Plant Disease Classification Using Texture-Based Methods through Leaf Images

Suleiman Abdulrashid Usman

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
May 2019
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Ali Hakan Ulusoy
Acting Director

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

Prof. Dr. Hadi Işık Aybay
Chair, Department of Computer Engineering

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

Assoc. Prof. Dr. Önsen Toygar
Supervisor

Examining Committee

1. Assoc. Prof. Dr. Önsen Toygar

2. Asst. Prof. Dr. Yıltan Bitirim

3. Asst. Prof. Dr. Kamil Yurtkan

ABSTRACT

Plant products have been a major source of food for animals, raw materials for industry and source of revenues to governments. In view of this, careful attention is needed for quality and quantity of plant products. Biotic and abiotic factors contribute immensely in hampering agricultural produce. In this research, computer vision techniques such as texture-based algorithms namely Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Binarized Statistical Image Features (BSIF) are employed in plant disease identification and classification. Nine different popular plant species are used with symptoms on leaf images to extract features to develop a novel system. We propose an approach that employs Decision-Level Fusion which is used to incorporate different algorithms' strengths for a robust and more accurate system. The proposed method is also compared with Scale Invariant Feature Transform (SIFT) and its derivatives such as Dense Scale Invariant Feature Transform (DSIFT) and Pyramid Histogram of Visual Words (PHOW). The experiments are conducted on PlantVillage database that includes healthy and infected plant leaf images of tomato, apple, cherry, corn, grape, peach, pepper, potato and strawberry plants. Consequently, the diverse nature of the database and the high accuracy of the proposed system show that Decision-Level Fusion of texture-based features extracted from plant leaves are good in detecting and classifying plant diseases.

Keywords: Plant disease identification, Computer vision, Texture-based features, Decision-Level Fusion

ÖZ

Bitkisel ürünler, hayvanlar için başlıca besin kaynağı, sanayi için hammedde ve devletler için gelir kaynağıdır. Bu yüzden bitkisel ürünlerin niteliği ve niceliğiyle ilgili özen gösterilmesi gerekir. Biyotik (canlı) ve abiyotik (cansız) faktörler, tarımsal ürünlerin bozulmasına büyük ölçüde sebep olurlar. Bu çalışmada, Yerel İkili Örüntü (LBP), Gradientlere Yönelik Histogramlar (HOG) ve İkili İstatistiksel Görüntü Öznitelikleri (BSIF) gibi bilgisayarla görü yöntemlerinden dokuya-bağlı algoritmalar, bitki hastalıklarının tanımlandırılması ve sınıflandırılmasında kullanılmıştır. Yaprak resimlerinin üzerinde hastalık semptomları olan sekiz değişik ve popüler bitki türünün özniteliklerini kullanarak yeni bir yaklaşım geliştirilmiştir. Karar-Seviyesi Kaynaşımını kullanan önerilen yaklaşım, farklı algoritmaların gücünü birleştirerek daha kuvvetli bir sistem yaratmıştır. Önerilen yaklaşım, Ölçeklemeden Bağımsız Öznitelik Dönüşümü (SIFT) ve türevlerinden olan Yoğun Ölçeklemeden Bağımsız Öznitelik Dönüşümü (DSIFT) ve Görsel Kelimelerin Piramit Histogramları (PHOW) yöntemleriyle de karşılaştırılmıştır. Deneyler; domates, elma, kiraz, mısır, asma, şeftali, biber, patates ve çilek bitkilerinin sağlıklı ve hastalıklı yaprak görüntülerini içeren PlantVillage veritabanı üzerinde yapılmıştır. Sonuç olarak, veritabanının doğal çeşitliliği ve önerilen yöntemin yüksek performansı, bitki hastalıklarının saptanması ve sınıflandırılmasında bitki yapraklarından çıkarılan dokuya-bağlı özniteliklerin Karar-Seviyesi Kaynaşımı ile birleştirilmesinin iyi sonuç verdiğini göstermiştir.

Anahtar Kelimeler: Bitki hastalıklarının tanımlanması, Bigisayarla görü, Dokuya-bağlı yöntemler, Karar-Seviyesi Kaynaşımı.

DEDICATION

This work is dedicated to all those going hungry world-wide as result of wars, climate change and disasters. We have not forgotten you, your pains are ours and you are always in our thoughts.

ACKNOWLEDGMENT

I express my sincere gratitude to Allah S. W. T for the blessings He gave us of good health, sound knowledge and means to pursue knowledge from another corner of the globe and most importantly the blessing of Iman.

My unswerving gratitude goes to my supervisor Assoc. Prof. Dr. Önsen Toygar for her guidance, advice, patience and understanding with me during the period of this research. My sincere gratitude also to the departmental staff of Computer Engineering for their concern in me and encouragements.

My appreciation goes to my parents and family for their support, prayers and for their handmade task of up-bringing me with all that is needed to be a complete human morally, psychologically, economically and socially. Your efforts will never go in vain.

I express my thanks to my friends who became my family in Cyprus and Turkey. I feel at home being with you, it's really an honour to be with you. My trust in you is undying.

To the people I met during the course of my masters, I am thankful to you all, one way or the other you are part of this journey and this success is for us all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BSIF	Binarized Statistical Image Features
DSIFT	Dense Invariant Feature Transform
GW	Gaussian Wavelet
HOG	Histogram of Oriented Gradients
KNN	K-Nearest-Neighbour
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
PCA	Principal Component Analysis
PHOW	Pyramid Histogram of Visual Words
RAM	Random Access Memory
RGB	Red Green Blue
SIFT	Scale Invariant Feature Transform
SURF	Speeded-up Robust Features
SVM	Support Vector Machine
VFD	Volumetric Fractal Dimension

Chapter 1

INTRODUCTION

As part of the United Nations sustainable development goal to achieve zero hunger worldwide, it becomes imperative on nations, researchers and scientists in devising means to feed 815 million people who are hungry worldwide by investing efforts and resources in food security to boost agricultural capacity worldwide for sustainable food production and zero hunger world [1]. As shown by [2], 70% to 80% of annual lose in crops produce is caused by diseases in plants.

Plant disease manifest in various parts of the plant such as roots, stems and leaves. The leaves are the most obvious to eyes due to their color and surface area making them a preference by farmers and experts to access the health of their crops through it [3]. Currently, with the advancement in computer vision and sophistication of computer algorithms in various aspects of human life, new and efficient methods are used in the identification, detection and diagnosis of different ailments in crops. These new methods are cheaper, easier, more efficient and more reliable than the traditional methods in which experts and farmers use experience and intuition to identify plant diseases.

This thesis compares the performance of the proposed method and established disease detection researches using different image texture-based algorithms for plant disease detection on PlantVillage dataset [42]. PlantVillage dataset is an online

platform for crop health information and resources for researchers and farmers alike. It has over 50,000 pathology expert curated images of over 150 crops with their diseases exceeding 1,800. The dataset consists of different plant leaves with various disease ailments. The plants studied in this thesis are apple, cherry, corn, grape, peach, pepper, potato, strawberry and tomato.

1.1 Statement of the Problem

Among the 5.6 billion people that depend on agriculture worldwide [5], most of which are subsistence farmers rely on late advent of disease on plants. Commercial agricultural farmers that employ plant pathologist and other experts for monitoring, identification and detection of plant disease in large fields and plantations require continuous and laborious effort from all parties involved. These will inevitably be expensive, time consuming, prone to mistakes and inefficiencies [6]. This results in large amount of loses of food and cash crops to disease infections thereby causing food shortages and at worst famines in different parts of the globe.

The need for an automatic disease detection in plants becomes paramount and necessary. In this thesis, we employed the use of computer vision and machine learning techniques to facilitate plant disease monitoring, inspection, detection and classification of the detected diseases. The experiments were carried out on PlantVillage datasets which at the end, the performance of the proposed system is compared with the state-of-the-art systems used in plant disease detection and classification.

1.2 Significance of the Study

This study employs different computer vision techniques in detecting and classifying plant disease using leaf images. This will significantly reduce the amount of loss in agricultural produce to disease infections by effective disease recognition and identification of the infected plant through its leaves. It will reduce the burden on experts and farmers in identifying diseases in plants. It will make disease identification easier and faster with less technical know-how. It demonstrates the application and successes of computer vision in plant disease detection and it will help in timely identification of disease manifestations in plants [7] for pre-emptive necessary action.

1.3 Plant Diseases

Diseases affecting plants arise either from Abiotic or Biotic factors. The former comprises of climatic conditions, polluted water bodies, inadequate or excess water for plants usage, chemicals in air and soil, nutrients deficiency while the latter is usually caused by weeds, pests and pathogens (chromistans, fungi, viruses, nematodes, phytoplasmas and bacteria). In these, diseases caused by biotic pathogens pose the most detrimental threats to the survival of plants and the quality of the produce due to their communicability, existence in different forms and difficulty in eradication. Black rot, blight, canker, mildew, spots, rusts, wilting are the most common types of pathogenic caused diseases in plants [2, 3, 8]. When plants come in contact with any of these disease agents, symptoms become apparent on leaves as signs of infections. Factors of disease development are summarized in Table 1.1 as abiotic and biotic factors.

Table 1.1: Factors of disease development

Abiotic factors	**Biotics factors**
Weather conditions	Bacteria
Chemicals	Fungi
Burning of chemicals	Viruses
Spring frosts	Chromista
Hail	Bacteria

1.4 Literature Review on Plant Disease Detection

This section gives a review of plant disease detection according to cash crops, food crops and fruits.

1.4.1 Cash Crops

This is the group of farm produce mainly grown for sale in exchange for cash. It is the main source of employment, revenue in many developing countries for the population and their governments. Most important cash crops include coffee, tea, cocoa, cotton, tobacco, oilseeds, sugar cane, oil palm and rubber [9]. The most studied cash crop is cotton which has been a vital raw material in textile industries. Camargo and Smith [10] reported a study to identify the best set of features to enhance the efficiency of disease identification using leaf images where texture, shape, gray-level and connectivity features are extracted to find good representatives of an image and its inherent disease characteristics irrespective of transformations undergone by the image such as scaling, rotation and translation to train a support vector machine (SVM) classifier. Features of 117 images were extracted to train the classifier in classifying the diseases into any of the three actual classes. Forward and backward feature selection was used to identify a set of the best 45 features which gave an accuracy of 93.1%.

Another cash crop is soybean which is used for its nutritional value in molybdenum, copper, protein and oil content which makes it a good industrial ingredient. Pires et al. [11] used local descriptors and bag of visual words for disease recognition in soybeans. They used five local descriptors namely Scale Invariant Feature Transform (SIFT) [40], Dense Scale Invariant Feature Transform (DSIFT), Pyramid Histogram of Visual Words (PHOW), Histogram of Oriented Gradients (HOG) and Speeded-up Robust Features (SURF) to extract features from leaf images used in training a SVM classifier on 1200 leaf images to classify diseases into 3 classes. A variant of SIFT namely Pyramid Histogram of Visual Words (PHOW) was also employed which is SIFT applied on dense multiple scale and uniform spacing. This algorithm achieved the highest classification rate of 96.25% on color leaf images.

Prajapati et al. [12] used shape and texture (Gray level co-occurrence matrix) features extracted from diseased leaves using Otsu segmentation method by thresholding the hue component of the image to remove noise and background which are of less interest in disease detection. Green components were also removed where only the diseased part of the leaf is left behind. This segmentation helps in extracting strong features that discriminate different cotton diseases using mean and standard deviation of color. The extracted features are used to learn a SVM classifier to be used to classify 190 leaf images.

1.4.2 Food Crops and Fruits

Haiguang et al. [13] applied different neural networks to recognize two structurally similar grape diseases namely grape downy mildew and grape powdery stripe rust using color, shape and texture as features input to a backpropagation neural network, radial basis function network, generalized regression neural network and

probabilistic neural network. Seven set of features were extracted from the dataset of 50 leaves for training and 35 leaves for testing using different number of neurons, different settings of the neural networks and various combination of features. A 100% prediction accuracy was achieved using back propagation neural network, generalized regression neural network and probabilistic neural network while radial basis function prediction accuracy is 94.29%.

Disease quantification by segmentation is the use of plant leaf parts to extract color and texture features to determine the severity of disease in plants. Disease quantification by using thresholding [15] is used by Wijekoon et al. [14] to determine the severity of disease and nutritional deficiencies using a freely available software package Scion by monitoring color changes in leaves due to damage caused by fungal infections on plant tissues. The experiments were conducted on clover leaves infected with spots, bean leaves infected with blight, corn leaves infected with blight, wheat leaves infected with powdery mildew, scrab infected potatoes, alfalfa black stems and yellow blotch infected alfalfa leaves. The disease severity reported from the experiments shows strong correlation with hand-crafted disease severity drawing from an expert examination.

Xu et al. [16] created a disease classification system using percent intensity histogram, percent differential histogram, Fourier transform (spectrum energy and time frequency spectrum analysis) and wavelet packet. Features were extracted from color, period character of disease and texture to identify and classify nitrogen and potassium deficient tomato plants. Understanding that nutrient deficiency manifest first on the leaf and on texture of the diseased plant. Genetic algorithm was used to optimize the set of extracted features to measure nutrients deficiency on a scale of

60-80%. Classifiers based on binary trees with each node using fuzzy k – nearest neighbour or backpropagation neural network depending on the availability of the extracted features to classify normal and abnormal leaves on one hand and nitrogen and potassium deficient leaves on the other hand. Classification accuracies of 92.5%, 85% and 82.5% was obtained for normal, nitrogen deficient leaves and potassium deficient leaves respectively and a 10 day earlier disease detection before an expert was achieved.

Romualdo et al. [17] implemented an artificial visual system to determine nitrogen deficiency in leaves which is an important element in chlorophyllic synthesis at different stages of corn life cycle most especially in its young stage to administer corrective measures using Volumetric Fractal Dimension (VFD), Gabor Wavelet bi-dimensional Gaussian function (GW) and VFD with canonical analysis. The corn plants were cultured in a green-house to different nitrogen levels, then small window openings were used to extract texture features to form vectors from the leaves to be used to train Naïve Bayes classifier using 960 images for 24 classes and 40 images in each class for training and 10 for testing. GW at the bottom of colored leaves proved to give more accurate estimation of nitrogen deficiency with accuracy of 82.5% at early stage, 87.5% at later stage and 98% in the middle part of the leaf at silking stage of the plant.

Ilic et al. [18] used statistical and prediction techniques to predict cherry fruit infection with monilinia laxa and coccomyces hiemalis (fungi) to find the most important factor among rain, temperature, humidity and time. The authors emphasized on the importance of factors resulting in cherry infection which will provide farmers with an accurate pre-emptive remedy on the forthcoming infection at

an early time. Using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) for pre-processing. Independent variables numbering up to 724 were used to train four prediction models and 500 test variables were used for testing the accuracy of the trained model. Specifically, Linear Discriminant Model, Quadratic Discriminant Model, Pseudo Linear Discriminant Model and Compact Classification Tree Model achieved the most accurate results on the test parameters. Additionally, the most significant factor in plant disease infection is rainfall as proved in the research. The research aims at prescribing the appropriate and right amount of chemical to be used in a farm which will in return give less hazardous fruits.

1.5 Plant Disease Symptoms

The symptoms of plant diseases are evident on different parts of a plant. However, leaves are found to be the most commonly observed parts for infection detection. Researchers have thus attempted to automate the process of plant disease detection and classification using leaf images. Several works utilized computer vision technologies effectively and contributed a lot to this domain. The physical features, characteristics, symptoms and causes of several plant infections that were extracted from leaf images used in this thesis are described below.

Apple scab is caused by Venturia inaequalis, which infects young leaves more often than old leaves [25, 27]. Lesions appear on leaves with an olive color in a circular shape of up to 1cm in diameter [26]. Apple black spot symptoms on leaves start as purple specks on the leaf surface with tendency to become round, brownish with increase in diameter as the disease grows [28]. Apple cedar '*Gymnosporangium juniperi-virginianae*' is a fungal disease that infects apple plants with the pathogen coming from juniper plants. Spores from juniper plants land on apple leaves which

leads to cedar on the plant. It appears as bright orange-yellow spot which makes it obvious to the naked eye, in acute cases it leads to the defloration of the plant leaves [29]. Figure 1.1 shows apple leaves with the three types of described diseases and a healthy leaf.

Black rot of grapes is caused by *Guignardia bidwellii* which is a disease affecting grape leaves, roots and fruits. It appears on leaves as reddish brown lesions and distinct spots counting in tens. In severe cases it becomes a blotch of tannish brown surrounded by black margin [30]. Grapevine measles also known as Esca measles is a grape disease that affects both fruits and grapes leaves which is caused by *Phaeomoniella chlamydospora* or *Phaeoacremonium* spp fungi. Early symptoms on leaves include the chlorosis in white cultivars and redness in red cultivars and finally necrosis (dead of the tissue) as the final stages of the disease. The measles dechlorify the infected leaves completely rendering it impossible to photosynthesize [31]. Figure 1.2 shows leaf images of the three grape diseases described and a healthy leaf.

Maize blight is a fungal disease caused by *Puccinia sorghi* which manifests on both the upper and lower parts of leaf surfaces of the infected plant as lesions. Its distinct visual features are reddish-brown in color, ranging from oval to an elongated shape of small jagged swellings on the leaves [32, 33]. Figure 1.3 shows the described maize diseases of early blight, late blight on a leaf image and a healthy leaf.

Potato early blight manifests on fresh potato leaves as black or brown lesions with size ranging from 1mm to 10mm. Large blights lead to chlorotic and dehiscleness of the leaf [34]. Late blights are characterised with large lesions which can lead to the

death of the leaves within days [35]. Figure 1.4 shows potato leaves with early blights, late blights and a healthy one.

Cherry mildew is a white cotton-like powder on cherry leaves that is infected by mildew which is caused by *Podosphaera* clandestine fungi [36]. Figure 1.5 shows a mildew infected cherry leaf and a healthy leaf.

Peach bacterial spots are recognized on peach leaves as small lesions, dark in color and normally clustered at the curvy end of the leaf. They often turn yellow as the disease manifestations progress. Peach bacterial spots are caused by *Xanthomonas campestris pathovar Pruni* [37]. Figure 1.6 shows a peach leaf infected with bacterial spot and a healthy peach leaf.

Pepper bacterial spot is caused by the bacterium *Xanthomonas campestris* pathovar *vesicatoria*. Symptoms appear on monitored area of leaves as slightly raised brown spots or lesions. As the spots grow, they converge together and form a necrotic area which eventually lead to the yellowing of the leaf increasing the chance of necrosis [38]. Figure 1.7 shows a bacterial infected pepper leaf and a healthy pepper leaf.

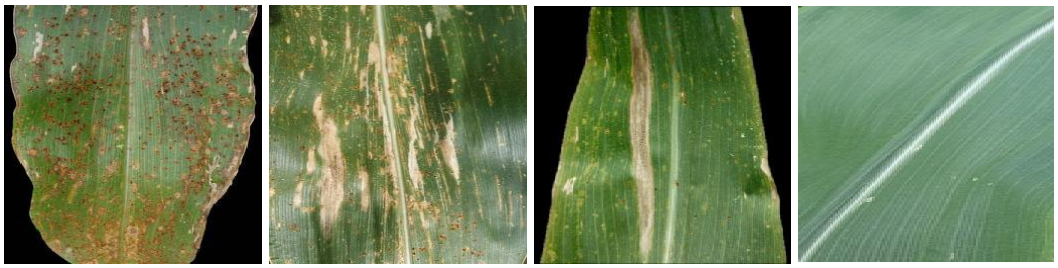
Strawberry leaf scorch is irregular dark purple lesion on strawberry leaves, as the infection progresses, the scorch accumulates all over the upper body of the leaf [39]. Figure 1.8 shows a sample of an infected and healthy strawberry leaves with the disease distinctively recognisable.

(a) (b) (c) (d)

Figure 1.1: Apple leaf (a) with apple scab (b) with black rot (c) with Cedar apple rust (d) healthy apple leaf.

(a) **(**b) (c) (d)

Figure 1.2: Grape leaves (a) with black rot (b) with black measles (c) with leaf blight (d) healthy grape leaf.



(a) (b) (c) (d)

Figure 1.3: Corn leaf (a) with Cercospora leaf spot (b) with common rust (c) with northern leaf blight (d) healthy corn leaf.

(a) (b) (c)

Figure 1.4:Potato leaf (a) with early blight (b) with late blight (c) healthy potato leaf.

(a) (b)

Figure 1.5: Cherry leaf (a) with powdery mildew (b) healthy cherry leaf.

(a) (b)

Figure 1.6: Peach leaf (a) with bacterial spot (b) healthy pepper leaf.

(a) (b)

Figure 1.7: Pepper leaf (a) with bacterial spot (b) healthy pepper leaf.

(a) (b)

Figure 1.8: Strawberry leaf (a) with leaf scorch disease (b) healthy strawberry leaf.

1.6 The Work Done in this Study

In this study, plant disease detection, identification and classification was carried out on different species of plants using image texture-based feature extraction algorithms. Considering the economic significance of disease identification and prevention, different regulations were enacted as early as 1901 until date to ensure food security in different parts of the world. This study identifies the significance of technological development over the decades to simplify disease identification and classification that was carried out previously by eye inspection, soil analysis, plant tissue analysis and field analysis to now use of computer vision technological development [19]. The study involved the supervised extraction of relevant features

from set of images to train the disease identification system which will be used to identify and classify unseen set of leaf images by the trained system. Twenty three different diseases from nine different plants are used in this study. The experiments are conducted in four different stages namely, acquisition, pre-processing, feature extraction and classification [3].

The algorithms employed in this thesis are Histogram of Oriented Gradients (HOG) [43], Local Binary Patterns (LBP) [44], and Binarized Statistical Image Features (BSIF) [45] to extract textural features from leaf images. These algorithms are considered as base algorithms in this thesis. Additionally, Scale Invariant Feature Transform (SIFT), Dense Scale Invariant Feature Transform (DSIFT) and Pyramid Histogram of Visual Words (PHOW) were used for comparison with the performance of the base algorithms.

Texture-based features were extracted to train a supervised machine learning model using leaf images infected with different diseases and healthy classes of leaves. Respective different set of leaves used in the training are used to test the system's performance to determine its accuracy. The system is tuned progressively until a satisfying performance or highest possible accuracy is achieved.

The rest of the thesis is organized as follows. The pre-processing employed and the architecture of plant disease detection system are explained in Chapter 2. In Chapter 3, feature extractors used are detailed and the proposed method is presented in Chapter 4. Chapter 5 is devoted to the experimental analysis. Finally, conclusion and deductions are presented in Chapter 6.

Chapter 2

PRE-PROCESSING AND ARCHITECTURE OF PLANT DISEASE DETECTION SYSTEM

2.1 Introduction

The approach of disease detection employed in this thesis has four principal steps namely acquisition, pre-processing, feature extraction and classification [23]. PlantVillage dataset is used in conducting the research. The database comprises of healthy and different infected plant leaf images of fruits and vegetables such as apple (*Malus pumila sp*), cherry (*Prunus sp*), corn (*Zea mays*), grape (*Vitis sp*), peach (*Prunus persica*), pepper (*Piper sp*), potato (*Solanum tuberosum*), and strawberry (*Fragaria ananassa*). As in most automatic plant disease detection techniques from subject images, acquisition, pre-processing, feature extraction and classification/matching are the sequential steps employed in automating disease detection. In general, classification of plant diseases involves two phases: training phase and testing phase. In the training phase, the images undergo the steps explained above to extract features to be used to create classes for each disease. During testing, an image is pre-processed and projected onto the trained classifier which will appropriately classify the leaf as healthy or to the category of the disease presented by the extracted leaf features. A general overview of a plant disease classification system is shown in Figure 2.1. These steps are described in the following subsections.

Figure 2.1: General overview of a leaf disease classification system.

2.2 Acquisition

This is an important step in the system's specification as the features to train the system are extracted from the acquired images. Having good set of images with clear distinct characteristics will make the extracted features distinctive from that of other classes which therefore increases the discriminating power of the classifier. Popular datasets used in plant disease detection include IPM dataset, PlantVillage dataset and APS dataset. Other sources of images include self-acquired images using hyperspectral imaging system. To obtain images with good characteristics such as consistent illumination and orientation, images are taken in a laboratory or using a sampling box [3].

2.3 Pre-processing

Different pre-processing techniques exist depending on the identification system to be used or created [4]. In this thesis, the pre-processing method employed is cropping the leaf images to further expose regions of interest of the leaf images, to remove unwanted background regions, remove less needed parts of the image and increase the localization of important features. Leaves from different plants come in different sizes and shapes, a standard will be needed that is appropriate to extract relevant features from the images, therefore resizing the images to a common scale is done.

Additionally, conversion of color images to grayscale for proper implementation of algorithms is necessarily implemented.

2.4 Feature Extraction

Features are inherent distinct characteristics extracted from an image which represent the image distinctiveness from other images and to represent the image content in a more concise and less dimensional form. Global and local characteristics of an image such as shape, spatial information, color, edges, contrast and entropy are extracted to form features of an image. Feature fusion is another technique where different significant features are combined to produce more robust features of an image [24]. Different feature extraction techniques and algorithms exist in image processing and the most prominent among local feature extractors are Histogram of Oriented Gradients (HOG) [43], Binarized Statistical Image Features (BSIF) [45], Local Binary Patterns (LBP) [44], Speeded-Up Robust Features (SURF), Scale Invariant Feature Transform (SIFT), Dense Scale Invariant Feature Transform (DSIFT) and Pyramid Histogram of Visual Words (PHOW). DSIFT is a variant of SIFT where key points are extracted from each pixel of an image and PHOW is built on DSIFT to extract features from an image in an increasing size of grid scales [3].

The aforementioned algorithms are used in this thesis as texture-based algorithms. They extract texture-based features from a grayscale image. These algorithms extract features which are distinct in characteristically representing the image they are extracted from. The extracted features can be improved by applying mathematical formulae and vector manipulations to enhance the features and make them suitable for solving different problems in computer vision and machine learning. The

extracted features are used to train a machine learning model that will be deployed for disease identification, recognition and classification upon its deployment.

2.5 Classification and Matching

Classification is the final stage of the plant disease recognition system. It is the assignment of a candidate leaf into the most appropriate category it belongs to based on the created model in the training phase. Different classes exist which represent each of the disease types involved in developing the system like common rust, northern leaf blight, cercospora, black rot, black measles, grape isariopsis and healthy class. Different machine learning techniques are used to implement the disease classification among which include Support Vector Machines (SVM), K-Nearest-Neighbour (K-NN), K-means Clustering, Linear Discriminant Analysis (LDA), Fuzzy Logic and Neural Networks. These classifiers fall into supervised, unsupervised or semi-supervised machine learning techniques [3]. In this thesis, Nearest Neighbour Classifier with Manhattan Distance measure is used to calculate the similarity between the trained model and test image. The value of the calculated distance determines the appropriate class the test image belongs to and consequently a test image is assigned to a class with the highest similarity. At the end of the classification process, a decision is taken that states the type of the disease in which the plant leaf image involves.

Chapter 3

FEATURE EXTRACTION

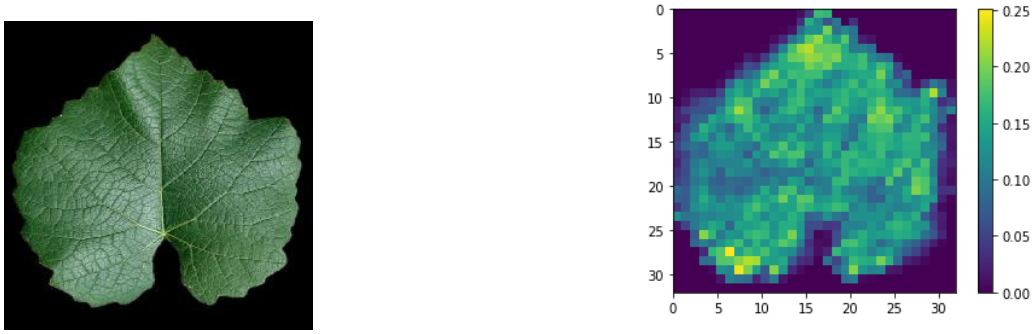
3.1 Introduction

This chapter presents the detailed mechanism of the algorithms used in this thesis. Three algorithms were used to extract features from images selected from PlantVillage database which were used in training and testing the proposed system. Histogram of Oriented Gradients (HOG) by Dala and Triggs [43], Local Binary Patterns (LBP) by Ahonen et al. [44] and Binarized Statistical Image Features (BSIF) by Kannala and Rahtu [45] were used to extract distinctive features from the datasets. Decision-level fusion strategy was used to incorporate the strengths of the individual feature extractors for a more robust system. The following sections give the mechanism employed by each of the employed algorithms.

3.2 Histogram of Oriented Gradients

In Histogram of Oriented Gradients method, local objects of an image are described by the distribution of local intensity gradients or edge directions of objects in an image. A dense grid of uniformly spaced cells of an image is formed where histogram of gradients orientation over the pixels of each cells are counted. The histograms are concatenated to form a feature descriptor after applying overlapping contrast normalization, which is the calculation of intensity of a larger region of cells called blocks and using the calculated value to normalize the individual cells within the image to improve the feature invariance to shadowing and illumination. These steps make the extracted descriptors more accurate in representing the local objects

in the image. The relevant features are extracted from sudden edges at fine scales [43]. Figure 3.1 shows a leaf image and its extracted HOG features.



(a) Input image (b) Computed features

Figure 3. 1: (a) An image sample and (b) its computed HOG features.

3.2.1 Histogram of Oriented Gradients Algorithm

The histogram of oriented gradients algorithm is described in the following steps.

Step 1: Gamma/color normalization: To normalize color and gamma values, the input image to the feature extractor should be an image of either RGB or LAB colored or grayscale. Comparable performance is achieved by any of the above image forms.

Step 2: Gradient computation: Gradients are calculated using Gaussian smoothing and discrete derivational mask of [-1, 0, 1]. For color images, gradient computation is done for all the color channels and the highest normalized value is taken to be the gradient of the whole image.

Step3: Spatial/Binning: Using orientation binning of 0^o – 180^o for unsigned weighted votes and 0^o – 360^o for signed, weighted votes from each pixels are collected over

spatial regions (‘cells’) for edge orientation histogram based on the orientation of the gradient element centred to it. The cells are either radial in shape or rectangular. This introduces nonlinearity to the image feature.

Step 4: Normalization of descriptor blocks: Normalization of the features is necessary due to variations in the computed gradients as a result of variation in illumination and contrast in the image. This is done by grouping cells together to form a larger block and contrast normalization of the individual cells improves the efficiency of the descriptors and makes the computed gradient even.

Step 5: The above steps produce the final and most accurate descriptors of the image which can be used to train a machine learning model for classification.

3.3 Local Binary Patterns

Due to different variations that can have a significant influence in computer vision tasks which will be insignificant to mammal recognition, Local Binary Patterns algorithm was developed to be resistant to these distortions. It extracts texture features, combines the features from each cell of the image which was initially divided. The extracted features are then concatenated (making it a global descriptor) to form a vector representing the image that can be used to train a recognition model. LBP works by thresholding a 3 by 3 neighbourhood pixels of an image with the centre value, a one value is obtained if the neighbour is greater than the centre pixel value and a zero value otherwise. It is improved by incorporating different number of neighbours in the circular neighbourhood represented as (P, R) where P is the number of points and R is the radius from the centre. Another improvement is the use of uniform patterns. Uniform patterns are formed when the binarized outputs contain

one or two transitions from the thresholding with the centre. The resulted histogram will contain edge information, spike spots, even areas as a representation of an image. Three levels (pixel, regional and regional histogram) of spatial information is concatenated to the pixel thresholded features for more robust representation [44]. LBP algorithm steps are described in the following subsection.

3.3.1 Local Binary Patterns Algorithm

Local Binary Patterns algorithm can be explained step by step as follows.

Step 1: The pixels of an image are used in the creation of the feature vector.

Step 2: Taking 3 by 3 neighbourhood pixels of an image, the middle value is thresholded and compared with the 8 neighbours surrounding it. Binary values are computed where a 1 is given if its value is greater than the thresholded middle value and a 0 otherwise.

Step3: As an improvement from **step 2**, the neighbourhood is taken to be circular with radius R and number of pixels P. Different number of neighbours P are taken using the range of the radius R for the LBP feature calculation. This is demonstrated on Figure 3.2.

Step 4: Further development is the uniform pattern where extracted binaries are uniform if they have at most a bit transition from 0 to 1 or vice versa which is called a circular pattern. A pictorial representation is shown on Figure 3.2 (a) and (b), where a white circle represents a value smaller than the middle pixel value and a dark circle represents a value greater than the middle pixel value. Whenever the neighbouring pixel values are all greater or smaller than the middle thresholded

value, those patterns are featureless, that is the image in that particular position does not contain any relevant information. Figure 3.3 shows an image pixel values and its LBP binarized output.

An edge is represented by LBP as a sudden change from set of initial pixel values which is known as 'uniform' patterns. Figure 3.4 shows LBP representation of detected edge. Figure 3.5 shows a gray image sample before LBP features are extracted. The extracted features from the sample image are shown on Figure 3.5.

The extracted features are fed into a machine learning algorithm for training to be used for classification. Simple machine learning models like k-nearest neighbour gives a good result.

Local Binary Patterns are calculated as follows:

$$LBP_t(x,y) = \sum_{n=1}^{N} LBP_t^n\,(x,y) \times f^n, \forall t \in [1,c] \qquad (3.1)$$

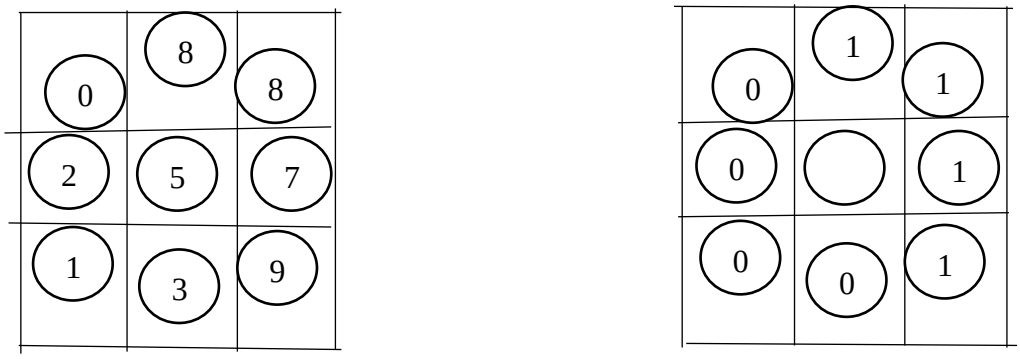
where

$$LBP_t^n\,(x,y) = \left\{ \begin{matrix} 1, & l_t^n\,(x,y) \geq l_t\,(x,y) \\ 0, & otherwise \end{matrix} \right\} \qquad (3.2)$$

and f^n is a weighing function defined as

$$f^n = (2)^{n-1}, \forall n \in [1,N] \qquad (3.3)$$

Figure 3.2: Uniform LBP (R, P) = 1,8 with two different output patterns



(a) (b)

Figure 3.3: (a) A thresholded middle value (b) LBP binarized output

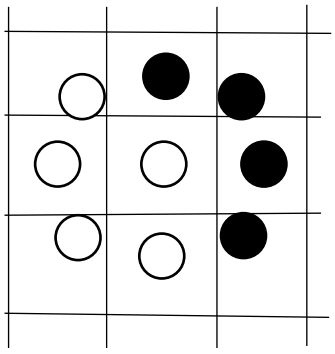


Figure 3.4: A detected edge by LBP

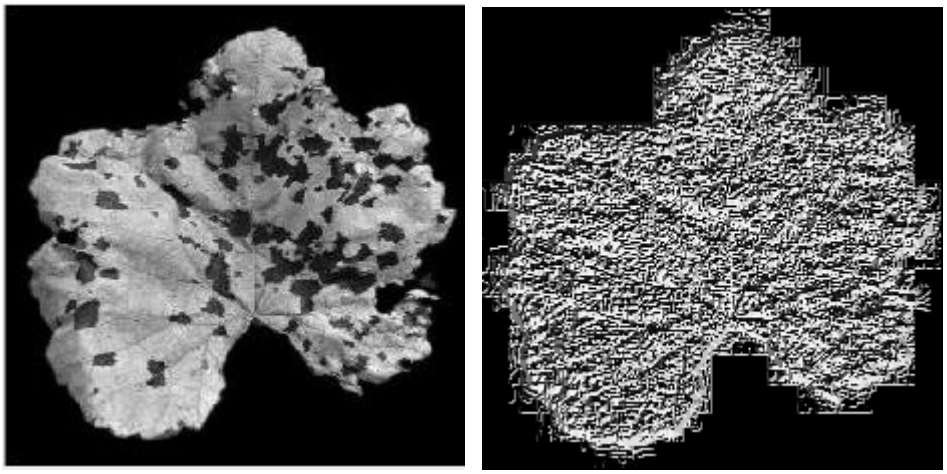


Figure 3.5: Grayscale Image and its extracted Local Binary Patterns

3.4 Binarized Statistical Image Features

Binarized Statistical Image Features is an image descriptor that is created from texture characteristics of an image. The method is based on using set of filters learned from natural images by Independent Component Analysis. The pixels in each neighbourhood of an image give output of binary strings. Different filters give different set of code. The output code is binarized and used as descriptors of an image.

BSIF is an improvement of LBP and local based quantization. In this method, image is represented as histogram of pixel's binary code. The code is generated by component analysis and binarizing the coordinates in the image by thresholding like in LBP. In BSIF, filters generated from small number of images are used to binarize the pixels neighbours of an image, different filters are used for different application of the algorithm and they produce different length of bit string features. The binarized output codes represent the texture features of the candidate image produced by the intensity pattern of its neighbours. A feature vector representation of an image is obtained by dividing the image into 8 by 8 segments. The descriptors are computed in each region and concatenated to produce a global descriptor of the image. The

features can be used to train a machine learning model for identification and recognition [45]. BSIF algorithm is explained step by step in the following subsection.

3.4.1 Binarized Statistical Image Features Algorithm

Binarized Statistical Image Features algorithm steps are described below:

Step1: Given an image patch X of size lxl pixels and a linear filter w_i of the same size, b_i as the binarized feature, the filter response s_i is obtained from

$$s_i = \sum_{u,v} w_i(u,v) \times (u,v) = w_i^t x \qquad (3.4)$$

where w_i and x_i are vectors, the features b_i in binary form are obtained if $s_i > 0$ equates 1 and 0 otherwise. Given number of linear filters N, they can be used all at the same time by concatenating them to form a matrix of size $N \times l^2$ to compute all the different descriptors by binarizing output of equation (3.4) from the different filters.

To obtain a high performing set of filters, statistical independence of equation (3.4) output is maintained by applying image restoration [21].

Step 2: To reduce the dimension of the filters, whitening of the training filters are done using Principal Component Analysis to determine the principal components of the natural image patches, where the principal components are taken and divided by their standard deviation to obtain whitened data samples.

Step 3: To obtain the final filter, standard Independent Component Analysis algorithm is applied on the whitened data samples to obtain the orthogonal matrix which is used as the filter for BSIF descriptors extraction.

Step4: Feature descriptors are extracted from an image using the developed filters in step 1, 2 and 3 for texture extraction.

Chapter 4

PROPOSED METHOD

The proposed method employed in this thesis principally includes pre-processing, feature extraction and finally matching and classification. The training and test images are used to implement the proposed system including images of each type of disease and healthy class. Figure 4.1 shows a graphical depiction of the sequential steps. Detailed mechanism of the adopted stages is given below.

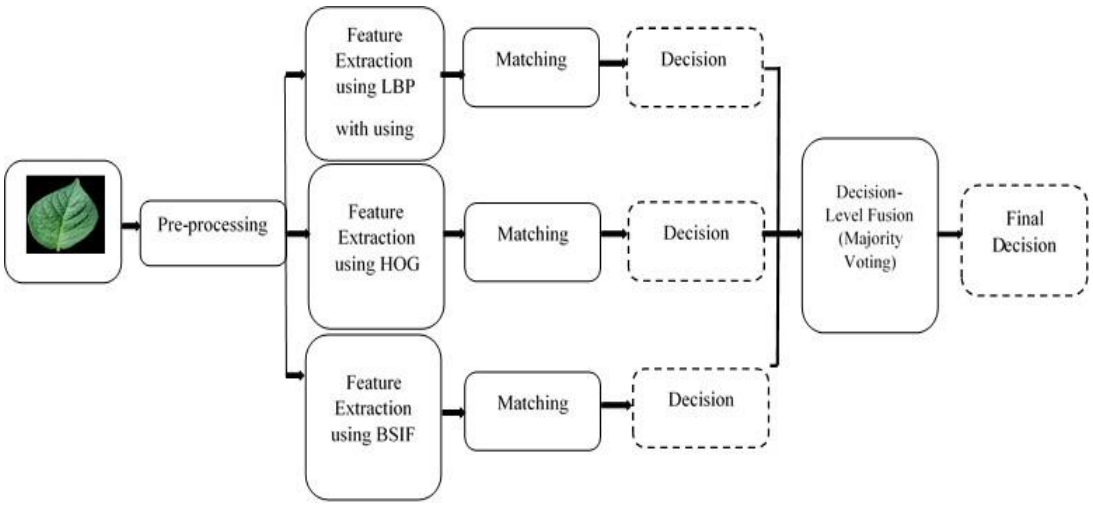


Figure 4.1: Block Diagram of the proposed method

4.1 Pre-processing Stage of the Proposed Method

Diseases on plants come in different forms, shapes and colors. Some disease do manifest on the whole leaf surface while others are spots of few radius. To extract features using the employed algorithms, disease regions need to be exposed further for effective disease symptoms extraction as features. Therefore cropping is used to

expose the disease regions and the images are resized to a proper scale as prerequisite for proper algorithmic implementations. Converting the images into grayscale to improve the performance of texture-based algorithms and speed in extracting features is also implemented. The system implements the steps on both training and test images.

4.2 Feature Extraction Stage of the Proposed Method

Three sets of features are extracted from the processed leaf images using HOG, LBP and BSIF texture-based feature extraction methods. HOG extracts histogram orientation gradients in X and Y direction from leaf image pixels while LBP extracts binary patterns by dividing leaf images into equal number of cells. Features are then extracted from pixel values and further converted to binary values to reduce dimensions and to simplify further processing without losing features distinctiveness. BSIF extracts texture features from an image using learned filters created from natural images using Independent Component Analysis. In this study, all algorithms are used to extract disease symptoms from leaf images.

4.3 Matching Stage of the Proposed Method

At the matching stage, the extracted features are used to create a system model that assigns test images to the appropriate class they belong to using the learned model. Different and varying number of classes exist depending on the number of diseases in a dataset in addition with the healthy class. Nearest Neighbour classifier is used as a machine learning technique with Manhattan Distance measurement due to its simplicity in implementation. Similarity measurement is taken between the trained model and test image. The distance output from the classifier determines the class of the test image based on the extracted features from the leaf objects in the image.

Finally, a test image is placed either in any of the diseased class or in the healthy class.

4.4 Decision-Level Fusion

The outcomes from each of the employed algorithms are taken and fused together using Decision-level fusion with Majority Voting method to complement each algorithm as implemented by Usman et al. in [24]. Majority voting method takes the decisions of each technique and concludes the final decision of the system according to the majority of the decisions. With this, a robust, effective and accurate system is formed with high reliability.

Chapter 5

EXPERIMENTS AND RESULTS

5.1 Introduction

This chapter gives the implementation details of the algorithms employed in this thesis. It starts from setting up the datasets for training the system to how features were generated by each algorithm and finally presents the performance evaluation of the proposed system.

5.2 Experimental Setup

The experimental setup used in the thesis comprises of 3200 images from nine different plant species. The dataset is divided into distinct groups for each plant species namely, datasets A, B, C, D, E, F, G, H, and dataset I for apple, grape, corn, potato, cherry, peach, pepper, strawberry and tomato respectively. Apple, corn and grape have three unhealthy classes each, accordingly they are represented as A1, A2, A3, B1, B2, B3, C1, C2 and C3 for each of the unhealthy class. Potato has two unhealthy classes, it is therefore divided into datasets D1 and D2 while cherry, peach, pepper and strawberry have single unhealthy class each ; E1, F1, G1, H1. From each of the unhealthy classes, fifty training images were taken making it a total of 750 training images and 750 for testing images from each respective dataset. Fifty images were used for each plant species as healthy datasets represented as A4, B4, C4, D3, E2, F2, G2, H2 summing up to 400 healthy images for training and also 400 for testing. Tomato dataset (I) has 9 unhealthy classes with one healthy class each

having 50 images in each class summing up to 1,000 training and test images. The breakdown is shown in Tables 5.1 through 5.9 for each of the datasets.

Table 5.1: Apple dataset: Number of train and test images for apple dataset

Apple datasets (A)	Number of training images	Number of test images
Apple scab (A1)	50	50
Black rot (A2)	50	50
Cedar apple rust (A4)	50	50
Healthy (A4)	50	50
Total	200	200

Table 5.2: Grape dataset: Number of train and test images for grape dataset

Grape datasets (B)	Number of training images	Number of test images
Grape black rot (B1)	50	50
Esca (Black measles) (B2)	50	50
Grape leaf blight (lsariopsis leaf spot) (B3)	50	50
Healthy (B4)	50	50
Total	200	200

Table 5.3: Corn dataset: Number of train and test images for corn dataset

Corn datasets (C)	Number of training images	Number of test images
Maize common rust (C1)	50	50
Maize northern leaf blight (C2)	50	50
Cercospora leaf spot (C3)	50	50
Healthy (4)	50	50
Total	200	200

Table 5.4: Potato dataset: Number of train and test images for potato dataset

Potato datasets (D)	Number of training images	Number of test images
Early blight (D1)	50	50
Late blight (D2)	50	50
Healthy (D3)	50	50
Total	150	150

Table 5.5: Cherry dataset: Number of train and test images for cherry dataset

Cherry datasets (E)	Number of training images	Number of test images
Powdery mildew (E1)	50	50
Healthy (E2)	50	50
Total	100	100

Table 5.6: Peach dataset: Number of train and test images for peach dataset

Peach datasets (F)	Number of training images	Number of test images
Bacterial spot (F1)	50	50
Healthy (F2)	50	50
Total	100	100

Table 5.7: Pepper dataset: Number of train and test images for pepper dataset

Pepper datasets (G)	Number of training images	Number of test images
Bacterial spot (G1)	50	50
Healthy (G2)	50	50
Total	100	100

Table 5.8: Strawberry dataset: Number of train and test images for strawberry dataset

Strawberry datasets (H)	Number of training images	Number of test images
Leaf scorch (H1)	50	50
Healthy (H2)	50	50
Total	100	100

Table 5.9: Tomato dataset: Number of train and test images for tomato dataset

Tomato datasets (I)	Number of training images	Number of test images
Bacterial spot (I1)	50	50
Early blight (I2)	50	50
Late blight (I3)	50	50
Leaf mold (I4)	50	50
Septoria leaf spot (I5)	50	50
Spider mites (I6)	50	50
Target spot (I7)	50	50
Tomato mosaic (I8)	50	50
Yellow leaf curl virus (I9)	50	50
Healthy (I10)	50	50
Total	500	500

Three texture-based feature extraction algorithms are implemented on the aforementioned datasets namely HOG, LBP and BSIF. HOG is applied on each training dataset to extract its features as follows: an image is converted to grayscale and HOG algorithm is iterated over the blocks, cells and pixel values of the image. At the pixels, gradients are taken from the pixel values in X and Y directions, using a Gaussian filter with sigma = 0.5 and at angles between the range of 0^0 and 180^0. HOG descriptors are computed over each pixel of an image and binned using Bilinear Interpolation. Features from each block are normalized using L1-norm and the whole feature vector of the image is normalized with L2 norm.

In LBP, the images are first converted to grayscale. Both the training and test images are divided into cells of 48 by 48 pixels. Each block is passed to the LBP algorithm to compute the descriptors using circular symmetric pattern over each pixel using (P, R) approach. The neighbours used in these experiments are P = 8 for pixels and R = 1 for radius, i.e $LBP_{8,1}^{u,2}$ is used in calculating the LBP descriptors. The length of features obtained from the algorithm depends on the size of each cell from which the LBP approach is applied to compute its features. The features from each block are

concatenated to form a single feature vector of the whole cells in an image. In this research, different cell sizes were tried and the best performance was obtained with the cell size of 48 by 48 pixels of the image.

In BSIF, using the pre-learned filters as described in the algorithm, the experimental images are first converted to grayscale and a 7 by 7 filter of 12 bits is used to generate the code descriptors from the image. An image is slid over the 7 by 7 filter to generate the descriptors from each cell to give feature vector of size 1 by 4096 for each image.

The experiments were undertaken using a Windows 7 professional OS with 3GB RAM and Pentium dual core CPU @ 2.00GHz.

5.3 Experimental Results

This section gives the experimental approach taken in developing the proposed system and the reasons the approach is taken.

5.3.1 Preliminary Experiments

Six algorithms are implemented to analyse their accuracy on plant disease classification. The experiments were conducted using Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), Binarized Statistical Image Features (BSIF), Scale Invariant Feature Transform (SIFT), Dense Scale Invariant Feature Transform (DSIFT) and Pyramid Histogram of Visual Words (PHOW) [41]. The base algorithms in this thesis, namely Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Binarized Statistical Image Features (BSIF) are tested on each dataset and performed better in classifying the leaf images to the appropriate class, thereby achieving higher accuracy than Scale Invariant Feature

Transform (SIFT), Dense Scale Invariant Feature Transform (DSIFT) and Pyramid Histogram of Visual Words (PHOW) except in peach and strawberry datasets where DSIFT and SIFT achieved higher accuracy than any of the first three algorithms. The results are shown on Table 5.10.

Finally, the systems' accuracy is calculated using equation (5.1) below to access its performance for each dataset.

$$Accuracy\ (\%) = \frac{Number\ of\ test\ images\ correctly\ classified}{Total\ number\ of\ test\ images} \times 100 \quad (5.1)$$

Table 5.10: Classification results (%) using all methods

Dataset	**HOG**	**LBP**	**BSIF**	**SIFT**	**DSIFT**	**PHOW**
Dataset A5 (Apple)	59	73.5	68	57	44	42
Dataset B5 (Grape)	79.5	71.5	89	72.5	68.5	65.5
Dataset C5 (Corn)	70	72.5	76.5	56	63	67
Dataset D4 (Potato)	54.67	59.3	70.67	64	54.67	54.67
Dataset E3 (Cherry)	97	98	100	99	98	98
Dataset F3 (Peach)	71	73	97	88	90	82
Dataset G3 (Pepper)	58	67	86	76	57	65
Dataset H3 (Strawberry)	58	72	86	88	78	65
Dataset I11 (Tomato)	38.6	50.4	60.2	31	34	32.4

Another factor considered is the computation time. The computation times for each method employed are calculated which aims at selecting methods with low computation time coupled with high accuracy. Low computation time is an important factor in all automation processes and a system with low computation time will give almost an instant result when deployed. Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) and Binarized Statistical Image Features (BSIF) have

the lower computation time in both feature extraction and testing than the other three algorithms namely Scale Invariant Feature Transform (SIFT), Dense Scale Invariant Feature Transform (DSIFT) and Pyramid Histogram of Visual Words (PHOW). The computation time for training and testing are calculated for each of the six methods and the results are demonstrated on Table 5.11.

Table 5.11: Computation times of all method

Algorithm	**Training time(sec)**	**Single Image Test Time (sec)**
Proposed method	16.499	0.184
HOG	5.267	0.062
LBP	2.975	0.049
BSIF	8.257	0.075
SIFT	22.315	0.121
DSIFT	143.727	0.661
PHOW	117.258	0.614

According to the above preliminaries, we decided to choose HOG, LBP and BSIF for the proposed method and for further experiments on plant disease classification. These methods achieve good results compared to the other methods and their computation times are less than the computation time of the other three methods namely SIFT, DSIFT and PHOW.

5.3.2 Experiments with Proposed Method

The experiments were conducted using HOG, LBP, BSIF and the proposed method using datasets described above under experimental setup. For the apple dataset A5, the proposed method achieved the highest accuracy on all the dataset than the other algorithms except on cedar apple rust with 88% accuracy by LBP's and the proposed method having 86% accuracy. The proposed method has a total highest accuracy of 76.8% for all the classes. On grape dataset, the proposed method obtained highest accuracy in two datasets, while HOG and BSIF obtained highest accuracy in a single

class each and 78.6% accuracy is achieved totally by the proposed method which is higher than the other algorithms. On corn dataset, BSIF achieved the highest classification in two of the three classes while the proposed method and HOG achieved the highest accuracies in one class each. Totally, the proposed method achieved the highest classification of 78.5%. On potato dataset, BSIF achieved the highest accuracies in the two of the three datasets, totalling 70.67% followed by the proposed method with 62% accuracy. Cherry dataset is accurately classified by both the proposed method and BSIF without a single misclassification. On peach and pepper dataset, BSIF achieved the highest classification in the two classes of each dataset. On strawberry dataset, leaf scorch class is 100% recognized by the proposed method while BSIF has the highest total recognition of 88% and 86% accuracy for the proposed method. On tomato dataset with 10 classes, HOG and LBP performed poorly, thereby dragging down the performance of the proposed method. The influence made BSIF overtake the proposed method recognition, nevertheless, a comparable performance is achieved by BSIF and the proposed method achieved 60.2% accuracy and 57% accuracy, respectively. The results are presented on Tables 5.12 through 5.20.

Table 5.12: Classification result (%) for apple dataset

	Method			
Apple Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset A1(Apple scab)	42	50	68	60
Dataset A2(Black rot)	56	80	74	82
Dataset A3(Cedar apple rust)	80	88	84	86
Dataset A4(Healthy)	58	76	46	78
Dataset A5(Total)	59	73.5	68	76.5

Table 5.13: Classification result (%) for grape dataset

	Method			
Grape Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset B1 (Grape black rot)	52	56	78	74
Dataset B2(Esca (Black measles))	100	66	90	96
Dataset B3(Grape leaf blight (lsariopsis leaf spot))	78	86	90	92
Dataset B4(Healthy)	88	78	98	98
Dataset B5 (Total)	79.5	71.5	89	90

Table 5.14: Classification result (%) for corn dataset

	Method			
Corn Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset C1 (Maize common rust)	82	82	90	88
Dataset C2 (Maize northern leaf blight)	40	60	68	62
Dataset C3 (Cercospora leaf spot)	58	64	70	70
Dataset C4 (Healthy)	100	84	78	94
Dataset C5 (Total)	70	72.5	76.5	78.5

Table 5.15: Classification result (%) for potato dataset

	Method			
Potato Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset D1 (Early blight)	44	34	78	44
Dataset D2 (Late blight)	50	62	66	64
Dataset D3 (Healthy)	70	82	68	78
Dataset D4 (Total)	54.67	59.3	70.67	62

Table 5.16: Classification result (%) for cherry dataset

	Method			
Cherry Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset E1 (Powdery mildew)	98	98	100	100
Dataset E2 (Healthy)	96	98	100	100
Dataset E3 (Total)	97	98	100	100

Table 5.17: Classification result (%) for peach dataset

	Method			
Peach Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset F1 (Bacterial spot)	58	58	98	64
Dataset F2 (Healthy)	84	88	96	90
Dataset F3 (Total)	71	73	97	77

Table 5.18: Classification result (%) for pepper dataset

	Method			
Pepper Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset G1 (Bacterial spot)	48	64	94	82
Dataset G2 (Healthy)	68	70	78	92
Dataset G3 (Total)	58	67	86	87

Table 5.19: Classification result (%) for strawberry dataset

	Method			
Strawberry Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset H3 (Leaf scorch)	22	58	88	100
Dataset H3 (Healthy)	94	86	84	72
Dataset H3 (Total)	58	72	86	86

Table 5.20: Classification result (%) for tomato dataset

	Method			
Tomato Leaf Dataset	**HOG**	**LBP**	**BSIF**	**Proposed method**
Dataset I1 (Bacterial spot)	44	32	62	58
Dataset I2 (Early blight)	16	20	38	38
Dataset I3 (Late blight)	18	36	42	36
Dataset I4 (Leaf mold)	30	30	44	38
Dataset I5 (Septoria leaf spot)	30	44	68	54
Dataset I6 (Spider mites)	38	72	52	66
Dataset I7 (Target spot)	16	26	34	22
Dataset I8 (Tomato mosaic)	60	50	66	60
Dataset I9 (Yellow leaf curl virus)	36	94	100	98
Dataset I10 (Healthy)	98	100	96	100
Dataset I11 (Total)	38.6	50.4	60.2	60.2

Figures 5.1 to 5.9 give the confusion matrices for all the test datasets used in testing the proposed model. Figure 5.1 is the confusion matrix of apple dataset which has the highest correct classification for Cedar rust disease. Figure 5.2 shows confusion matrix for grape where black rot and healthy class have identification and classification above 90% each. Figure 5.3 presents the confusion matrix of corn dataset where healthy corn leaf has the highest classification accuracy. Figure 5.4 presents confusion matrix for Potato dataset with three classes, the highest classification accuracy is on early blight class. The leaves in cherry dataset are all correctly classified this is shown on figure 5.5. Bacterial spot of peach and pepper datasets have the highest accuracy due to the disease distinctiveness from a healthy leaf. Figure 5.6 and 5.7 shows the respective confusion matrices. Strawberry dataset has high accurate recognition on healthy leaves with no misplacement into the class,

although not all were recognized, figure 5.8 shows the confusion matrix. For tomato dataset, yellow curl disease class has a 94.2% accuracy, followed by tomato mosaic and tomato target spot with 78.9% and 78.6% respectively. Bacterial spot, early blight and late blight accuracy value fell below 50%. This is as a result of imprecision in the distinctiveness of the diseases from other diseases within the dataset. The configure matrix is shown on figure 5.9. Overall, the true classes have the highest accuracy value.

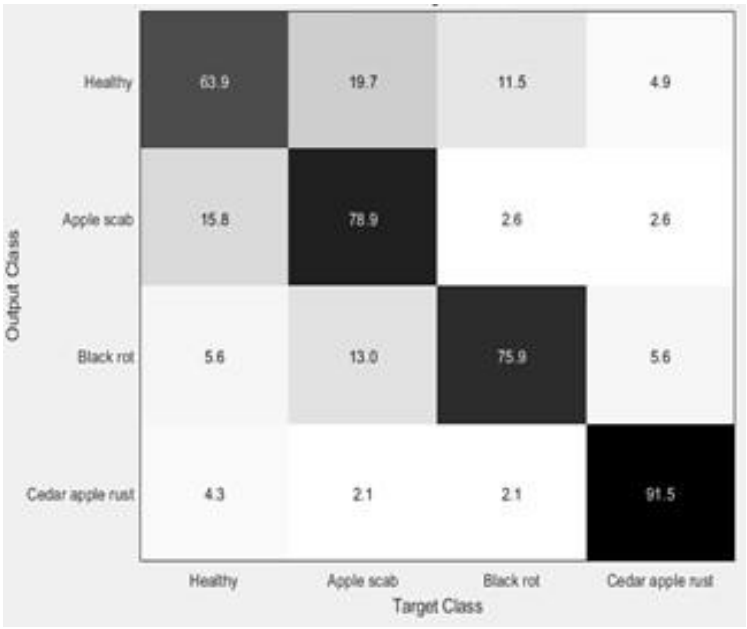


Figure 5.1 : Confusion matrix for dataset A5 (Apple)

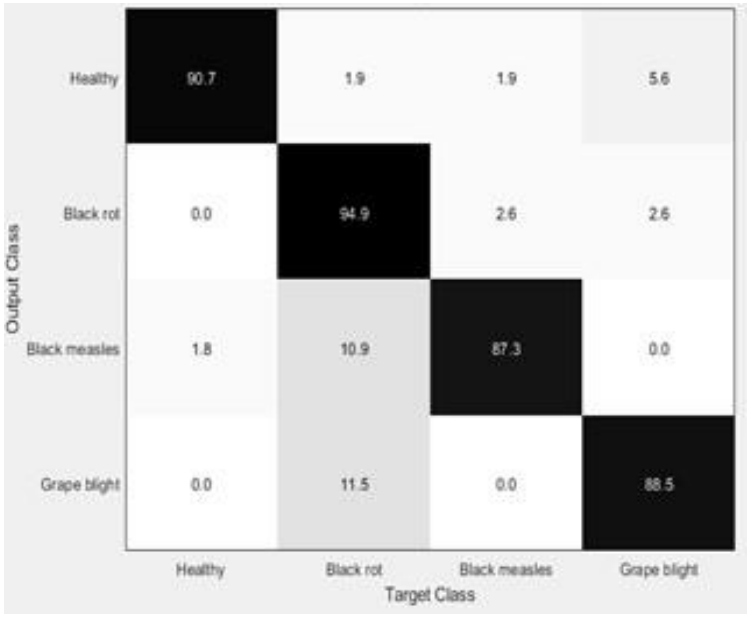


Figure 5.2 : Confusion matrix for dataset B5 (Grape)

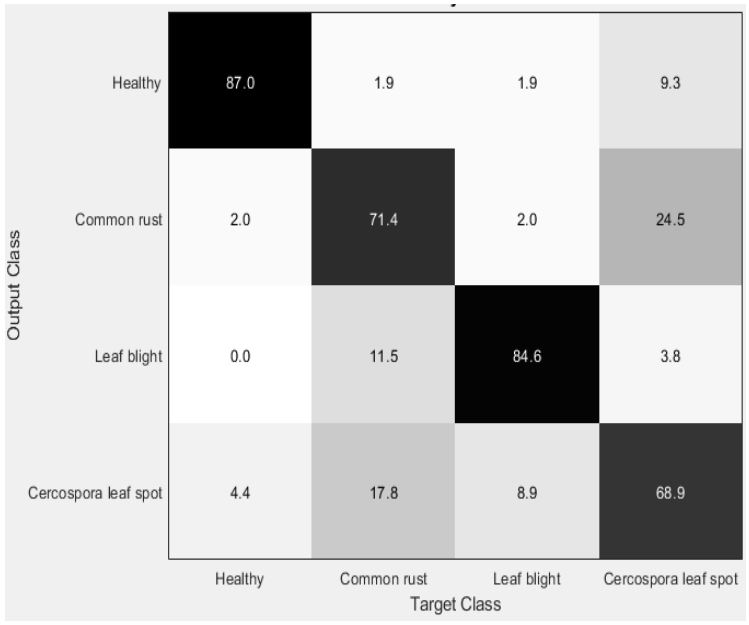


Figure 5.3 : Confusion matrix for dataset C5 (Corn)

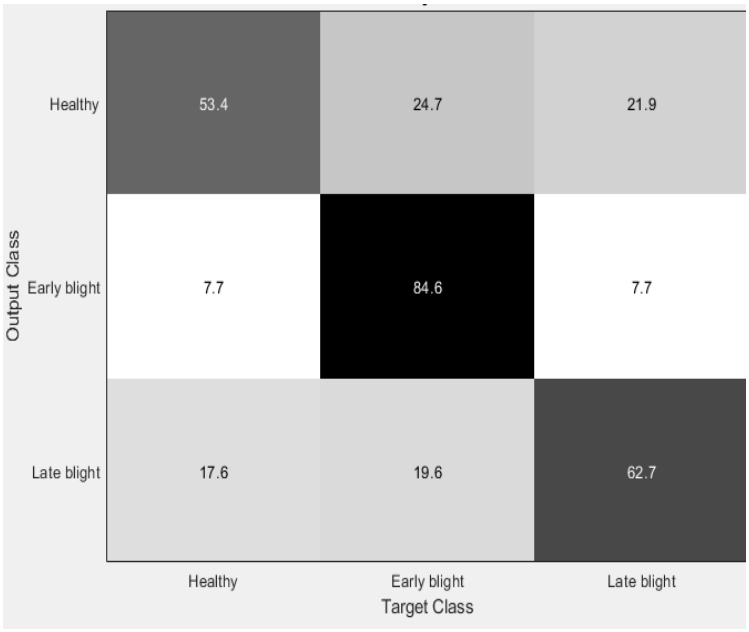


Figure 5.4 : Confusion matrix for dataset D4 (Potato)

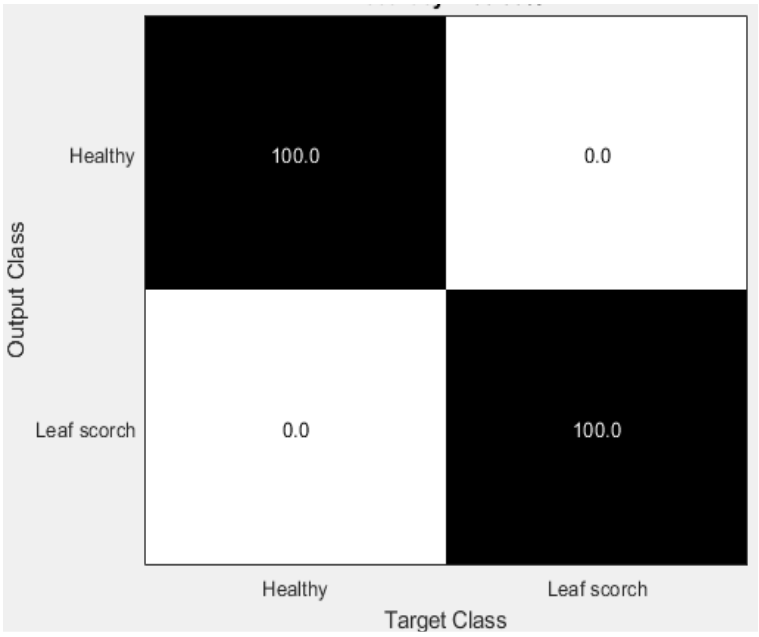


Figure 5.5 : Confusion matrix dataset E3 (Cherry)

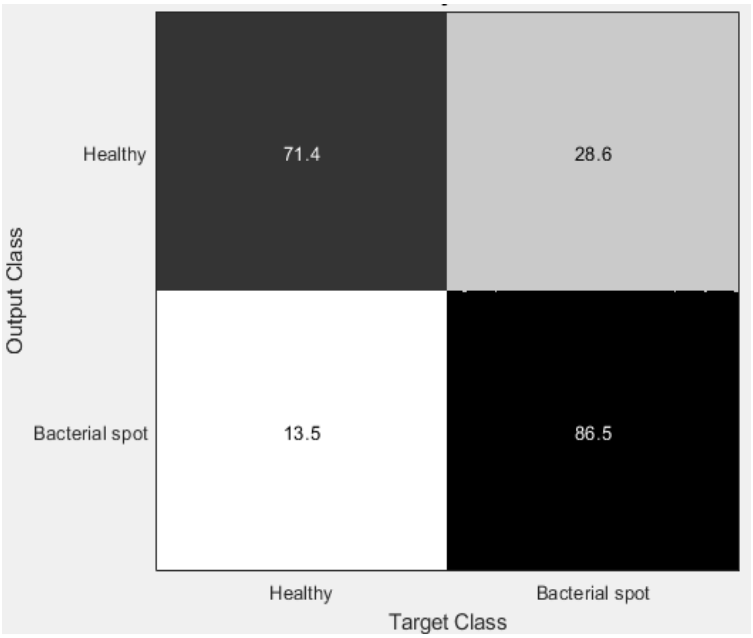


Figure 5.7 : Confusion matrix for dataset F3 (Peach)

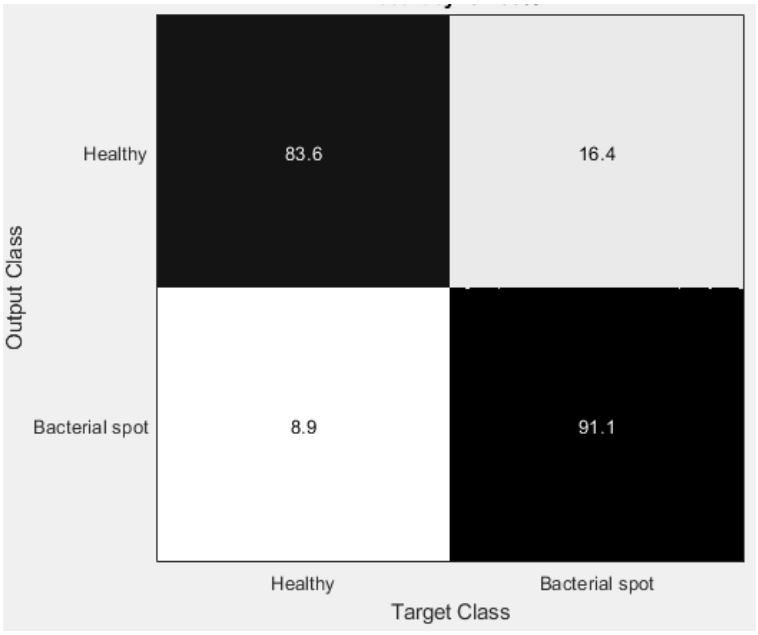


Figure 5.7 : Confusion matrix for dataset G3 (Pepper)

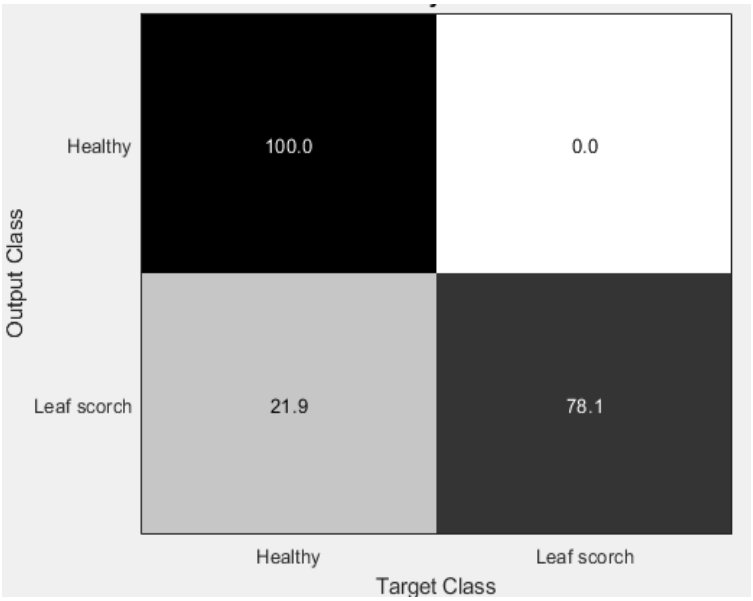


Figure 5.8 : Confusion matrix for dataset H3 (Strawberry)

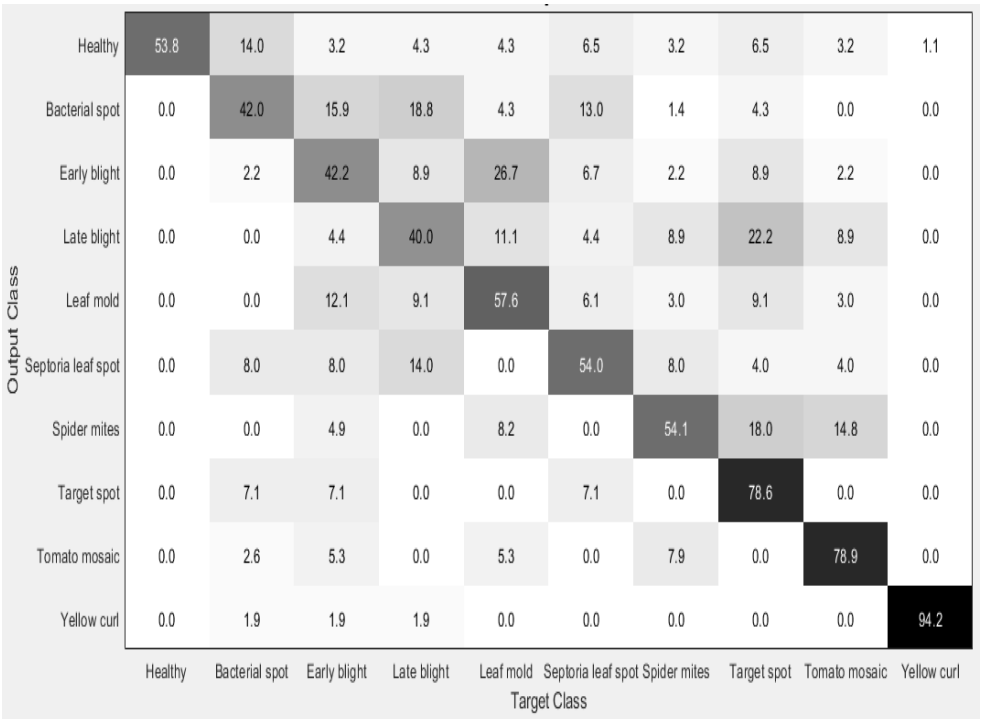


Figure 5.9 : Confusion matrix for dataset I11 (Tomato)

The histogram below give us the overall classification accuracy of the proposed system on all the test datasets in Figure 5.10. Dataset E3 of cherry and dataset C5 of grape have 100% and 90 % accuracies, while Dataset I11 of tomato with 10 classes has an accuracy of 57% which is a good achievement of the proposed system despite high number of classes.

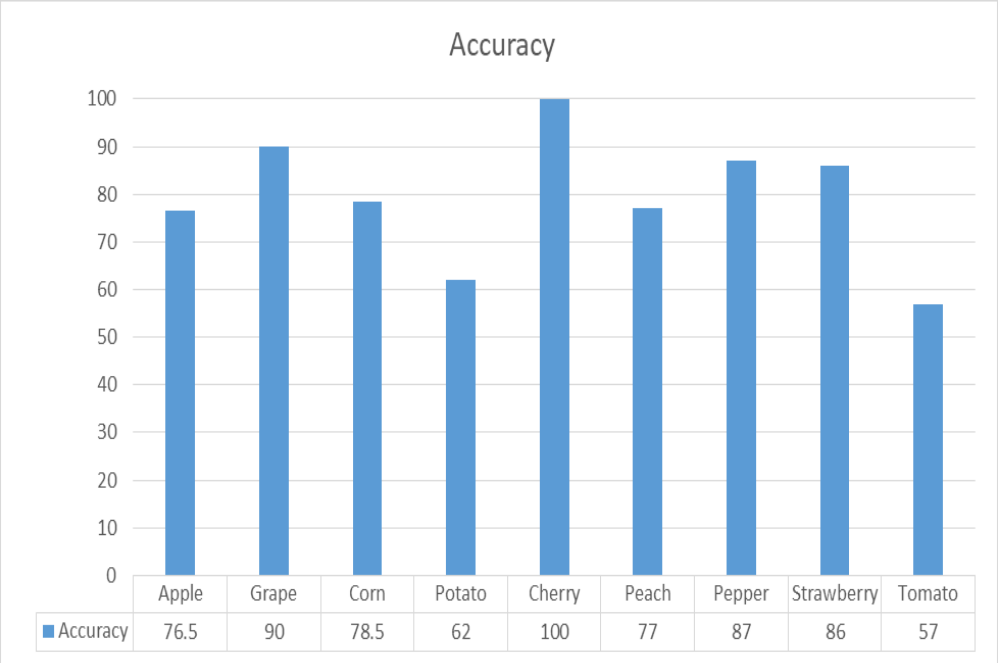


Figure 5.10: Accuracies for all datasets

5.4 Discussion on Experimental Results

The experimental results indicate that the proposed method achieved the highest accuracies in classifying plant diseases with a reliable final decision better than the individual methods employed.

From the results presented above, different algorithms performed relatively better on different diseases presented on the leaves. The physical features of the leaves, such as shape, has influence on the disease recognition. This effect is seen clearly on

datasets with large surface area like in cherry and grape datasets. Higher recognition rates are obtained on datasets with large surface leaf images. The algorithms' performance also varies within the same dataset. For a more robust system, decision-based fusion is used to improve the whole systems' performance as the system votes for the most probable decision from each algorithm. This improved the reliability of the system significantly as shown in the above tables.

Chapter 6

CONCLUSION

In this thesis, texture-based feature extraction algorithms were used on plant leaf images exhibiting various disease symptoms to extract features to develop a plant disease recognition and classification system. Histogram of Oriented Gradients, Local Binary Patterns and Binarized Statistical Image Features were used as feature extractors to extract features from the leaf images. A new method is proposed using decision-level fusion which has proved to be more accurate than the other algorithms used. Scale Invariant Feature Transform, Dense Scale Invariant Feature Transform and Pyramid Histogram of Visual Words performances were also compared with the aforementioned algorithms and the proposed method, where the proposed method appears to be performing better than the other algorithms in accurate disease detection.

The system is ready to be deployed in a field for plant disease detection and using it will make disease recognition easier, reduce over dependency on inadequate experts, reduce plant produce loss, help in early disease recognition for proper action to be taken and it's a proof to the ubiquity of computer vision in different fields of human life. For further research, more plant diseases should be incorporated with a more comprehensive system. A portable plant disease detection system for smart devices is also encouraged.

REFERENCES

[1] Goal 2: Zero Hunger - United Nations Sustainable Development. (n.d.). Retrieved from https://www.un.org/sustainabledevelopment/hunger/

[2] Moore, D., Robson, G. D., & Trinci, A. P. (2018, July 01). 21st Century Guidebook to Fungi, SECOND EDITION, by David Moore, Geoffrey D. Robson and Anthony P. J. Trinci. Retrieved from http://www.davidmoore.org.uk/21st_Century_Guidebook_to_Fungi_PLATINUM/Ch14_01.htm

[3] Kaur, S., Pandey, S., & Goel, S. (2018). Plants Disease Identification and Classification Through Leaf Images: A Survey. *Archives of Computational Methods in Engineering*,*26*(2), 507-530. doi:10.1007/s11831-018-9255-6.

[4] Yan-li, A. O. (2015). Introduction to digital image pre-processing and segmentation. *Seventh International Conference on Measuring Technology and Mechatronics Automation*, pp.588-593.

[5] Zavatta, G., Zavatta, G., Perrone, T., Figus, C., Perrone, T., & Figus, C. (2014, October 27). Agriculture Remains Central to the World Economy. 60% of the Population Depends on Agriculture for Survival. Retrieved from http://www.expo2015.org/magazine/en/economy/agriculture-remains-central-to-the-world-economy.html

[6] Singh, V. and Misra, A. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1), pp.41-49.

[7]Barbedo, J. G., Koenigkan, L. V., & Santos, T. T. (2016). Identifying multiple plant diseases using digital image processing. *Biosystems Engineering,147*, 104-116. doi:10.1016/j.biosystemseng.2016.03.012

[8] Cooperative Extension Service University of Kentucky College of Agriculture, Food and Environment, Lexington, KY, 40546 PPA-46 Plant Diseases

[9] Cash crops and food security : contributions to income, livelihood risk and agricultural innovation. January 2014, Project: Food security, Thom Achterbosch, Siemen Van Berkum, Gerdien W. Meijerink, D. Oudendag

[10] Camargo, A., & Smith, J. (2009). Image pattern classification for the identification of disease causing agents in plants. *Computers and Electronics in Agriculture,66*(2), 121-125. doi:10.1016/j.compag.2009.01.003

[11] Pires, R. D., Gonçalves, D. N., Oruê, J. P., Kanashiro, W. E., Rodrigues, J. F., Machado, B. B., & Gonçalves, W. N. (2016). Local descriptors for soybean disease recognition. *Computers and Electronics in Agriculture,125*, 48-55. doi:10.1016/j.compag.2016.04.032

[12] Prajapati, B. S., Dabhi, V. K., & Prajapati, H. B. (2016). A survey on detection and classification of cotton leaf diseases. *2016 International Conference on*

Electrical, Electronics, and Optimization Techniques (ICEEOT). doi:10.1109/iceeot.2016.7755143

[13] Wang, H., Li, G., Ma, Z. and Li, X. (2012). Application of neural networks to image recognition of plant diseases. 2012 International Conference on Systems and Informatics (ICSAI2012).

[14] Wijekoon, C., Goodwin, P., & Hsiang, T. (2008). Quantifying fungal infection of plant leaves by digital image analysis using Scion Image software. *Journal of Microbiological Methods,74*(2-3), 94-101. doi:10.1016/j.mimet.2008.03.008

[15] Barbedo, J. G. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus,2(1). doi:10.1186/2193-1801-2-660

[16] Xu, G., Zhang, F., Shah, S. G., Ye, Y., & Mao, H. (2011). Use of leaf color images to identify nitrogen and potassium deficient tomatoes. *Pattern Recognition Letters,32*(11), 1584-1590. doi:10.1016/j.patrec.2011.04.020

[17] Romualdo, L., Luz, P., Devechio, F., Marin, M., Zúñiga, A., Bruno, O., & Herling, V. (2014). Use of artificial vision techniques for diagnostic of nitrogen nutritional status in maize plants. *Computers and Electronics in Agriculture,104*, 63-70. doi:10.1016/j.compag.2014.03.009

[18] Ilic, M., Ilic, S., Jovic, S., & Panic, S. (2018). Early cherry fruit pathogen disease detection based on data mining prediction. *Computers and Electronics in Agriculture,150*, 418-425. doi:10.1016/j.compag.2018.05.008

[19] Lloret, J., Bosch, I., Sendra, S., & Serrano, A. (2011). A Wireless Sensor Network for Vineyard Monitoring That Uses Image Processing. *Sensors,11*(6), 6165-6196. doi:10.3390/s110606165

[20] Pires RDL, Gonc¸alves DN, Orueˆ JPM, Kanashiro WES, Rodrigues JF, Machado BB, Gonc¸alves WN (2016) Local descriptors for soybean disease recognition. Computer Electron Agric 125:48–55

[21]. Wen, Z., Fraser, D., & Lambert, A. (2009). Image Restoration Using Natural Image Statistics. *Frontiers in Optics 2009/Laser Science XXV/Fall 2009 OSA Optics & Photonics Technical Digest*. doi:10.1364/srs.2009.stuc7.

[22] R.C. Gonzales, R.E. Woods. Digital Image Processing Addison-Wesley Publishing Company, New York (1993)

[23] Romualdo, L., Luz, P., Devechio, F., Marin, M., Zúñiga, A., Bruno, O., & Herling, V. (2014). Use of artificial vision techniques for diagnostic of nitrogen nutritional status in maize plants. *Computers and Electronics in Agriculture,104*, 63-70. doi:10.1016/j.compag.2014.03.009

[24] Usman, S. A., Toprak İ., and Toygar Ö., Classification of Corn and Grape Plant Diseases Through Leaf Images, *6th International Symposium on Engineering, Artificial Intelligence and Applications (ISEAIA2019),* Girne, Northern Cyprus (March 6-8, 2019).

[25] Li, B., & Xu, X. (2002). Infection and Development of Apple Scab (Venturia inaequalis) on Old Leaves. *Journal of Phytopathology,150*(11-12), 687-691. doi:10.1046/j.1439-0434.2002.00824.x

[26] Apple scab. (n.d.). Retrieved from http://www.apsnet.org/edcenter/intropp/lessons/fungi/ascomycetes/pages/applescab.aspx

[27] Li, B. and Xu, X. (2002). Infection and Development of Apple Scab (Venturia inaequalis) on Old Leaves. *Journal of Phytopathology*, 150(11-12), pp.687-691.

[28] Black Rot of Apple. (n.d.). Retrieved from https://articles.extension.org/pages/60351/black-rot-of-apple

[29] Cedar Apple Rust: Symptoms, Treatment and Control. (n.d.). Retrieved from https://www.planetnatural.com/pest-problem-solver/plant-disease/cedar-apple-rust/

[30] Black Rot of Grapes. (n.d.). Retrieved from http://www.missouribotanicalgarden.org/gardens-gardening/your-garden/help-

for-the-home-gardener/advice-tips-resources/pests-and-problems/diseases/fruit-spots/black-rot-of-grapes.aspx

[31] Grapevine Measles. (n.d.). Retrieved from https://articles.extension.org/pages/64365/grapevine-measles

[32] Technical Difficulties. (n.d.). Retrieved from https://www.pioneer.com/home/site/us/agronomy/crop-management/corn-insect disease/common-rust-of-corn/

[33] Common Rust of Corn. (2016, April 05). Retrieved from https://ohioline.osu.edu/factsheet/plpath-cer-02

[34] Early blight of potato and tomato. (n.d.). Retrieved from https://www.apsnet.org/edcenter/intropp/lessons/fungi/ascomycetes/Pages/PotatoTomato.aspx

[35] Late blight of potato and tomato. (n.d.). Retrieved from http://www.apsnet.org/edcenter/intropp/lessons/fungi/oomycetes/pages/lateblight.aspx

[36] Washington State University. (n.d.). Retrieved from http://treefruit.wsu.edu/crop-protection/disease-management/cherry-powdery-mildew/

[37] Shane, B., Sundin, G., Michigan State University Extension, Michigan State University Extension, & Department of Plant Pathology. (2018, October 03). Management of bacterial spot on peaches and nectarines. Retrieved from https://www.canr.msu.edu/news/management_of_bacterial_spot_on_peaches_and_nectarines

[38] VEGETABLE CROPS. (n.d.). Retrieved from http://vegetablemdonline.ppath.cornell.edu/factsheets/Pepper_BactSpot.htm

[39] Louws, F., & Ridge, G. (n.d.). Leaf Scorch of Strawberry. Retrieved from https://content.ces.ncsu.edu/leaf-scorch-of-strawberry

[40] Lowe, D. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*. doi:10.1109/iccv.1999.790410

[41] Bosch, A., Zisserman, A., & Munoz, X. (2007). Image Classification using Random Forests and Ferns. *2007 IEEE 11th International Conference on Computer Vision*. doi:10.1109/iccv.2007.4409066

[42] An open access repository of images on plant health to enable the development of mobile disease diagnostics David P. Hughes Marcel Salathé

[43] Dalal, N. and Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection, pp.886 - 893.

[44] Ahonen, T., Hadid, A. and Pietikäinen, M. (2004). Face Recognition with Local Binary Patterns. Lecture Notes in Computer Science, pp.469-481.

[45] Kannala, J. and Rahtu, E. (2012). BSIF: Binarized Statistical Image Features. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012),* pp.1363-1366.