DETECTION AND MEASUREMENT OF PADDY LEAF DISEASE SYMPTOMS USING IMAGE PROCESSING

Abstract

Plants are one of the major resources to avoid the global warming in the world. But the plants are affected by the diseases like Blast, Canker, Black spot, Brown spot, Bacterial leaf Blight and Cotton mold .The objective of this paper is to recognize the leaf diseases. For examples some the disease is Blast Disease (BD), Brown spot Disease (BPD), Narrow Brown spot disease (NBSD), which stops the growth and protection of the paddy. Disease can infect at different stages of growth and all parts of the plants as the leaf neck and the node. The list of the disease can be caused by bacteria, fungus etc. The methodology was designed to remove the noise automatic, error by human and minimizing the time taken to mensurate the affect of leaf disease. And it also increases the accuracy. In this project it survey, k-means techniques for leaf detection and identification. Three classification techniques are performed to find the disease name and rate of accuracy. The three techniques are SVM, ANN and Fuzzy. Results are compared and identified the best classification technique.

CHAPTER 1
INTRODUCTION
1.1 Image Processing

Image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it.

Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging.

Closely related to image processing are computer graphics and computer vision. In computer graphics, images are manually made from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from natural scenes, as in most animated movies. Computer vision, on the other hand, is often considered high-level image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans).

Image processing is a method to convert an image into digital form and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which input is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them.

It is among rapidly growing technologies today, with its applications in various aspects of a business. Image Processing forms core research area within engineering and computer science disciplines too.

Image processing basically includes the following three steps.

- Importing the image with optical scanner or by digital photography.
- Analyzing and manipulating the image which includes data compression and image enhancement and spotting patterns that are not to human eyes like satellite photographs.
- Output is the last stage in which result can be altered image or report that is based on image analysis.

The purpose of image processing is divided into 4 groups. They are:

1. Visualization - Observe the objects that are not visible.
2. Image sharpening and restoration - To create a better image.
3. Image retrieval - Seek for the image of interest.
4. Measurement of pattern – Measures various objects in an image.
4. Image Recognition – Distinguish the objects in an image.

1.2 Types of Image Processing:

The two types of methods used for Image Processing are Analog and Digital Image Processing.

1.2.1 Analog Image Processing

Analog image processing is any image processing task conducted on two-dimensional analog signals by analog means (as opposed to digital image processing). If the pictorial representation of the data represented in analog wave formats that can be named as analog image. Eg: television broadcasting in older days.

. Analog or visual techniques of image processing can be used for the hard copies like printouts and photographs. Image analysts use various fundamentals of interpretation while using these visual techniques. The image processing is not just confined to area that has to be studied but on knowledge of analyst. Association is another important tool in image processing through visual techniques. So analysts apply a combination of personal knowledge and collateral data to image processing.

An analog or analogue signal is any continuous signal for which the time varying feature (variable) of the signal is a representation of some other time varying quantity, i.e., analogous to another time varying signal. For example, in an analog audio signal, the instantaneous voltage of the signal varies continuously with the pressure of the sound waves. It differs from a digital signal, in which a continuous quantity is represented by a discrete function which can only take on one of a finite number of values. The term analog signal usually refers to electrical signals; however, mechanical, pneumatic, hydraulic, human speech, and other systems may also convey or be considered analog signals.

An analog signal uses some property of the medium to convey the signal's information. For example, an aneroid barometer uses rotary position as the signal to convey pressure information. In an electrical signal, the voltage, current, or frequency of the signal may be varied to represent the information.

Any information may be conveyed by an analog signal; often such a signal is a measured response to changes in physical phenomena, such as sound, light, temperature, position, or pressure. The physical variable is converted to an analog signal by a transducer. For example, in sound recording, fluctuations in air pressure (that is to say, sound) strike the diaphragm of a microphone which induces corresponding fluctuations in the current produced by a coil in an electromagnetic microphone, or the voltage produced by a condensor microphone. The voltage or the current is said to be an "analog" of the sound.

An analog signal has a theoretically infinite resolution. In practice an analog signal is subject to electronic noise and distortion introduced by communication channels and signal processing operations, which can progressively degrade the signal-to-noise ratio (SNR). In contrast, digital signals have a finite resolution. Converting an analog signal to digital form introduces a constant low-level noise called quantization noise into the signal which determines the noise floor, but once in digital form the signal can in general be processed or transmitted without introducing additional noise or distortion. Therefore, as analog signal processing systems become more complex, they may ultimately degrade signal resolution to such an extent that their performance is surpassed by digital systems. This explains the widespread use of digital signals in preference to analog in modern technology. In analog systems, it is difficult to detect when such degradation occurs. However, in digital systems, degradation can not only be detected but corrected as well.

The primary disadvantage of analog signals is that any system has noise – i.e., random unwanted variation. As the signal is copied and re-copied, or transmitted over long distances, or electronically processed, the unavoidable noise introduced by each step in the signal path is additive, progressively degrading the signal-to-noise ratio, until in extreme cases the signal can be overwhelmed. This is called generation loss. Noise can show up as "hiss" and inter modulation distortion in audio signals, or "snow" in video signals. This degradation is impossible to recover, since there is no sure way to distinguish the noise from the signal; amplifying the signal to recover attenuated parts of the signal amplifies the noise (distortion/interference) as well. Digital signals can be transmitted, stored and processed without introducing noise. Even if the resolution of an analog signal is higher than a comparable digital signal, after enough processing the analog signal to noise ratio will be lower.

However, unlimited processing capacity to refine a digital signal and the resources required should not be taken for granted. In other words, processing capacity available for a digital signal is itself an analog signal - it reflects non-discrete factors "in the real world" such as quantity and quality of computers, connections between them, sufficient electric power, shutdowns, hacker attacks, running costs and everything else that can affect performance and availability. In comparison, analog signals are normally affected by a limited range of factors, whose degrading effect is, for the most part, already represented by the noise. So while an aircraft's altimeter may be thrown off if the plane suddenly enters an air zone with different pressure (an emergency scenario), it will give correct altitude most of the time. Subtle pressure changes are already covered by the small random vacillations (noise) an altimeter can be configured to ignore - in the simplest case, the trembling of the gauge arrow. Because a digital signal by definition does not vary with another continuous quantity that could be compensated for, all digital interference that does occur is to some extent unpredictable. Electrically, analog noise can be diminished by shielding, good connections and several cable types such as coaxial or twisted pair.

1.2.2 Digital image processing

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and signal distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modelled in the form of multidimensional systems.

Digital Processing techniques help in manipulation of the digital images by using computers. As raw data from imaging sensors from satellite platform contains deficiencies. To get over such flaws and to get originality of information, it has to undergo various phases of processing. The three general phases that all types of data have to undergo while using digital technique are Pre- processing, enhancement and display, information extraction.

Digital image processing allows the use of much more complex algorithms, and hence, can offer both more sophisticated performance at simple tasks, and the implementation of methods which would be impossible by analog means.

In particular, digital image processing is the only practical technology for:

- Classification
- Feature extraction
- Pattern recognition
- Projection
- Multi-scale signal analysis

Some techniques which are used in digital image processing include:

- Pixelation
- Linear filtering
- Principal components analysis
- Independent component analysis
- Hidden Markov models
- Anisotropic diffusion
- Partial differential equations
- Self-organizing maps
- Neural networks
- Wavelets

A digital signal is a physical signal that is a representation of a sequence of discrete values (a quantified discrete-time signal), for example of an arbitrary bit stream, or of a digitized (sampled and analog-to-digital converted) analog signal. The term digital signal can refer to either of the following:

1. any continuous-time waveform signal used in digital communication, representing a bit stream or other sequence of discrete values
2. a pulse train signal that switches between a discrete number of voltage levels or levels of light intensity, also known as a line coded signal or baseband transmission, for

example a signal found in digital electronics or in serial communications, or a pulse code modulation (PCM) representation of a digitized analog signal.

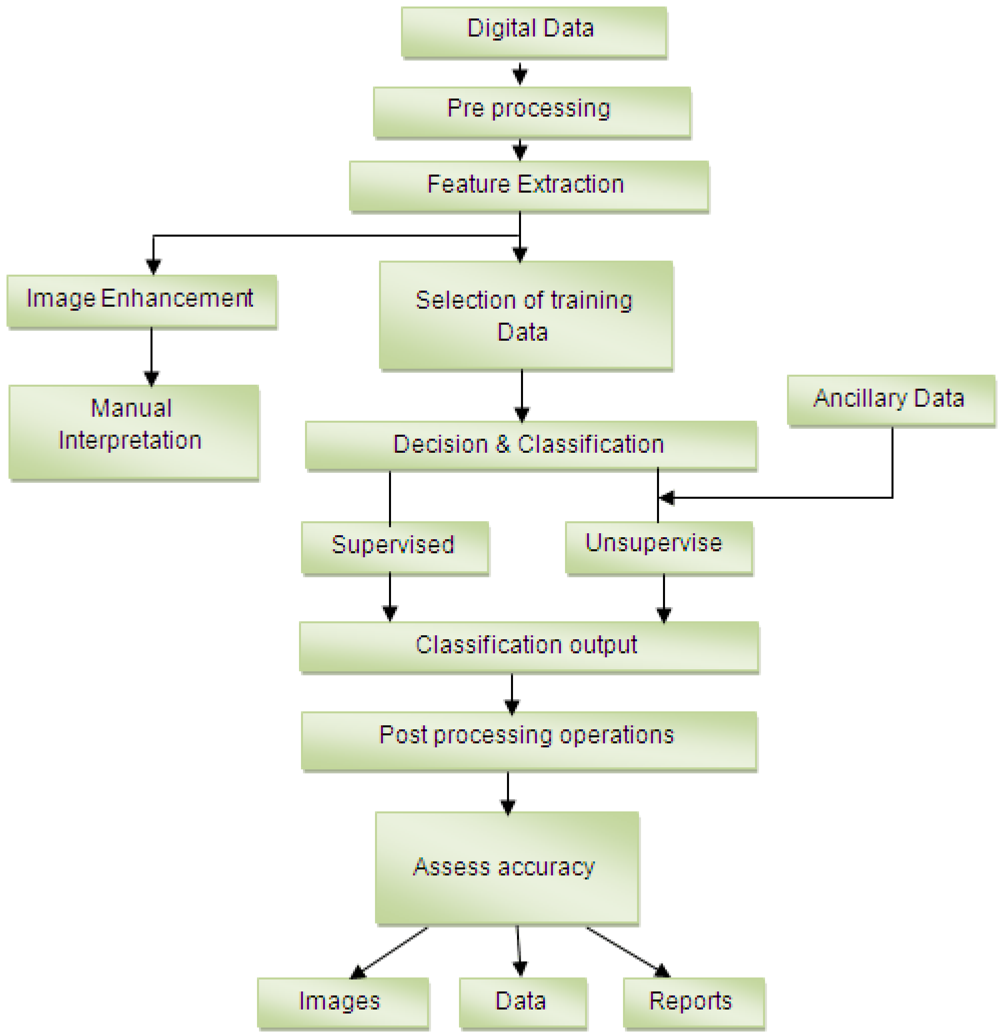


Figure 1.1 Digital Signal Processing

1.3 Image segmentation

In image segmentation, one challenge is how to deal with the nonlinearity of real data distribution, which often makes segmentation methods need more human interactions and make unsatisfied segmentation results. Medical image segmentation plays an instrumental role in clinical diagnosis. An ideal medical image segmentation scheme should possess some preferred properties such as minimum user interaction, fast computation, and accurate and robust segmentation results.

Image segmentation is an image analysis process that aims at partitioning an image into several regions according to a homogeneity criterion. Image segmentation is a very complex task, which benefits from computer assistance, and yet no general algorithm exists. It has been a research field in computer science for more than 40 years now, and the early hope to find general algorithms that would achieve perfect segmentations independently from the type of input data has been replaced by the active development of a wide range of very specialized techniques. Most of the existing segmentation algorithms are highly specific to a certain type of data, and some research is pursued to develop generic frameworks integrating these techniques.

Segmentation can be a fully automatic process, but it achieves its best results with semi-automatic algorithms, i.e. algorithms that are guided by a human operator. This concept of semi-automatic process naturally involves an environment in which the human operator will interact with the algorithms and the data in order to produce optimal segmentations. The simplest example of the need of a human intervention during the task of segmentation results from the specificity of -the existing algorithms: depending on the type of input data, the operator will have to carefully pick the best adapted algorithm, which most of the time cannot be done in an automatic way. The subjective point of view of the human is required.

1.4 Colour Image Processing

The use of colour in image processing is motivated by two principal factors. First, colour is a powerful descriptor that often simplifies object identification and extraction from a scene. Second, humans can discern thousands of colour shades and intensities, compared to about only two dozen shades of gray. This second factor is particularly important in manual image analysis.

1.4.1 Background

- Humans can perceive thousands of colors, and only about a couple of dozen gray shades (cones/rods)
- Divide into two major areas: full color and pseudo color processing
- Full color – image is acquired with a full-color sensor like TV camera or color scanner

  Pseudo colour – Assign a colour to a range of monochrome intensities
- 8-bit color vs 24-bit color

1.4.2 Colour fundamentals

- Color spectrum/prism
- White light divided into different colors
- Colors blend into each other smoothly
- Color – Perceptual result of light in the visible region of spectrum as incident on the retina 400 nm to 700 nm
- White light is result of reflected light balanced across all visible wavelengths

1.4.3 Colour models

- ➢ RGB colour model
- ➢ CMYK colour model
- ➢ HSI colour model

RGB: The RGB colour model relates very closely to the way we perceive colour with the **r**, **g** and **b** receptors in our retinas. RGB uses additive colour mixing and is the basic colour model used in television or any other medium that projects colour with light. It is the basic colour model used in computers and for web graphics, but it cannot be used for print production.

The secondary colours of RGB – cyan, magenta, and yellow – are formed by mixing two of the primary colours (red, green or blue) and excluding the third colour. Red and green combine to make yellow, green and blue to make cyan, and blue and red form magenta. The combination of red, green, and blue in full intensity makes white.

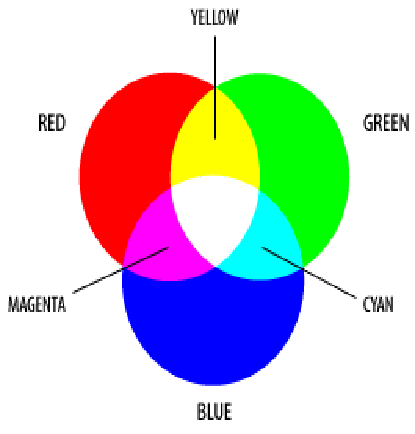


Figure 1.2 Representation of RGB colour model

CHAPTER 2
LITERATURE SURVEY

This chapter briefly reviews, explains and discusses on existing literature review related with the current project which is “Paddy Disease Detection System Using Image Processing” that will be developed later. This chapter comprises three sections. The first section describes the overviews of paddy. The subsections are the definition, type of paddy disease, paddy symptom and paddy management. The second section is the review of some existing system that used same techniques and methods. The third section discusses the review on technique and method used by the system. The subsections are image acquisition, image segmentation and artificial neural network.

2.1 Paddy Overviews

In this section, firstly presents a definition of paddy. After that, this subsection briefly discusses on type of paddy disease, symptoms and management of paddy disease.

2.1.1 Definition of Paddy

Paddy also known as rice is the starchy seeds of an annual southeast Asian cereal grass (*Oryza sativa*) that are cooked and used for food. This cereal grass that is widely cultivated in warm climates for its seeds and by-products [8]. Rice is one of the most utilized food plants and widely grown originated in ASIA. Rice is an important crop worldwide and over half of the world population relies on it for food. Many people in the world including Malaysia eat rice as staple food.

2.1.2 Paddy Diseases, Symptoms and Management

There are many factors that make paddy rice production become slow and less productive. One of the main factors is paddy disease. The table below will show you type of paddy disease, the symptom of paddy disease and the management of paddy disease. This researches focus on three types of diseases, which are paddy blast, brown spot disease and narrow brown spot disease. The following paddy diseases are shown below:

A. Paddy Blast

Paddy Blast Symptoms

- Disease can infect paddy at all growth stages and all aerial parts of plant (Leaf, neck and node).
- Among the three leaves and neck infections are more severe.
- Small specks originate on leaves - subsequently enlarge into spindle shaped spots(0.5 to 1.5cm length, 0.3 to 0.5cm width) with ashy center.

- Several spots coalesce -> big irregular patches

Fig. 2.1 : Blast disease on leaf

Management

- Avoid excess N - fertilizer application
- Apply nitrogen in three split doses.
- Removes weed hosts from bunds.
- Use of tolerant varieties (Penna, Pinakini, Tikkana, Sreeranga, Simphapuri, Palghuna, Swarnamukhi, Swathi, Prabhat, Co 47, IR - 64, , IR - 36, Jaya)
- Burning of straw and stubbles after harvest
- Dry seed treatment with *Pseudomonas fluorescens*talc formulation @10g/kg of seed.
- Stagnate water to a depth of 2.5cm over an area of 25m2 in the nursery. Sprinkle 2.5 kg of *P. fluorescens* (talc) and mix with stagnated water. Soak the root system of seedlings for 30 min and transplant.
- Spray *P. fluorescens* talc formulation @ 0.5% from 45 days after transplanting at 10 day intervals, three times.
- Seed treatment at 2.0 g/kg seed with Captan or Carbendazim or Thiram or Tricyclazole.
- Spraying of Tricyclazole at 1g/lit of water or Edifenphos at 1 ml/lit of water or Carbendazim at 1.0 gm/lit.
- 3 to 4 sprays each at nursery, tillering stage and panicle emergence stage may be required for complete control.

Nursery stage

- Light infection - Spray Carbendazim or Edifenphos @ 0.1 %.

Pre-Tillering to Mid-Tillering

- Light at 2 to 5 % disease severities - Apply Edifenphos or Carbendazim @ 0.1 %. Delay top dressing of N fertilizers when infection is seen. Panicle initiation to booting
- At 2 to 5% leaf area damage spray Edifenphos or Carbendazim or Tricyclazole @ 0.1 %.

Flowering and after

At 5 % leaf area damage or 1 to 2 % neck infection spray Edifenphos or Carbendazim or Tricyclazole @ 1 g /lit of water.

B. Paddy Brown Spot Disease

Symptoms

- Initial lesions are water-soaked to greenish gray and later become grayish white with brown margin.
- Lesions on leaf sheaths near waterline
- Presence of sclerotia
- Lesions may coalesce death of the whole leaf
- Partially filled or empty grains

Fig 2.2: Paddy Brown Spot Disease

Management

- Apply FYM 12.5 t/ha or green manure 6.25 t/ha to promote antagonistic microflora.

- Soil application of *P. fluorescens* @ 2.5 kg/ha mixed with 50 kg FYM after 30 days of transplanting.
- Foliar spraying of *P. fluorescens* @0.2% at boot leaf stage and 10 days later.
- Avoid flow of irrigation water from infected to healthy field.
- Carbendazim (1 g/lit), Propiconazole (1ml/lit) may be applied.
- Spraying of infected plants with fungicides, such as Benomyl and Iprodione, and antibiotics, such as Validamycin and Polyoxin, is effective against the disease
- Reduce Nitrogen dosage and skip top dressing

C. Narrow Brown Spot Disease

Symptoms:

- Short, narrow, elliptical to linear brown lesions
- usually on leaf blades but may also occur on leaf sheaths, pedicels, and glumes or rice hulls
- Lesions about 2-10 mm long and 1 mm wide
- Lesions narrower, shorter, and darker brown on resistant varieties
- Lesions wider and lighter brown with gray necrotic centers on susceptible varieties
- Leaf necrosis may also occur on susceptible varieties
- Lesions occur in large numbers during the later growth stages

Fig 2.3 : Narrow Brown Spot Disease

Why Occurs?

- The disease is observed on rice crops grown on soil deficient in potassium.

- Temperature ranging from 25-28° C is favorable for the optimum growth of the disease. Susceptibility of the variety to the fungus and the growth stage of the rice crop are other factors that affect the development of the disease. Although rice plants are susceptible to the fungus at all stages of growth, they are more susceptible from panicle emergence to maturity, thus, becoming more severe as rice approaches maturity.

Management

- Cultural practices, such as the use of potassium and phosphorus fertilizers, and planting of early maturing cultivars early in the growing season, are recommended to manage the narrow brown leaf spot.
- The use of resistant varieties is the most effective approach to manage the disease. However, the resistant varieties and lines are only grown in United States and India.
- Spraying of fungicides such benomyl, propicanazole,carbendazim, propiconazole, and iprodione, when the disease is observed in the field is effective.

2.2 Grape Leaf Disease Detection from Color Imagery using Hybrid Intelligent System

A study conducted by A. Meunkaewjinda, P. Kumasawat, K. Attakitmongcol and A. Srikew on Grape Leaf Disease Detection from Color Imagery using Hybrid Intelligent System [2]. Vegetable and fruits are the most important export agricultural products of Thailand. In order to obtain more value added products, a product quality control is essentially required. Many studies show that quality of agricultural products may be reduced from many causes. One of the most important factors of such quality is plant disease.

Consequently, minimizing plant diseases allows substantially improving quality of the products. This system shows an automatic plant disease diagnosis using multiple artificial intelligent techniques. The system can analyze or diagnosis plant leaf disease without maintain any expertise once the system is trained.

2.2.1 Method

There are several steps must be achieve of this system which are grape leaf color segmentations, grape leaf disease segmentation and analysis and classification of the disease.

A Grape leaf color segmentation

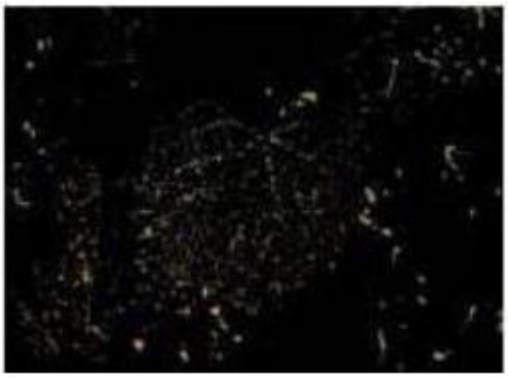
The grape leaf color segmentation is pre-processing module which segments out of any irrelevant background information. A self-organizing features map together with a back-propagation neural network is deployed to recognize colors of grape leaf. This information is used to segment grape leaf pixels within the image. The input is enhanced by using anisotropic diffusion technique [12] to preserve information of the affected pixels before extracting grape color from the background.

(a) Original image (b) Grape leaf with color extraction

Fig. 2.4 : Example of grape leaf color extraction

B Grape leaf disease segmentation

The grape leaf disease segmentation is performed using modified selforganizing features map with genetic algorithms for optimization and support vector machines for classification. Using modified self-organizing feature map (MSOFM) [13], the clustering process does not require any training nor predefined number of color groups.



(a) Original image (b)Grape leaf disease color extraction

Fig 2.5: Example of grape leaf disease segmentation

C Analysis and Classification of Disease

The resulting segmented image is filtered by Gabor wavelet [14] which allows the system to analyze leaf disease color features more efficient. The support vector machines are then again applied to classify types of grape leaf diseases. The image can be able to categorize the image of grape leaf into three classes:

i) Scab disease

ii) Rust disease

iii) No disease

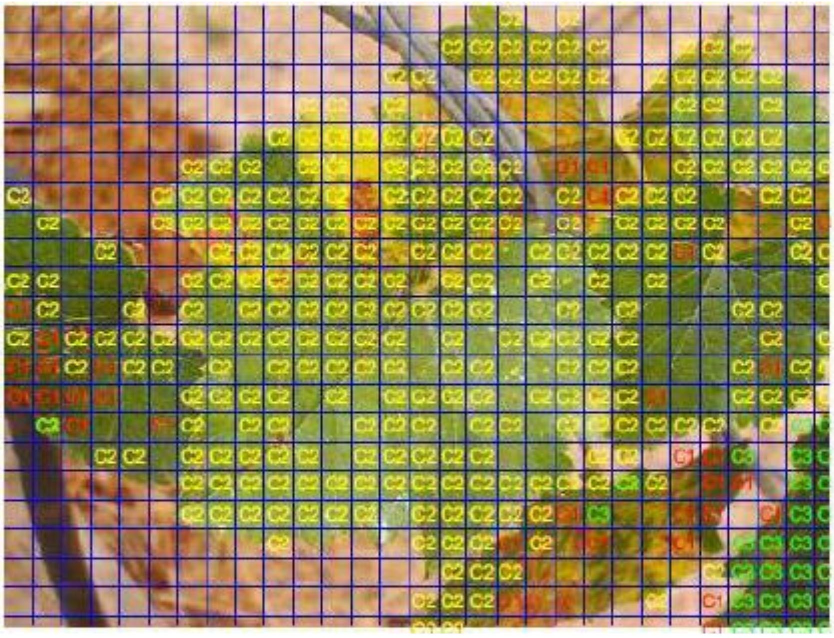


Fig 2.6 Example of Rust disease

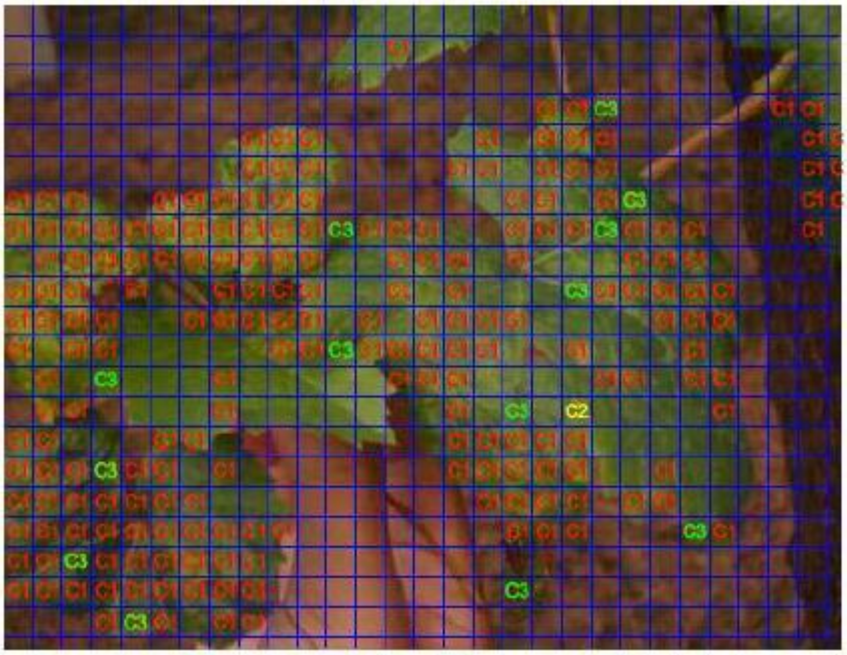


Fig 2.7 Example of SCAB disease

2.3 A Framework for Detection and Classification of Plant Leaf and Stem Disease

A study conducted by Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad on a framework for Detection and Classification of Plant Leaf and Stem Diseases. Studies show that relying on pure naked-eye [15] observation of experts to detect such diseases can be prohibitively expensive, especially in developing countries. Providing fast, automatic, cheap and accurate image processing-based solutions for that task can be of great realistic significance.

2.3.1 Method

The proposed framework is image processing-based and is composed of the following main steps;

a) The images at hand are segmented using the K-Means technique

b) The segmented images are passed through a pre-trained neural network.

A Clustering method

K-means clustering is used to partition the leaf image into four clusters in which one or more clusters contain the disease in case when the leaf is infected by more than one disease.The k-means clustering algorithms tries to classify *objects* (pixels in our case) based on a set of features into K number of classes. The classification is done by minimizing the *sum of squares* of distances between the objects and the corresponding cluster or class *centroid*. On this experiment, the K-means clustering is set to use squared Euclidean distances.

B Feature Extraction

The method followed for extracting the feature set is called the *Color Co-occurrence Method* or CCM method in short. It is a method, in which both the color and texture of an image are taken into account, to arrive at unique features, which represent that image.

- Co-occurrence Methodology for Texture Analysis
- Normalizing the CCM matrices
- Texture features identification

As a testbed, D. Al Bashish (2001) use a set of leaf images taken from Al-Ghor area in Jordan. There are five diseases which effect on the plants;

a) Cottony mold

b) Early scorch

c) Ashen mold

d) Late scorch

e) Tiny whiteness.

The experimental results indicate that the proposed approach can significantly support accurate and automatic detection of leaf diseases. The developed Neural Network classifier that is based on statistical classification perform well and could successfully detect and classify the tested diseases with a precision of around 93%.

CHAPTER 3

TECHNIQUES AND MEASURING LEAF DISEASES

3.1 Image Processing

Rao (2014), is defined as the process of improving and enhancing the raw images that are taken through digital cameras, sensors, and many other sophisticated means such as Satellite, space probes and aircrafts for various applications. When someone suggested thousands of years ago that “a picture speaks a thousand words”, probably the idea of computing was limited to basic number crunching. Above adage still has significance to computing with images. Most researchers in computer vision and image processing aim at deriving effective and better tools as well as proper approaches that give different ideas on the same image by providing means to comprehend not only the content of the image but also give meaning, and significance of the image. There is no way image processing can be compared or matched with the human eye in terms of accuracy, but it can outperform it easily on observational consistency, and ability to carry out detailed mathematical operations. Again, image processing can be used to compute and find solutions to simple or structured tasks by providing reliable, consistent and cheap results. Unlike some years back, researches conducted based on image processing in recent years have been broadened widen to cover a large range of information ranging from simple and basic pixel based low-level operations to high-level analysis that now includes the use sophisticated tools including techniques like artificial intelligence for the purposes of interpretation and understanding of the image. These new and modern techniques for processing images are being developed to get a better meaning and understanding of images based on the relationship between its components, its context, and its history if it is a part of a sequence, and a priori knowledge gained from a range of images.

3.2 Uses of Image Processing

(Jayamala K. Patil and Rai Kumar, 2011) identified about five main uses of image processing in respect to agricultural plants and their fruits. In their research, they identified that image processing is very useful to:

1. Detect plant leaves, stems and fruits that are affected by diseases.
2. Quantify the areas affected by disease in plant leaves, fruit and stems.
3. Detect the shape of the area of the leaves, fruits and stem that has been affected.
4. Determine the colour of the affected areas and finally
5. Find out the size and shape of fruits.

(K.M.M. Rao, 2014)also identified some uses of Image Processing as applied and used in various applications such as Material Science, Agriculture,Remote Sensing, Medicine,Document processing, engineering, Non-destructive Evaluation, Forensic Studies, Textiles, Military,Printing Industry, Film industry and Graphic design.

3.3 Image Processing Methods

Rao (2014) presented two main methods of image processing. In his paper, he presented Analog image processing as the first method. This method refers to the changes, modification and adjustment of image through electrical means. A typical example of this method is the image produced by the television. The television transmits signals in a form of voltage which varies in amplitude to represent brightness through the image. The writer continues with the second method which he identified as digital method of processing image. In this case, supposed image will be change or converted to digital form through a device known as scanner digitizer for further processing.

Image Processing Techniques

3.3.1. Image segmentation

(Ballard and Brown, 1982) defined image segmentation as the process of dividing or breaking an image into different parts based on certain characteristics. The parts usually conform to something that human beings can easily separate, view and analyze as individual objects. The digital computer as we know does not have the ability to recognize objects intelligently on their own; this is why different researchers have come out with different approaches and methods to segment images. Images are usually segmented depending on the various characteristics and features found in the image. These features may include colour information which is used to create histogram, information about pixels that indicate boundaries and texture information.

In the case of computers, image segmentation can be defined as the process of dividing digital image into several parts or components (a set of pixels that can also be referred to as super pixels). The purpose of segmenting any image is to simplify (make the image easier or less complicated) and change the representation of the image into something meaningful for the purposes of easier analysis. Image segmentation is usually used to detect and discover characteristics, objects and boundaries such as curves and lines in images. Image segmentation can simply be defined as the process of assigning some description to every pixel in an image to identify pixels with similar and unusual characteristics. The outcome of the segmentation process depends on a set of components that together the entire image or a set of outcome deduced from the image. Each of the pixels in a particular region is similar with respect to some characteristic or computed property, including intensity, texture and colour. Neighboring regions are significantly different with respect to the similar characteristics. When applied to a pile of images, typical in medical imaging, the resulting outline after image segmentation can be used to create3Dimensional reconstructions with the help of a written algorithm.

3.3.2 Image Thresholding

Image Thresholding refers to the process of creating a binary image (bitonal) by setting a starting point which serves as the base value of the pixel intensity of the original image. Thresholding technique is usually performed on grayscale images; however, thresholding may be applied to original (true colour) image. The threshold value of every image can either be set manually or automatically using a specific software or application. In this case, all pixels that fall below that set threshold value are converted to black which represents bit value of zero whilst any other pixels above the threshold value are changed to white representing a bit value of one. The thresholding can also be described as the process of breaking an image apart in order to get foreground values as well as background values (black and white). Thresholding can be simple or complex depending on the threshold value. Thresholding is said to be simple if there is only one threshold value set for all the pixels in the image for no matter the difference or variations in contrast. On the other hand, complex and sophisticated thresholding (adaptive thresholding) takes number of regions of the image and setthe threshold value accordingly. It should be noted that, quality cannot be compromised in thresholding especially when dealing with scanning images with Optical Character Recognition (OCR) systems. The simplest form of segmentation is by means of thresholding. This is because you only need to define a threshold then examine all pixels in the image by comparing them with the threshold value. All pixels found above the threshold value are considered as foreground, whilst the pixels located below the threshold value are considered as background. Oftenly, the threshold represents the intensity or the colour value of the image. Other forms of thresholding permit different or variable threshold values throughout the image. In this case, the threshold is permitted to undergo through several changes throughout the entire image. Thresholding is said to be a primitive technique because it works for most operations that need segmentation.

3.3.3 Image thresholding algorithms

There are so many researchers who have proposed different algorithms as far as image thresholding is concerned. This portion of the research seeks to discuss and analyze some of these algorithms as proposed by some researchers. In fact, choosing correct an appropriate algorithm is a difficult thing to do. This is due to the variation resulting from different algorithms since they assume differently about the content of the image. (Ridler and Calvard, 1978) conducted a research into plant diseases using image thresholding technique. They proposed an algorithm that uses iterative clustering approach to estimate disease severity on leaves. In their approach, they used an approximated threshold such as mean image intensity as the initial threshold value. Based on the initial value, they grouped the pixels in the image by assigning white to all pixels found above the threshold all below it are assigned black respectively. The threshold is repeated several times and re-approximated as a mean of the two class means. (Tsai,1985) proposed an algorithm that is used to determine thethreshold of an image by preserving the First three moments in the input image.(Otsu, 1979) proposed a paper that presented an algorithm that uses discriminant analysis approach to detect varied colours in plant leaves. This approach uses the zeroth and the first order cumulative moments of the histogram for calculating the threshold values. (Kapur et.al, 1985) presented an algorithm that measures the extent of deficiency of an image using the thresholding technique through an algorithm they proposed. This algorithm takes the threshold image as two different groups of events with each group represented by a Portable Document Format (pdf). This approach goes on to bring down the level of disorders that exist in the two pdfs thereby achieving a common threshold value. (Parker,1996)described two approaches which implemented the use of the entropy of the intensity histogram according to two definitions. These approaches were based on the method proposed by (Haung and Wang, 1995) and (Yager, 1997)of fuzziness. In another paper published by Rosin in 2011, he also proposed an algorithm that fits a straight line from the top or peak of the intensity to the last of it. The highest point of deviation that exists between the histogram and the line is usually found at the corner which is selected as the threshold value. The suitable and perfect algorithm fits a normal distribution to the intensity of the zxhistogram.

3.4 Measuring Disease severity on Leaves

(Pradnya R. Narvenar et al, 2014) presented a paper on the methods of detecting and analyzing leaf diseases using SGDM matrix method. In their research, several methods were revealed concerning methods of measuring leaf diseases.(Libo Liu and Ouomin Zhou, 2009)conducted a research on the methods of identifying leaf diseases in rice based on the characteristics of the colours of leaf lesion area using the thresholding method. (Vane Zhang, 2005) conducted a research on cucumber in order to determine the methods of diagnosing the nutritional status of green crops usingmachine vision technology. The result showed a mutual relationship betweenthe green components andothercolour componentsof the leaves with the nitrogen which could be used rapidly asdiagnosis of crop diseases indicator under the same conditions. (Chunhua Hu et.al, 2004) used two different statisticstical features of the ROB system to distinguish between the deficiency cucumber blades, using features of the Ohta system. The system was used to identify the colour of different cucumber leaves and obtain the Hue (H) relative percentagehistogramof Hue Saturation Value (HSV) colour system in order to calculate the extent of deficiency in leaves. This algorithm was believed to have been successful.(P. Revathi, et.al, 2012) detected spot disease on cotton leaf using a technique known asIEDS (Image Edge detection Segmentation). In the IEDS technique, the captured images are first processed for enrichment. Afterwards,the image was segmented using Red, Green, and Blue (RGB) colour feature to obtain targeted regions (disease area). They finally extracted certain features including image boundary,image shape, texture and colour to identify diseases and control the pest on the leaves.

Methods of Quantifying Leaf Disease Severity

There are several methods used by various researchers to quantify leaf disease severity. Some of these methods are discussed in this paper. Plant leaf disease can be estimated and quantified either by the area been affected or by the extent of which the disease has been rooted (how deep the affection is) on the leaf that can be estimated through features like colour and texture. Most of the techniques and algorithms used to quantify disease severity include a segmentation step to separate the symptoms in order to extract features and properly process the features to achieve an estimate for the severity of the disease. This paper classifies the strategies and methods used by other researcher based on the specific approach used to estimate the disease severity. Such strategies and approaches include thresholding methods, colour analysis, Fuzzy Logic, Neural networks, Knowledge Base Systems (KBS), Support Vector Machines (SVM) and third party application packages.

3.4.1 Thresholding Method

(Lindow and Webb,1983) quantified plant leaf disease using the thresholding approach. They captured the images under red light using digital video cameras. The purpose of the red light was to highlight the necrotic areas of the image. The images were later converted into digital form. The tests were performed using leaves from four different crops namely tomatoes, bracken fern, sycamore and California buckeye. Simple thresholding method was used to identify the necrotic regions. Their algorithm was further used to apply a correction factor to equally balance the pixels from the healthy regions based on the variations in the colour features of the leaves so that it will be easier to identify all of the pixels that were wrongly counted as part of the diseased areas so that they can be classified to their proper and correct regions. (Tucker et.al, 1997) used a thresholding algorithm to identify and estimate leaf diseases in some plants including sunflower and oat. In the first stage, the algorithm was used to segment the images with varied threshold values depending on how the disease is considered (blight or rust). The pixels obtained in the segmentation process were associated into groups depicting the regions affected by the disease. Based on the attributes and characteristics deduced from the lesion areas, they further classified the lesions into two categories; either category a or category b depending on the exhibited characteristics. This approach was reported to be one of the best methods ever used to estimate leaf diseases. However, it recorded some wrong results due to inadequate lighting system at the time of capturing the images. (Martin et.al, 1998) suggested a new approach of determining and quantifying the symptoms caused by the maize streak virus. Their method was based on an earlier research conducted by Lindow and Webb in 1993. They used a software package to visually compare the results. The results showed that both the commercial off the shelf and costumed applications had almost the same functionalities and efficiency; hence the two methods through computers produced the same output in terms of accuracy and precision as compared to the visual approach. (Weizheng et.al, 2008) proposed a new approach of detecting and estimating leaf disease on soybean using multiple algorithms. They used an algorithm based on two-step algorithms. The first threshold was intended to separate the leaf portion from the background image. After that, the image is then converted into HIS colour space where the Sobel operator is used to estimate the diseased area. This is followed by the second threshold where the threshold is applied to the results obtained from the first threshold. Finally, the objects in the binary image will be discarded filling all the holes enclosed by white pixels. The resulting objects showed the diseased regions. (Llret et.al, 2011) proposed a new system to and monitor the conditions of health in vineyards. They used cameras to acquire the images by spreading the cameras all over the field. The reason for doing that was to detect measure and quantify diseased leaves. They used five different steps to achieve the desired results. The first stage aims at estimating the size of the leaf which is an important step due to varied distance between the cameras and the plants. The second step is the application of thresholding technique to separate the healthy leaves from the diseased ones from the ground by representing the images in Red, Green and Blue (RGB) and HSV forms. This is followed by a set of structural operations which aims to reduce noise but maintaining important and useful features without eliminating them. The fourth step is used to detect and separate the ground from actual diseased leaves. Finally, the diseased ratio is estimated based on the results of the entire image ratio. The system gives a warning that the plant requires some attention if any fault is detected based on the ratio.

3.4.2 The use of Color Analysis

Color Analysis in plants is the process of analyzing the color of plant leaves using color patterns. This method was first developed by John Wolfgang von Geothe (1749-1832), a German Professor philosopher. He used color analysis to determine the differences between different colors and created a color psychology out of his findings. Since then, color analysis has become popular in many fields of academia. Johannes Itten, (1888-1967), a Swiss expressionist painter, teacher, writer and theorist also used colour analysis in most of his art works throughout his career as a professional. This made color analysis more popular and interesting. Currently, color analysis has become a field of study and is more useful in diverse ways including image processing.

(Pagola et.al, 2009) suggested a new system to estimate the level of nitrogen deficiency in plant leaves associated with the barley plant. The proposed system was based through different colour media by applying the Principal Component Analysis approach (PCA) to measure the amount of chlorophyll content in the leaves. To achieve correct and accurate results, the authors performed four different tests using different strategies with the intention of emphasizing on the relevant and ignoring the negative effects of the less important areas that are photosynthetically inactive. Such regions include the veins and areas with spot on the leaves. They stated in their conclusion that, their method had high relationship with the other approaches that are largely adopted based on non-destructive hand-held chlorophyll meters.

Contreas-Medina et.al, 2011) developed a system that could be used to detect and estimate at least five types of symptoms associated with plant leaf diseases. The proposed system comprises five different independent modules. First, the red and green components of the image were combined using Chlorosis algorithm with the intention of determining the yellowness of the leaf which is a symptom of Chlorosis. That is followed by a Necrosis algorithm to differentiate between the leaves and the background through the blue component in the image. The third stage uses the blue to segment and calculates the level of deformation in the leaf through deformation algorithm. The next step applies thresholding on the blue component to calculate the ratio of the diseased area using white spot algorithm. Finally, several types of the venations in the leaves were identified through mosaic algorithm, canny edge detector, and other morphological operations.

3.4.3 Using Fuzzy logic

Fuzzy logic in computing is a procedure based on degrees of truth instead of depending the on Boolean logic which deals with the usual true or false which is the method used by modern computers. Dr. Lotfi Zadeh, a Lecturer of University of California in the early 1960s first advanced the idea of fuzzy logic when working on the computer's understanding of natural language. Fuzzy logic has been so beneficial in computing in many ways including artificial intelligence and image processing.

(Sekulska-Nalewajko et.al, 2011) developed a system to identify and estimate symptoms of some disorders in pumpkin and cucumber leaves. The authors used flatbed scanners to acquire the images for the experiment. They first separated the leaves from the plants, treated and stained them before they were captured. They then implemented their ideas based on functions present in the matlab toolbox. They initially separated the leaves using thresholding algorithms. That was followed by converting the image HSV colour space from RGB space. The brightness component of the image was discarded. Then, the pixels were grouped into two main clusters using Fuzzy c-means algorithm. This grouping represents the healthy and diseased regions. The authors in their conclusion stated that, their system is a simple and efficient way of estimating plant diseases as compared by other methods that according to them need several operations to achieve the desired results.

(Zhou et al, 2011) in a related research conducted by Zhou et.al in 2011, a new method was proposed to examine the degree of grass hopper infestation in rice crops. They focused on the stem of rice plant because the-hoppers extremely manifests in the stem. In the algorithm they first extract the region of interest then fractal-dimension value features are extracted using the box-counting dimension method. These features are used to derive a regression model. They finally classify the regions into four (no infestation region, mild infestation region, moderate infestation region and severe infestation region) using a fuzzy C-means algorithm.

3.4.4 Knowledge-Based System

Knowledge Base System is an aspect in artificial intelligence that uses database of a subject to solve complex problems. It is basically an application that uses artificial intelligence and expert system technique in problem solving.

(Boissard et al, 2008) presented a paper that describes a method to estimate the severity of plant disease in some plants as a method of detecting early pest in the pants. They used two separate knowledge-based systems (KBS) to estimate the number of insects. The first system was based on Classification Knowledge Base System (C-KBS), which takes results from the numerical image processing operations and translate the numerical results into higher level concepts which are then examined to assist the algorithm to choose and keep only the regions that contain insects. Their second system was also based on Supervision Knowledge Base System (S-KBS) which served as a means of selecting the appropriate image processing tools to be applied as well as the constrains to be used in order to collect and feed the most meaningful information to the first system. Even though they achieved some good results, they concluded by stating that their proposal had some problems which were not mentioned and addressed.

3.4.5 Third party image processing packages

(Bock et.al 2008) proposed a system to detect and quantify the severity of Foliar Citrus Canker in the leaves of grape fruit. They used software known as Assess V1.0 to perform analysis on the image. Finally, they used Image Analysis Software to quantify plant disease. (Peressotti et al,2011) used a software package known as ImageJ to detect and estimate downy mildew sporulation in grapevine. They developed a special application (macro) for ImageJ, which was used to gradually change adjust colour balance and contrast found in the image before presenting the image to the user. Users can then use that to test for so many different threshold values to segment the image until the desired results is obtained. They concluded by reporting that, their method has similar results as other methods and that of the visual assessment.

3.4.6 Using Neural Network

Neural Network is an Artificial Intelligences system that seeks to imitate the human brain. Neural network was first developed in the early 1950s by Bernard Widrow of Stanford University. Neural network s is used in various fields in computing most especially in voice recognition systems, image recognition systems, robotics, medicine etc.

(Pydipati et al, 2005) detected and grouped some diseases in citrus plants using two different approaches. They gathered features like shape, colour, texture and size and created four different smaller groups of those features and classified them by two approaches. They used Mahalanobis minimum distance classifier in their first approach together with a principle known as Nearest Neighboring Principle to detect the diseases. The second step was the classification stage where they used Radial basis Functions (RBF) Neural Network Classifiers together with the back propagation algorithm to group the plants based on similar characteristics. They concluded by stating that both classification approaches produced equal results.

(Sanyal et al, 2007) detected and classified six different mineral deficiencies in rice plant using neural network. They first used an algorithm to obtain some characteristic features like colour and texture from the plants. The features were individually submitted to their own specific Multilayer Perception (MLP) using Neural Network. The two networks have a common hidden layer, but the number of neurons in the hidden layer is different depending on the feature submitted to it; (40 for texture and 70 for colour). The results produced by the two networks were combined for the purpose of the final classification. They also used a similar method to detect leaf disease in 2008 but in this case, their aim was to identify two kinds of diseases including blast and brown spots that affect rice crops.

3.4.7 The use of Support vector machines (SVM)

Support vector machines are Learning models associated with learning algorithms that analyze data with the intention of classifying and analyzing objects. They are discriminative classifiers based on decision planes that define decision boundaries.

(Youwen et al. 2008) used an algorithm based a statistic pattern recognition to identify some diseases that affect cucumber leaves. The algorithm was used to segment the leaves into healthy and diseased regions. That is followed by extracting colour, shape and texture features from the image. The final classification is done by feeding the features into the SVM. They concluded by stating that, Support Vector Machine produce better results than Neural networks based on their experimentation. (Camargo and Smith 2009) also tried to detect and measure the extent of defect in cotton plants. Images were captured from the stem, leaves and the fruits of the cotton plant for the experiment. They segmented the images using a technique they had already developed earlier which was described in this paper under Thresholding. They then extracted several features from the diseased regions. Those features are then used to feed the SVM for detection and estimation of the diseased portions of the images. (Jian and Wei,2010) presented a paper that uses the SVM to detect cucumber leaf diseases. They used the simple thresholding method to segment the healthy and diseased regions of the leaves. That was followed by extracting features like colour, shape and texture from the image for further processing. Those features are fed into the SVM with Radial Basis Function (RBF) as kernel, which performs the final classification.

3.4.8 The use of Fuzzy classifier

(Hairuddin et al, 2011) used the fuzzy classifier to detect some deficiencies in oil palm plant. They captured images and segmented them based on specific characteristics mainly similarities in colour. However, they did not make available the procedure for doing that in their paper. . After the segmentation process, they obtained features like colour, and texture and processed them further by feeding them to the fuzzy classifier which suggests the amount of fertilizers to apply on the plants to correct the deficiencies instead of revealing the type of deficiency detected in the plant. Unfortunately, the technical details available in this paper are shallow, making it difficult to get a clear understanding about their system and approach.

CHAPTER 4
PROPOSED METHODOLOGY

4.1 Proposed Flow:

In this system, for diagnosing leaf disease of Blast, brown spot and narrow brown spot. It involves several techniques such as image acquisition, image segmentation, preprocessing feature extraction and image classification.

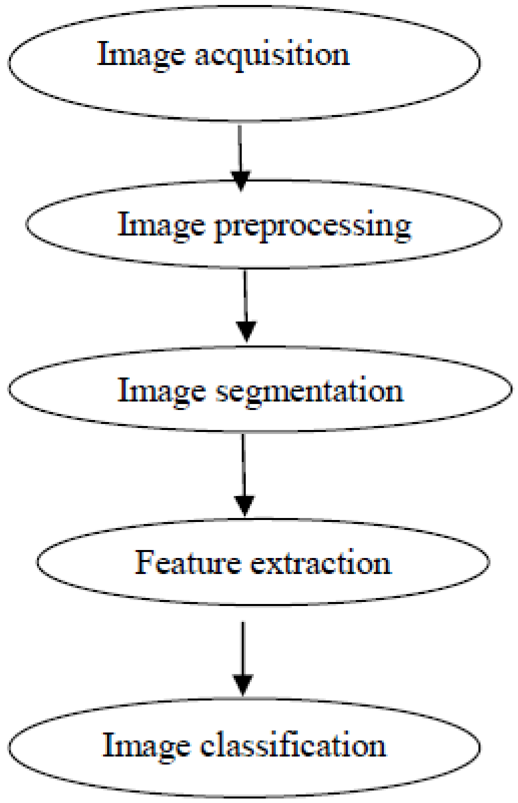


Image acquisition

The RGB Color images of leaf are captured using smart phones or digital camera, with the pixel size 768*1024 for the clear image. The digitized images are 225KB size of each image. Those images are cropped into smaller images with the dimensions of 109*310 pixels. Images are stored in BMP format by using mat lab image processing library.

Image preprocessing

Image preprocessing task involves the image enhancement. dimension of the leaf image compromising 109*310 pixels are used. In order to achieve high accuracy, the RGB images are converted into gray scale images. To increase the contrast using various contrast enhancement techniques like histogram equalization, contrast adjustment.Occurrence matrix is generated from the input images according to the probability distribution.

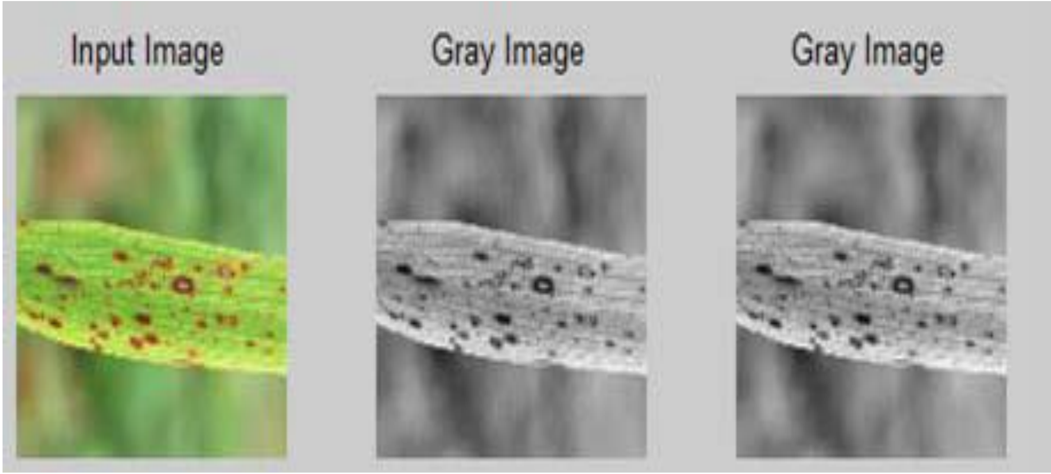
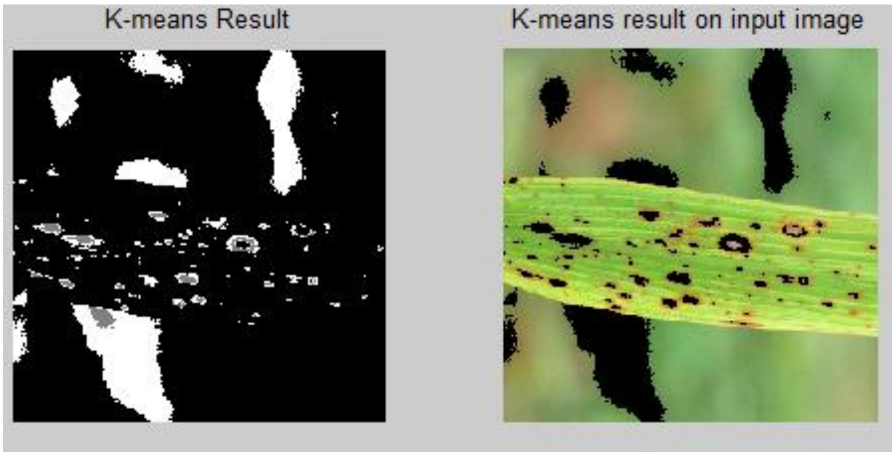
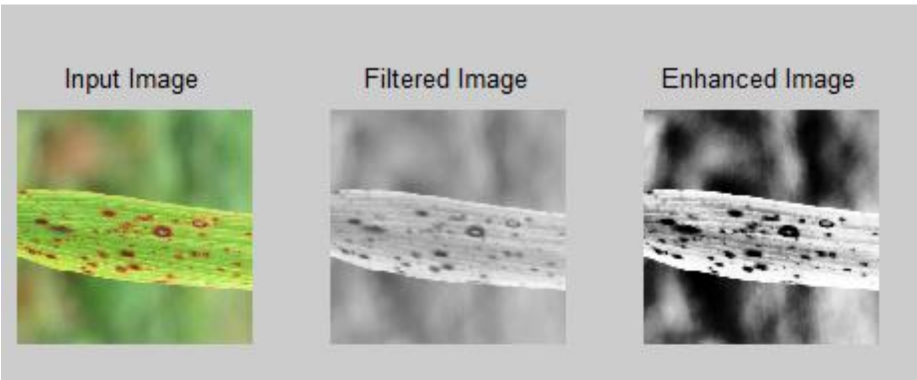


Image segmentation

In image segmentation, the noises of the image that affect the image quality are detected. K means algorithm is applied to remove the noise and unnecessary spots. The binary image with noise is converted into free of noise image. The noise free images are the filtered images. Then the filtered image is converted into enhanced image using K means algorithm which give high quality image for detecting the leaf disease.





Feature extraction

As the leaf disease consists of several types of disease blast, brown spot and narrow brown spot that had different lesion shape and lesion color.

Shape feature extraction

Shape is one of the important parameter of the image. Breadth and length of the image are significant characteristic to describe the shape. A simple approach is to measure the breadth and height of the image is to measure the count of the object pixel.

Color feature extraction

Color plays an important role in image processing. Digital image processing produce quantitative color measurement that are useful for the work of inquring the lesion for early diagnosis. The pixel in the color images are commonly represented in RGB format, where RGB are RED GREEN BLUE values respectively from the color images capturing device,.

Image classification

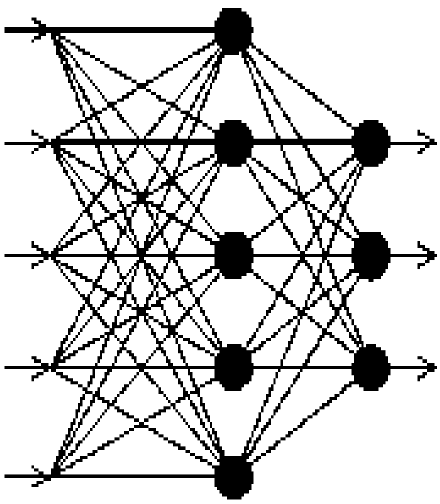
Different types of classification features like SVM, artificial neural network [ANN], fuzzy classification are proceed. Based on lesion type, boundary color, spot color and paddy leaf color, of the leaf paddy disease which is recognized using ANN or fussy logic method. SVM [support vector classification] classification is only classified into two stages as the leaf is defected or not defected. But using ANN and FUZZY classification ,It can identify the disease of the paddy plant.

4.2 ANN Classifier

Architecture of Artificial Neural Network:

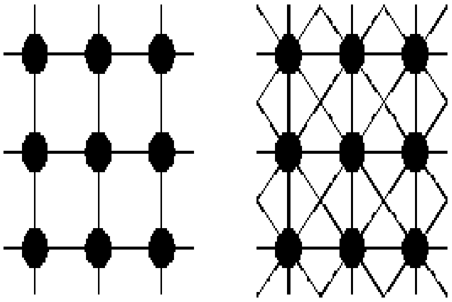
A. Feed-forward networks

Feed-forward ANNs allow signals to travel one way only; from input to output. There is no feedback (loops) i.e. the output of any layer does not affect that same layer. Feed-forward ANNs tend to be straight forward networks that associate inputs with outputs. They are extensively used in pattern recognition. This type of organization is also referred to as bottom-up or top-down.



B. Feedback networks

Feedback networks (figure 1) can have signals traveling in both directions by introducing loops in the network. Feedback networks are very powerful and can get extremely complicated. Feedback networks are dynamic; their 'state' is changing continuously until they reach an equilibrium point. They remain at the equilibrium point until the input changes and a new equilibrium needs to be found. Feedback architectures are also referred to as interactive or recurrent, although the latter term is often used to denote feedback connections in single-layer organizations.



C. Network layers

The commonest type of artificial neural network consists of three groups, or layers, of units: a layer of "**input**" units is connected to a layer of "**hidden**" units, which is connected to a layer of **"output**" units.

The activity of the input units represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units.

The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units. This simple type of network is interesting because the hidden units are free to construct their own representations of the input. The weights between the input and hidden units determine when each hidden unit is active, and so by modifying these weights, a hidden unit can choose what it represents.

We also distinguish single-layer and multi-layer architectures. The single-layer organization, in which all units are connected to one another, constitutes the most general case and is of more potential computational power than hierarchically structured multi-layer organizations. In multi-layer networks, units are often numbered by layer, instead of following a global numbering.

4.3 SVM Classifier

Support Vector Machine (SVM) is an algorithm that was developed for pattern classification but has recently been adapted for other uses, such as finding regression and distribution estimation. It has been used in many fields such as bioinformatics, and is currently a very active research area in many universities and research institutes which include the National University of Singapore (NUS) and Massachusetts Institute of Technology (MIT).

Although the SVM can be applied to various optimization problems such as regression, the classic problem is that of data classification. The basic idea is shown in figure 4.1. The data points are identified as being positive or negative, and the problem is to find a hyper-plane that separates the data points by a maximal margin.

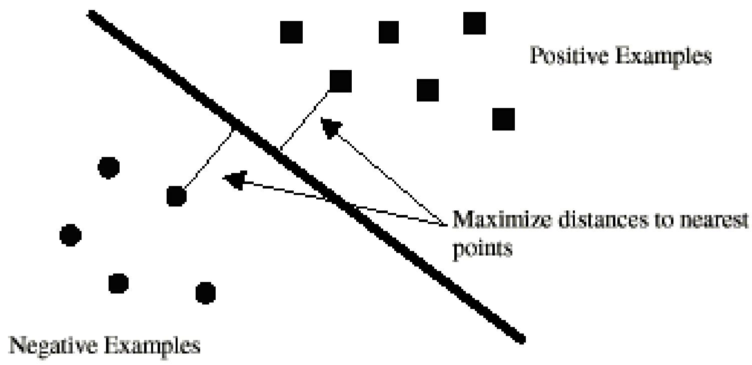


Figure 4.1: Data Classification

The above figure only shows the 2-dimensional case where the data points are linearly separable. The mathematics of the problem to be solved is the following:

$$\min_{\vec{w},b} \frac{1}{2}||w||,$$

$$s.t \quad y_i = +1 \Rightarrow \vec{w}\cdot\vec{x}_i + b \geq +1$$
$$y_i = -1 \Rightarrow \vec{w}\cdot\vec{x}_i - b \leq -1$$
$$s.t \; y_i(\vec{w}\cdot\vec{x}_i + b) \geq 1, \quad \forall i$$

The identification of the each data point x_i is y_i, which can take a value of +1 or -1 (representing positive or negative respectively). The solution hyper-plane is the following:

$$u = \vec{w}\cdot\vec{x} + b$$

The scalar b is also termed the bias.

A standard method to solve this problem is to apply the theory of Lagrange to convert it to a dual Lagrangian problem. The dual problem is the following:

$$\min_{\alpha} \Psi(\vec{\alpha}) = \min_{\alpha} \tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j (\vec{x}_i \cdot \vec{x}_j) \alpha_i \alpha_j - \sum_{i=1}^{N} \alpha_i \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \quad \forall i$$

The variables α_i are the Lagrangian multipliers for corresponding data point x_i.

A Brief Survey of SVM Algorithms

For small problems, traditional algorithms from optimization theory exist. Examples include conjugate gradient decent, and interior points methods. For larger problems, these algorithms do not work well because of the large space requirements to store the kernel matrix. Often they are slow because they do not make use of the characteristics of real-world SVM problems, for example that the number of support vectors are usually sparse.

Existing algorithms can be classified into three classes:

1. Algorithms where the kernel components are evaluated and discarded during learning

These methods reportedly slow and require multiple scans of the dataset.

2. Decomposition and chunking methods

The main idea of these algorithms is that a small dataset is selected heuristically for local training. If the result of the local training does not give a global optimum, the dataset is reselected or modified and is trained again. The process iterates until a global optimum is achieved. The SMO belongs to this class.

3. Other methods

There are currently many new methods under research and development.

Some of the SVM algorithms can be parallelized, but very few parallel SVM algorithms exist. In [5], a possible parallelization of a conjugate method using Matlab*P is described.

Definition:

More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training-data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function K(x,y) selected to suit the problem. The hyper planes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyper planes can be chosen to be linear combinations with parameters α^i of images of feature vectors x_i that occur in the data base. With this choice of a hyper plane, the points x in the feature space that are mapped into the hyper plane are defined by the relation:

$\sum_i \alpha_i k(x_i, x) = constant$

Note that if becomes small as grows further away from, each term in the sum measures the degree of closeness of the test point to the corresponding data base point. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points mapped into any hay a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

CHAPTER 5
SOFTWARE DESCRIPTION

5.1 Introduction

MATLAB is a high-performance language for technical computing. It integrates computation, visualization, and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical notation. MATLAB stands for matrix laboratory, and was written originally to provide easy access to matrix software developed by LINPACK (linear system package) and EISPACK (Eigen system package) projects. MATLAB is therefore built on a foundation of sophisticated matrix software in which the basic element is array that does not require pre dimensioning which to solve many technical computing problems, especially those with matrix and vector formulations, in a fraction of time.

MATLAB features a family of applications specific solutions called toolboxes. Very important to most users of MATLAB, toolboxes allow learning and applying specialized technology. These are comprehensive collections of MATLAB functions (M-files) that extend the MATLAB environment to solve particular classes of problems. Areas in which toolboxes are available include signal processing, control system, neural networks, fuzzy logic, wavelets, simulation and many others.

Typical uses of MATLAB include: Math and computation, Algorithm development, Data acquisition, Modeling, simulation, prototyping, Data analysis, exploration, visualization, Scientific and engineering graphics, Application development, including graphical user interface building.

5.2 Basic Building Blocks of MATLAB

The basic building block of MATLAB is MATRIX. The fundamental data type is the array. Vectors, scalars, real matrices and complex matrix are handled as specific class of this basic data type. The built in functions are optimized for vector operations. No dimension statements are required for vectors or arrays.

5.2.1 MATLAB Window

The MATLAB works based on five windows: Command window, Workspace window, Current directory window, Command history window, Editor Window, Graphics window and Online-help window.

a. Command Window

The command window is where the user types MATLAB commands and expressions at the prompt (>>) and where the output of those commands is displayed. It is opened when the application program is launched. All commands including user-written programs are typed in this window at MATLAB prompt for execution.

b. Work Space Window

MATLAB defines the workspace as the set of variables that the user creates in a work session. The workspace browser shows these variables and some information about them. Double clicking on a variable in the workspace browser launches the Array Editor, which can be used to obtain information.

c. Current Directory Window

The current Directory tab shows the contents of the current directory, whose path is shown in the current directory window. For example, in the windows operating system the path might be as follows: C:\MATLAB\Work, indicating that directory “work” is a subdirectory of the main directory “MATLAB”; which is installed in drive C. Clicking on the arrow in the current directory window shows a list of recently used paths. MATLAB uses a search path to find M-files and other MATLAB related files. Any file run in MATLAB must reside in the current directory or in a directory that is on search path.

d. Command History Window

The Command History Window contains a record of the commands a user has entered in the command window, including both current and previous MATLAB sessions. Previously entered MATLAB commands can be selected and re-executed from the command history window by right clicking on a command or sequence of commands. This is useful to select various options in addition to executing the commands and is useful feature when experimenting with various commands in a work session.

e. Editor Window

The MATLAB editor is both a text editor specialized for creating M-files and a graphical MATLAB debugger. The editor can appear in a window by itself, or it can be a sub window in the desktop. In this window one can write, edit, create and save programs in files called M-files.

MATLAB editor window has numerous pull-down menus for tasks such as saving, viewing, and debugging files. Because it performs some simple checks and also uses color to differentiate between various elements of code, this text editor is recommended as the tool of choice for writing and editing M-functions.

f. Graphics or Figure Window

The output of all graphic commands typed in the command window is seen in this window.

g. Online Help Window

MATLAB provides online help for all it's built in functions and programming language constructs. The principal way to get help online is to use the MATLAB help browser, opened as a separate window either by clicking on the question mark symbol (?) on the desktop toolbar, or by typing help browser at the prompt in the command window. The help Browser is a web browser integrated into the MATLAB desktop that displays a Hypertext Markup Language (HTML) documents. The Help Browser consists of two panes, the help navigator pane, used to find information, and the display pane, used to view the information. Self-explanatory tabs other than navigator pane are used to perform a search.

5.3 MATLAB Files

MATLAB has three types of files for storing information. They are: M-files and MAT-files.

5.3.1 M-Files

These are standard ASCII text file with 'm' extension to the file name and creating own matrices using M-files, which are text files containing MATLAB code. MATLAB editor or another text editor is used to create a file containing the same statements which are typed at the MATLAB command line and save the file under a name that ends in .m. There are two types of M-files:

1. Script Files

It is an M-file with a set of MATLAB commands in it and is executed by typing name of file on the command line. These files work on global variables currently present in that environment.

2. Function Files

A function file is also an M-file except that the variables in a function file are all local. This type of files begins with a function definition line.

5.3.2 MAT-Files

These are binary data files with .mat extension to the file that are created by MATLAB when the data is saved. The data written in a special format that only MATLAB can read. These are located into MATLAB with ‘load’ command.

5.4 MATLAB System:

The MATLAB system consists of five main parts:

5.4.1 Development Environment:

This is the set of tools and facilities that help you use MATLAB functions and files. Many of these tools are graphical user interfaces. It includes the MATLAB desktop and Command Window, a command history, an editor and debugger, and browsers for viewing help, the workspace, files, and the search path.

5.4.2 MATLAB Mathematical Function:

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigen values, Bessel functions, and fast Fourier transforms.

5.4.3 MATLAB Language:

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

5.4.4 Graphics:

MATLAB has extensive facilities for displaying vectors and matrices as graphs, as well as annotating and printing these graphs. It includes high-level functions for two-dimensional and three-dimensional data visualization, image processing, animation, and presentation graphics. It also includes low-level functions that allow you to fully customize the appearance of graphics as well as to build complete graphical user interfaces on your MATLAB applications.

5.4.4 MATLAB Application Program Interface (API):

This is a library that allows you to write C and FORTRAN programs that interact with MATLAB. It includes facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.

5.4 Some Basic Commands:

pwd prints working directory

Demo demonstrates what is possible in Mat lab

Who lists all of the variables in your Mat lab workspace?

Whose list the variables and describes their matrix size

clear erases variables and functions from memory

clear x erases the matrix 'x' from your workspace

close by itself, closes the current figure window

figure creates an empty figure window

hold on holds the current plot and all axis properties so that subsequent graphing

commands add to the existing graph

hold off sets the next plot property of the current axes to "replace"

find find indices of nonzero elements e.g.:

d = find(x>100) returns the indices of the vector x that are greater than 100

break terminate execution of m-file or WHILE or FOR loop

for repeat statements a specific number of times, the general form of a FOR

statement is:

```
FOR variable = expr, statement, ..., statement END

for n=1:cc/c;

magn(n,1)=NaNmean(a((n-1)*c+1:n*c,1));

end
```

diff difference and approximate derivative e.g.:

DIFF(X) for a vector X, is [X(2)-X(1) X(3)-X(2) ... X(n)-X(n-1)].

NaN the arithmetic representation for Not-a-Number, a NaN is obtained as a result of mathematically undefined operations like 0.0/0.0

INF the arithmetic representation for positive infinity, a infinity is also produced by operations like dividing by zero, e.g. 1.0/0.0, or from overflow, e.g. exp(1000).

save saves all the matrices defined in the current session into the file, matlab.mat, located in the current working directory

load loads contents of matlab.mat into current workspace

save filename x y z saves the matrices x, y and z into the file titled filename.mat

save filename x y z /ascii save the matrices x, y and z into the file titled filename.dat

load filename loads the contents of filename into current workspace; the file can be a binary (.mat) file

load filename.dat loads the contents of filename.dat into the variable filename

xlabel(‘ ’) : Allows you to label x-axis

ylabel(‘ ‘) : Allows you to label y-axis

title(‘   ‘)   : Allows you to give title for plot

subplot() : Allows you to create multiple plots in the same window

5.6 Some Basic Plot Commands:

Kinds of plots:

plot(x,y) creates a Cartesian plot of the vectors x & y

plot(y) creates a plot of y vs. the numerical values of the elements in the y-vector

semilogx(x,y) plots log(x) vs y

semilogy(x,y) plots x vs log(y)

loglog(x,y) plots log(x) vs log(y)

polar(theta,r) creates a polar plot of the vectors r & theta where theta is in radians

bar(x) creates a bar graph of the vector x. (Note also the command stairs(x))

bar(x, y) creates a bar-graph of the elements of the vector y, locating the bars according to the vector elements of 'x'

Plot description:

grid creates a grid on the graphics plot

title('text') places a title at top of graphics plot

xlabel('text') writes 'text' beneath the x-axis of a plot

ylabel('text') writes 'text' beside the y-axis of a plot

text(x,y,'text') writes 'text' at the location (x,y)

text(x,y,'text','sc') writes 'text' at point x,y assuming lower left corner is (0,0) and upper right corner is (1,1)

axis([xmin xmax ymin ymax]) sets scaling for the x- and y-axes on the current plot

5.7 ALGEBRIC OPERATIONS IN MATLAB:

Scalar Calculations:

+ Addition

- Subtraction

* Multiplication

/ Right division (a/b means a ÷ b)

\ left division (a\b means b ÷ a)

^ Exponentiation

For example 3*4 executed in 'matlab' gives ans=12

4/4 gives ans=0.8

Array products: Recall that addition and subtraction of matrices involved addition or subtraction of the individual elements of the matrices. Sometimes it is desired to simply multiply or divide each element of an matrix by the corresponding element of another matrix 'array operations”.

Array or element-by-element operations are executed when the operator is preceded by a '.' (Period):

a .* b multiplies each element of a by the respective element of b

a ./ b divides each element of a by the respective element of b

a .\ b divides each element of b by the respective element of a

a .^ b raise each element of a by the respective b element

5.8 Matlab Working Environment:

5.8.1 Matlab Desktop

Matlab Desktop is the main Matlab application window. The desktop contains five sub windows, the command window, the workspace browser, the current directory window, the command history window, and one or more figure windows, which are shown only when the user displays a graphic.

The command window is where the user types MATLAB commands and expressions at the prompt (>>) and where the output of those commands is displayed. MATLAB defines the workspace as the set of variables that the user creates in a work session.

The workspace browser shows these variables and some information about them. Double clicking on a variable in the workspace browser launches the Array Editor, which can be used to obtain information and income instances edit certain properties of the variable.

The current Directory tab above the workspace tab shows the contents of the current directory, whose path is shown in the current directory window. For example, in the windows operating system the path might be as follows: C:\MATLAB\Work, indicating that directory "work" is a subdirectory of the main directory "MATLAB"; WHICH IS INSTALLED IN DRIVE C. clicking on the arrow in the current directory window shows a list of recently used paths. Clicking on the button to the right of the window allows the user to change the current directory.

MATLAB uses a search path to find M-files and other MATLAB related files, which are organize in directories in the computer file system. Any file run in MATLAB must reside in the current directory or in a directory that is on search path. By default, the files supplied with MATLAB and math works toolboxes are included in the search path. The easiest way to see which directories are soon the search path, or to add or modify a search path, is to select set path from the File menu the desktop, and then use the set path dialog box. It is good practice to add any commonly used directories to the search path to avoid repeatedly having the change the current directory.

The Command History Window contains a record of the commands a user has entered in the command window, including both current and previous MATLAB sessions. Previously entered MATLAB commands can be selected and re-executed from the command history window by right clicking on a command or sequence of commands.

This action launches a menu from which to select various options in addition to executing the commands. This is useful to select various options in addition to executing the commands. This is a useful feature when experimenting with various commands in a work session.

5.8.2 Using the MATLAB Editor to create M-Files:

The MATLAB editor is both a text editor specialized for creating M-files and a graphical MATLAB debugger. The editor can appear in a window by itself, or it can be a sub window in the desktop. M-files are denoted by the extension .m, as in pixelup.m.

The MATLAB editor window has numerous pull-down menus for tasks such as saving, viewing, and debugging files. Because it performs some simple checks and also uses color to differentiate between various elements of code, this text editor is recommended as the tool of choice for writing and editing M-functions.

To open the editor , type edit at the prompt opens the M-file filename.m in an editor window, ready for editing. As noted earlier, the file must be in the current directory, or in a directory in the search path.

5.8.3 Getting Help:

The principal way to get help online is to use the MATLAB help browser, opened as a separate window either by clicking on the question mark symbol (?) on the desktop toolbar, or by typing help browser at the prompt in the command window. The help Browser is a web browser integrated into the MATLAB desktop that displays a Hypertext Markup Language(HTML) documents. The Help Browser consists of two panes, the help navigator pane, used to find information, and the display pane, used to view the information. Self-explanatory tabs other than navigator pane are used to perform a search.

CHAPTER 6
CODE AND RESULTS

6.1 Code

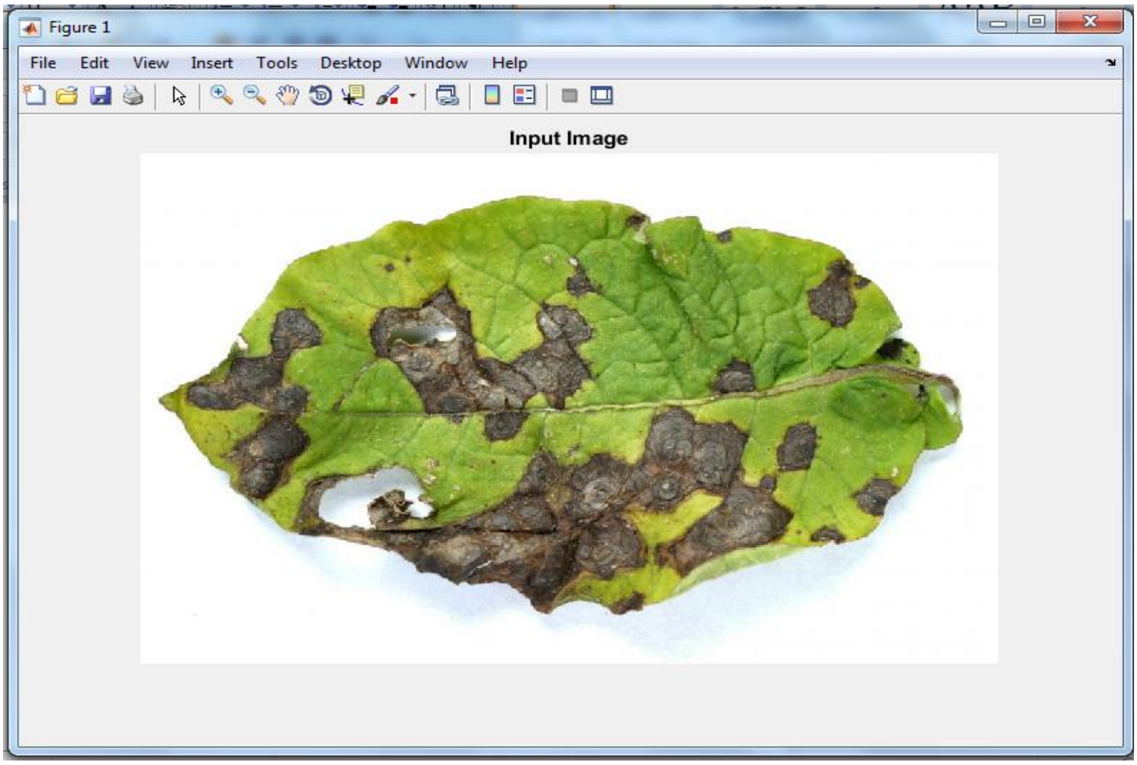
```
clc
clear
close all
warning off
 length_t = [];
width_t = [];
his_t = [];
area_t = [];
addpath('/leafs')
datab = imageSet('leafs')
for i = 1:datab.Count
[file path] = uigetfile('*.bmp','Select Image');
if path==0
    msgbox('Image not selected')
    return
end
data(i).name = file;
data(i).num = i;
I1 = imread(strcat(path,file));
figure,imshow(I1)
title('Input Image')
I1 = imresize(I1,[256 256]);
 I = rgb2gray(I1);
figure,imshow(I)
title('Grey Image')
I = imresize(I,[256 256]);
 I_filt = imadjust(I);
figure,imshow(I_filt)
title('Fitered Image')
```

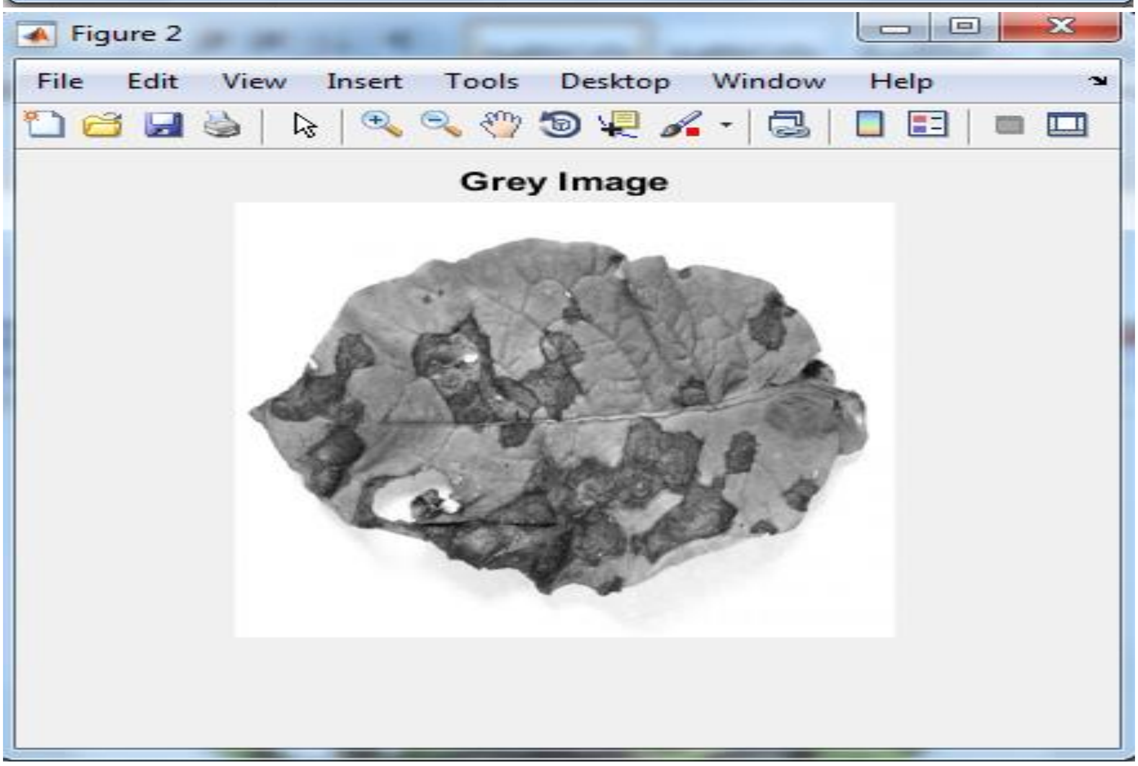
```
I_adhis = adapthisteq(I_filt);
figure,imshow(I_adhis)
title('Enhanced Image')
 [~,I_seg] = kmeans(I_filt,2);
figure,imshow(I_seg)
title('Segmented Image')
ff = imfuse(I1,I_seg);
figure,imshow(ff)
title('Fused Image')
 length = sum(I_seg(:,round(size(I_seg,2)/2)));
width = sum(I_seg(round(size(I_seg,1)/2),:));
length_t = [length_t;length];
width_t = [width_t;width];
area = sum(I_seg(:));
area_t = [area_t;area];
 figure,
imhist(I)
title('Histogram of an image')
figure,
imhist(I_adhis)
title('Histogram of an Enhanced image')
his = imhist(I);
his_t = [his_t;his'];
title('Color information')
end
 query = input('Enter which image to test:');
 %% ANN Classifier
Inputs_ann = [len_f;wid_f;color_f;area_f];
Targets_ann = [1:15];
x = Inputs_ann;
t = Targets_ann;
 % Choose a Training Function
% For a list of all training functions type: help nntrain
% 'trainlm' is usually fastest.
```

```
% 'trainbr' takes longer but may be better for challenging problems.
% 'trainscg' uses less memory. Suitable in low memory situations.
trainFcn = 'trainscg';  % Scaled conjugate gradient backpropagation.
 % Create a Pattern Recognition Network
hiddenLayerSize = 10;
net = patternnet(hiddenLayerSize, trainFcn);
 % Setup Division of Data for Training, Validation, Testing
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
 % Train the Network
[net,tr] = train(net,x,t);
 % Test the Network
y = net(x);
e = gsubtract(t,y);
error_performance = perform(net,t,y);
tind = vec2ind(t);
yind = vec2ind(y);
percentErrors = sum(tind ~= yind)/numel(tind);
fprintf('Detection using ANN : %s\n',data(query).name(1:end-4))
%% SVM classifier
 % SVM classification
TrainingSet=Inputs_ann;
TestSet=Inputs_ann(:,query);
GroupTrain=[1;2;3;4;5;6;7;8;9;10;11;12;13;14;15];
 mdl = fitcecoc(TrainingSet',GroupTrain);
label_svm = predict(mdl,TestSet');
fprintf('Detection using SVM : %s\n',data(label_svm).name(1:end-4))
```

6.2 Results

Consider an input with disease. The process of proposed model find out the correct disease image after classifying with ANN and SVM





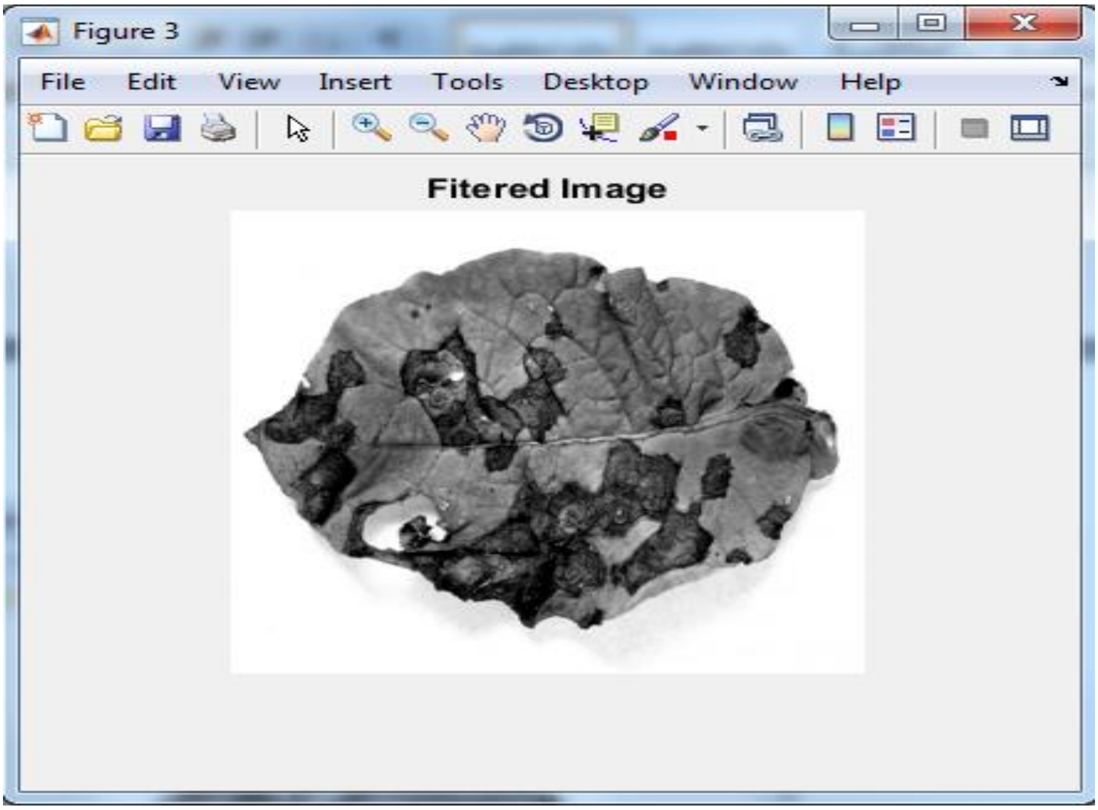


Figure 3
File Edit View Insert Tools Desktop Window Help
Fitered Image

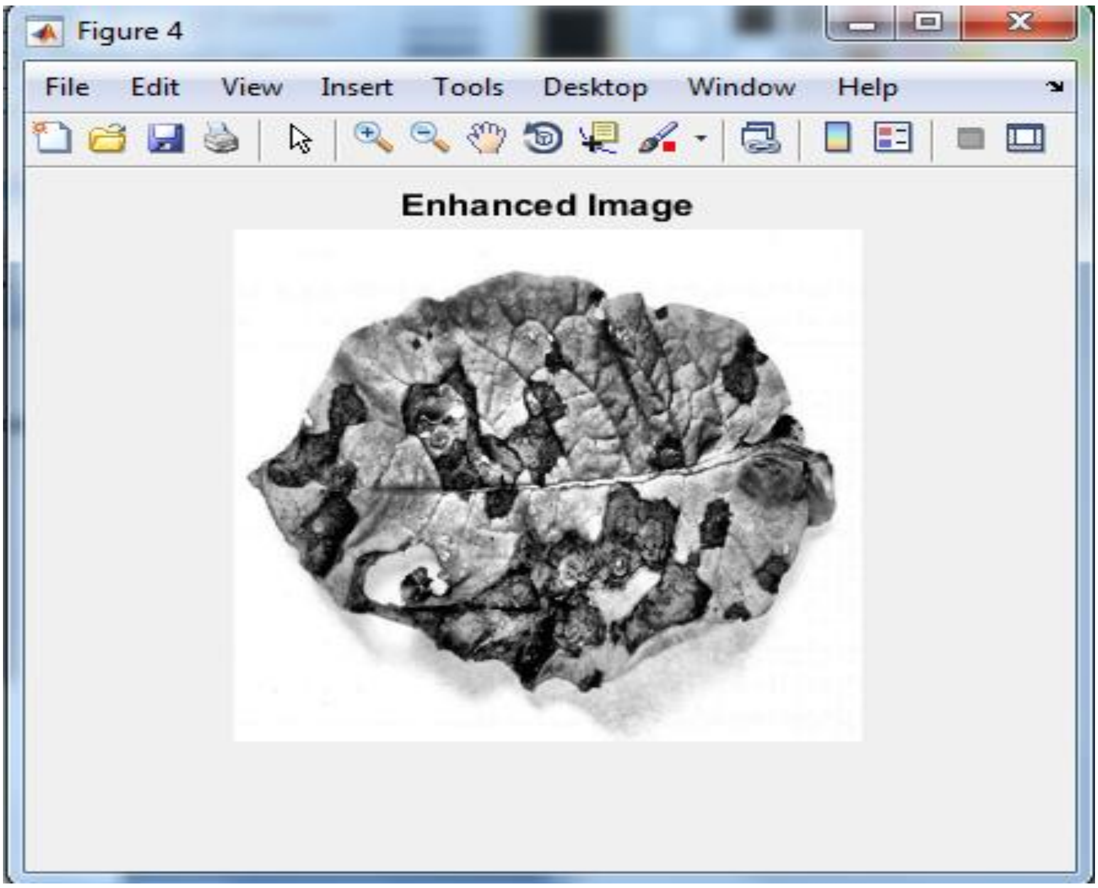


Figure 4
File Edit View Insert Tools Desktop Window Help
Enhanced Image

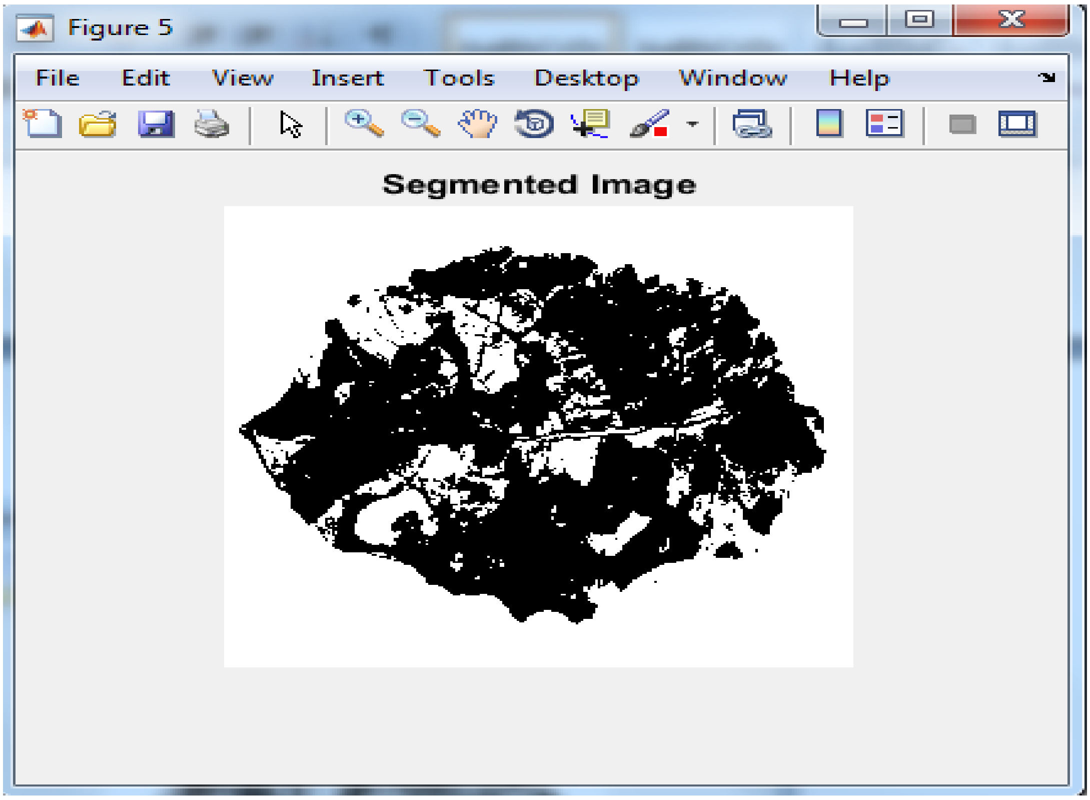


Figure 5
File Edit View Insert Tools Desktop Window Help
Segmented Image

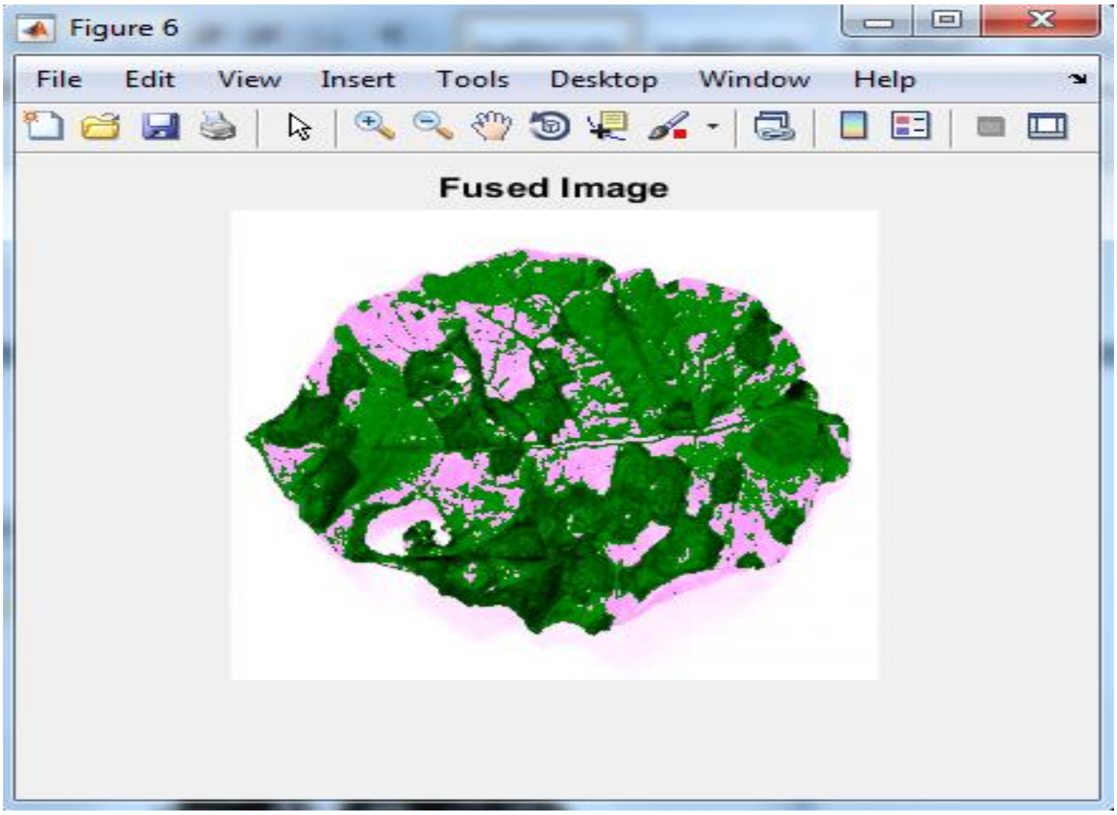


Figure 6
File Edit View Insert Tools Desktop Window Help
Fused Image

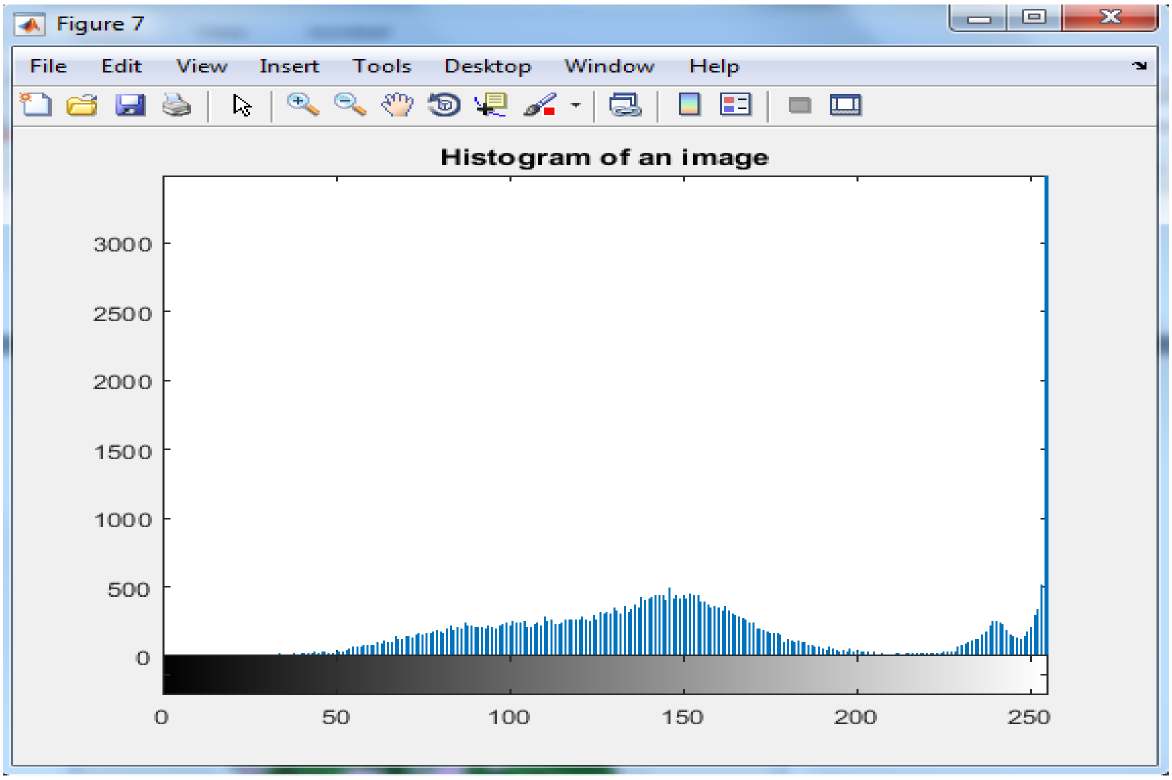


Figure 7
File Edit View Insert Tools Desktop Window Help
Histogram of an image
3000
2500
2000
1500
1000
500
0
0 50 100 150 200 250

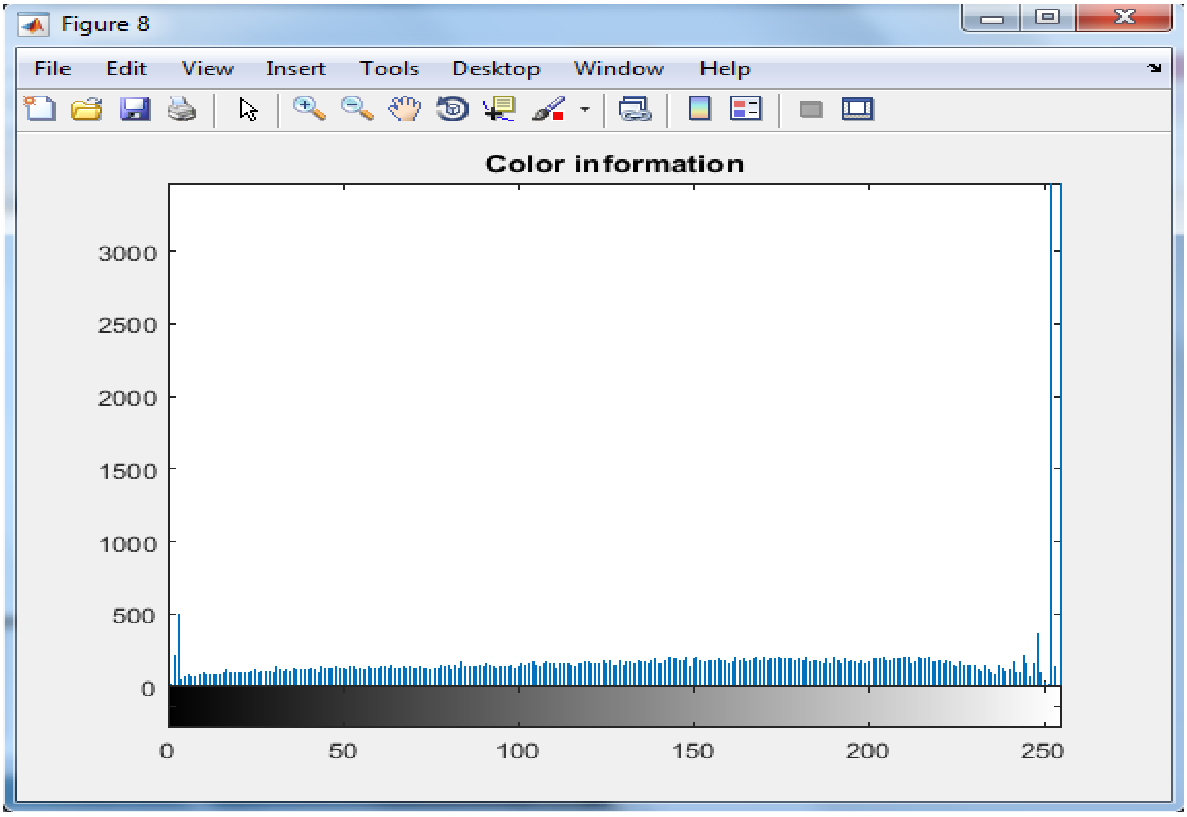
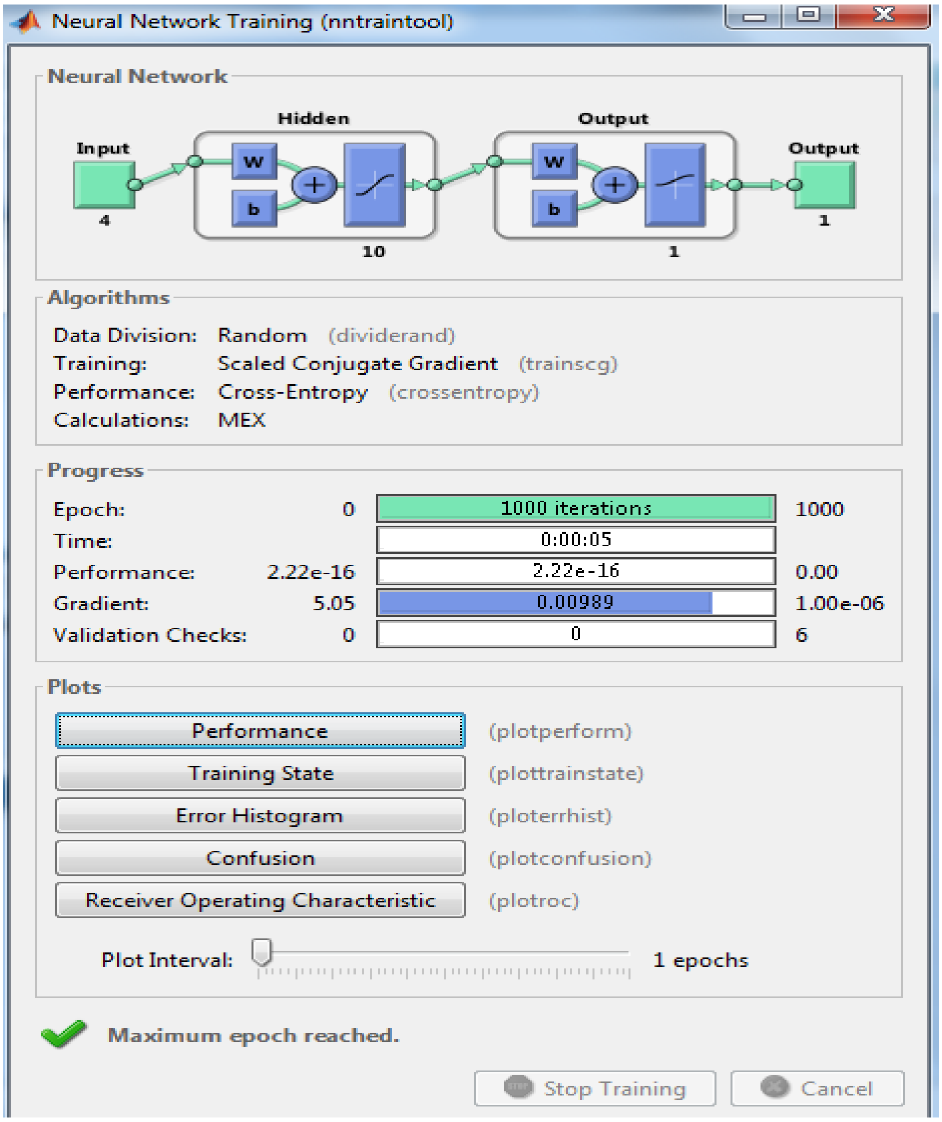
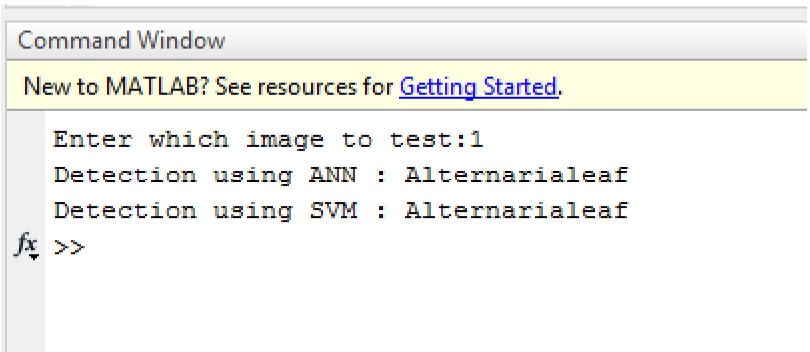


Figure 8
File Edit View Insert Tools Desktop Window Help
Color information
3000
2500
2000
1500
1000
500
0
0 50 100 150 200 250


Neural Network Training (nntraintool)
Neural Network
Hidden
Output
Input
W
b
4
10
1
Output
1
Algorithms
Data Division: Random (dividerand)
Training: Scaled Conjugate Gradient (trainscg)
Performance: Cross-Entropy (crossentropy)
Calculations: MEX
Progress
Epoch: 0 1000 iterations 1000
Time: 0:00:05
Performance: 2.22e-16 2.22e-16 0.00
Gradient: 5.05 0.00989 1.00e-06
Validation Checks: 0 0 6
Plots
Performance (plotperform)
Training State (plottrainstate)
Error Histogram (ploterrhist)
Confusion (plotconfusion)
Receiver Operating Characteristic (plotroc)
Plot Interval: 1 epochs
Maximum epoch reached.
Stop Training
Cancel


Command Window
New to MATLAB? See resources for Getting Started.
Enter which image to test:1
Detection using ANN : Alternarialeaf
Detection using SVM : Alternarialeaf
>>

CHAPTER 7

CONCLUSION

In this project we have proposed a new histogram based concept of detecting damaged leaf. From histogram we extract the difference between the intensity among the original paddy leaf and the diseases affected leaf. A system for identifying the diseases like Blast, Brown spot and Narrow brown spot are detected. It is mainly based on the mat lab application using k-means algorithm This project evaluates the techniques in digital image processing for detecting, diagnosing, recognizing of crop leaf diseases k-means clustering algorithm is used for automatically the disease for more accuracy. As more no. of image samples are produced accordingly, three is more scope of identifying the various errors during the simulation. The primary result of the proposed methodology indicates a strong and systematic way of assessing disease intensity by plant pathology more precisely. The result of the preliminary test shows the better result of disease extraction. The classification techniques like SVM, ANN are performed and SVM is proved to be the best.

REFERENCES

[1] Nonik Noviana kurniawati & Salwani Abdullah, "Texture Analysis for Dignosing Paddy Disease(2009)."

[2] Arnal barbedo & Jayme Garcia, Digital Image Processing Techniqe for Detecting, Quantifying and Classifying Plant Diseases(2008).

[3] Shenweizheng & wcyachun, "Grading method of Leaf Spot Disease based on Image Proccessing(2008)."

[4] Manoj Mukherjee & Titan Pal, "Damaged Paddy Leaf Detection using Image Processing(2012)".

[5] P.Tamije seivy & kowsalya.N, "Paddy Crop Disease Finder using Probabilistic Approach.(2016)".

[6] Alhouse, Mark L.G. & Chein-I Chang. 2010. Image Segmentation by Local Entropy Methods. Proc. International Conference on Image Processing 3:61-64

[7] R. Pydipati, T.F. Burks, W.S. Lee, "Identification of citrus disease using color texture features and discriminate analysis", Computers and Electronics in Agriculture 52

[8] J.B. Cunha. 2003. Application of Image Processing Techniques in the Characterization of Plant Leafs. Proc. IEEE Intl' Symposium on Industrial Electronics.

[9] K.J. Runtz. 1991. Electronic Recognition of Plant Spesies for Machine Vision Sprayer Control Systems. Proc. IEEE Western Canada Conference on Computer, Power, and Communication System in a Rural Environment (WESCANEX'91)

.[10] S.Dave and K. Runtz. 1995. Image Processing Methods for Identifying Species of Plants. Proc. IEEE Western Canada Conference on Computer, Power, and Communication System in a Rural Environment (WESCANEX'95),

[11]. Rick Cartwright and Fleet Lee, 10 – Management of Rice Diseases, pp.-87

[12].http://agritech.tnau.ac.in/crop_protection/crop_prot_crop%20diseases_cereals_paddy.html#3

[13]. Kobayashi, T., Kanda, E., Kitada, K., Ishiguro, K., Torigoe, Y., 2001. Detection of rice panicle Blast with multispectral radiometer and the Potential of using airborne multispectral Scanners. Phytopathology, 91(3):pp. 316-323.

[14]. A.Senthil Rajan , Image Processing Techniques for Diagnosing Paddy Disease Proceedings of the World congress on Engineering 2012 Vol II WCE 2012, July 4-6, 2012, London, U.K.

[15].http://www.goepel.com/en/industrial-vision-solutions/tom-line/tom-line-software/image-pre-processing.html

[16]. R. Pydipati, T.F. Burks, W.S. Lee, “Identification of citrus disease using color texture features and discriminate analysis”, Computers and Electronics in Agriculture 52

[17]. http://www.forestryimages.org/browse/hostimages.cfm?host=6108&cat=16

[18]. Sanjeev S Sannakki1, Vijay S Rajpurohit, V B Nargund, Arun Kumar R, Prema S Yallur , “Leaf Disease Grading by Machine Vision and Fuzzy Logic”, Int. J. Comp. Tech. Appl., Vol 2 (5), 1709-1716, 2011.

[19]. Yan Li, Chunlei Xia and Jangmyung Lee, “Vision-based pest detection and automatic spray of greenhouse plant,” IEEE International Symposium on Industrial Electronics, July 5-8 2009.

[20]. Paul Boissard, Vincent Martin and Sabine Moisan, “Acognitive vision approach to early pest detection in greenhouse crops,” Computers and Electronics in Agriculture, pp. 81-93, 2008.

[21]. Alhouse, Mark L.G. & Chein-I Chang. 2010. Image Segmentation by Local Entropy Methods. Proc. International Conference on Image Processing 3:61-64.

[22]. S.C. Scardaci et al., Rice Blast: A New Disease In California, Agronomy Fact Sheet Series, Department of Agronomy and Range Science University of California, Davis.

[23]. Robert W. Herdt, Research Priorities for Rice Biotechnology, G.S. Khush and G.H.

[24]. B.Cunha. 2010. Application of Image Processing Techniques in the Characterization of plant Leaf s. Proc. IEEE Intl’ Symposium on Industrial Electronics.