation of color space Hue, Saturation and Intensity (HSI) the following Equations 1, 2 and 3 were used to convert image components from RGB to model of HSI [13] . after completed this process, every pixel map is used to build a new matrix of color co_occurrence, and that found three matrices, for every one of the components Hue(H), Saturation (S) and Intensity (I) pixel maps. The color space (HSI) is the most popular space of color because it is like a color perception of human [14]. Electromagnetic radiation it's called visible light Its wavelength ranges from 450 to 700 nanometers because the visual system of human visual system can see . Hue is often related to the light wavelength and the light amplitude. the saturation(S) component measures the color level in the space of Hue Saturation Intensity space (HSI) [15]. Also, can easily transform the Color spaces from space to other space [17].

$$" \quad Hue(H)=\begin{cases} "2\text{-}A*COS\left\{\dfrac{"[(R\text{-}G)+(R\text{-}B)]"}{\sqrt[2]{(R\text{-}G)^2+(R\text{-}G)(G\text{-}B)}}\right\}, B>G" \\ "A*COS\left\{\dfrac{"[(R\text{-}G)+(R\text{-}B)]"}{\sqrt[2]{(R\text{-}G)^2+(R\text{-}G)(G\text{-}B)}}\right\}, B\leq G" \end{cases} \qquad (3.1)$$

The HSI color space is mapped in a cylindrical coordinates system

With

$$\theta = "cos^{-1}\left(\frac{1/2[(R-G)+(R-B)]"}{[(R-G)^2-(R-B)(G-B)]"^{1/2}}\right)"$$

to avoid divide by zero, we must do an exception by adding a small number to the denominator[17].

$$(H)=\begin{cases}\theta, & \text{if } B\leq G\\ 360-\theta, & \text{if } B>G\end{cases}$$

The saturation component given by :

$$Saturation\ (s) = 1 - "\frac{3.min\ .(R,G,B)"}{(R,G,B)} \quad (3.2)$$

The (3 min) represent the Minimum elements of the value of color pixels array [17].

$$Intensity"\ (I)=(R+G+B)/3" \quad (3.3)$$

3.2 IMAGES SMOOTHING.

When the images collection, may be some noise introduced because of the light or camera light(flash). This noise can affect on the color texture of the image therefore will affect the accuracy of the detection of the disease.

To reduce the noise effect or remove unnecessary spot, we need to use Image smoothing technique. for this purpose we used median filter, as shown in figure 3.3.

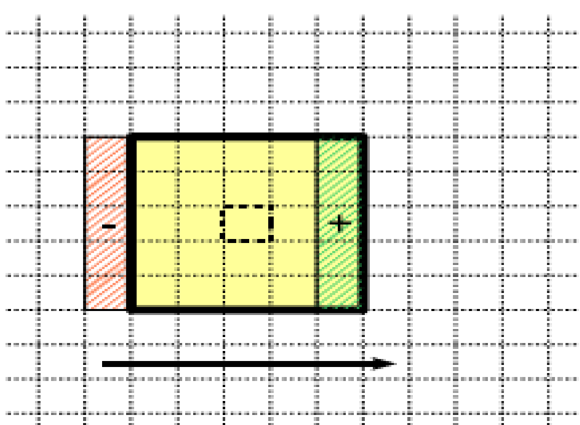

Figure 3.3: Description of the filtering process [21].

Median filter is nonlinear in nature and suitable kind of higher order statistics filter, which is a collection of the numerical value is used to half the values of pixels in the mask matrix, which are equal or little than to the median, and the middle is bigger than or equal to the median replaces the middle pixel value, by the median of the matrix mask in grayscale levels, to perform this filter, the start window is translated and reduce all the pixels around by this window. After that, the median filter is calculated and this value will be the center value of a pixel. when the numbering of elements is odd in the M*N window, the value in the middle is inserted as the new value, else the number of elements in M*N window is even, Average of the two intermediate values will be the median value.

3.3 SEGMENTATION OF DISEASE AREA

After the image filtering, we need to recognize the disease spot region of interested (ROI) by using an Otsu method.

The most important thing to select the gray level threshold to determine the disease spot area from the leaf. when the histogram range has sharp deep between peaks, the bottom of the deep valley we can select as a threshold. But the problem is when the valley is wide. In that case, can't use this technique because The objects cannot be separated or selected inside the image from the background. therefore we will use the method of Otsu for disease detect. To make enhanced contrast and segmented image ROI(Region Of Interest) as in figure 3.4.

Figure 3.4: plant leaf segmentation to find Disease spots.

3.4 CLUSTERING USING K-MEANS TECHNIQUE

K-means clustering algorithm needs two steps to be implemented. The first stage begins by creating a matrix to converts the color space independently. This matrix used to determine the color and produce the same matrix of color pixels whatever is the device used to display. Then, we create a structure of color conversion which determines the conversion of the color space. After that, we applied the matrix of the transformation of the independent space of color, to converts the value of color in the leaf image to the specified color space in the structure of color transformation.

This transformation of color space can specify many factors to the transformation. The color space in matrix-based pixels is the resulting color area on the device used to display it. If the result color taking a matrix of pixels with the values of rgb pixels will be changed like the contrast and brightness on the display devices.

Then the color system rgb dependent the value of color space. The algorithm of k-means works for classifying the objects from the background of the images, classifying the pixels of color in our research, although, the features dataset to k of the selected classes.

the disease classification is work to minimizing the squares values of distances in the objects inside the image and the identical cluster or the centroid of class. Even though, the cluster of k-means is used to split the plant image into three cluster in which one cluster or more clusters contain the infected spot, so that the image of the disease is more discoverable through the region of interest (roi).

Use image texture analysis and learning algorithms techniques like k-means clustering, svm classification algorithm give more efficient to detect the disease. The classification using support vector machine(svm) will provide a relatively short time and low cost for users to know the disease that afflicts their own farms and plants.

3.5 FEATURE EXTRACTION

Feature Extract is an extract the color feature from image texture analysis, to extract the features from the color and texture of the processed image, The technique of analysis leaf image selected for this project it is (CCM) or the Color "Co_occurrence matrix style. this method use both the texture and color of the leaf image , to arrive at special features to make the Dataset, which represent the selected image to classify, use the features of the processed image in the visible spectrum light, provides additional characteristic specifications and make the disease detection in the plant leaves more easy to establish the CCM methodology, we must convert the RGB images of plant leaves disease to representation the (HSI) space of color. when the (HSI) Space representation process is completed, A matrix will be formed from a color co-occurrence by use each pixel map, and three CCM matrices will be configure , one for each pixel maps of the H, S and I. After the conversion processes, when we determine the value of the feature set for H and S, we left (I) since it does not give us important information that we need , in addition , the use of GLCM function in Matlab its very important to create new matrix of the co-occurrence gray-level; and set the gray levels number to (8) level , and finally, offset is given a" 0" value. (GLCMs) the method of Gray level co-occurrence matrix, this method computes how much pairs of the gray level of the pixels is a vector in this case isolated by a specific distance and oriented by a specific direction when scanning lines of pixels from top-to-bottom and from left-to-right.

The most important criteria of feature extraction are determining the texture and space. the color space is a space has a multi-dimension, in which multi-dimensions is a different component of the color. often the spaces of color often consist of three dimensions. An example of a color space is Red Green Blue(RGB), which refer to a vector have in each pixel a three-element, which gives the color density of the three main colors. The covered space by Red, Green, and Blue values, specifying visible colors completely, which are considered as vectors in the three-dimensional RGB space color. therefore, these color space give us the first step for extract color features from the processed images, The co-occurrence method is used in the extraction of color features. The important step is the Separation of ingredients of the RGB from the real images. The second step is to determine the value of the features and normalizing the co-occurrence matrices: after using the GLCM function to create a gray-level matrix, normalize the feature result using the equation given below.

$$p(i,j) = \frac{p(i,j,1,0)}{\sum_{i=0}^{Ng-1}\sum_{j=0}^{Ng-1} p(i,j,1,0)} \tag{3.4}$$

When, p(i,j) is the matrix of image attribute p(i,j,1,0), consider as the intensity co-occurrence matrix, and N_g represents the total number of density levels.

Dataset (Features Extracted from the images)_Is a model learns to perform classification tasks directly from images)),

The Dataset's the Features that extracts from the image represented as a single feature , this features use to classify and compare between the Characteristics of colors degree in the texture of the image . we use this features as input in the Artificial Neural Network (ANN) system to classify the Disease that occurred in the leaf on the plants.

The input consist of two files , the first file is the Training_Dataset) and the second is (Accuracy_Dataset)

3.5.1 Training Dataset is Contain Training_Features and Training_ Label

3.5.1.1 The training_features

The Training_features consist of matrix the element of this matrix we taken from the features that extracted from plant leaves infected images these image selected carefully to cover wide verity of images in the same disease to detect more infected cases of leaves , the Training_features matrix size (125*13) the (125) rows each row represent one training leaf image, the (13) column represent the features of this images , the training_features in this Dataset classified to five classes each class consist of twenty five rows and it represent one type of disease , the last class is for healthy leaves and the other four classes for four different types of disease.

3.5.1.2 Training_label

Training_label,consist of matrix size (1*125) from (0 to 4) each number in (25) column to represent the Disease class (Disease Name) and the last set number (4) represent the healthy leaves class.

3.5.2 Accuracy Dataset is also Contain Training_Features and Training_Label

3.5.2.1 The training_features

The Training_features consist of matrix size (125*13) the element of this matrix contain the features that extracted from leaves infected images its similar to the matrix of the training feature in Training Dataset, this Training_features matrix use to give more Accuracy of recognition disease result by reiteration the test many times of the feature extracted from the input image that we want to discover the disease type in it, with the feature of the images of infected leaves that we use to train, that make more efficiency to perform the results of the detect disease program.

3.5.2.2 The training_label

The Training_label,consist of matrix size (125*1) from (0 to 1) to determine the class of the disease .

which a model learns to perform classification tasks directly from images,

these features in this project will be identified by a thirteen features this features is :

a- Contrast

b- Correlation

c- Energy

d- Homogeneity

e- Mean

f- Standard_Deviation (S.D)

g- Entropy

h- RMS

i- Skewness

j- Variance

k- Smoothness

l- Kurtosis

m- IDM

a- Contrast :

Whis process, take a specified value if the pixel values more than the specified value it will be displayed as white pixel , if the pixel values less than specified value its will display as black pixel, and the values less than white and more than black, in the middle of the two values are shown as light shades, The outcome is a linear graphic of a subset of pixel magnitude to the full set of grayscale, from white to black, resulting in an image of more higher contrast.

The code used in matlab to determine the contrast is:-

```
I = imadjust(I,stretchlim(I));
```

$$\text{contrast} = \sum_{n=0}^{N_g-1} n^2 \sum_{|i-j|} P_d\,(i,j)" \qquad (3.5)$$

b- Correlation coefficients:

We can obtained the correlation , by calculating the matrix of gray level co-occurrence on desired leaf spot image. Combined with color extraction, Correlation determine the linear relation with two variables ,if there is no correlation between features, there is no relation between the values of the features to decrease or increase in tandem. The uncorrelated features not necessarily independent, if it is have a nonlinear relationship.

$$"\text{correlation}=\frac{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1} ijP(i,j)-I_2^{\,2}}{I_2}" \qquad (3.6)$$

c- Energy:

The energy of pixels represents the brightness of the image

the equation for Calculating energy using MATLAB[16]

$$Energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P^2 d\,(i,j) \text{ "} \quad (3.7)$$

d- Homogeneity:

We can obtained the homogeneity, by calculating the matrix of gray level co occurrence on identified region of interest in leaf spot image. Combined with color extraction, they experimented on detecting diseases on Peas leaves.
Used the angle (E) angular moment is to find out the homogeneity of the image and we can calculate it as the equation in:

$$\text{"Homogeneity} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_d(i,j)}{1+|i-j|} \text{ "} \quad (3.8)$$

e- Mean:

Mean or average of matrix elements computes the mean of all values in array returns the average of the elements of array Read a grayscale image into the workspace as a in Array object, then calculate the mean of the pixel intensity values.

The mean intensity level (Mean) represent the brightness of the image and this Equation below can be derived from the matrix of co-occurrence.

$$\text{"}Mean = \sum_{i=0}^{N_g-1} ip(i) \text{ "} \quad (3.9)$$

f- Standard_Deviation (S.D):

S.D returns the pixels A of the matrix along the dimension of the first array whose, doesn't equal his size.

When A is a matrix have random variables columns and its rows are noted, then the row vector includes S.D conforming to every column.

Where , N is represent all the number of pixels, x_i is represent the pixel value

Standard Division(σ)="1/N $\sum_{i=1}^{N} \sqrt{(x_i - \mu)^2}$ (3.10)

g- Entropy:

The Entropy is a feature extract from the training images to determine the randomness by a statistical tables that can we used to classify the texture of the leaf image in which we want to detect the changes that have occurred in its texture.

Entropy uses two bins for arrays (logic array) in default and 256 bins for double arrays, or uint8, uint16. the entropy converts each class which not logical to uint8 for the histogram count determination so that the values of the pixels are separate and correspond directly to the value of bin.

" $Entropy = -\sum_{i,j} P_{(i,j)} \, logP_{(i,j)}$" (3.11)

Where (p) represents the histogram counts to normalized from imhist Tips

H= row vector of calculated entropies (in bits)

X = data to be analyzed [18]

And we can computed the sum and the difference of entropies (se and de) using Equations (12) and (13) respectively.

$$"se=\sum_{k=0}^{2(N_g-1)} p_{x+y}(k) \ln p_{x+y}(k)" \quad (3.12)$$

$$de=\sum_{k=0}^{N_g-1} p_{x-y}(k) \ln p_{x-y}(k)" \quad (3.13)$$

This feature (entropy) is the measure of the amount of order , we can computed as the Equation below[17]

$$"Entropy=\sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P(i,j)\ln P(i,j) " \quad (3.14)$$

h- RMS (Root Mean Square):

Root mean square level (RMS), we can compute the true root mean square of the(RMS) block magnitude value of the signal. The true RMS magnitude value of the input signal is computed over an operating average window of one cycle of the particular essential frequency:

Because this RMS block uses the window of operating average, we must entire complete one cycle of simulation before the output gives the correct magnitude value. For the first cycle of simulation, the output is held to this determined initial RMS magnitude value.

A = mean2(rms(img)) .

Returns the Root-mean-square level (RMS) level of the average input, im (to matrix of disease spot image). If im is a column or row vector, A is a real-valued scalar. For matrices, A contains the RMS levels calculated with each element the first array dimension of img with greater size than 1.

,if im is an N-by-M matrix with N > 1, the result will be a 1-by-M row vector including the RMS levels of the (im) columns.

The code use in program is :

RMS = mean2(rms(seg_img)),

i- Skewness:

A = skewness(img) returns the sample skewness of img. For the value of image pixels as a vectors , skewness(img) is the skewness of the elements of img. For matrices of image, skewness(img) is the row vector of image matrix representing the skewness of each column. To Number-dimensional arrays, the skewness running along the first non singleton dimension of img.

j- Variance:

The variance use to returns the variance value of elements of the array along the dimension of the first array if the size of this array does not equal 1.

And we can defined the variation of the image intensity by the variance color and texture features (var) and is calculated as

$$\text{"variance} = \sum_{i=0}^{N_g\text{-}1} (i - E)^2 \; p_x(i) \text{ "} \tag{3.15}$$

k- Smoothness:

Smooth response data it's smooth the data of image matrix in the column vector y using an average filter. Results are returned in the column vector of the matrix.

l- Kurtosis:

Kurtosis returns the sample kurtosis of the matrix elements of identified region of interest of image . For the kurtosis vectors, (I) is the kurtosis of the elements located in the vector I. For matrices kurtosis(I) returns the specimens of kurtosis for every column in I.

m- Interval Data Message (IDM):

The Interval Data Message use to Automatic Meter Reading to read utility meters by processing Interval Data Message (IDM) signals emitted by meters. And it use to recorded data from a file, in this dataset project the value of IDM is often equal one .

Texture feature reduction from the practical experimentation there are thirteen color texture features which are exist in all of the sample images. Then, these thirteen features is useful and important to the classification and detect the disease in plant with. thirteen features of color and texture have been considered as first-level for plant disease detection feature reduction. The Texture and color feature decrease is done established on the value of delta and threshold. The threshold value is taken empirically determined and rejected Any values feature below the threshold. we chose The threshold established on an average of less feature value and upper feature value.

The table(1) below is for the feature extracted from 25 specimen of plant leaves infected with the (Tuta Absulota)

Table 3.1 : Feature extracted from 25 specimen of plant leaves infected with the (Tuta Absulota) plant Disease, the size of the images is (255,255) with white background .

Image No	IDM	Skewness	Kurtosis	Smoothness	Variance	R.M.S	Entropy	S.D	Mean	Homogeneity	Energy	Correlation	Contrast
1	255	2.06056	5.65185	1	4908.78	6.67462	2.56342	71.5164	31.4413	0.915426	0.646189	0.821141	1.51354
2	255	2.35669	7.53018	1	2153.67	5.86388	2.06101	52.0009	21.914	0.942452	0.641181	0.829524	0.80579
3	255	2.85153	7.99131	1	3380.96	4.82926	1.54498	61.5219	22.1904	0.928927	0.713612	0.749331	1.68971
4	255	1.57107	3.99931	1	3760.93	8.06285	3.03014	63.4702	34.8272	0.889673	0.463159	0.782868	1.63496
5	255	3.44043	13.7609	1	2614.18	5.03356	1.83998	53.5975	16.6826	0.971433	0.788079	0.937974	0.292785
6	255	1.50923	3.54822	1	5940.48	7.58592	2.46804	78.3963	41.2832	0.92242	0.556189	0.869237	1.48316
7	255	1.87589	5.27785	1	3819.64	7.33878	2.72977	64.1122	32.1653	0.856677	0.502454	0.627127	2.866389
8	255	2.63972	12.0527	1	1449.55	8.50176	3.17152	39.4951	21.3389	0.914991	0.456711	0.739435	0.647212
9	255	1.25048	4.45881	1	2343.73	11.0073	4.44831	50.3462	40.3482	0.880799	0.246139	0.775791	0.979642
10	255	2.4332	7.11408	1	4379.38	6.56242	2.67755	70.7727	27.6376	0.955758	0.747963	0.86876	1.14206
11	255	2.42639	7.11843	1	4221.11	5.20057	1.71721	70.3451	26.5612	0.961127	0.733788	0.914341	0.733563
12	255	1.32478	3.61031	1	3239.15	8.61031	3.12101	60.4884	37.7754	0.935435	0.456183	0.881035	0.72546
13	255	0.993062	2.3658	1	4685.89	9.3227	3.66739	71.9565	49.3273	0.869934	0.371917	0.809481	1.75545
14	255	2.83317	9.39974	1	3447.13	6.70089	2.74016	60.6624	22.1469	0.969281	0.787905	0.927721	0.456832
15	255	2.20696	7.07745	1	2607.58	7.62144	2.64019	52.4995	24.8006	0.881168	514337	0.679142	1.5631
16	255	1.87484	5.6789	1	2574.9	7.80496	2.98861	56.6078	30.1591	0.908676	0.498909	0.771532	1.09401
17	255	3.3535	13.138	1	2420.86	4.94013	186114	53.3088	17.1196	0.967692	0.77104	0.878062	0.567708
18	255	1.51228	3.78638	1	4866.19	9.44123	3.92405	72.8584	42.9946	0.89406	0.450801	0.839608	1.41552
19	255	1.33188	4.7879	1	2092.84	11.517	4.97136	50.0507	41.4315	0.874152	0.232989	0.739277	1.07624
20	255	2.74331	9.1425	1	3024.13	5.40792	1.9707	60.2896	22.025	0.968565	0.743548	0.92183	0.46996
21	255	1.32478	3.61031	1	2329.15	8.64073	3.12101	60.4884	37.7754	0.935435	0.456183	0.881035	0.72546
22	255	2.54016	9.61061	1	1353.87	7.56262	2.6858	37.7557	17.399	0.932575	0.5574	0.784803	0.509283
23	255	3.3323	13.9782	1	3003.21	4.9094	1.67267	35.3604	11.6987	0.961207	0.755024	0.762723	0.511979
24	255	2.78778	10.59556	1	1291.26	6.04454	2.06025	37.77157	14.9002	0.953581	0.66718	0.770119	0.560202
25	*255*	*3.00852*	*12.6712*	*1*	*1133.35*	*6.94602*	*2.68595*	*35.3823*	*14.7553*	*0.946389*	*0.619841*	*0.748791*	*0.53638*

Table 3.2 : feature extracted from 25 specimen of plant leaves infected with the (Anthracnose) plant Disease, the size of the images is (255,255).

No	IDM	Skewness	Kurtosis	smooth	Variance	R.M.S	Entropy	S.D	Mean	Homogeneity	Energy	Correlation	Contrast
1	255	3.44043	13.7609	1	2614.18	5.0335	1.8399	53.5975	16.6826	0.971433	0.788079	0.93797	0.29278
2	255	1.50923	3.54822	1	5940.48	7.5859	2.4680	78.3963	41.2832	0.92242	0.556189	0.869237	1.48316
3	255	1.87589	5.27785	1	3819.64	7.3387	2.7297	64.1122	32.1653	0.856677	0.502454	0.627127	2.86638
4	255	2.63972	12.0527	1	1449.55	8.50176	3.17152	39.4951	21.3389	0.914991	0.456711	0.739435	0.64721
5	255	1.25048	4.45881	1	2343.73	11.0073	4.44831	50.3462	40.3482	0.880799	0.246139	0.775791	0.97964
6	255	2.4332	7.11408	1	4379.38	6.5624	2.6775	70.7727	27.6376	0.955758	0.747963	0.86876	1.14206
7	255	2.42639	7.11843	1	4221.11	5.20057	1.71721	70.3451	26.5612	0.961127	0.733788	0.914341	0.73356
8	255	1.32478	3.61031	1	3239.15	8.61031	3.12101	60.4884	37.7754	0.935435	0.456183	0.881035	0.72546
9	255	0.993062	2.3658	1	4685.89	9.3227	3.66739	71.9565	49.3273	0.869934	0.371917	0.809481	1.75545
10	255	2.83317	9.39974	1	3447.13	6.70089	2.74016	60.6624	22.1469	0.969281	0.787905	0.927721	0.45683
11	255	2.20696	7.07745	1	2607.58	7.62144	2.64019	52.4995	24.8006	0.881168	514337	0.679142	1.5631
12	255	1.87484	5.6789	1	2574.9	7.80496	2.98861	56.6078	30.1591	0.908676	0.498909	0.771532	1.09401
13	255	3.3535	13.138	1	2420.86	4.94013	186114	53.3088	17.1196	0.967692	0.77104	0.878062	0.56770
14	255	1.51228	3.78638	1	4866.19	9.44123	3.92405	72.8584	42.9946	0.89406	0.450801	0.839608	1.41552
15	255	1.33188	4.7879	1	2092.84	11.517	4.97136	50.0507	41.4315	0.874152	0.232989	0.739277	1.07624
16	255	2.74331	9.1425	1	3024.13	5.40792	1.9707	60.2896	22.025	0.968565	0.743548	0.92183	0.46996
17	255	1.32478	3.61031	1	2329.15	8.64073	3.12101	60.4884	37.7754	0.935435	0.456183	0.881035	0.72546
18	255	2.54016	9.61061	1	1353.87	7.56262	2.6858	37.7557	17.399	0.932575	0.5574	0.784803	0.50928
19	255	2.07154	6.03604	1	3003.21	7.04093	2.44004	57.042	26.3292	0.992332	0.565237	0.76074	1.46074
20	255	3.23192	12.269	1	2674.68	5.3293	1.5706	55.0748	17.7597	0.96622	0.76595	0.92814	0.3559
21	255	1.72821	5.2596	1	3459.14	9.0220	3.3682	60.4048	35.4486	0.91291	0.42671	0.87840	0.7366
22	255	1.92539	5.61502	1	3015.26	7.6567	2.6399	56.8561	28.3741	0.88108	0.52029	0.81522	1.0172
23	255	1.52948	3.55295	1	5583.09	6.9761	2.6951	82.2267	45.1042	0.91714	0.56076	0.89178	1.2538
24	255	1.86415	4.83567	1	4500.25	6.4434	2.5963	71.5037	33.9344	0.94134	0.62928	0.94140	0.4746
25	255	1.765	4.7665	1	4667.25	7.108	1.8399	53.5975	16.6826	0.97143	0.68807	0.96797	0.2927

3.6 CLASSIFICATION USING " SUPPORT VECTOR MACHINE"(SVM).

We can apply ("support vector machine" SVM) when the dataset has contained just a two of classes. the SVM classifies dataset by taking the best hyperplane that split all points of data in one class from, those of the other class, The texture and color features have been used to identify the disease with the sample of the diseases images. Algorithms for extraction the features of color and texture, Which has been sophisticated, and applied to train "support vector machine" algorithm and the classifier of (ANN)artificial neural network. The research of Jegadiesh D.Pojari, Rejesh Yakkundimath and Abdulmonaf. Syedhosain Biadgy (the Classification Based of ANN and SVM of Plant Diseases) has resulted depend on a reduced dataset feature for detection of images of infected leaves. and the results are the SVM classifier is more efficiency for classification and identification of leaf diseases detection.

Classification Algorithm (SVM) "Support Vector Machine" Classification is the Algorithm of supervised Machine Learning which can be applied for the classification or regression challenges. And it is generally used in classification problems. In the Algorithm, we draw each dataset subject as a point in the matrix of N-Dimensional space (where N is the numbering of features we have) with the magnitude value of each feature being the value of specific coordinate. And, we apply the classification by determining the hyperplane (decision boundary) that differentiate the two classes very well(as illustrate in figure 3.5 below).

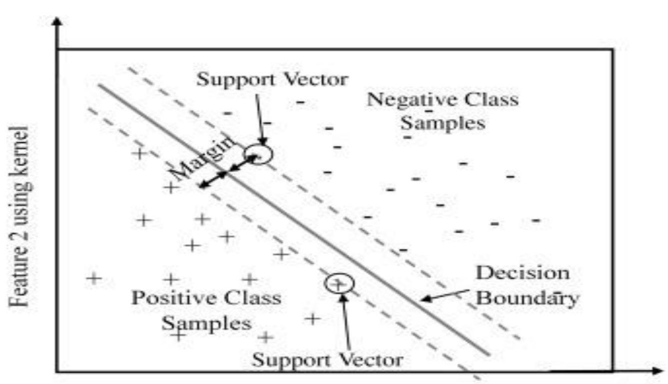

Figure 3.5: Explains how the support vectors work with the separated dataset points.[19]

Support vectors are the points of the dataset nearest to the comma; These data points are on the edge of the tile. With the Positive signal(+) refers to the points dataset of type 1, and the negative signal (-) refers to the points dataset of type 2

To increase the accuracy and efficiency of plant leaves disease detection in this research, we will use image analysis techniques with support vector machine (SVM) classification algorithms, the use of SVM will reduce the time required and will be easier to use for users and farmers to detect the disease that affect in their own plants.

3.6.1 Some advantage of the Support Vector Machine.

1- High accuracy prediction.
2- Robust, if there is an error in the training examples.
3- The SVM represents a geometric interpretation with different solutions.
4- for the most given pattern, the Support Vector Machine doesn't consider the noise in the image as part of the pattern as in ANN.
5- The SVM calculation complexity does not depend on the input size dimensions, as in the Neural Network.

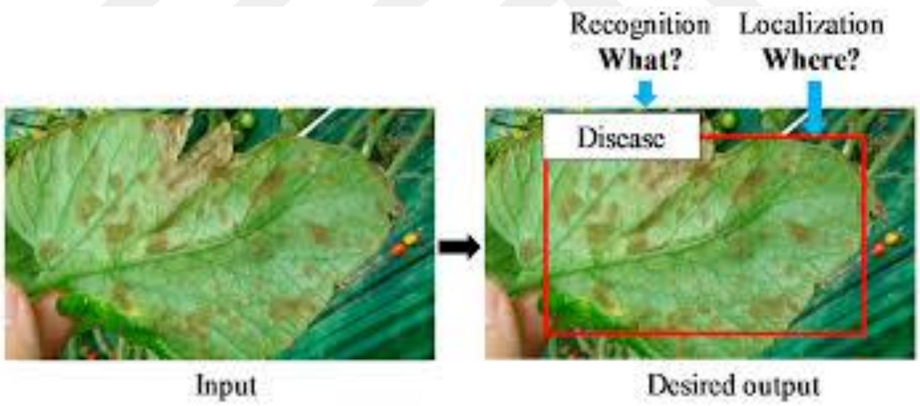

Figure 3.6: SMV Recognition area**.**

3.6.2 The Shortcomings and disadvantages of SVM are:

1- The SVM classifier involves a long time for training.

2- The Weights parameters (learned function) is not easy to understand in SVM.

4. THE WORK PART IMAGE PROCESS STEPS.

We will take some steps to make a clearer image by reducing the noise to determine the affected area to facilitate the process of detecting the disease.

The value of color intensity in (RGB) and the grayscale image is from 1 to 0 The operations in these steps will increase the contrast of the output image. by decrease the intensity in the high-density pixels. and increase the intensity in the low-density pixels to make the image more smooth, more clear and low noise.

We will applied the code below to leaf image infected by (Tuta Absoluta) Disease.

4.1 IMAGE RESIZE.

We will use this code below to resize the image to (256) pixel row and(256) pixel Column , To unify all the input images in the same size.

```
I=imresize(I,[256,256]);
```

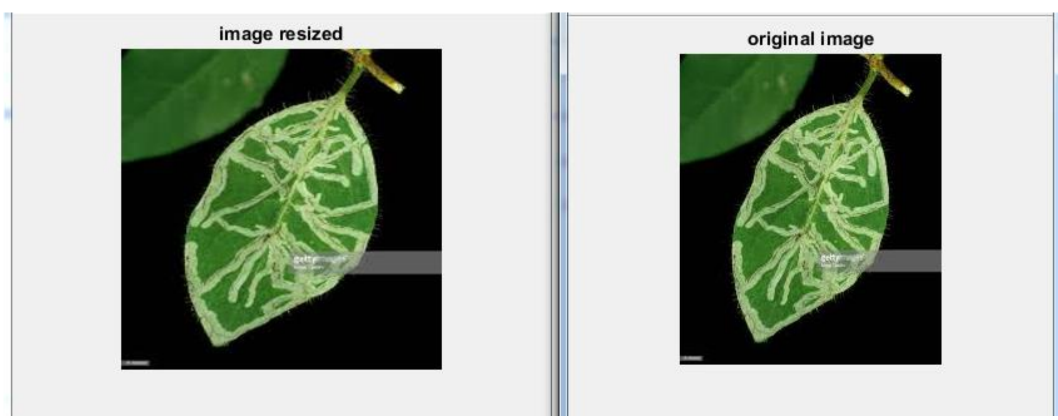

Figure 4.1: Image Resize process.

4.2 IMAGE ADJUSTMENT

We will make some change to the image to get more enhance in the disease area and to identify the shape and the sizes in a spot of the plant leaf containing the disease. we will apply

I=imadjust (im) To make enhancements to the image.

Since the value of color intensity in (RGB) and the grayscale image is from 1 to 0 the adjust process will give more smooth to the leaf image.

J = imadjust(I) this code make the intensity values in a grayscale image I to new output value in J. By default, adjust saturates the bottom 1% and the top 1% of all pixel values. This operation increases the contrast of the output image J. that make the recognition of disease area in the plant leaf easier

We can perform contrast adjustment to the suitable value using a GPU (requires Parallel Computing Toolbox™). This syntax is equivalent to imadjust(I,stretchlim(I)).

J = imadjust(I,[low_in high_in])

Maps intensity values in I to new values in J such that values between low_in and high_in map to values between 0 and 1.

example

J = imadjust(I,[low_in high_in],[low_out high_out]) maps intensity values in I to new values in J such that values between low_in and high_in map to values between low_out and high_out.

J = imadjust(RGB,[low_in high_in],)

maps the values in true color image RGB to new values in J. You can apply the same mapping or unique mappings for each color channel.

```
I=imread('im2.jpg');
% I=imresize(I,[256,256]);
I1=imadjust(I,stretchlim(I));
```

```
figure, imshow(I);title('original image');
figure, imshow(I1); title ('enhanced image');
```

Figure 4.2:Image Enhancement Process.

4.3 USE GLOBAL IMAGE THRESHOLD USING OTSU'S METHOD.

level = graythresh(I) this code is determine the global level threshold, this method use to convert the pixel in image from intensity to a(0 or 1) black and white image with imbinarize. The graythresh function uses Otsu's method, which chooses the threshold to minimize the intraclass variance of the black and white pixels [20].

[level,EM] = graythresh(I) returns the effectiveness metric, EM, as the second output argument. The effectiveness metric is a value in the range [0,1] that indicates the effectiveness of the thresholding of the input image. The lower bound is attainable only by images having a single gray level, and the upper bound is attainable only by two-valued images.

```
I=imread('im2.jpg');
I=imresize(I,[256,256]);
I2_otsu=im2bw(I1,graythresh(I1));

figure, imshow(I);title('original image');
figure,imshow(I2_otsu); title('otsu');
```

to get the new enhanced as shown in figure 13. below

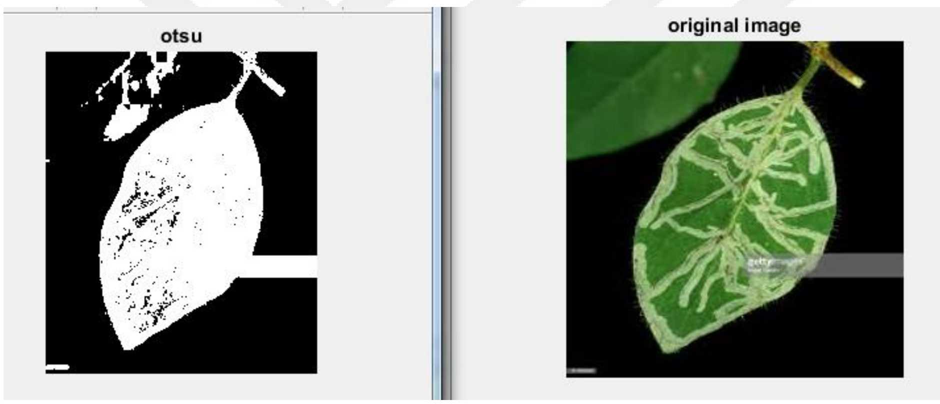

Figure 4.3: Apply Threshold Using Otsu's Method.

4.4 USE HSI FUNCTION (HUE, SATURATED, INTENSITY)

Is a function converts the RGB to HIS . Display the Hue image, Saturation image and the Intensity image. The input image is supposed to be the size "M-by-N-by-3", where the third dimension represents three levels of images: blue, red and green, in this format. If all RGB image components are equal, HSI conversion is not defined. The input image can be of a double class (with a range of values [0, 1]), uint8, or uint16.

By use the code hsi = rgb2hsv(rgb)

```
"I=imread('im2.jpg');"
imshow(I);title('original image')
I1=imresize(I,[256,256]);

I2 = im2double(I);
r = I2(:, :, 1);
g = I2(:, :, 2);
b = I2(:, :, 3);

% Implement the conversion equations.

"num = 0.5*((r - g) + (r - b));"
"den = sqrt((r - g).^2 + (r - b).*(g - b));"
"theta = acos(num./(den + eps));"

H = theta;
H(b > g) = 2*pi - H(b > g);
H = H/(2*pi);

num = min(min(r, g), b);
den = r + g + b;
den(den == 0) = eps;
S = 1 - 3.* num./den;

H(S == 0) = 0;

I = (r + g + b)/3;

% collect all three outcome into an hsi image.

hsi = cat(3, H, S, I);
"figure,imshow(hsi); title('apply HIS functrion')"
```

We will get Image with apply HIS Function.

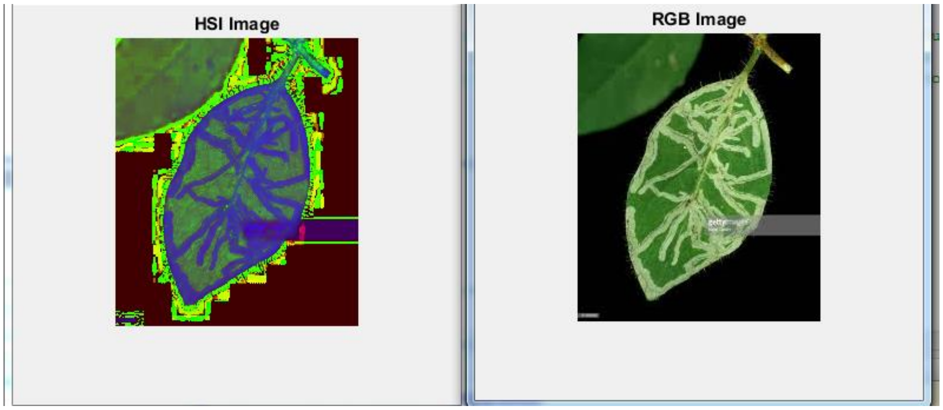

Figure 4.4: Apply HIS function to the RGB image.

by use this code

```
im=imread('im2.jpg');
im=im2double(im);
r=im(:,:,1);
g=im(:,:,2);
b=im(:,:,3);
th=acos((0.5*((r-g)+(r-b)))./((sqrt((r-g).^2+(r-b).*(g-b)))+eps));
Hue=th;
Hue(b>g)=2*pi-Hue(b>g);
Hue=Hue/(2*pi);

imshow(im); title('original mage');
figure; imshow(Hue); title ( 'hue image' )
```

we will get image apply with a Hue function only

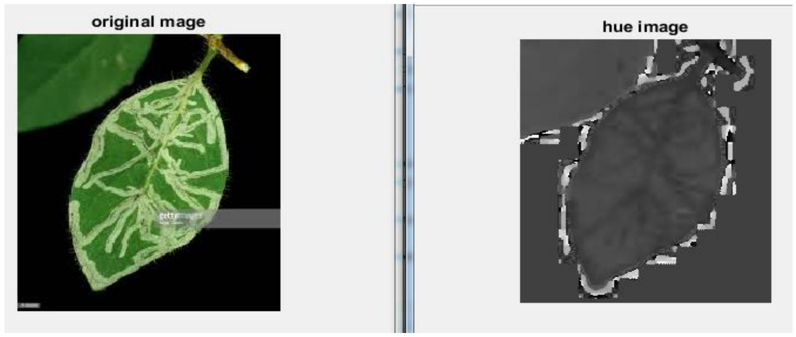

Figure 4.5: The Hue part image from the HIS .

And use this code

```
Saturation=1-3.*(min(min(r,g),b))./(r+g+b+eps);
Intensity=(r+g+b)/3;
hsi=cat(3,H,Saturation,Intensity);
HE=H*2*pi;
HE=histeq(HE);
HE=HE/(2*pi);
SE=histeq(Saturation);
IE=histeq(Intensity);

figure; imshow(Saturation); title ( 'Saturation image' )

figure; imshow(Intensity); title ( 'Intensity image' )
```

We will get an image with the changes that occur from the application of Saturation and Intensity function image.

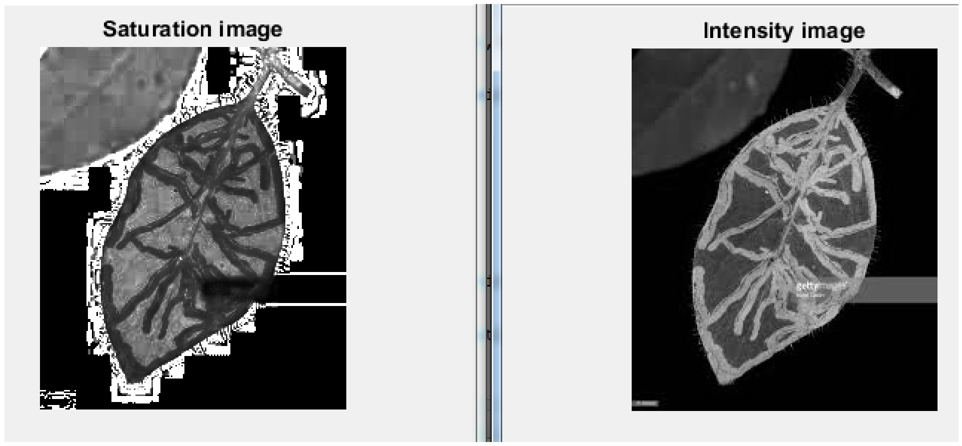

Figure 4.6:The Intensity part image from the HIS.

4.5 USE THE FUNCTION "MAKECFORM".

The color conversions structure defined by makecform pass as an argument to applying the function of Makecform. the function of "makecform" deal with the transformations between elements of the family of independent color spaces Created by the "Commission Internationale de l'Éclairage" ("International Commission on Illumination"). and the "makecform" function also supports transformations from and to the CMYK and RGB spaces of color. To conversions the color space .

This code used to make cform:

" cform = makecform('srgb2lab');"

" lab_he = applycform(I,cform);"

4.6 USE K MEAN CLUSTER.

We can use this K Mean cluster code to get image labeled by cluster index to determine the spot disease area in the plant leaf to label every pixel in the image by using the result of K means. we have to Classify the colors in a*b* color space using K means clustering. because we have (3) colors in the image we will make 3 clusters. and Measure the distance using Euclidean Distance Metric.

In this research we used the K Means clustering to transformation the plant leaf Image from Color Space (RGB) to the L*a*b* Space. The color space layers "L*a*b*" where "L*" represent the glossiness layer, " a*" and "b*" represent the color_layer. all of the color information is contained in the "a*" and 'b*' layers.

The code below will Label every pixel in the image using results from K mean, and this Image is Labeled by the Cluster Index.

```
I=imread('tota.jpg');

imshow(I); title('original image');

cform = makecform('srgb2lab')

label_he = applycform(I,cform)

ab = double(label_he(:,:,2:3));

n rows = size(ab,1);

n cols = size(ab,2 );

ab = reshape(ab,nrows*ncols,2);

nColors = 3;

"[cluster_idx cluster_center] = kmeans(ab,nColors,'distance','sqEuclidean... ,'

'                          Replicates',3);"

pixel_labels = reshape(cluster_idx,nrows,ncols);

figure,imshow(pixel_labels,[]), title('Image Labeled by Cluster Index);
```

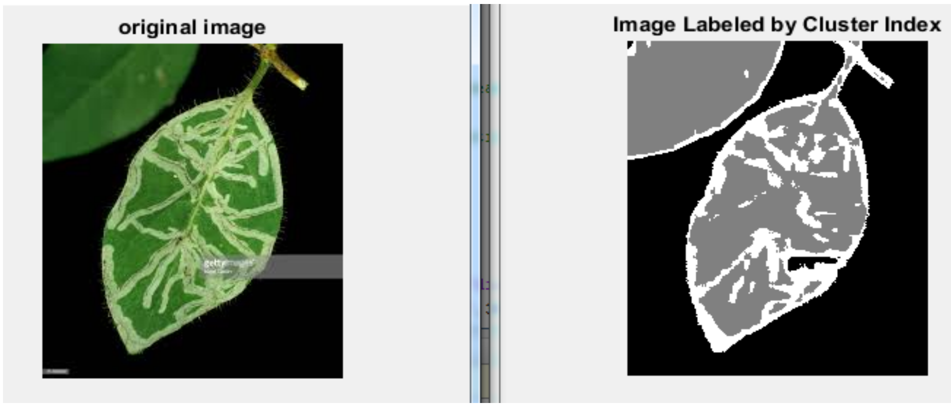

Figure 4.7: Image labeled by cluster index using K mean.

When we apply all these steps functions to the image before entered it to the K Mean cluster its will give us a more suitable enhanced image, and the color map of the effective area more obviously.

```
clc;

clear all;

I=imread('im2.jpg');
I=imresize(I,[256,256]);
I1=imadjust(I,stretchlim(I));
 figure, imshow(I);title('original image');
figure, imshow(I1); title ('enhanced image');

otsu=im2bw(I1,graythresh(I1));
figure,imshow(otsu); title('otsu');

  HIS=rgb2hsv(I1);
 figure, imshow(HIS); title('HIS');

cform=makecform('srgb2lab');

% Apply the colorform
lab_he = applycform(I1,cform);

ab = double(lab_he(:,:,2:3));
nrows = size(ab,1);
ncols = size(ab,2);
ab = reshape(ab,nrows*ncols,2);
nColors = 3;
[cluster_idx cluster_center] = kmeans(ab,nColors,'distance','sqEuclidean', ... 'Replicates',3);

pixel_labels = reshape(cluster_idx,nrows,ncols);
figure,imshow(pixel_labels,[]), title('Image Labeled by Cluster Index');
```

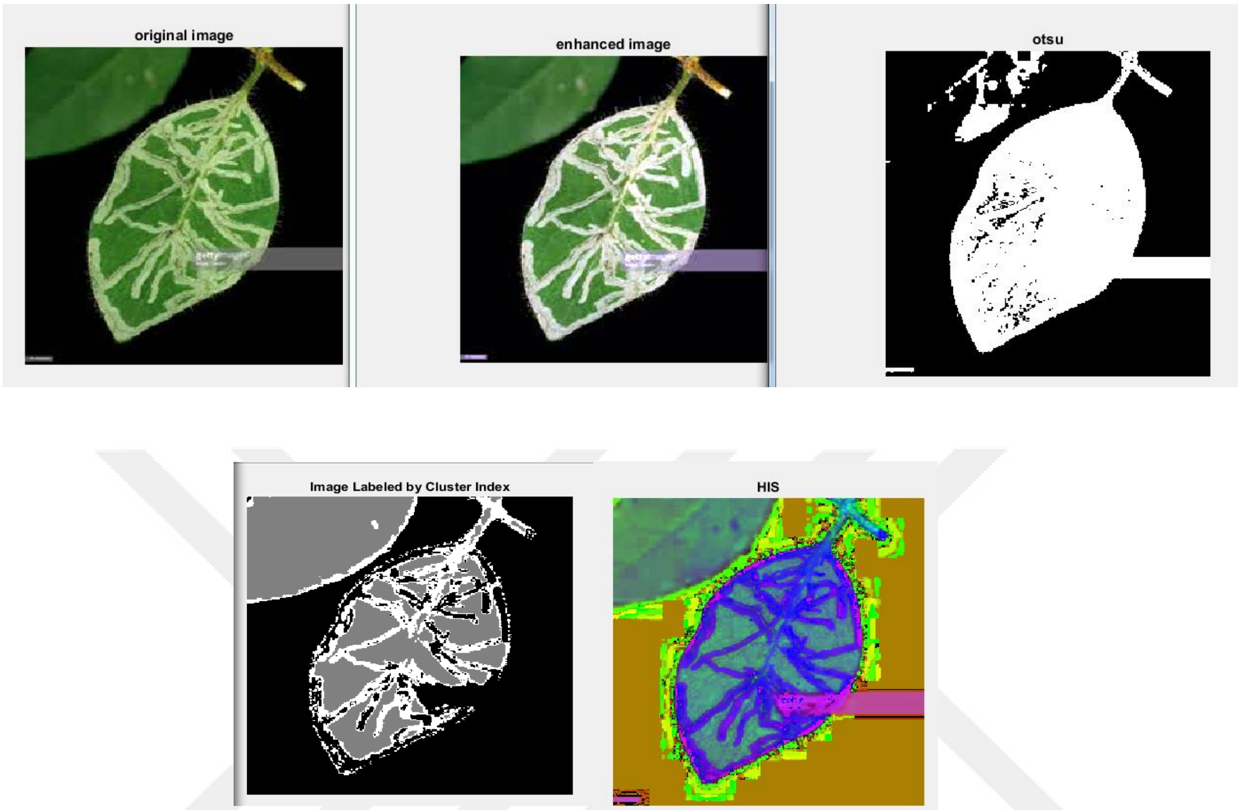

Figure 4.8: Result image after apply image processing steps.

4.7 BWCONNCOMP

The function of " bwconncomp" (black and white connection component) is classifying and labeling the pixels of a scanned image and groups the pixels of this image into components by pixel connectivity. Using the code (bwconncomp) to Find connected components in binary image this code find the connected components CC found in the binary image (black and white) . bwconncomp uses default connectivity of 8 for two dimensions, 26 for three dimensions.

In our reserch, we determine 6 require connectivity to connected components for a two-dimension image of spot infected leaf image to calculate the area of the infected leaf spot in .pixels.

We use this code in the project :-

"cc = bwconncomp(seg_img,6); "

4.8 REGIONPROPS (MEASURE PROPERTIES OF IMAGE REGIONS)

Returns the measurements of the set of properties specified by properties for each 8-connected component in the binary image(in plant leaf), and define "BW. stats" is a structure array that includes a structure for each object in the image. and we can use "regionprops" in contiguous regions and discontinuous regions. we use this code to find the leaf of the plant from the total image, to determine the total area of the plant leaf and to calculate the percentage of disease spot area in the leaf. We use this code in the program.

```
Leaf data = regionprops(kk,'basic')
```

4.9 GLCM (GRAY-LEVEL CO-OCCURRENCE MATRIX)

This code creates a gray-level co-occurrence matrix (GLCM) from the image I. Another name for a gray-level co-occurrence matrix is a"gray-level spatial dependence matrix". "graycomatrix" generates GLCM by computing the number of times the gray pixel (grayscale intensity)value " i "horizontally adjacent next to the pixel with the value j. (we can specify other pixel spatial relationships using the "Offsets" parameter.) Each element (i,j) in glcm specifies the number of times that the pixel with the value "i" occurred horizontally adjacent to a pixel with value j.

When applying "glcm" code and by repeating this process on the infected leaf image, we can compute the number of infected disease spots and the regions of disease in the leaf, by the code below.

```
" glcms = graycomatrix(img)"
```

5. DISCUSSION

In this research, we have been able to improve the performance of the default to detect four types of diseases that afflict plants in addition to the identification of healthy leaves in plants without the disease. In addition to detecting a new plant disease (Tuta Absoluta) which appeared recently in many countries and caused significant damage to agricultural crops, by creating a new dataset from the features that extracted from plant leaves infected to detect the disease.

In this final project, we present the methods related to the plant diseases detection affecting plants and extract the features through the knowledge of the chromatic and texture effect that occurs in the shape and color of the plant leaf when the disease, and through certain differences in the plant's leaf caused by each type of disease compared to the healthy leaf plants can detect the type of disease, After performing the image processing by Matlab program, image resizing, image filtering, to remove impurities and noise, contrast enhancement, and using the global image threshold using the " OTSO" method. This method is used to convert pixel in the image from intensity to black and white image(0 or 1). The gray threshold function uses Otsu's method, which selects the threshold to reduce the intraclass variance of the white and black pixels, and the use of the function" HSI" is the function of converting RGB to HIS. Display image hue, image saturation, image density, image segmentation using K-Min to three sections where we choose which image shows the disease area more clearly as well as naming each pixel in the color matrix in the image, as well as solutions based on color and image analysis algorithms to detect Areas of high variability in color and variation in plant leaf texture.

This project evaluated the obtained performance from three directions:

First, the main problem of noise and distortions in the image due to the lack of clarity and light produced by camera flash or any other source of light or shadow that gives dark areas in the image and thus the background color of the image will be close to the color degree of the leaves, which reduces the contrast and makes the process of detecting the area affected by the disease more difficult. Several methods have been carefully developed and tested, revealing their benefits and disadvantages in order to improve the approach obtained.

Second, similarities in the physiological changes that occur in the tissue and color of the plant leaf when infected in more than one disease, that means there is more than one disease occur changes and similar symptoms on the color and texture of tree leaves, which leads to the difficulty to detect the disease from other diseases which have similar symptoms in the plant.

Third, features extraction, extract the features from the different shape of the plant leaves and the color of all plant leaves is almost dark green and have a symptoms similar to different diseases, which makes extracting the features of many diseases in the same database is difficult Because of a similarity in the values of the attributes entered in the database, may be more than one disease have the same value of the same feature, that make confuse in the dataset system.

6. CONCLUSION

This paper introduced the efficient of use Image Processing Using MATLAB and Machine Learning to Detect Diseases that Occur On Plant Leaf this application is easier to use from the farmer, it will help farmers to check infected plants without any manual supervision. Data obtained from farmers will help to recognize the type of disease in a particular region and thus take proper action. The Disease area are various in color if it compared with the green color of a healthy plant leaf. So we transformed the color of RGB image to use for more suitable segmentation of the disease spots in a plant. we will comparison In this paper the effect of, HSI in the operation of disease spot in plant detection. then, we can calculate the threshold by applying the Otsu method on the component of the color to recognition the disease area and the type of disease in plant leaves. An algorithm which avoids the background noise, plant type and the difference of disease spot color was developed and various experiments were carried out on plants leaves, this software will provide short time and low cost for users to know the disease that afflicts their own plants.

Some constraints in work are the Irregular plant leaf images collected cannot be directly classified by the automation techniques such as images whose background color is similar to the intensity of the leaf color. And that's because: (a) Because of the noise in the image, lack of clarity in the features which occur in the poor contrast of the original leaf image and (b) The large dimensions of the input image size cause large data size and thus complexity in the system.

6.1. FUTURE WORK SCOPE

1- Securing food through the development and increasing the production of agricultural crops.

2- Preserving the green vegetation cover and reducing the area of Desertified land area as well as reducing the temperature of the atmosphere, especially in hot and dry areas.

3- This application can be able used easily by Anyone, even illiterate Because it deals with digital images of plant leaves so it gives scope for all people interested in agriculture.

4- Reduces human effort because it works automatically and will help farmers and specialists in non-specialized in agriculture, especially those working in the field of modern agriculture in greenhouses in the verification of plant health and quality of crops without any supervision or manual control.

5- the Dataset on plant diseases obtained from farmers will help competent authorities or government to identify the disease occurring in large areas, and therefore take appropriate measures for treatment and control, to avoid the occurrence of epidemics may cause economic disasters.

REFERENCES

[1] Shen Weizheng, Wu Yachun, Chen zhanliang and Wei Hongda3, "Grading Method of Leaf Spot Disease Based on Image Processing", International Conference on ComputerScience and Software Engineering, IEEE 2008, pp. 491-494.

[2] Libo luo and Guomin Zohu, "Extraction of the Rice Leaf Disease Image Based on BP Neural Network", International Conference on Computational Intelligence and Software Engineering (CiSE 2009), IEEE

[3] Nunik Noviana Kurniawati, Siti Norul Huda Sheikh Abdullah, Salwani Abdullah, Saad Abdullah, "Investigation on Image Processing Techniques for Diagnosing Paddy Diseases", International Conference of Soft Computing and PatternRecognition, 2009 IEEE, pp. 272-277.

[4] Geng Ying, Li Miao, Yuan Yuan and Hu Zelin, "A Study on the Method of Image Pre-Processing for Recognition of Crop Diseases", International Conference on Advanced Computer Control, 2008 IEEE, pp. 202-206.

[5] http://theseedsite.co.uk/monocots2.html, Accessed on april 28, 2012.

[6] http://www.backyardnature.net/mondiclf.htm, Accessed on April 28, 2012.

[7] Santanu Phadikar and Jaya Sil, "Rice Disease Identification using Pattern Recognition Techniques", Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008), Khulna, Bangladesh, IEEE, pp. 420-423.

[8] Di Cui, Qin Zhang , Minzan Li, Glen L. Hartman and Youfu Zhao, "Image Processing Methods for Quantitatively Detecting Soybean Rust from Multispectral Images", Published by Elsevier Ltd, Biosystems Engineering 107(2010), pp. 186-193.

[9] Song Kai, liu zhikun,Su hang,Guo chunhong ,"A Research of Maize Disease Image Recognition of Corn Based on BP Networks", Third International Conference on Measuring Technology and Mechatronics Automation, 2011 IEEE, pp. 246-249.

[10] Prof. Sanjay B. Dhaygude, Mr.Nitin P.Kumbhar, "Agricultural plant Leaf Disease Detection Using Image Processing", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 1, January 2013

[11] Taha Omar Ali Ben Husin Biological control of tomato leaf miner Tuta absoluta using entomopathogenic nematodes.

[12]Retinal disease detection from the site https://shodhganga.inflibnet.ac.in/bitstream/10603/10111/9/09_chapter%203.pdf

[13] Santanu Phadikar and Jaya Sil, "Rice Disease Identification using Pattern Recognition Techniques", Proceedings of 11th International Conference on Computer and Information Technology (ICCIT 2008), Khulna, Bangladesh, IEEE, pp.420-423.

[14]SVM and ANN Based classification of Plant Disease and ANN Based Classification of Plant Diseases Using Feature Reduction Technique Jagadeesh D.Pujari, Rajesh Yakkundimath and Abdulmunaf. Syedhusain Byadgi

[15] Stone, M. C. (August 2001). "A Survey of Color for Computer Graphics". Course at SIGGRAPH 2001.

[16] Ford, A. and Roberts, A., (2010) Color Space Conversions, August 11, 1998. http://www.poynton.com/PDFs/coloureq.pdf, viewed on May2010.

[17] Fast and Accurate Detection and Classification of Plant Diseases**,** H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh Department of Information Technology Al-Balqa' Applied University, Salt Campus, Jordan**.**

[18] A field guide for identification and scoring methods of diseases in the mountain crops of Nepal , Hira Kaji Manandhar, Ram Devi Timila, Sarala Sharma, Sharada Joshi, Shrinkhala Manandhar, Suk Bahadur Gurung, Sajal Sthapit, Epsha Palikhey, Anju Pandey, Balkrishna Joshi, Gyanu Manandhar, Devendra Gauchan, Devra Ivy Jarvis and Bhuwon Ratna Sthapit. Bioversity International 2016.

[19] Machine Learning and Event-Based Software Testing: Classifiers for Identifying Infeasible GUI Event Sequences.

[20] Otsu, N., "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, 1979, pp. 62-66.

[21] Color Transform Based Approach for Disease Spot Detection on Plant Leaf" Piyush Chaudhary, Anand K. Chaudhari2, Dr. A. N. Cheeran3 and Sharda Godara, Electrical Department, Veermata Jijabai Technological Institute, Mumbai, India Accenture Services Pvt. Ltd., Pune, Maharastra, India.