

**LEAF DISEASE DETECTION USING IMAGE PROCESSING BY
COMPARING TWO DIFFERENT ALGORITHMS**

BY

**JOHANA KABIR TISHA
ID: 151-15-5339**

**REDWAN AHMED
ID: 151-15-5504**

AND

**SOURAV NARAYAN KOER
ID: 151-15-5338**

This Report Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Rezwana Sultana Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Israt Ferdous Lecturer
Department of CSE
Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY DHAKA,
BANGLADESH**

MAY 2019

APPROVAL

This Project/internship titled "**LEAF DISEASE DETECTION USING IMAGE PROCESSING BY COMPARING TWO DIFFERENT ALGORITHMS**", submitted by Redwan Ahmed, ID No: 151-15-5504, Johana Kabir Tisha, ID No: 151-15-5339 and Sourav Narayan Kocer, ID No: 151-15-5338 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 03-05-2019.

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Dr. Md. Ismail Jabiullah
Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

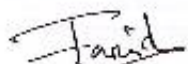
Internal Examiner



Dr. Sheak Rashed Haider Noori
Associate Professor & Associate Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Dewan Md. Farid
Associate Professor


Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

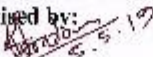
We hereby declare that this project has been done by us under the supervision of **Rezwana Sultana, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:




Rezwana Sultana
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:




Israt Ferdous
Lecturer
Department of CSE
Daffodil International University


Submitted by:



Johana Kabir Tisha
ID: 151-15-5339
Department of CSE
Daffodil International University



Redwan Ahmed
ID: 151-15-5504
Department of CSE
Daffodil International University



Sourav Narayan Koer
ID: 151-15-5338
Department of CSE
Daffodil International University

ACKNOWLEDGMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound our indebtedness to **Rezwana Sultana, Lecturer**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “*Machine Learning*” to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project. We would like to express our heartiest gratitude to Rezwana Sultana, Israt Ferdous, Syed Akhter Hossain Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University. We would like to thank our entire course mate at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Leaf diseases cause many significant damages and losses to the farmers around the world. Appropriate measures on disease identification should be introduced to prevent the problems and minimize the losses. Technical approaches using machine learning and computer vision are actively researched to achieve intelligence farming by early detection of leaf disease. An analyzer is obviously desirable to aid the farmers in diagnosing what sorts of diseases a leaf has. This dissertation presents the research,

design, and implementation of an analyzer which can automatically identify the leaf diseases based on its appearance with some computer vision and machine learning technique. Many experiments and evaluations on different segmentations, feature extractions, and classification methods were done to find the most effective approach. The target group of the user is those who request a free and quick diagnosis of common leaf disease at any time of the day.

TABLE OF CONTENTS

| CONTENT | PAGE NO |
|--------------------------------|----------------|
| Approval | i |
| Declaration | ii |
| Acknowledgments | iii |
| Abstract | iv |
| List of Figures | viii |
| List of Table | ix |
| CHAPTER | |
| CHAPTER 1: INTRODUCTION | 1-3 |

| | |
|-------------------------|-----|
| 1.1 Introduction | 1 |
| 1.2 Background | 1-2 |
| 1.3 Problem Statement | 2 |
| 1.4 Research Questions | 2 |
| 1.5 Expected Output | 3 |
| 1.6 Scope of Research | 3 |
| 1.7 Thesis Organization | 3 |

CHAPTER 2: BACKGROUND 4-11

| | |
|--|-------|
| 2.1 Introduction | 4 |
| 2.2 Leaf Diseases, Symptoms and Management | 4-6 |
| 2.3 Case Study On Existing System | 6 |
| 2.3.1 Investigating on Image Processing Techniques for Diagnosing Leaf Diseases | 6-7 |
| 2.3.2 Plant Disease Detection Using Image Processing | 8 |
| 2.4 Literature Review | 8 |
| 2.5 Theoretical Study | 8 |
| 2.5.1 The EM Algorithm for Gaussian Mixture Models | 9 |
| 2.5.4 Support Vector Machine | 9 |
| 2.6 Research Summary | 10-11 |
| 2.7 Challenges | 11 |

CHAPTER 3: IMPLEMENTATION OF PROPOSED METHODOLOGY 12-24 THE

| | |
|-------------------------|-------|
| 3.1 Introduction | 12 |
| 3.2 Work Flow | 12 |
| 3.3 Segmentation | 12-13 |
| 3.3.1 Leaf Segmentation | 13-16 |
| 3.4 Feature Extraction | 17 |
| 3.4.1 Color histogram | 17-18 |

| | |
|--|-------|
| 3.5 Learning and Classification | 18-19 |
| 3.5.1 Support Vector Machine | 19 |
| 3.6 Disease Detection | 20-21 |
| 3.7 Graphical User Interface Development | 22-24 |

CHAPTER 4: EXPERIMENTAL ANALYSIS **25-28**

| | |
|---------------------------------------|-------|
| 4.1 Introduction | 25 |
| 4.2 Experimental Setup | 25 |
| 4.3 Performance Measures – Definition | 25-26 |
| 4.4 Result Analysis | 26 |
| 4.5 Classification Result | 27 |
| 4.5.1 Disease Free | 27 |
| 4.5.2 Powdery Mildew | 27 |
| 4.5.3 Septoria | 27 |
| 4.5.4 Rust | 27 |
| 4.6 Overall Classification Result | 28 |
| 4.7 Conclusion | 28 |

CHAPTER 5: CONCLUSION **29-30**

| | |
|-----------------------|-------|
| 5.1 Summary | 29 |
| 5.2 System Limitation | 29-30 |
| 5.3 Future Work | 30 |

REFERENCES **31-32**

APPENDIX **33**

| | |
|---------------------------------|----|
| Appendix A: Research Reflection | 33 |
| Appendix B: Related Issues | 33 |

PLAGIARISM **34**

LIST OF FIGURES

| FIGURES | PAGE NO |
|--|----------------|
| Figure 1.1.1: Some variation of leaf | 1 |
| Figure 2.3.1.1: The methodology for diagnosing leaf diseases | 7 |
| Figure 2.3.2.1: Basic steps for plant disease detection and classification | 8 |
| Figure 2.5.4.1: Support Vector Machine | 10 |
| Figure 2.6.1: The three kinds of leaf disease included in this study | 11 |
| Figure 3.2.1: User Interaction Flow Diagram | 12 |
| Figure 3.3.1: The three kinds of leaf disease included in this study | 12-13 |
| Figure 3.3.1.1: Test picture | 14 |
| Figure 3.3.1.2: Visualization of the changes of clustering for each pixel during the EM process with the increasing EM iterations. | 14-15 |
| Figure 3.3.1.3: Leaf Segmentation Example | 16 |
| Figure 3.3.1.4: Processing segmented image with a function to get rid of some irrelevant Colour | 16 |
| Figure 3.4.1.1: This figure shows the color histogram of some examples of the three kinds of disease. | 17-18 |
| Figure 3.5.1: Some sample recognition results over the leaves in these four categories | 20-21 |
| Figure 3.6.2: Recognition results for two non-leaf images | 21 |
| Figure 3.7.1: The Graphical User Interface | 22 |
| Figure 3.7.2: Menu Bar | 23 |
| Figure 3.7.3: The Graphical User Interface | 23 |
| Figure 3.7.4: The Graphical User Interface (after detecting disease) | 24 |
| Figure 3.7.5: The Graphical User Interface | 24 |

| | |
|--|----|
| Figure 3.7.6: The Graphical User Interface | 25 |
| Figure 3.7.7: The Graphical User Interface | 25 |

LIST OF TABLES

| TABLES | PAGE NO |
|---|----------------|
| Table 2.1: The Following Leaf Disease are Shown Below | 4-6 |
| Table 3.1: Final Parameter Estimates of the Means and the Proportions for the Test Picture. | 15 |
| Table 4.4.1: Confusion matrix of proposed system | 26 |
| Table 5.11: Overall Learning Evaluation | 28 |

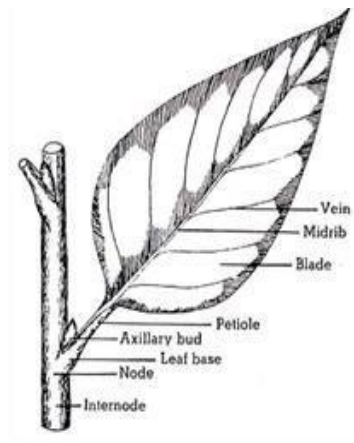
CHAPTER 1 Introduction

1.1 Introduction

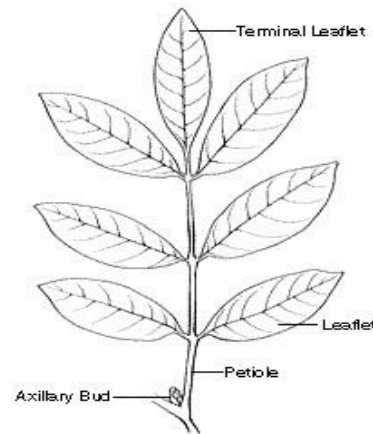
This chapter discusses an overview of the study conducted. The title is “Leaf Disease Detection using Image Processing”. It consists of background, problem statements, research questions, expected output and the scope of the study. The background briefly describes the identification of thesis and related issues. Problem statements describe the problems that arise and make the selected projects to be undertaken. The research questions and expected output are the goals list for the research to be achieved. The scope of the study discusses the limitations of the research. Lastly, thesis organization gives a summary of the sequence for each chapter in the thesis.

1.2 Background

A leaf is an organ of a vascular plant and is the essential level furthest point of the stem. The leaves and stem together shape the shoot. Leaves are everything considered implied as foliage, as in "fall foliage". Framework of a clear leaf.



Simple Leaf



Compound Leaf Fig

1.1.1: Some variation of leaf

1. Peak
2. Fundamental Vein
3. Discretionary Vein
4. Lamina
5. Leaf Margin

6. Petiole

A leaf is a thin, dorsiventrally smoothed organ, when in doubt borne over the ground and focused for photosynthesis. In numerous leaves, including the create foliage of Eucalyptus, palisade mesophyll is accessible on the opposite sides and the leaves are said to be isobilateral. Most leaves have unquestionable upper surface (adaxial) and cut down surface (abaxial) that fluctuate in shading, fuzziness, the amount of stomata the aggregate and structure of epicuticular wax and distinctive features.

Some types of leaf disease:

- (a) Anthracnose
- (b) Rust
- (c) Septoria
- (d) Powdery Mildew
- (e) Downy mildew

1.3 Problem Statement

Most of the farmer faces many problems to harvest their crops because they had been attacked by many diseases. when the leaf had been attacked, the others areas had been exposed to be infected. Thus, it will decrease farmer's income and lead to significant losses to the farmer. Currently, the farmer determines the type of disease manually. The errors might occur in order to determine the type of diseases because of inexperience. It may also occur for lack of good chemist. Farmer also has to spend a lot of time to detect the type of disease. It also takes a time as the farmers manually check the disease since the crops field is in a wide area.

1.4 Research Questions

The common research questions in this thesis are Which segmentation technique is best for this situation? How to Reduce the image into an appropriate size? And How can we detect leaf diseases?

1.5 Expected Output

The development of the application has met its specification successfully. For the given leaves with three categories of disease, the application can recognize all them.

1.6 Scope of Research

From existing research and gathering knowledge we fixed our scopes range. Existing research most of them using one segmentation technique. In this research, we can use two different segmentation technique and four Classification technique. After the procedure, it is possible to detect leaf diseases more perfect and accurate.

1.7 Thesis Organization

The dissertation consists of six chapters representing the work of the thesis. The first part of the report contains the background information and problem statement to do the thesis. It is a necessity to study whether a thesis is worthy to be done. It also concisely describes the overall work included in the project. Related work on the previous research related to the thesis and the basic knowledge required to read the dissertation are discussed in the second part. The thesis reader should be a generous computer science student. however, the content of the report includes too much-specialized knowledge in machine vision and statistics. Therefore, it is necessary and helpful to talk about some basic knowledge about our thesis we focused on. The third part illustrates how this system was designed by some design proposals and the fourth part mainly discusses the implementation of the thesis including various algorithm designs and the implementation process. The fifth part mainly concerns the evaluation of the thesis. It not only talks about the generalization of the segmentation method for other categories of leaves but also includes the comparison of these different implemented pattern recognition method with 10-fold cross-validation. The final part summarizes what we have done and what we will do in the future if there are some future works.

CHAPTER 2 Background Study


2.1 Introduction



This chapter briefly reviews, explains and discusses on existing literature review related with the current project which is “Leaf Disease Detection System Using Image Processing” that will be developed later. This chapter comprises three sections. The first section describes the overviews of Leaf Diseases, Symptoms and Management. The subsections are the definition, type of Leaf disease, Leaf symptom, and leaf management. The second section is the review of some existing system that used the same techniques and methods. The third section discusses the Literature Review.

2.2 Leaf Diseases, Symptoms and Management

There are many factors that make crops production become slow and less productive. One of the main factors is leaf disease. The table below will show you some types of leaf disease, the symptom of leaf disease and the management of leaf disease. These researches focus on three types of diseases, which are Powdery Mildew, Septoria, and Rust.

Table 2.1: Description of Leaf Disease

| No | Disease Name and Picture | About Disease |
|----|---|--|
| 1 | <p>Powdery Mildew</p>  | <p><u>Description:</u></p> <p>Powdery Mildew is a fungal disease. It is caused by a variety of closely related fungal species, each with a limited host range. Powdery mildew starts on young leaves as raised blister-like areas that cause leaves to curl, exposing the lower leaf surface. Infected leaves become covered with a white to gray powdery growth.</p> |

| | | |
|---|---|---|
| | | <p><u>Distribution:</u></p> <p>Low soil moisture combined with high humidity levels at the plant surface favors this disease.</p> |
| 2 | <p>Septoria</p>  | <p><u>Description:</u></p> <p>Septoria leaf spot usually appears on the lower leaves after the first crops sets. Spots are circular, about 1/16 to 1/4 inch in diameter with dark brown margins and tan to gray centers with small black fruiting structures. Characteristically, there are many spots per leaf. This disease spreads upwards from oldest to youngest growth. If leaf lesions are numerous, the leaves turn slightly yellow, then brown, and then wither. Fruit infection is rare.</p> <p><u>Distribution:</u></p> <p>Septoria is caused by fungus.</p> |
| 3 | <p>Rust</p>  | <p><u>Description:</u></p> <p>Rust diseases represent the most economically significant fungal diseases in wheat and other cereal crops worldwide, and they are widely distributed across wheat growing regions. Rust diseases possess the ability to spread and travel long distances by dispersal of windborne spores and can rapidly develop under optimal weather conditions.</p> |

| | | |
|--|--|--|
| | | <p><u>Distribution:</u></p> <p>In temperate zones it is destructive on winter wheat because the pathogen overwinters.</p> |
|--|--|--|

2.3 Case Study On Existing System

Medical image processing for diagnosing disease becomes popular these years, because it can not only give a rapid and reliable result to access patient condition but also reduce the cost significantly in comparison to the traditional diagnosis.

2.3.1 Investigating on Image Processing Techniques for Diagnosing Leaf Diseases

A study conducted by Nunik Noviana Kurniawati, Siti Norul Huda, Sheikh Abdullah, Salwani Abdullah, Saad Abdullah from Universiti Kebangsaan, Malaysia aims to develop a prototype system to automatically and correctly detect and classify the Leaf diseases with Powdery Mildew (PM), Septoria, and Narrow Brown Spot Disease (NBSD) using image processing technique as an alternative or supplemental to the traditional manual method.

The methodology for diagnosing leaf diseases can be simplified as Fig. 2.2.1.1. This process involves several tasks, such as image acquisition and collection, image segmentation and pre-processing, shape feature extraction and color feature extraction, and leaf diseases classification based on lesion type, boundary color, spot color, and broken leaf color.

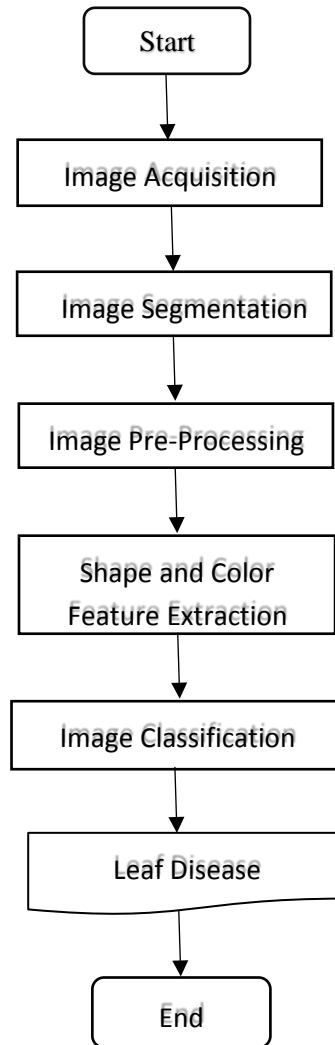


Fig 2.3.1.1: The methodology for diagnosing leaf diseases

2.3.2 Plant Disease Detection Using Image Processing

Health monitoring and disease detection on the plant is a necessity for sustainable agriculture. It is very difficult to monitor the plant diseases manually. It requires a tremendous manual task on each plant, expertizes in the plant diseases, and also require the excessive processing time. Hence, image processing can be used for the detection of plant diseases. Disease detection involves steps like image acquisition, image pre processing, image segmentation, feature extraction, and classification.

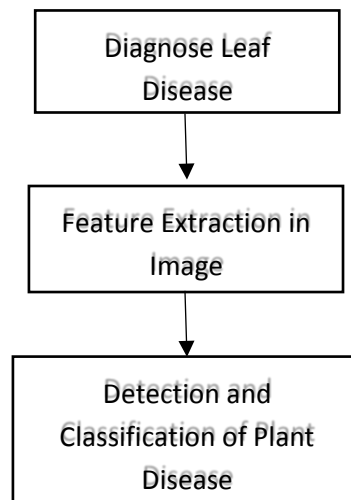


Fig 2.3.2.1: Basic steps for plant disease detection and classification

2.4 Literature Review

In the past decades, image processing and pattern recognition for plant disease detection have been an active research area. The choice of segmentation and feature extraction technique of image processing technique and classification learning method are extremely vital to get a correct result. Some previous paper works on leaf disease detection have been discussed below:

- (a) 2017 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR). They worked with SVM Classifiers to get the result.
- (b) Detection of plant leaf diseases using image segmentation and soft computing techniques. They worked with SVM, ANN, PNN, SELF ORG MAP and Fuzzy logic. [15]

2.5 Theoretical Study

The segmentation part of the project implemented the algorithm of the Expectation Maximization (EM) in Gaussian Mixture Model. The EM algorithm is an iterative algorithm for fitting the probability distribution models via maximum likelihood (ML) estimation of the incomplete data which are the observations. In this section, we will focus on a brief description of the EM algorithm.

2.5.1 The EM Algorithm for Gaussian Mixture Models

We define the EM (Expectation-Maximization) algorithm for Gaussian mixtures as follows. The algorithm is an iterative algorithm that starts from some initial estimate of θ (e.g., random), and then proceeds to iteratively update θ until convergence is detected. Each iteration consists of an E-step and an M-step.

E-Step: Denote the current parameter values as θ . Compute ω_{ik} for all data points x_i , $1 \leq i \leq N$ and all mixture components $1 \leq k \leq K$. Note that for each data point x_i the membership weights are defined such that $\sum_{k=1}^K \omega_{ik} = 1$. This yields an $N \times K$ matrix of membership weights, where each of the rows sums to 1.

M-Step: Now use the membership weights and the data to calculate new parameter values. Let $N_k = \sum_{i=1}^N \omega_{ik}$, i.e., the sum of the membership weights for the k th component—this is the effective number of data points assigned to component k .

Specifically,
$$\alpha_k^{\text{new}} = \frac{1}{N_k}, 1 \leq k \leq K \dots\dots(2.1)$$

These are the new mixture weights.

$$\mu_k^{\text{new}} = \left(\frac{1}{N_k}\right) \sum_{i=1}^N \omega_{ik} \cdot x_i \quad 1 \leq k \leq K. \dots\dots(2.2)$$

2.5.2 Support Vector Machine

In this section, a memory based learning method called Support Vector Machine (SVM) will be introduced and used in the training of our project which uses all the training data for each prediction using kernel methods. Because of the complexity of the SVM with lots of mathematic formulas, it has to use a big chapter to introduce it. Hence in this section, just a simple description will be described. SVM train the machine by processing the data into a higher dimension and then separating the data categories with a hyperplane. With a non-linear mapping from original feature dimension to a sufficiently higher dimension, data from two categories can be always separated. A simple diagram in Figure 2.4.4.1 shows an illustration of two dimensions data, and the nearest training patterns marked as solid points are called support vectors.

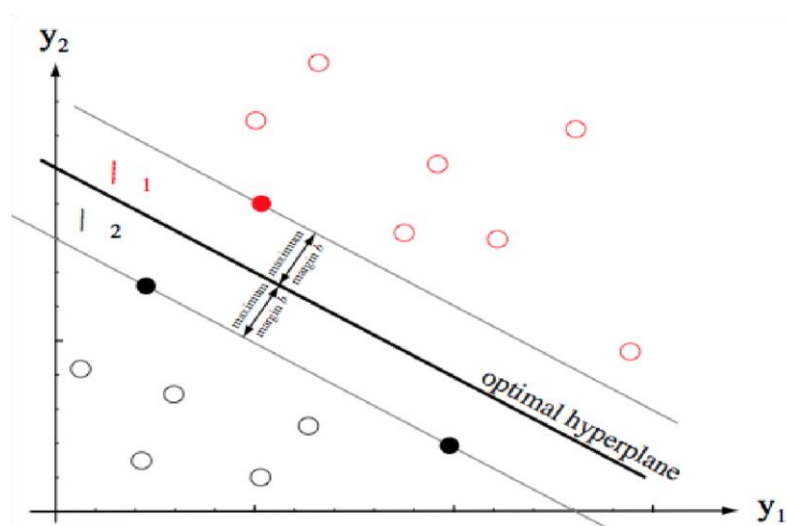


Figure 2.5.2.1: Support Vector Machine

2.6 Research Summary

This study describes a computer vision system to classify a plant image by the disease spot category that it contains from colored image analysis. Three kinds of common leaf diseases Powdery Mildew, Septoria and Rust were used to be analyzed as a sample shown in Figure 2.6. The provided images have the different complex background and light intensity. At first, the first challenge is how to identify and segment the leaf from the different complex backgrounds, some approaches were tried and tested. And then the disease area needed to be partitioned out from the segmented leaf. At last, Expected

Maximization (EM) segmentation was used to segment the leaf from the background and Otsu is for the disease spot segmentation from the leaves. After segmentation, what features should be the most appropriate to be extracted as discriminators when the disease does not follow a well-defined shape or color domain pattern. Additionally, the light intensity is another tricky problem to choose the most proper features for differentiating these three diseases. Hue Color Intensity histogram and texture features were extracted to discriminate the diseased spot in our application. What classifier is the most appropriate for the disease spot identification is another important study in this project. Support Vector Machine (SVM) were implemented and evaluated to compare the pros and cons of them and which one is the best one for this situation. The features on the whole leaf or only disease spot points are also compared and evaluated to decide which one can be the best discriminator. At last an additional computer-based program was implemented to allow the user to use a computer in the window to analyze.



(a) Powdery Mildew



(b) Septoria



(c) Rust

Figure 2.6.1: The three kinds of leaf disease included in this study

2.7 Challenges

It is challengeable to implement such an application to identify the disease in such complex background, particular for us without image processing background before. Since no similar application exists in a software store and many researchers are studying this topic, it is impossible for an undergraduate student to build a perfect application about that. However, we worked hard and tried our best to learn and research some image processing and machine learning technique so that the application was implemented which achieved the goals in the specification. On the other hand, with few requirement and constraints, the goal is clear and the system design is flexible.

CHAPTER 3 Implementation of the Proposed Methodology

3.1 Introduction

The purpose of this chapter is to discuss the detailed approach and system for the thesis. Method, technique or approach that has been used while designing and implementing the thesis are included in this chapter. This chapter also explains about the justification of method or approach used and hardware and software necessity.

3.2 Work Flow

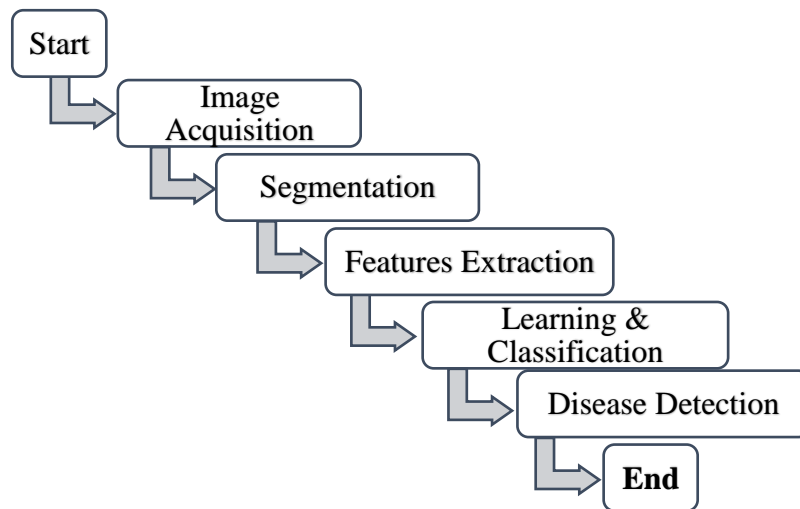


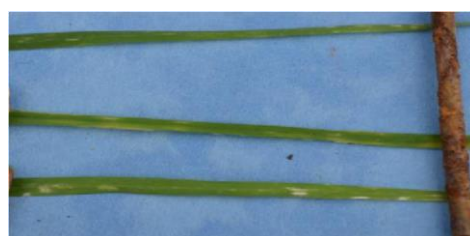
Figure 3.2.1: Work Flow of proposed methodology

3.3 Segmentation

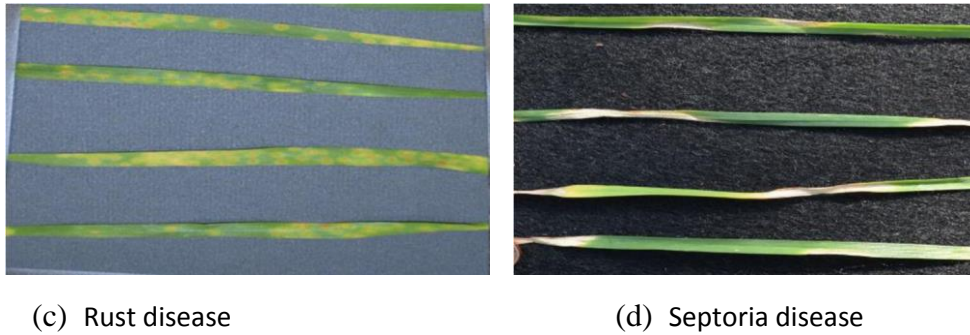
Segmentation is one of the most difficult parts of this project. The provided image is very large and the background is complex with some leaf and branches, so it is not easy to segment the diseased spot out in an interactive application for a short time. In addition, the aim of the project is to produce an application, thus the segmentation has to be automatic rather than by human manipulation.



Natural leaf



(b) Powdery Mildew disease



(c) Rust disease

(d) Septoria disease

Figure 3.3.1: The three kinds of leaf disease included in this study

Image Segmentation approach can be generally split into two categories, which are edge detection or pixel classification. Both of them have different kinds of advantages and disadvantages. For the edge detection, some unclosed regions may appear in the segmented image, and pixel classification can bring some isolated classifications. Pixel classifications are usually unsupervised clustering of the pixels. There is some wellknown clustering algorithm. One of them is the k-means algorithm which is based on the recursion of two operations to classify a set of data into k categories. The algorithm assigns each point to the nearest mean to be labeled, and then calculate the means for labeled points of each category. These two operations iterate again and again until the error function converges. However, the first k means are randomly initialized, which could cause different results for the same data set. For the leaf segmentation, we think Expectation Maximization (EM) segmentation in Gaussian Mixture Model should be suitable, which is a color segmentation method based on Maximum Likelihood (ML) estimation for the pixel clustering. The EM algorithm in Gaussian Mixture Model is explained in Chapter 2, so for this part, we just care about the EM segmentation and its implementation.

3.3.1 Leaf Segmentation

EM segmentation

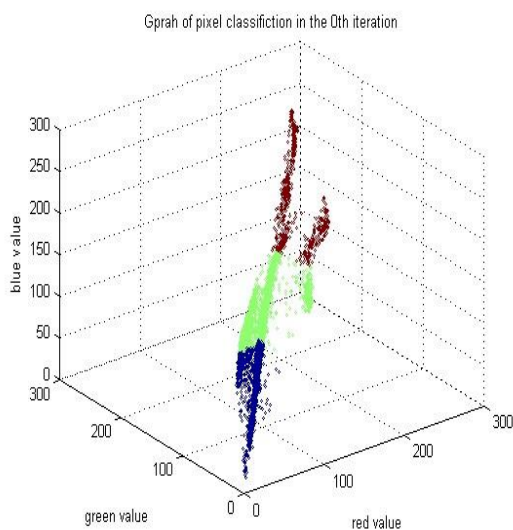
The EM color segmentation segments the image into different images based on the iterative EM algorithm. An RGB image consists of many pixels, and every pixel consists of three elements which represent for red, green and blue respectively. It can also be viewed as red, green and blue three planes combined together. According to the Expectation Maximization algorithm introduced in Chapter 2, an algorithm was implemented to estimate the prior possibility of the mixture of Gaussians gave a set of

pixel data, and then each pixel can be classified into the corresponding category by possibility comparison. To visualize the process, the function was tested by giving a leaf picture (Figure 3.3.1.1) and 3 categories, and the classification result during each iteration was plotted. Since the quality of the given pixel is too high, we have to resize the image into a smaller one for further classification. For a quick and clear plot, the image was resized into the one width 100; As the data shown in Figure 4.3, the classification becomes more and more accurate

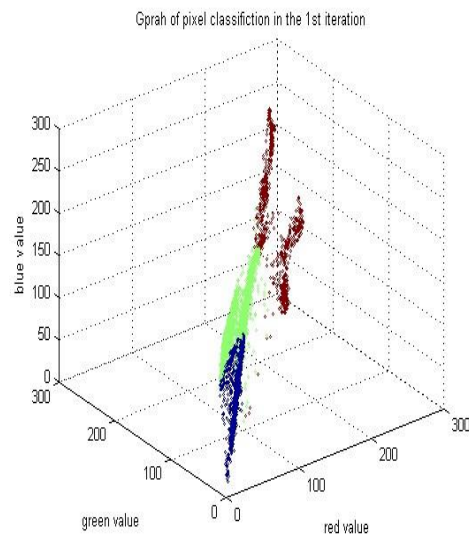


Figure 3.3.1.1: Test picture

Visualization of the changes of clustering for each pixel during the EM process, which was implemented with MatLab. The red, green points represent the pixel data clustered into three kinds of categories.



(a) 0th iteration



(b) 1th iteration

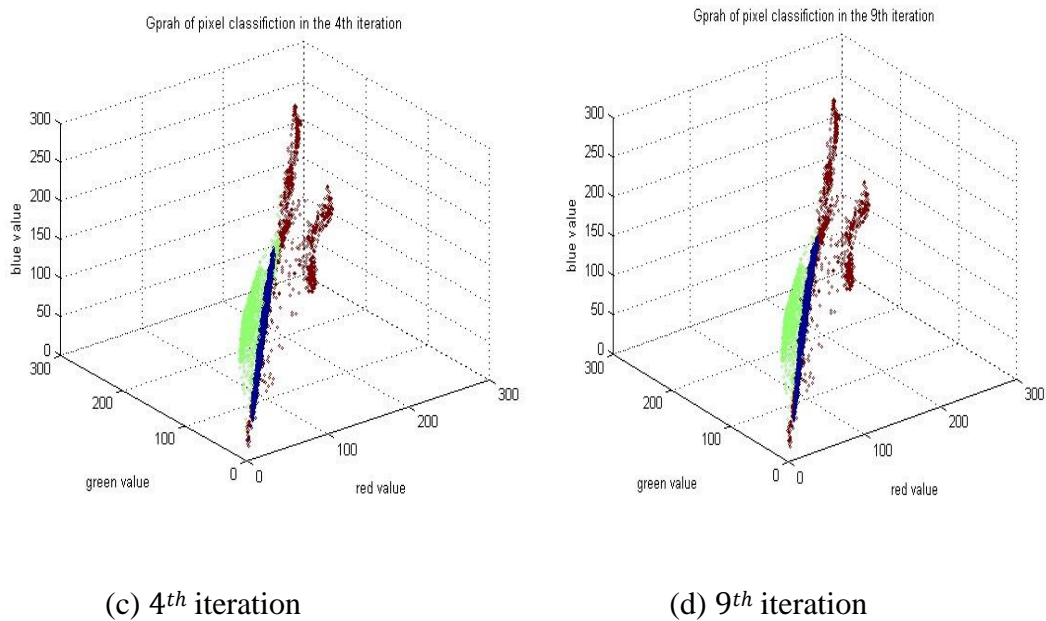


Figure 3.3.1.2: Visualization of the changes of clustering for each pixel during the EM process with the increasing EM iterations.

In order to save the segmentation time, stopping criteria should be introduced. If the error rate is less than a certain value (0.1), the iteration will stop. The final parameter estimates of the means and the proportions of the clusters are shown in Table 3.1

Table 3.3: Final Parameter Estimates of the Means and the Proportions for the Test Pcture.

| Case | Cluster#1 | Cluster#2 | Cluster#3 |
|------------|-----------|-----------|-----------|
| R | 105.6604 | 105.8797 | 182.5012 |
| G | 107.3948 | 134.7149 | 167.3033 |
| B | 110.1285 | 91.4143 | 156.2868 |
| Proportion | 0.8158 | 0.0895 | 0.0947 |

After that, we need to partition the color image into some colored segments with the segmentation labels for each pixel. Figure 3.2.1.3 shows an example of a segmented color image. (a) is the original image, and (b),(c),(d) are segmented color images from (a) with EM algorithm. The segmentation result for this example is good.

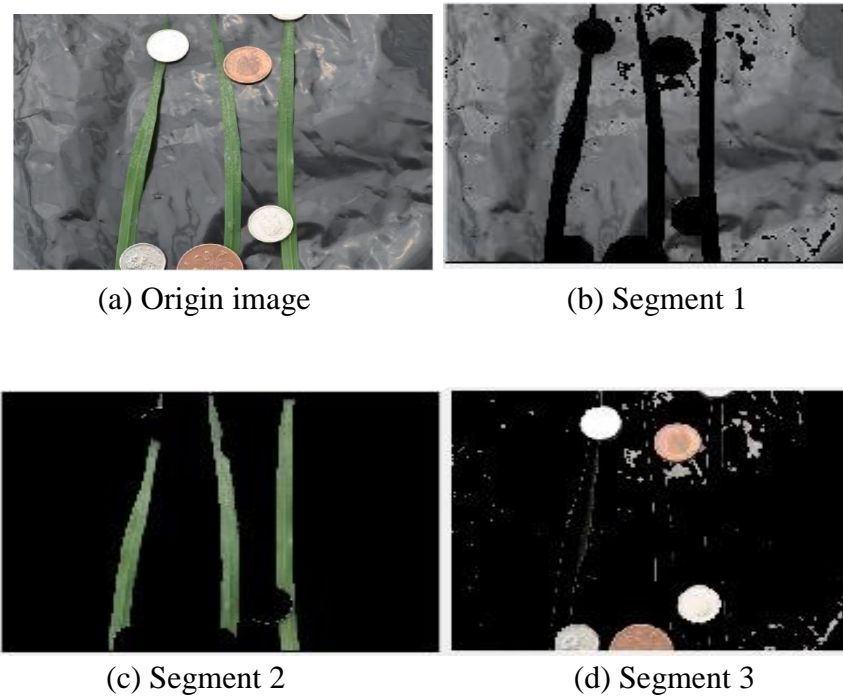


Figure 3.3.1.3: Leaf Segmentation Example

Post processing

Although the EM segmentation can perform well for most images, there are still some serious problems for some particular images. Therefore, some post-processing technique had been taken for a better segmentation result. Look at Figure 3.2.1.4 (a) is one of segmentation result images, however, some coins and blue edges can still not be removed. For some defects in the image, we can use the color threshold to get rid of some defects. Thus a keeping green and yellow method were added to post-process the image by a color threshold. Picture (b) is the result of the method, which looks good.

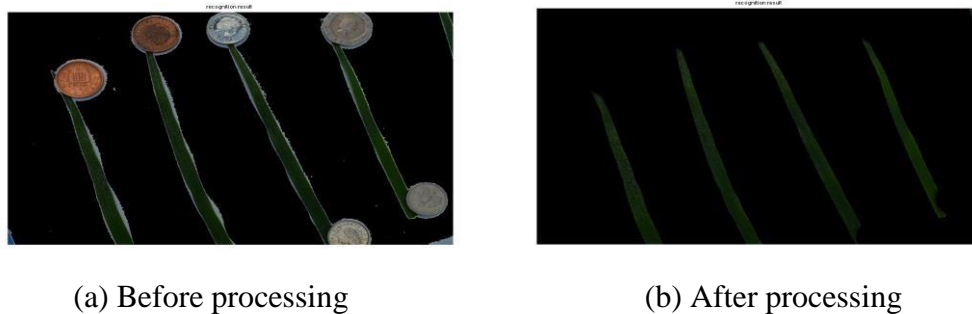


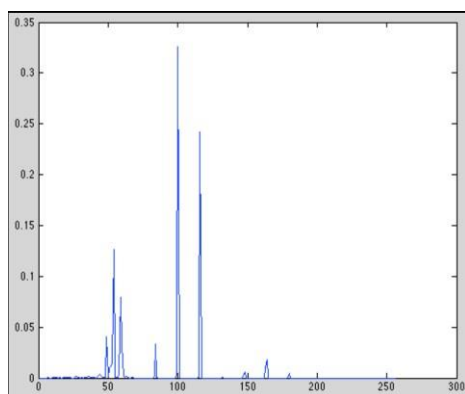
Figure 3.3.1.4: Processing segmented image with a function to get rid of some irrelevant Colour

3.4 Feature Extraction

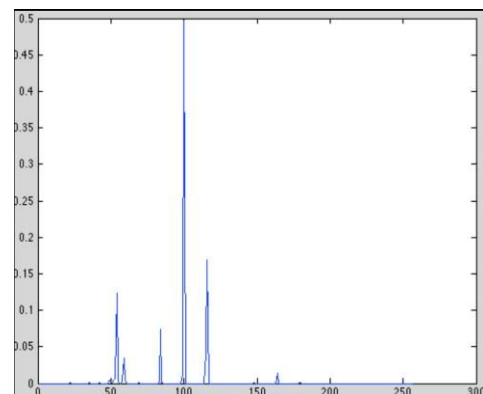
In order to identify the leaf disease, we should choose some appropriate features as a descriptor to distinguish different kinds of leaf disease. Inappropriate or excessive features would lead the classification to overfit and long search time. Hence it is critically important to choose a good descriptor from the various diseases. After lots of experiments, it was found that the combination of color histogram and Tamura Texture can be a good feature descriptor. We will discuss about that in the following two sections.

3.4.1 Color histogram

Color is often the most expressive among all the visual features. The different disease has different color distribution, and the same disease should have a similar color. Thus we choose color histogram as a descriptor to discriminate the leaf diseases. We choose to use HSV (Hue, Saturation, and Value) color space to calculate the histogram. The program was implemented with the method introduced by Manjunath B.S. As the figures shown in Figure 3.3.1.1, they are the results of the color histograms of the same examples. It can be easily found that the color histogram of the leaves with the same disease is similar, but different diseases are totally different. Hence HSV color histogram should be an appropriate descriptor.



(a) Powdery Mildew1



(b) Powdery Mildew2

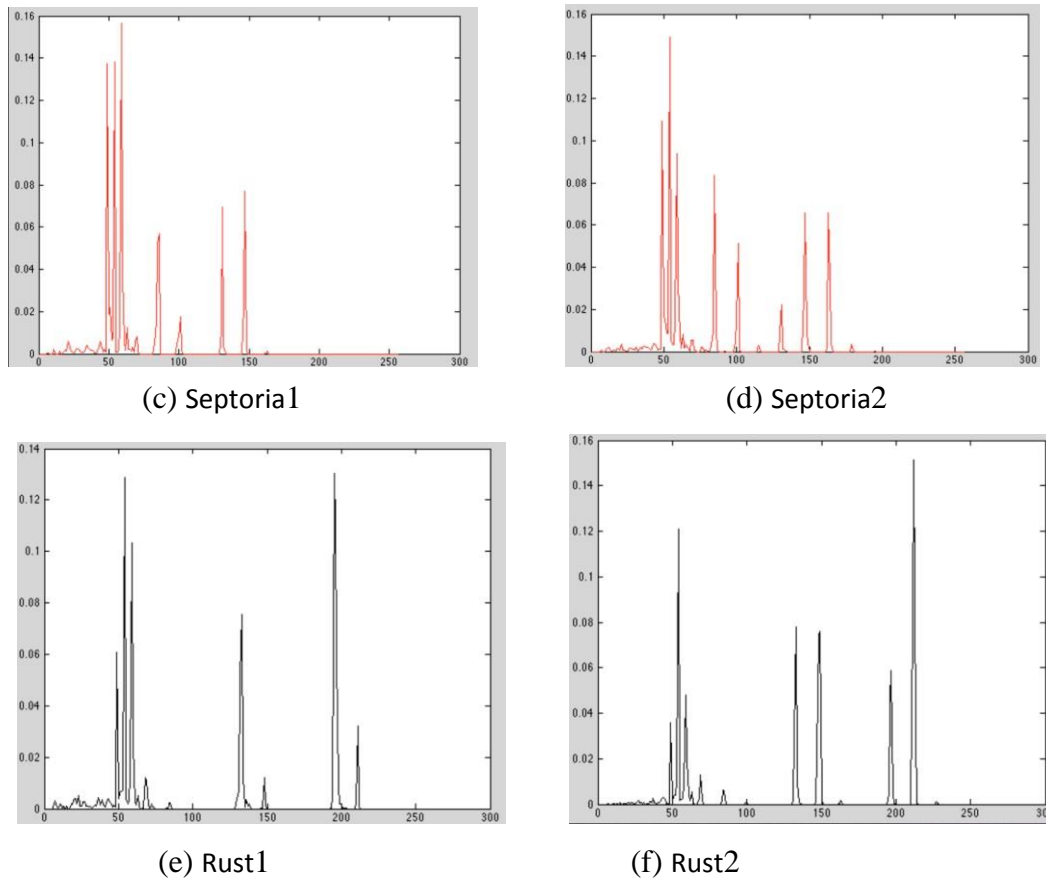


Figure 3.4.1.1: This figure shows the color histogram of some examples of the three kinds of disease.

Figure (a) and (b) in blue line are the color histograms of two different leaves with the Powdery Mildew disease. Similarly, (c) and (d) in red line are the histograms of Septoria, and (e) and (f) in black line of the Rust.

3.4.2 Tamura's Texture Features

Texture feature is another commonly used and very important feature in many image analysis and computer vision applications. Image texture represents the spatial arrangement of intensities and color in the image. Characterizing the textures falls into two approaches which are Statistical and Structural. For this project, one statistical method calls Tamura texture was used as discriminate to classify the diseases. Tree properties of Tamura's texture were implemented in this project which is Coarseness, Contrast, and Directionality.

Finally, we store the color histogram and Tamura texture features in a $1 * M$ vector for each segmented image, where M is the number of features. Then we put the features vectors together in an $N * M$ matrix, where M represents a segmented disease images features and N represents the total number of image examples. There is also an $N * 1$ vector which represents the labels of the corresponding disease examples. These labels are numbered from 1 to 4, which represents disease free, Powdery Mildew, Septoria, Rust leaf respectively.

3.5 Learning and Classification

In order to ease the implementation, verification, and documentation, the system could be divided into three different components based on its functionalities. The first part is the GUI as a client which provide a user interface to interact with the functions of the application. The user can choose to take a picture or select an image from the gallery as the leaf image to be diagnosed. Once the users decide which leaf picture he/she wants to use, the application will send the image to the second part. The second part is the server implemented with java servlet which aims to communicate between client and main functions. The servlet will open a socket on a port for each connection and receive an input image and call the main function on the third part, finally sent the resulting image back to the client. The third part is the main component of the application which is used to recognize the category of the disease and implemented by Matlab, and it is also the most difficult and central program. The diagnosis program design would be roughly introduced in the next section. After the recognition, the resulting image would be saved in the server for returning to the client.

3.5.1 Support Vector Machine

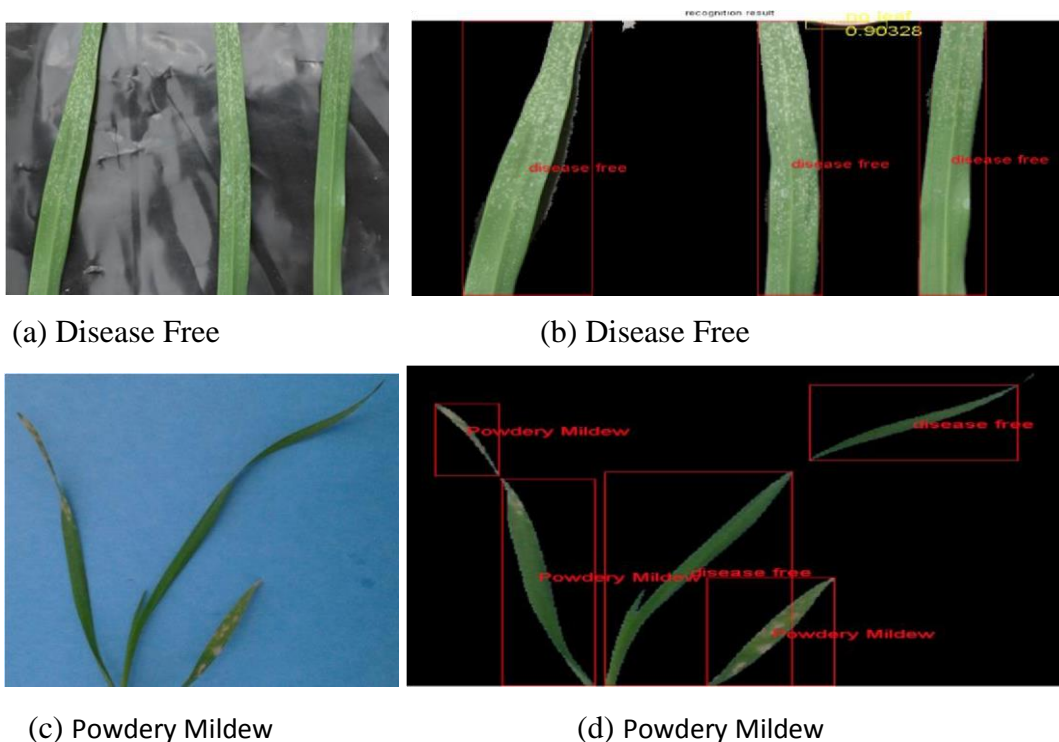
In terms of the implementation of the support vector machine, the MatLab Built-in function SVM-train was used to train the classifier for every two classes. Kernel functions are provided by the SVM-train including the linear kernel, quadratic kernel, polynomial kernel, Gaussian Radial Basis Function kernel with a default scaling factor, sigma, of 1 and Multilayer Perceptron kernel. All the kernel functions were used to train and classify with testing on the feature matrix to find which one is better. At last, the linear kernel gets the best performance result which would be shown in the Evaluation Chapter below.

Multiple Class classification

The pairwise classification was used by the support vector machines. In terms of the training process, each two of the leaf diseases will create an SVM struct. For this project, we know that there are four leaf disease classes, so we should build $(4-1)*4/2 = 6$ SVM struct. When we test a leaf disease data, we shall evaluate the 6 SVM and then classify according to which of the classes gets the highest numbers of the vote. A vote means a classifier put a pattern into the class.

3.6 Disease Detection

After the implementation of the classifier, it is not difficult to build a recognition program to classify which category every leaf belongs to. Firstly, segmentation function should be used to segment the meaningful area out and then feature extraction function was called to get the distinguished features from the segment storing in a vector. Successively the feature vector should be put into the classifier to predict its category. Finally, the predicted label will be drawn on the leaf. Figure 3.6.1 shows some sample results of the leaves over these four categories.



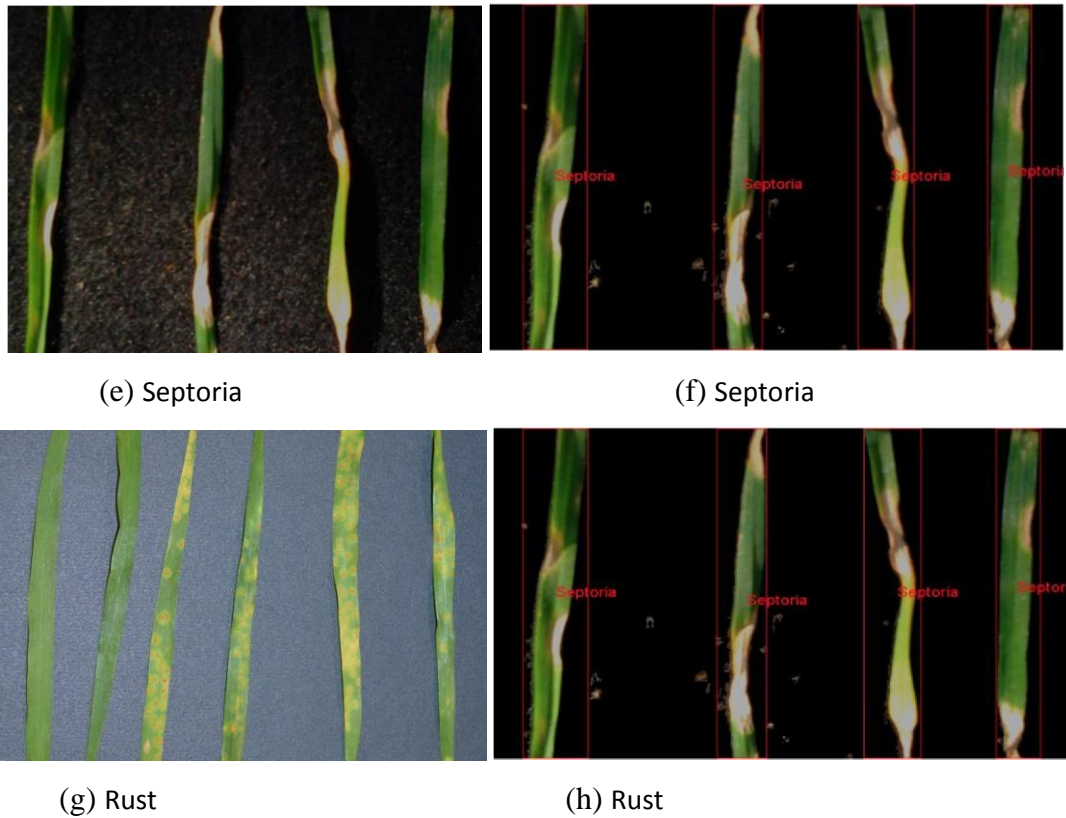


Figure 3.6.1: Some sample recognition results over the leaves in these four categories

Sometimes wrong pictures for example without a big enough leaf in the image may be entered into the program, it is necessary to handle this situation at that time. Figure 3.6.1 shows the two kinds of catch methods used in this program for this kind of exception. After the EM segmentation, if none of the segmented partition has an overwhelming green proportion, the result as sample 1 will be returned with showing 'no leaf in the picture' at the top of the image. If the segmented part is eligible but the pixels of the objects in the partition don't satisfy certain leaf condition, this kinds of object will be marked as no leaf in yellow as the sample 2 shown in Figure 3.6.2 (b).



Figure 3.6.2: Recognition results for two non-leaf images

3.7 Graphical User Interface Development

The user can choose an image from the disk on the computer and do some operations on the resulting image which will be introduced below. Figure 3.7.1 and 3.7.2 show the graphical user interface of the software. After the user opened the software, a welcome page will be shown as Figure 3.7.1.

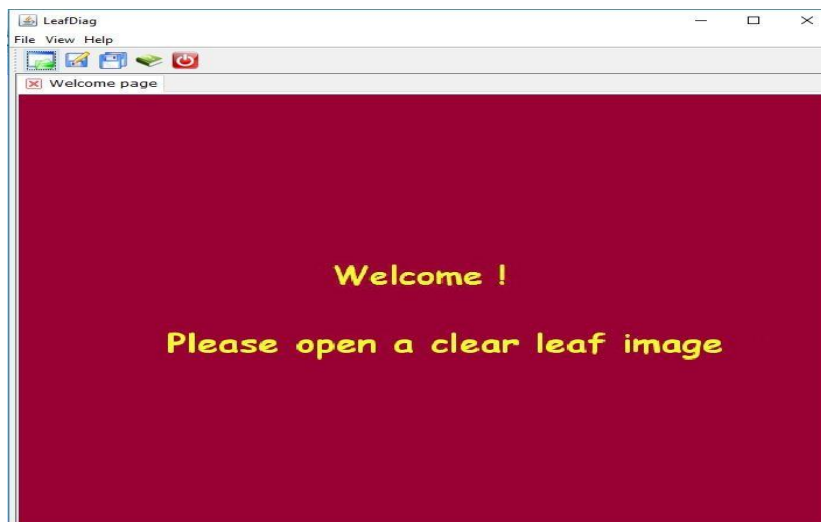
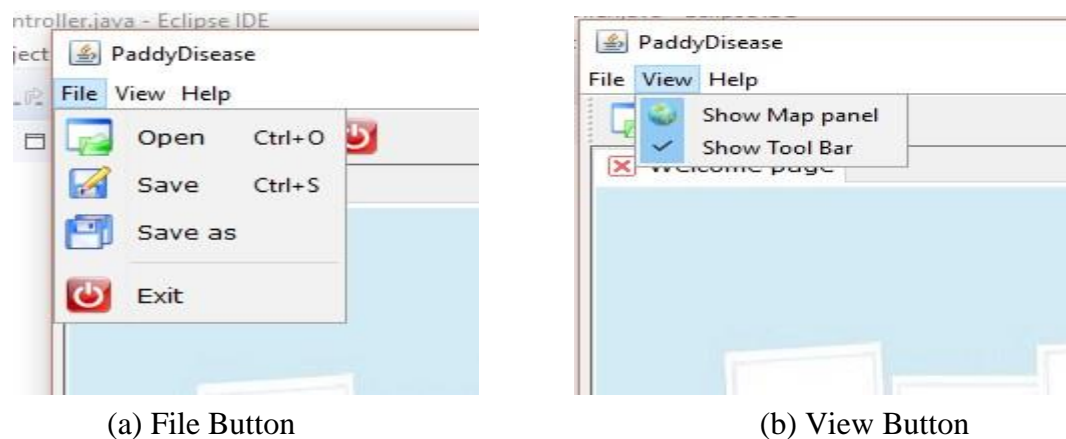


Figure 3.7.1: The Graphical User Interface

The menu bar is at the top of the software shown in Figure 3.7.2 (a). If the user clicks the file button in the menu, the list in Figure 3.7.2 (a) will appear. Then the user can choose to open, save or exit the software. If the user clicks the view button, two options are provided which make the user show or hide the panel and tool bar by their preference as shown in Figure 3.7.2 (b).



(a) File Button

(b) View Button

Figure 3.7.2: Menu Bar

The toolbar is also available shown in Figure 3.7.3 which is very convenient to the user and contain all the main functions. If the user doesn't want to show the toolbar, he can make it invisible by the setting of view preference.



Figure 3.7.3: The Graphical User Interface

At the start of using this software, the user can click the button in the menu, or tool bar or type Ctrl + O to open an image from the computer. Once an image is chosen, the Matlab program will start to diagnose what disease the leaf in the image has. And the result will be shown in a new tab panel with leaving the previous ones as

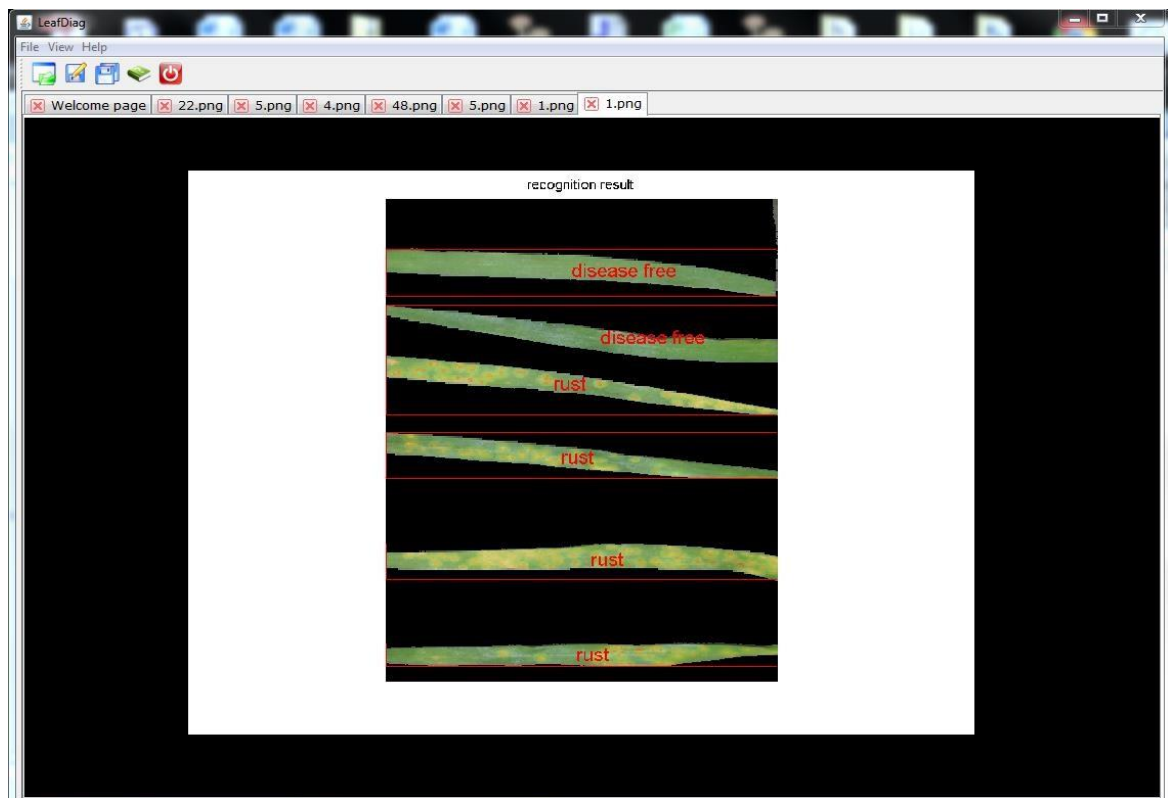


Figure 3.7.4: The Graphical User Interface (after detecting disease)

The tabbed pane can also allow the user to drag the tabs to in any order if they like as shown in Figure 3.7.5.



Figure 3.7.5: The Graphical User Interface

The user can use the mouse to click and drag the resulting image to any position they want, and can even use mouse wheel to zoom in and out the image to the bigger or smaller size. When user zoom in or zoom out the map, four red rectangle arcs will show around the zoom area and the zoom center will be around the mouse position which is shown in Figure 3.7.6.

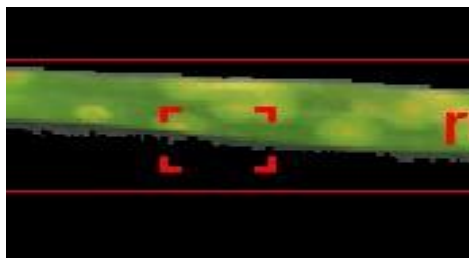


Figure 3.7.6: The Graphical User Interface

If the user clicks the right button of the mouse, a pop menu will occur to allow the user to zoom or center the image in another manner as shown in Figure 3.7.7.

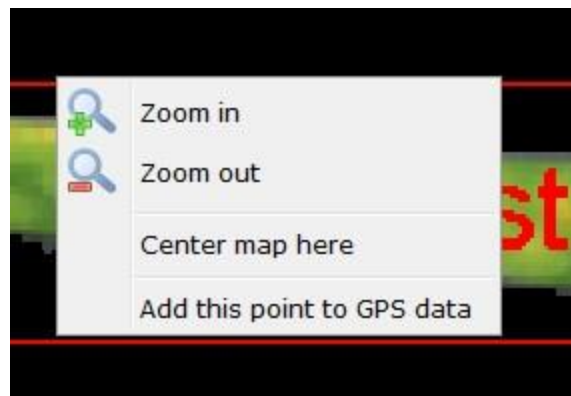


Figure 3.7.7: The Graphical User Interface

CHAPTER 4 Experimental Analysis

4.1 Introduction

There are 274 samples of leaf image used as sample image in the segmentation phase of this development. The wheat images samples had gone through the phases as discuss in the chapter 4. This chapter will briefly describe about the output result of each phases.

4.2 Experimental Setup

The experiments and related analysis are done in this chapter. The experiments and analysis processes are done on a computer with Core-i5 processor having 4 cores with each core having 2.5GHz Speed. Also the system had 8GB of RAM, and 2GB of internal intel HD video memory. For software, MATLAB Version 9.5.0.944444(Release R2018b), NetBeans IDE 8.2 and GlassFish Server Open Source Edition 4.1.

4.3 Performance Measures – Definition

Confusion Matrix – Confusion matrix is to summarize the performance of a classification technique. Classification accuracy can be misleading for an unequal number of observations in each class or more than two classes in dataset. Calculating confusion matrix gives better idea of what classification model is getting right and what types of errors it is making.

True Positives (TP) - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes. E.g. if actual class value indicates that this leaf is Powdery Mildew and predicted class tells the same thing.

True Negatives (TN) - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no. E.g. if actual class says this leaf is not Powdery Mildew and predicted class tells you the same thing.

False positives and false negatives, these values occur when your actual class contradicts with the predicted class.

False Positives (FP) – When actual class is no and predicted class is yes. E.g. if actual class says this is not Powdery Mildew but predicted class tells that this leaf is Powdery Mildew.

False Negatives (FN) – When actual class is yes but predicted class in no. E.g. if actual class value indicates that this is Powdery Mildew and predicted class tells that leaf is Other.

Accuracy – Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. One may think that, if we have high accuracy then our model is best. Yes, accuracy is a great measure but only when you have symmetric datasets where values of false positive and false negatives are almost same. Therefore, one has to look at other parameters to evaluate the performance of the model.

4.4 Result Analysis

Let's indicate Disease Free, Powdery Mildew, Septoria and Rust, 4 classes. And predicted class and actual class are on row and column. So, the table of confusion matrix is as below:

Table 4.4.1: Confusion matrix of proposed system

| | | Actual Class | | | |
|-----------------|----------------|--------------|----------------|----------|------|
| | | Disease Free | Powdery Mildew | Septoria | Rust |
| Predicted Class | Disease Free | 74 | 2 | 0 | 0 |
| | Powdery Mildew | 2 | 65 | 8 | 0 |
| | Septoria | 1 | 9 | 59 | 1 |
| | Rust | 0 | 0 | 2 | 51 |

4.5 Classification Result

4.5.1 Disease Free

Total 77 leaves which are disease free are used for experiment. In those leaves, 3 images were misclassified and the other 74 images are correctly classified.

Accuracy of Disease Free = 96.1%

4.5.2 Powdery Mildew

Total 76 leaves which are affected by Powdery Mildew disease are used for experiment. In those leaves, 11 images were misclassified and the other 65 images are correctly classified.

Accuracy of Powdery Mildew = 85.5%

4.5.3 Septoria

Total 69 leaves which are affected by Septoria disease are used for experiment. In those leaves, 10 images were misclassified and the other 59 images are correctly classified.

Accuracy of Septoria = 85.5%

4.5.4 Rust

Total 52 leaves which are affected by Rust disease are used for experiment. In those leaves, 1 image were misclassified and the other 51 images are correctly classified.

Accuracy of Rust = 98.1%

4.6 Overall Classification Result

Table 4.6.1: Overall Learning Evaluation

| Classifier | Wheat Leaf | Correct | Incorrect | Recognition Accuracy (%) | Avg. Recognition Accuracy |
|------------|----------------|---------|-----------|--------------------------|---------------------------|
| SVM | Disease Free | 74 | 3 | 96.1% | 90.9% |
| | Powdery Mildew | 65 | 11 | 85.5% | |
| | Septoria | 59 | 10 | 85.5% | |
| | Rust | 51 | 1 | 98.1% | |

4.7 Conclusion

In this chapter, a system for diagnosis the leaf disease has been developed using the MATLAB application. The image processing techniques are applied to find the features and neural network is used to build a trained model. In testing phase, the model is tested with the test images. Before testing each image is passed through leaf color analysis to find out the normal leaf images. The accuracy of testing around 90.9 % percent which is good for practical use.

CHAPTER 5 Conclusion

This chapter summarizes the features and evaluation of the software that is an overall success in the project specification. System limitation will be talked about to show what more functions could be added and which part of the application can be optimized. Finally, future work will be discussed on the possible extension of the project if further development needs to be done.

5.1 Summary

The development of the application has met its specification successfully. For the given leaves with three categories of disease, the application can recognize all them. The segmentation program can successfully segment the leaves and diseases spot out with few defects. The machine learning methods based on the extracted features work very well. The result can demonstrate that both features extracted and the classification method are appropriate for this application. Furthermore, the result will be much higher if more training samples are given in the future. In addition, the implementation of the Java software is also tested with JUnit Test and all of them are passed. The operation is easy for the user to start to use the application. The user evaluation results categorize the software as high user-friendly. Overall the project is successful for identifying the plant's disease based on its leaf appearance. It can be a central prototype to be used by the agricultural industry to improve the production of crops or by gardeners to recognize the disease and avoid it next time.

5.2 System Limitation

Although the system has many positive aspects and successfully achieved all the goals on the specification, it could still be enhanced to be a publishable application. In terms of the segmentation, the leaves cannot be segmented into very good result if the background is too complex which is nearly an unavoidable problem. However, the application can add a new function to allow the user to cut their interest part out, then the result should be acceptable. The application can also make the user input the EM segmentation parameter if they would like to. For example, if the user chooses to partition the image into two sub-images, but the result is not very good, then the user can try to partition it into three. For the feature extraction, the more complicated feature can be tried in experiments. It is a complicated part, many papers discuss that in the different complex method. However, implementing and apply all the candidate algorithms in real time is not practical. The system only trained three categories disease of the same plant with a limited amount of images which can not lead generally used in reality. But the project is just a research in this research field, the recognition could be more generalized and better if more images are given and trained in the future.

5.3 Future Work

Many extensional functions can be added to the application which could make it is a real application in a Google Store or Apple Store. Other types of software such as website could also be implemented to make the program more convenient to attain. The application should allow the user to cut the leaf region out if the background is too big or too complex, which would very helpful to segment the leaf out successfully. More other advanced features can be implemented and tested to optimize the algorithm to get a better result. Other categories of leaves with different diseases should be trained via experiments to make the application applicable to as many categories of plants as possible. The application software can be built in some additional functions such as viewing histories, saving locations and so on.

REFERENCES

- [1] Planet Natural Disease Center, Available: << <https://www.planetnatural.com/pest-problem-solver/plant-disease/>>> [Last Accessed on 07-12-2018]
- [2] A. P. Dhawan, "A review on biomedical image processing and future trends," *Computer Methods and Programs in Biomedicine*, vol. 31, no. 3-4, pp. 141-183, March–April 1990.
- [3] A. M. P. K. K. A. and A. S. , "Grape leaf disease detection from color imagery using hybrid intelligent system," in *2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, Thailand, 2008*.

- [4] S. Phadikar and J. S. , "Rice disease identification using pattern recognition techniques," in 2008 11th International Conference on Computer and Information Technology, Khulna, Bangladesh, 24-27 Dec. 2008.
- [5] S. D. Khirade and A. P. , "Plant Disease Detection Using Image Processing," in 2015 International Conference on Computing Communication Control and Automation, Pune, India, 26-27 Feb. 2015.
- [6] S. J. D. Prince, in *Computer Vision: Models, Learning, and Inference*, London, Cambridge University Press, 2012, Jun 18, 2012, p. 580.
- [7] J. R. Parker, "Algorithms for Image Processing and Computer Vision, 2nd Edition", John, 2010.
- [8] C. M. Bishop, "Pattern Recognition and, Springer Science+Business Media, LLC", 2006.
- [9] B. S. Manjunath and Y. Deng, "segmentation of color-texture regions in images and video.," IEEE Computer Society, pp. 800-810, October 2001.
- [10] P. Karmakar, S. Wei Teng, D. Zhang, Y. Liu and G. L. , "Improved Tamura Features for Image Classification Using Kernel Based Descriptors," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), no. 21 December 2017, 29 Nov.-1 Dec. 2017.
- [11] E. Pekalska and B. Duin, "Pattern Recognition Tools," 15 July 2015. [Online]. Available: <http://37steps.com/4859/cross-validation/>. [Last Accessed on 08-12-2018].
- [12] J. Brownlee, "Code Machine Learning Algorithms From Scratch," 18 November 2016. [Online]. Available: << <https://machinelearningmastery.com/confusion-matrix-machinelearning/> >> [Last Accessed on 09-12-2018].
- [13] R. Joshi, "Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures," 9 September 2016. [Online]. Available: << <https://blog.exsilio.com/all/accuracy-precision-recallf1-score-interpretation-of-performance-measures> >> [Last Accessed on 10-12-2018].
- [14] S. Weizheng, W. Y. C. Zhanliang and W. Hongda, "Grading Method of Leaf Spot Disease Based on Image Processing," in 2008 International Conference on Computer Science and Software Engineering, Hubei, China, 2008. [Last Accessed on 15-12-2018]
- [15] S. Weizheng, W. Y. C. Zhanliang and W. Hongda, "Grading Method of Leaf Spot Disease Based on Image Processing," in 2008 International Conference on Computer Science and Software Engineering, Hubei, China, 2008. [Last Accessed on 19-12-2018].
- [16] Publisher: IEEE "2015 International Conference on Computing Communication Control and Automation" << <https://ieeexplore.ieee.org/abstract/document/7155951> >> [Last Accessed on 19-12-2018].

APPENDIX

Appendix A: Research Reflection

The purpose of this appendix is about research reflection. From Spring 2018 semester we started our journey to make this Application. The main feature of our application is detecting leaf disease. Also using our application people can easily identify leaf disease. In the very first days, we studied a lot of papers about leaf disease and others. We have studied a lot to detect the disease. On the other hand, it is challengeable to implement such an analyzer to identify the disease in such complex background, particular for us without image processing background before. Since no same application exists in software store and many researchers are studying this topic, it is impossible for an undergraduate student to build a perfect application about that. However, we worked hard and tried our best to learn and research some image processing and machine learning technique so that the application was implemented which achieved the goals in the specification. On the other hand, with few requirement and constraints, the goal is clear and the system design is flexible.

Firstly, we built a model for our project. Then we implemented our project step by step.

After many hard works and spending a lot of time finally we were able to reach our goal. The project "Leaf Disease Detection Using Image Processing" will be very helpful for every normal people and farmers. People will be satisfied very easily. So we believe that our "Leaf Disease Detection Using Image Processing" application will be very positive and effective for users.

Appendix B: Related Issues

We did not find any related application like this, but there are some similar applications such as 'Garden Compass Plant / Disease Identifier' (iTunes Store) for iPhone or 'A&L Plant Disease Diagnosis' (Google Play) for Android, most of them need to take a photo and submit it to a team of plant experts who will identify the symptom. However, the user have to take lots of time to wait for the result and pay some money for the application.

Turnitin Originality Report

Processed on: 01-Apr-2019 12:22 +06
 ID: 1103607346
 Word Count: 8450 Submitted: 1

151-15-5504 By Redwan Ahmed

| | |
|------------------------------------|--|
| Similarity Index 29% | Similarity by Source |
| | Internet Sources: 24% Publications: 9% Student Papers: 26% |

| |
|---|
| 5% match (Internet from 12-Jan-2016) http://umpir.ump.edu.my/6994/1/CD7686.pdf |
| 4% match (student papers from 13-Mar-2017) Submitted to University of Ghana on 2017-03-13 |
| 2% match (publications) Sunil Saumya, Jyoti Prakash Singh. "Detection of spam reviews: a sentiment analysis approach", CSI Transactions on ICT, 2018 |
| 2% match (student papers from 05-Apr-2018) Class: Article 2018 Assignment: Journal Article Paper ID: 941491107 |
| 1% match (student papers from 07-Apr-2018) Class: Article 2018 Assignment: Journal Article Paper ID: 942517073 |
| 1% match (student papers from 18-Dec-2012) Submitted to Universiti Malaysia Pahang on 2012-12-18 |
| 1% match (student papers from 31-Mar-2019) Submitted to Daffodil International University on 2019-03-31 |
| 1% match (Internet from 24-Jan-2019) https://thetomatoladyblog.com/author/elzbthc/ |
| 1% match (Internet from 22-Aug-2018) http://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/ |
| 1% match (Internet from 09-Jun-2017) https://ijircce.com/upload/2017/ncraet/20_EC-121-PLANT%20DIESEASE%20DETECTION%20USING%20IMAGE%20PROCESSING.pdf |
| 1% match (Internet from 17-Nov-2017) http://www.ics.uci.edu/~smyth/courses/cs274/notes/EMnotes.pdf |
| 1% match (Internet from 31-Dec-2015) http://dSPACE.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/908/Two%20-%20Step%20Verification%20System%20Of%20National%20Id.pdf?sequence=1 |
| 1% match (Internet from 03-Jan-2011) http://planetnatural.com/site/xdpy/kb/powdery-mildew.html |
| 1% match (Internet from 13-May-2009) http://www.ics.uci.edu/~smyth/courses/ics274/notes4.pdf |

1% match (student papers from 01-Apr-2019)
[Submitted to Daffodil International University on 2019-04-01](#)