Developing an eAgriculture application for identification of fungal disease in plants through leaf images

Thesis submitted in partial fulfillment of the requirement for the award of the degree of
DOCTOR OF PHILOSOPHY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted by
SUKHVIR KAUR
Registration No. 901503017

Under the guidance of

Dr. Shreelekha Pandey
Assistant Professor
Department of Computer Science & Engineering
Thapar Institute of Engineering and Technology
Patiala-147004, INDIA

Dr. Shivani Goel
Professor
Department of Computer Science Engineering
Bennett University
Greater Noida 201310, U.P., INDIA

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

COMPUTER SCIENCE AND ENGINEERING DEPARTMENT
THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY,
PATIALA - 147004

February 2020

CERTIFICATE

I, Sukhvir Kaur, Registration Number 901503017, hereby declare that the thesis entitled "**Developing an eAgriculture application for identification of fungal disease in plants through leaf images**" submitted to the Computer Science and Engineering Department at Thapar Institute of Engineering and Technology, Patiala, Punjab, India is an authenticated record of my own work for the award of the degree of "Doctor of Philosophy" under the supervision of Dr. Shreelekha Pandey and Dr. Shivani Goel. This report has not been submitted to any other institution for the award of any other degree.

Sukhvir Kaur

Sukhvir Kaur

Regn. No. 901503017

Place: Patiala

Date: 7-2-2020

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

Verified by:

Shivani Goel.

Dr. Shreelekha Pandey
Assistant Professor
Department of Computer Science & Engineering
Thapar Institute of Engineering and Technology
Patiala-147004, INDIA

Dr. Shivani Goel
Professor
Department of Computer Science Engineering
Bennett University
Greater Noida 201310, U.P., INDIA

ABSTRACT

Development of automatic disease detection and classification system is significantly explored in precision agriculture. In the past few decades, researchers have studied several cultures exploiting different parts of a plant. The symptoms of plant diseases are evident in any part of a plant, however leaves are found to be the most commonly observed one for infection identification. 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 in this domain. The work presented in this study summarizes the pros and cons of all such studies to throw light on various important research aspects. A discussion on commonly studied infections and research scenario in different phases of a disease detection system is presented. The performance of state-of-the-art techniques are analyzed to identify those that seem to work well across several crops or crop categories. Discovering a set of acceptable techniques, the study presents a discussion on several points of consideration along with the future research directions.

Based on the understandings gained during the survey, a computer vision based systems for plant disease detection using leaf images are developed. The main culture focused in this study is soybean due to its several benefits. A rule-based system using concepts of k-means is designed and implemented to distinguish healthy leaves from diseased leaves. The system works with a set of rules proposed in this study. Once a leaf is identified as unhealthy, it is classified into one of the three categories (*downy mildew, frog eye,* and *septoria leaf blight*) effectively by utilizing the framed rules. The efficacy of the system is proved by performing experiments separately on various color features, texture features and their combinations. Results are generated using thousands of images collected from PlantVillage dataset. Acceptable average accuracy values are reported for all the considered combinations which are also found to be better than existing ones. An attempt has also been made to discover the best performing feature set for leaf disease detection in soybean. The system is shown to efficiently compute the disease severity as well. Qualitative as well as quantitative measures are utilized to further prove suitability of the proposed system in detection, classification, and severity calculation.

Another area focused is to design a generalized framework that can detect a leaf image as healthy or unhealthy and in case of diseased identify its type. The basic idea in developing this system is to eliminate the rules completely. It is a two-stage framework and termed as semi-automatic system in this work. It continues with the same concepts as is utilized in rule-based system in addition to the concepts of two-/multi-class classifiers. Classifiers are trained on numerous features (texture and color). The framework may employ fusion technique in case of more than one classifier and is flexible to work with the best performing classifiers. The study discusses a fusion method as well to combine the results logically. The framework, i.e. semi-automatic system is tested on six different datasets formed with leaf images from legumes, vegetables, fruits, and commercial crops to verify the notion of generalization. Satisfying results are achieved in each of the considered cases.

Both the developed systems are compared with several existing systems in literature and are found to perform better on the following parts: image acquisition, segmentation, and number of training/testing images. The systems have shown agreeable performance on three cross-domain scenarios too. For better comparison, three papers are implemented and tested on PlantVillage dataset, here too the proposed systems outperforms. Lastly, on generalization characteristic of the semi-automatic systems results are good considering the system simplicity. However, there exist a few deep-learning based systems which are superior but in the absence of any standardized datasets the results presented here are acceptable. A web-application is also developed with the proposed semi-automatic system that serves as a good assistance to any naïve user too.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank the almighty God for giving me the opportunity, strength, ability, and, knowledge, to undertake this research study and to complete it. Without his blessings, this attainment would not have been possible.

I would like to express my sincerest gratitude to my supervisors, Dr. Shreelekha Pandey and Dr. Shivani Goel, for their valuable advice, encouragement, and support throughout the journey. With their constructive comments and continued supervision, I am able to make progress in my research work. At every stage of research, they always show me the correct direction. Also they shared with me their invaluable experiences to succeed in the life.

I am grateful to Director, Dean RSP, Head of Department, CSED and management of Thapar Institute of Engineering and Technology, who provided me all the necessary resources for completing my research work. I would like to thank doctoral committee members Dr. Maninder Singh, Dr. Rinkle Rani, Dr. Vijay Kumar, and Dr. Kulbir Singh for their constructive suggestions. I am also thankful to Dr. Deepak Garg for valuable suggestions for research proposal.

I would like to pay deepest gratitude to my parents Mr. Mohinder Singh and Mrs. Iqbal Kaur for their unconditional love, care and encouragement through all the ups and downs of my research journey. I would like to thanks all my family members Manmohan Singh Gill, Varinder Kaur for motivation.

I would also thanks to my friends Upasna Joshi, Amandeep Kaur, Manjit Kaur, Latika Tomar, and Aman Sharma, who make my journey pleasant. At the end, I would like to thank all whosoever has contributed in my research journey.

Sukhvir Kaur

(Sukhvir Kaur)

List of Publications

Published

1. **Sukhvir Kaur, Shreelekha Pandey, Shivani Goel**, Plants Disease Identification and Classification Through Leaf Images: A Survey, Archives of Computational Methods in Engineering, 2018, 26(2), p.507-530, DOI: https://doi.org/10.1007/s11831-018-9255-6 [**SCIE Indexed**, Impact Factor: 7.242, **Accepted: 11 Jan 2018, Published online: 19 Jan 2018**].
2. **Sukhvir Kaur, Shreelekha Pandey, Shivani Goel**, Semi-automatic leaf disease detection and classification system for soybean culture, IET Image Processing, 2018, 12(6), p.1038-1048, DOI: http://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2017.0822 [**SCI Indexed**, Impact Factor: 2.004, **Accepted: 28 Jan 2018, Date of Publication: 1 Feb 2018**].

Under Review

1. **Sukhvir Kaur, Shreelekha Pandey, Shivani Goel**, A generalized semi-automatic framework for leaf disease detection and classification. Optik [**SCI Indexed**, Impact Factor: 1.914].

Conference paper

1. **Sukhvir Kaur, Shreelekha Pandey, Shivani Goel,** An Automatic Leaf Disease Detection System for Legume Species, Journal of Biology and Today's World, Special Issue on: 'Role of IT in Bioinformatics and Neuro-informatics' 2017, 6 (6), 115-122 [**International Conference on Smart Computing and Communication (SmartTech-2017)**].

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
Adaboost	Adaptive Boosting
ANN	Artificial Neural Network
API	Application Program Interface
APS	American Phytopathological Society
BPNN	Back Propagation Neural Networks
CCD	Charge Coupled Device
CCR	Correct Classification Rate
CMC	Curvelet Modulus Correlation
CMYK	Cyan, Magenta, Yellow, and Key
CNN	Convolutional Neural Network
DCT	Discrete Cosine Transform
DNA	Deoxyribose Nucleic Acid
DSIFT	Dense Scale Invariant Feature Transform
DWT	Discrete Wavelet Transform
GA	Genetic Algorithms
GLCM	Gray Level Co-occurrence Matrix

GRNN	General Regression Neural Network
HMI	Human Mobile Interface
HOG	Histogram of Oriented Gradients
HSI	Hue, Saturation, and Intensity
HSV	Hue, Saturation, and Value
IIS	Internet Information Services
INIBAP	International Network for the Improvement of Banana and Plantain
IPM	Integrated Pest Management
IRKT	Improved Rotation Kernel Transformation
IRRI	International Rice Research Institute
K-NN	K-Nearest Neighbor
LDA	Linear Discriminant Analysis
LSDA	Locality Sensitive Discriminant Analysis
Matlab	Matrix Laboratory
MCR	MATLAB Compiler Runtime
MLP	Multilayer Perceptron
NB	Naive Bayes
NN	Neural Network
OLDP	Orthogonal Locally Discriminant Projection
PCA	Principal Component Analysis
PDI	Percentage Disease Index

PHOW	Pyramid Histograms of Visual Words
PNN	Probabilistic Neural Network
RBF-NN	Radial Basis Function Neural Network
RFT	Random Forest Tree
ROC	Receiver Operator Characteristic
ROI	Region of interest
RUSBoost	Random Undersampling Boosting
Sceboost	Symmetric Cross Entropy Based Boosting
SDK	Software Development Kit
SGDM	Spatial Gray Level Dependence Matrix
SIFT	Scale Invariant Feature Transform
SMO	Sequential Minimal Optimization
SOFM	Self Organizing Feature Map
SOM-NN	Self Organizing Map Neural Network
SRG	Seeded Region Growing
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
WPD	Wavelet Packet Decomposition

Chapter 1

INTRODUCTION

Agriculture, one of the main economical sources for any country, depends on the quality and quantity of farming products, especially plants. Several production improvement measures are thus employed to escalate yield of any crop. Irrespective of all the precautions, presence of any harmful symptom, caused due to abiotic and biotic factors significantly affect the production [1, 2]. These symptoms can easily be observed as well as controlled by examining several parts of a plant properly. Leaf, the most sensitive part of a plant is known to show disease symptoms at the earliest as compared to other parts. Since last few years researchers have put lots of efforts in automating disease detection process instead of adopting a time consuming manual technique. Advancements in machine vision and pattern recognition domains have actually motivated researchers to employ the technology in agriculture domain for automatic detection of leaf diseases caused by various biotic and abiotic stresses [3-12].

Depending upon the cause, a plant may get affected from a specific infection out of a range of diseases. This fact further complicates applicability of computer vision techniques in proper recognition of plant diseases [13, 14]. Different plants disease detection techniques are proposed and a survey of traditional and innovative techniques is also presented in literature [15, 16]. The popular traditional techniques include molecular, serological, and deoxyribose nucleic acid (DNA). Volatile organic compounds and imaging & spectroscopic techniques are also utilized innovatively to automate the detection process. Such innovative techniques are faster and do not need personnel monitoring. Research by Zhang and Meng (2011) reported an accuracy of 87.99% (using an imaging technique) and 86.87% (using human experts on screen) for automatic detection of *citrus canker* on leaves. Their study further supports the usage of image processing techniques to automatically detect plant diseases at an early stage [17].

This work thus attempts to develop frameworks utilizing innovative techniques to detect and classify plant infection using leaf images. For designing effective frameworks, a comprehensive study of disease types as well as several existing systems developed for different crops is done. Although the main focus of this work is legume species, particularly **soybean**, thus the developed frameworks are validated using leaf images belonging to this crop category. Leaves infected from *downy mildew*, *frog eye*, and *septoria leaf blight*, are considered during experimentations. Also, comparative results with the existing works are presented for the same category using several datasets. The work attempts to propose a generalized framework with respect to crop as well as disease categories. The heterogeneity of that framework is tested on different datasets such as legumes, vegetables, fruits, and commercial crops collected using Google image search engine along with the base dataset of **soybean** crop.

This chapter starts with Section 1.1 portraying the motivation to select **soybean** crop for this research work. Section 1.2 describes different types of diseases occurred commonly in plants followed by an introduction to image processing and machine learning in Section 1.3. Section 1.4 presents a general architecture of an image based plant disease detection system. Objectives and scope of the work are given in Section 1.5. Details of available as well as self collected image databases used for experimental analysis along with the experimental setup are discussed in Section 1.6. Performance measures are described in Section 1.7. Lastly, Sections 1.8 briefs organization of rest of the thesis chapters.

1.1. Motivation

Various studies are carried out to automatically detect disease(s) in different cultures. However in this research work, **soybean** crop which is listed in top ten staple foods feeding the world is primarily considered [18]. **Soybean**, also known as Glycine max crop, belongs to a family of legumes whose roots contain rhizobia bacteria which fixes nitrogen and thus facilitates plant growth [19]. Moreover, **soybean** is high in nutritional values and is mainly identified as an oilseed rather than a pulse. The top five **soybean** producing countries are United States (103.4 million tonnes), Brazil

(103.0 million tonnes), Argentina (57.0 million tonnes), China (12.2 million tonnes), and India (11.7 million tonnes) [20].

Soybean is usually attacked by ***fungal*** diseases which are caused by biotic factors [21]. There exists a range of ***fungal*** diseases (*rust*, *powdery mildew*, *sudden death syndrome*, etc.) which affects plant at different stages depending upon temperature, moisture etc. [22]. **Soybean** development occurs in two stages, vegetative and reproductive [23]. It has been reported that the yield losses due to *frog eye*, *septoria leaf blight*, and *downy mildew* are 10–60%, 8–34%, and 9–18%, respectively [24, 25]. Moreover diseases, like *frog eye* and *downy mildew*, affect crop quality frequently in initial stages by contaminating the young leaves. On the other hand, various diseases such as *septoria leaf blight* influence the plant growth not only in initial but later stages too [26]. Thus, it can be said that disease detection at an early plant stage would be of great help. Infection may affect any part of **soybean** but this study is concentrating solely on leaves as they are known to show symptoms at the earliest.

1.2. Types of Plant Diseases

There are two categories of plant diseases: biotic and abiotic. Those originated from living organisms are biotic [27]. Fungi, bacteria, and viruses are the main causes of different forms of biotic diseases. Abiotic diseases, on the other hand, are produced by non-living ecological circumstances such as hail, spring frosts, weather conditions, burning of chemicals, etc. Abiotic diseases are non-infectious, non-transmissible, less dangerous, and are mostly avoidable. This study thus considers the biotic diseases. Their categorization with a few common forms is shown in Fig. 1.1. A range of works exists for various fungal and bacterial diseases, but those under viral category are not focused much in literature [28, 29]. *Spots* (caused either through fungi or bacteria), *mildew*, and *rust* are the top three types which are most commonly considered for identification and classification. In addition, deficiency of nutrients is explored for automation. All these observed facts are further detailed in Chapter 2.

1.3. Image Processing and Machine Learning

Image processing is an important part in computer vision applications. In addition to some basic functionality, it mainly deals with functions that improve image quality by

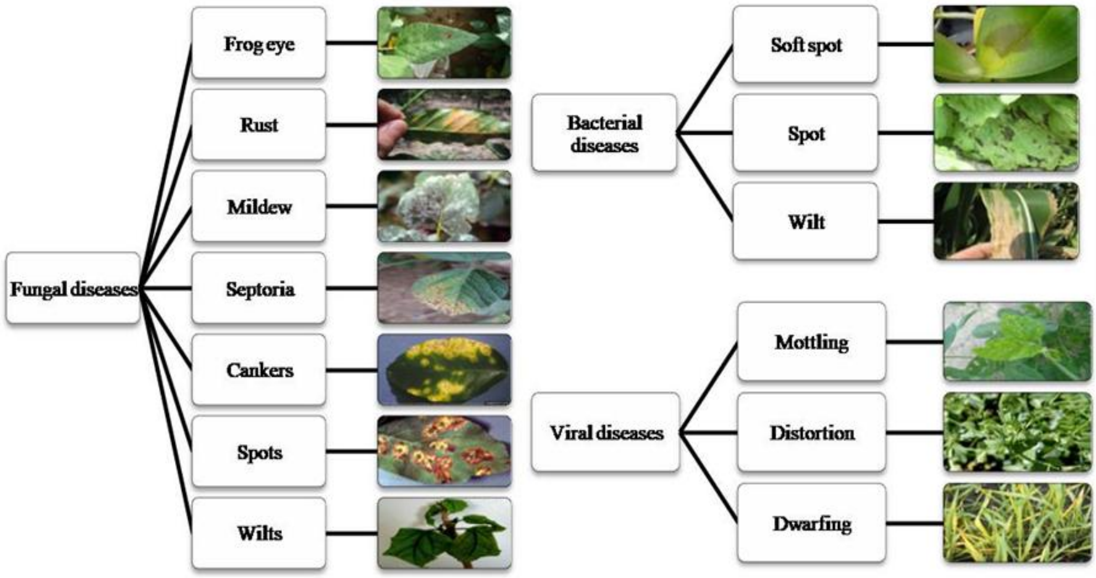


Fig. 1.1 Different types of plant biotic diseases and their types in various cultures.

removing defects (related to focus, motion, noise, lighting, etc.), by separating required image area from background by extracting significant information that enables fine decision making. These functions are broadly divided as low level, intermediate level, and high level functions [30]. Low level involves image acquisition and pre-processing to get an enhanced image via filtering. Segmentation falls in intermediate level that divides an image into various areas with strong correlation. High level performs interpretation and recognition using arithmetic classifiers and multilayer neural networks. These steps successively refine data for better processing in further modules.

Similar to image processing, machine learning algorithms are being used as an essential module in numerous real-time systems [31-34]. In general, machine learning methods are categorized as supervised and unsupervised [35]. The training set for supervised methods consists of input and the corresponding response values. Differing from this, unsupervised methods build inferences for absent labeled responses in the training set. A special class of supervised methods is a semi-supervised method which utilizes a mix of labeled and unlabeled training data. Graph based learning techniques are usually fall in this category. Lastly, there is a class of reinforcement learning algorithms which are popularly used in games [36]. They allow software agents to automatically find ideal behavior according to specific context, for performance maximization.

Combination of image processing with machine learning is immensely studied in almost all sorts of application like face detection, face recognition, remote sensing, biomedical, underwater images, character or handwriting recognition, security, speech-text mining, health-care, finance, marketing and many more decision support systems. Similarly, in agriculture domain, researchers have applied image processing and machine learning techniques to automate processes like crop production, crop classification, agricultural land soils classification, identification and differentiation of infected and non-infected leaf areas, unknown grain type identification [37-39]. Other specific processes include detecting leaf wilting, estimating plant nitrogen content, forecasting **cotton** leaf *curl* disease using weather based prediction model, detecting fungal disease on fruit crops, fruit skin defect identification, grading of fruits into different categories. Clearly, a huge range of frameworks and applications exist in literature. Many of such studies utilizing leaf images to automate disease detection are investigated in Chapter 2.

1.4. General Plant Disease Detection System

Fig. 1.2 shows a general architecture of a plant disease detection system. Its main modules are acquisition, pre-processing, segmentation, feature extraction, and classification (or recognition). The system has two phases: training and testing. The training phase starts with capturing an image of a specific part like leaves, stems,

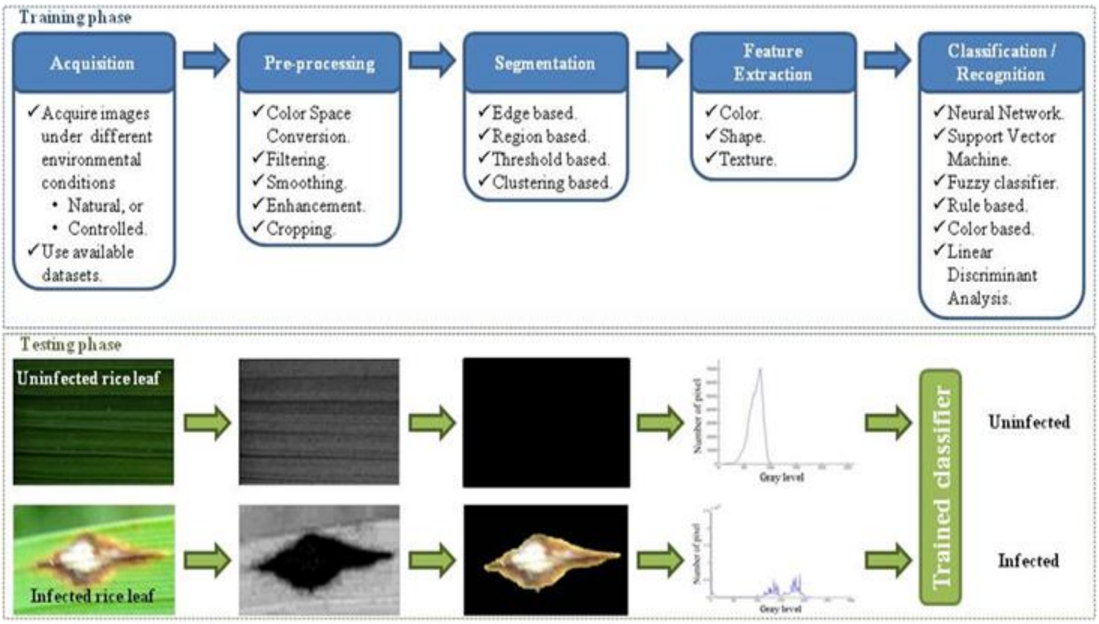


Fig. 1.2 General architecture of a plant disease detection system with imaging techniques. Testing phase images are from [40].

roots, and branches. The captured (or dataset) images may or may not be pre-processed to correct various geometric misinterpretation, grey level correction, noise reduction, and blur improvement. Segmentation module then separates the regions of interest from background and identifies two or more regions from an infected training leaf image. Lastly, feature vectors of the regions of interest are extracted and forwarded to the classifier for learning. In testing phase, a test leaf image passes through pre-processing, segmentation, and feature extraction modules. The trained classifier then identifies the test image as an infected or a healthy sample. Necessity of all the modules is discussed in separate subsections with a brief summary of different techniques utilized or proposed in literature for each module.

1.4.1. Acquisition

Image acquisition is important as system's performance depends greatly on image samples used for training. In this domain, researchers have used a few known datasets, namely IPM Images, PlantVillage Images, and APS Image database [41-44]. Images with theoretical details are also available at the website of University of Minnesota Extension [45]. Some of the works are privileged to access datasets of research centers like International Rice Research Institute (IRRI) and International Network for the Improvement of Banana and Plantain (INIBAP) [46, 47]. A few have observed single culture for a period of time instead of a full-fledged dataset [48, 49]. Scanned images are even used in a work or two [50, 51]. A range of studies have used self-collected image datasets taken either under controlled environmental conditions or in field with complex backgrounds. For effective control of illumination, lighting, and intensity, images are also acquired inside laboratories or a sampling box [52-55], like infected **soybean** leaves are placed on white base to remove background complexity [29]. On contrary, some works have captured images with complex background in the field [5, 56, 57].

The quality of image samples also depends camera type and its orientation. Most of the studies used digital cameras keeping optical axis perpendicular to the leaf plane, but a few have used specialized techniques. Charged coupled device (CCD) color cameras with different specifications are combined with software tools to capture RGB color images [56, 58, 59]. Android mobile is also used to capture a leaf at some fixed distance from its surface [60]. A multispectral CDD camera with a portable spectro-radiometer is also employed to acquire images of **soybean** leaves [61]. Recently a hyper-spectral imaging system is utilized to collect **tomato** leaf images [62]. It's obvious that images collected under controlled environment are easier to process. Similarly, equipments and techniques used for capturing provide different image details. The performance of a plant disease detection system thus varies with background of the acquired image as well as the capturing conditions [13].

1.4.2. Pre-processing

During pre-processing, images are improved by distortion removal to get enhanced features which eases further processing. Popular pre-processing techniques include color space conversion, cropping, smoothing, enhancement, etc. Depending upon quality of image dataset the functionality of pre-processing module varies. As per the literature survey most of the systems apply color space conversion followed by filtering and enhancement.

Being able to closely resemble the human color sensing properties hue, saturation, and value (HSV) is the most preferred color space [63-67]. HSI (I is for intensity), another color space from the same class is also popularly used [3, 68-71]. Works utilizing different color spaces like YCbCr, Hue-Max-Min-Diff, CIE 1976 L*a*b*, RGB, and CIE 1976 uniform chromaticity scale diagram also exists [72-74]. RGB color space is also converted to a new space H, I3a and I3b to facilitate automatic segmentation which is comparable to the one done using manual procedure [47].

After color space transformation, filters are applied for desired enhancements, like increased contrast and brightness. Noise occurrence is very general, thus plant disease detection systems popularly use median [46, 56, 75, 76] and rank [77] filters. Laplacian filter is used for sharpening [64]. Apart from these, techniques like histogram equalization and Gabor wavelets are also used for filtering and controlling varying lighting conditions [5]. Concept of anisotropic diffusion is recently presented for enhancement [65]. Other commonly used filters are spatial low pass filter, neighborhood mean, and frequency low-pass filter [78]. Cropping is also important if images are captured in an uncontrolled environment with complex backgrounds. Cropping can either be done automatically using functions [7, 9, 79] or manually [64].

1.4.3. Segmentation

Segmentation divides an image into clusters with robust correlation along with the objects of interest. Segmentation of pre-processed images leads to accurate feature extraction. Features of an effectively segmented image help in an easy identification of healthy or infected samples, for instance, number of peaks in the histogram [40]. Edges, thresholds, locality or color based segmentation techniques are shown to work well with plant disease detection systems. Edge based techniques, like Sobel operator and canny edge detectors are employed in range of studies [4, 52, 53, 77, 80, 81]. A few studies have also exploited genetic algorithms [5, 82] and Grab-cut segmentation [83]. Methods based on the concept of entropy and Otsu methods are popular threshold based segmentation techniques [66, 84-86]. A manual threshold setting technique is presented for an effective segmentation of disease spots in HSI color space [87]. An integration of seeded region growing (SRG) concept with local threshold is also explored for an automatic and efficient segmentation of leaf *spot* [88].

The infected leaf area shows significant color differences from its original color and this leads to the development of *spot* color based segmentation [56]. Also, k-means clustering is observed to be more accurate than Sobel, prewitt and canny edge operator based segmentation approaches [7, 9, 65, 75, 77]. But in the presence of noise, k-mediods based segmentation is found to be more robust [89]. Notion of Fermi energy is uniquely employed to perform segmentation based on color and grey level intensity values [6]. The method is shown to work better than Otsu and k-means based methods. A combination of saliency region threshold and k-means algorithm is also utilized to directly extract the diseased leaf area [90]. This approach is found better than mean shift and unsupervised optimal fuzzy C-means clustering. A study recommends simple k-means clustering over fuzzy C-means or expectation maximization for accurate leaf disease detection [91].

In conclusion, determination of threshold value is an important step in the process of segmentation. An incorrect threshold determination may infer inaccurate segmentation results which lead to an erroneous classification phase as well as a system [92].

1.4.4. Feature Extraction

Images are usually interpreted as color, texture, and shape feature vectors. Color is commonly defined in terms of moments and histograms. Several properties like, contrast, homogeneity, variance, and entropy, can be attached to texture features. Similarly, for shape features, roundness, area, eccentricity and concavity characteristics are identified. Literature for plants disease detection system reveals that texture is the best suitable feature [93]. But for heterogeneous datasets, a combination of two or more features works well.

The classical gray level co-occurrence matrix (GLCM) and its spatial variants are utilized to compute texture parameters like, energy, entropy, moment of inertia, etc, of an infected area [58, 73]. After color space conversion, a spatial gray level dependence matrix (SGDM) of the H image is also employed to extract several parameters [65, 68]. A hybrid feature combing two or more texture features based on discrete cosine transform (DCT), structure, Fourier transform, difference operators, and Wavelet packet decomposition (WPD) is built for efficient disease detection [55, 94]. Fourier spectrum based fractal descriptors from each lesion are also shown to give good results [64].

A few works have combined texture features with color (histograms or moments) as well as shape (area, perimeter, length, width, compactness, rectangularity, roundness, and elongation) features to detect the type of plant leaf diseases and are found to improve the system performance [52, 56, 58]. Some studies have eliminated texture and worked with a combination of color and shape features. Shape features are computed along with mean, median, standard deviation, Quartile 1, Quartile 3, and average brightness of RGB color space [6, 95]. Another research presented Eigen vector based extraction to detect cotton leaf diseases [96]. Recently, local descriptors such as speeded-up robust features (SURF), histogram of oriented gradients (HOG), scale-invariant feature transform (SIFT), dense SIFT (DSIFT), pyramid histograms of visual words (PHOW) are explored and compared for better detection as well as classification of **soybean** diseases [50]. The reported results show that usage of PHOW leads to a better system.

1.4.5. Classification or Recognition

Classification is the most important module in a plant disease detection system. This study focuses on systems that detect plant diseases using leaf images, thus classification here is defined as a process of categorizing plant leaf images based on identified diseases. In general images from a training set are used to first train the classifier; the trained classifier then classifies or recognizes test set images. In plant disease detection, a range of machine learning methods, are explored to identify disease types in several cultures. The classifier used must differentiate between a healthy and an unhealthy leaf image too [97].

Popular classification techniques explored in the domain of plant disease identification are shown in Fig. 1.3. In addition to classifiers, a few works lead to accurate identification using other techniques based on features, fuzzy logic, etc. Works under this category are also shown in Fig. 1.3. Discussion on different classifiers explored for identifying plant diseases in different cultures is presented in Chapter 2.

1.5. Objectives and Scope of the Work

The proposed work aims developing frameworks to detect and classify plant infection

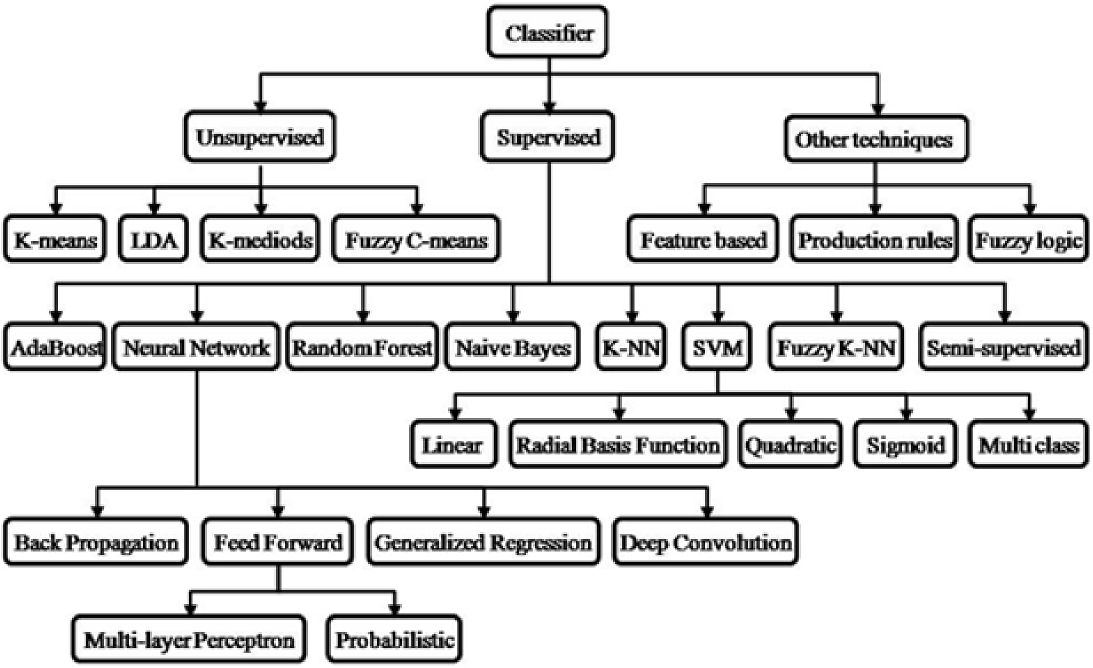


Fig. 1.3 Classifiers popularly explored in the domain of plants disease detection system.

using leaf images in soybean culture utilizing the existing innovative techniques. Such systems would be an asset to researchers as well as experts working in this domain. In an attempt to design effective frameworks, following are the objectives of this work:

- **To study and analyze the image processing techniques for identification of fungal disease in plants.** Fungal diseases are one of the most harmful diseases that may result in death of the plant. Moreover a large number of plant diseases (approximately 8,000) belong to this category [98]. Also, studying the related existing works is necessary to understand the current research scenario, specifically techniques utilized, in plant leaf disease detection and thus to identify the ways that would help in improving systems performance.
- **Design and implementation of an eAgriculture application for accurate identification of fungal disease in legume plants.** Most of the existing systems are not publically available. Accessibility to a system that can detect then classify a plant disease would definitely be a good addition in this domain.
- **Testing and validation of the developed application.** The system is tested on a subset of PlantVillage image dataset and on some other **soybean** datasets such as self collected, Shrivastava dataset, and IPM images. Also, validity of proposed framework is tested on different datasets collected from Google image search engine.

1.6. Experimental Setup

The proposed frameworks are implemented using MATLAB R2018a software on a PC with Intel Core i7 processor, 8 GB RAM, 2 TB HDD, and Windows 10 64-bit operating system. The dataset used to validate the proposed system is collected from PlantVillage [99]. A part of this dataset images consisting of three types of soybean leaf infections, namely *downy mildew* (105 images), *frog eye* (1,662 images), and *septoria leaf blight* (1,929 images) along with healthy leaves (1,079 images) are used for experimentation. As stated earlier, *frog eye* and *downy mildew* affect mostly during early leaf stages, hence single category corresponding to these two infections exist in the dataset. However, *septoria leaf blight* can affect **soybean** leaf at all the leaf stages and PlantVillage contains images corresponding to three stages (initial, intermediate, and last) of this disease. 1,929 images considered for *septoria leaf blight* contains approximately equal number of images for each of the three stages.

Fig. 1.4 shows a few representative images of each disease type along with normal leaves [26, 100]. Different disease classes of **groundnut**, **tomato**, **potato**, **apple**, **mango**, **sugarcane**, **cotton**, **chilli**, and **jute** are also considered to validate the applicability of the proposed framework on variety of cultures (a few representative images are shown in Fig.1.5).

(a) Normal leaf – Healthy soybean plant leaves.

(b) Frog eye – Caused by *cercospora sojina* fungus in young leaves as compared to mature ones. Initially appear as tiny spots which become large as ¼ inch diameter. The lesion center changes from gray to brown along with reddish-purple boundary [100].

(c) Downy mildew – Young plant leaves get infected by *peronospora manshurica* fungus that leads to seed quality degradation. Symptoms appear as irregular structure variation from pale green to light yellow with light brown color [26].

(d) Septoria leaf blight – Appears as small spots in early plant stages. As attack of *septoria glycine* fungus increases, these spots get surrounded by small yellow margin. Later, the whole leaf turns into yellow to brownish color [26].

Fig. 1.4 A few representative images from PlantVillage dataset for four categories.

Fig. 1.5 A few representative images from legumes, vegetables, fruits, and commercial crops collected using Google image search engine.

The proposed frameworks are also validated in a cross-domain scenario. Model trained using PlantVillage dataset are tested using diseased images from three different datasets, namely IPM images [41], a few images from [60], and a self-collected dataset[1]. Images in the self-collected dataset are personally collected from the fields of Punjab Agricultural University, India. IPM images contain leaf samples for the three disease categories considered in this study; however the other two have leaves only for *septoria leaf blight*.

1.7. Performance Measures

The effectiveness and applicability of plant disease detection systems are accessed using various performance measures. Popularly used one is accuracy and has also been utilized in this research work. It is also known as precision, classification rate, recognition rate, and success rate [40]. In addition, a few existing works evaluated performance using prediction time [72] and mean square error [95]. Higher accuracy

[1] Leaf samples were collected with the help of several experts from the College of Agriculture, Punjab Agricultural University, Ludhiana, Punjab.

values with smaller prediction time and lower mean square error prove superiority of one system over another.

1.8. Organization of the Thesis

The research work presented in the thesis is organized in five chapters. A brief narration of all the chapters is as follows:

Chapter 1 discusses the importance of plants in the economy of a country and depicts the rationale for selecting **soybean** crop. Basics of plant disease types, image processing, and machine learning concepts are introduced for better understanding of the proposed framework. It presents a general architecture of an image based system used for plant disease detection as well. Specifying objectives and extent of the work, the chapter underlines details of datasets used to perform experiments, experimental setup, and performance measures. Chapter ends describing the thesis organization.

Chapter 2 presents a discussion on various classifiers explored by the researchers for an accurate identification of plant diseases in different cultures. On the basis of the observations made, an attempt to identify the best known system, the most studied culture as well as the most popular classifier is made. Performance of state-of-the-art techniques are analyzed to identify those that work well across several crop categories.

Chapter 3 presents the proposed rule-based framework and overall description of the system. Procedures employed in each intermediate phase, pre-processing, segmentation, feature extraction, and classification, of the system are described.

Chapter 4 details the complete procedure adopted to eliminate the manual inspection required to form the rules in the proposed rule-based system. Followed are the proposed semi-automatic methodology and its overall system description.

Chapter 5 discusses several observations while describing and analyzing the experimental results of the systems proposed in Chapter 3 and Chapter 4. Also, a comparative performance analysis of the proposed systems with other existing works is provided. The chapter also includes implementation details of the developed Web application in this work.

Chapter 6 presents the concluding remarks of this research work. The complete summarization along with the review and accomplishments of objectives is given. Lastly, the possible future research directions are talked about.

Chapter 2

LITERATURE REVIEW

The existing heterogeneity in leaf images greatly affects the performance of classifiers to identify and classify infected leaves. This chapter thus presents the classifier analysis after grouping crops (or cultures) in various categories as is shown in Fig. 2.1. Initial sub-sections discuss the performance of different classifiers with respect to various cultures in each category. Many articles explore leaf diseases for a single culture and some focus on diseases irrespective of the culture. Latter works, presented in Section 2.5 under the head 'Assorted Cultures', use datasets consisting of heterogeneous cultures. Fig. 2.2 shows the

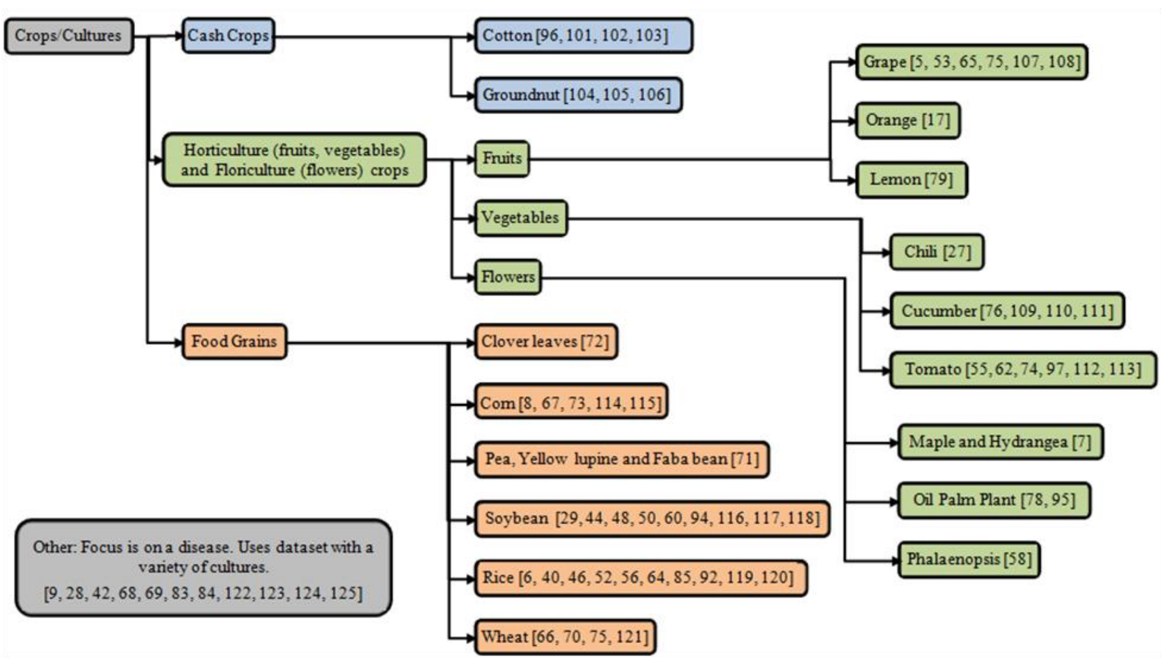


Fig. 2.1 Classification of crops followed in the preparation of the literature.

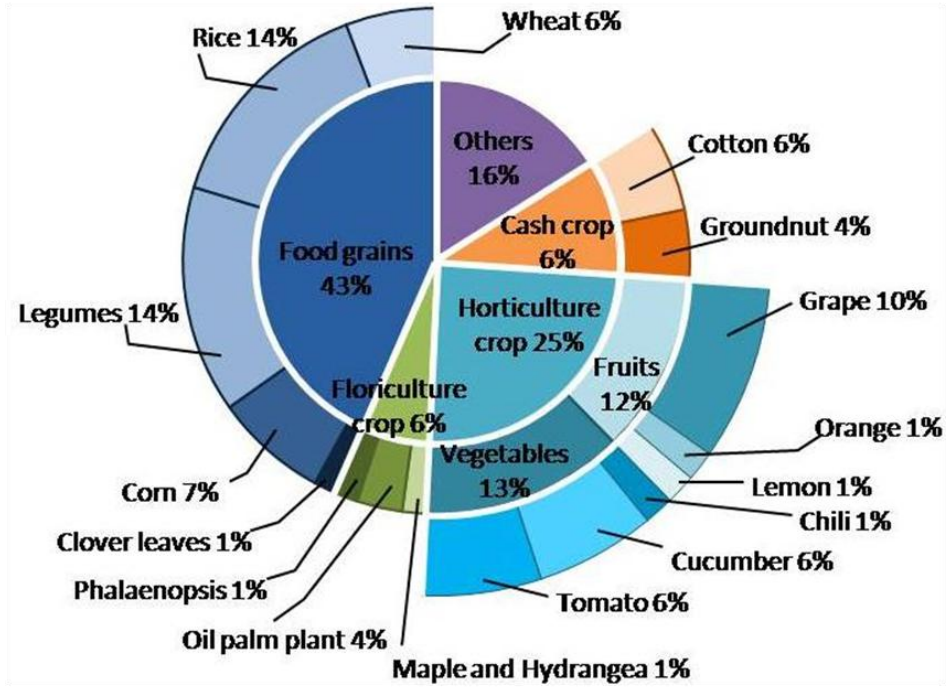


Fig. 2.2 Current state of crops explored during last 10 years in terms of percentage of research papers.

current state of research in different crops in terms of percentages of papers. Clearly, the crops under food grains category are studied the most and very few studies have focused on floriculture crops. A range of crops do exists which are still unexplored because they are less famous or/and their required images are unavailable. Moreover, the economy of a country depends largely on a well-known crop, and hence it is worthy to explore it. Overall analysis is presented in Section 2.6.

2.1 Cash Crops

Cotton has always been a very popular raw material in textile industries and its crisis is one of the major issues for any country to deal with. Maintaining its quality using an automatic system that detects various **cotton** diseases would be of great help. The literature survey shows that 80% to 90% of **cotton** diseases can be recognized just by observing the appearances of the leaves [101]. The study focusing on attainment of desired accuracy in a machine vision based recognition systems is conducted [102]. It trains SVM models on several combinations of features to identify the best classification [102]. The work obtains a subset of informative features using a dataset containing *spots*, *stains* or *strikes* infected leaf images. The study declares texture as the best discriminators (83%) and the worst is shape (55%). The most appropriate set of 45 features reports maximum classification accuracy of 93.1%. Another work presents a system that extracts features using Eigen vectors [96]. The image space is decomposed into sub-spaces then features are regularized and extracted in each of the sub-spaces separately. The system is trained using nearest neighborhood classifier and achieves 90% accuracy in detecting the red *spot*. The proposed Eigen vector based features improves success rate and can also be utilized to identify other **cotton** leaf infections. Another system employs back propagation neural networks (BPNN) with adaptive learning characteristics to detect *powdery mildew*, *downy mildew* and *leafminer* diseases [101].

Obtained results show appropriateness of neural networks in accurately identifying **cotton** leaf diseases. A slow snake segmentation based feed forward BPNN is developed to detect *myrothecium*, *bacterial blight*, and an *alternaria* disease in cotton leaves [103]. The proposed system employs Hu moments for training using Levenberg-Marquardt optimization and reports an average classification accuracy of 85.52%. Although, the system has long training and testing phase timings, its performance and robustness can be enhanced by incorporation of other features. BPNN is utilized to develop an automatic leaf disease detection system for another globally important cash crop, i.e., **groundnut** [104]. **Groundnut** that bridges vegetable oil deficit in most of the countries usually suffers from *cercospora*. The system uses color and texture features to detect four phases *cercospora*, *cercospodium personatum*, *phaeoisariopsis*, and *altemars* of this disease. Reporting a classification accuracy of 97.41%, the system further proves relevance of neural networks in automation of plant disease detection systems. Another work combines morphology operations with heuristics that are designed using specialist knowledge. The system measures early and late leaf *spots* caused by *Cercosporidium personatum* and *Cercospora arachidicola* fungi in **peanut** crop [105]. Trained using CMYK color channel only, this system is automatic, practical, and quick.

Moreover, it uses only two training images (one from each *spot*) and is tested with 124 early and 114 late leaf *spots* images. The results presented prove suitability of this computationally effective system. Production of any crop gets affected by deficiencies also; a study thus attempts to detect different stages of deficiencies in **groundnut** crop

Table 2.1 Summary of the cash crops.

[Reference] Year	Disease or Deficiency	Number of images: Training images	Testing images	Classifier	Performance measure (Accuracy)
Cotton					
[102] 2009	Spots, Stains and Strikes	Total images = 117		SVM	93.1%
[96] 2012	Red spot	100	50	Nearest neighborhood	90%
[103] 2015	Myrothecium, Bacterial blight, and Alternaria	N/A	N/A	Feed-forward BPNN	85.52%
[101] 2016	Powdery mildew, Downy mildew, and Leafminer	6	38	BPNN	Visual Examination
Groundnut					
[106] 2011	Different stages of deficiencies	Total images = 160		Feature based (geometric moments)	93%
[105] 2013	Early leaf spots	1	123	Color based	Visual Examination
	Late leaf spots	1	113	Color based	Visual Examination
[104] 2015	Four stages of Cercospora	360	40	BPNN	97.41%

using Geometric moments [106]. The system can assist farmers in deficiency detection as well as estimation of stage, which is difficult to be done with naked eyes. The proposed system reports an accuracy of 93% and can successfully be applied to other cultures.

Table 2.1 summarizes all the researches covered under the category of cash crops in this study. For **cotton**, SVM is shown to achieve maximum accuracy of 93.1% in detecting *spots*, *stains* and *strikes*. However, BPNN detects *cercospora* in **groundnut** with an accuracy of 97.14%. In summary, neural networks, i.e. BPNN can be considered as the most preferred classifier. Also spots are the commonly explored disease in case of cash crops. Researchers have reported an accuracy of more than 90% in correct identification and classification of *spots*.

2.2 Horticulture Crops

2.2.1 Fruits

The important commercial group in this category is citrus plants. Some popular citrus crops are the **tangerines**, **limes**, **oranges**, **grapefruits**, and **lemons**. These crops are mainly affected by *melanose*, *scab*, *canker*, *downy mildew*, *powdery mildew*, *greasy spots*, and *anthracnose*. Using normal and infected **grapefruit** leaves, both front and back, a generalized square distance classifier based system is presented to classify a test sample as normal, *melanose*, *scab*, and *greasy spot* [53]. Eight statistical classification models based on combinations of texture based HSI color features are compared. Model built using intensity features reduces classification accuracy with leaf fronts but not with the leaf backs. Moreover, model based on hue and saturation reports better results (95.8%) as compared to intensity based model (81%). Models using the HSI texture features or reduced hue and saturation features reported 100% accuracy. The models are computationally efficient, robust to light variations, and are best suited to examine citrus diseased leaves under laboratory conditions. The study suggests usage of cameras that can control lighting levels which in turn helps to reduce impact of low lighting conditions on hue and saturation. Another work focuses on automatic identification and classification of infected **grapefruit** leaf using multiple artificial intelligence techniques [5]. The developed system uses self organizing feature map (SOFM) and BPNN for pre-processing; modified SOFM, genetic algorithms (GA), and SVM for disease segmentation; and SVM again for classifying leaf samples as normal, *rust*, and *scab*. The system reports 97.8% accuracy. This work presents a complex but effective blend of SOFM, BPNN, modified SOFM, GA, and SVM. The final resulting features are filtered using Gabor for improved SVM performance. As a result acceptable classification accuracy (scab – 83.5% and rust – 82.5%) is observed. Contrary to this, another work uses a simple feed forward BPNN to detect downy and *powdery mildew* diseases [65]. The system is robust to lighting effect as it employs only hue color component. As a result 100% classification accuracy is observed on small dataset of 33 images.

The developed system can also be utilized to detect other leaf diseases like *anthracnose*. The study suggests replacing k-means by other segmentation technique to improve its appropriateness in accurate lesion extraction. Another system also attempts to detect *downy* and *powdery mildew* using a heterogeneous combination of PCA reduced color, texture, and shape features to train several neural networks (BPNN, RBF-NN, GRNN, and PNN) [75]. The developed system reports 100% fitting accuracy in each of the four cases. But among all, GRNN and PNN are found to be the best (94.29%) for fungal disease detection followed by BPNN (approximately 91%) and RBF-NN (80%).

Instead of focusing on leaf disease detection, an approach to discover potassium deficiency in six red **grapes** varieties viz. cabernet sauvignon, cabernet franc, merlot, malbec, shiraz, and tempranillo, is developed [107]. The study mainly compares the performance of histogram and k-NN based segmentation techniques. The study reveals inability of histogram based techniques to distinguish colors and hence found them suitable for grayscale images only. But for images with shadows or taken in less controlled environment conditions, k-NN based techniques are preferred. The presented approach can also be used to identify other deficiencies after some minor revisions, like addition of sample color classes to the database and designing of rules to categorize the symptoms. A two phase system combining techniques of image processing, K-means, and fuzzy set theory is developed to identify *downy mildew* [108]. The first phase utilizes k-means for feature reduction. Fuzzy value is then computed for each cluster feature with respect to the number of infected images. Features with fuzzy values larger than the predefined threshold are used for detection. If the average of the retained fuzzy values for an image is greater than or equal to some predefined threshold then the sample is infected. This fuzzy system reports a classification accuracy of 87.09% thus proves the effectiveness of fuzzy set theory in the domain of automatic detection of leaf diseases.

Similarly, the performance of k-NN, Naive Bayes (NB), LDA, and Random Forest Tree (RFT) is observed to automatically detect **lemon** leaf diseases [79]. Forty sample images from each category (*greasy spot*, *scab*, and *melanose*) are collected in addition to normal leaves. The classifiers are trained using texture features of 50% of images. The study observes the following arrangement of classifier in the increasing order of classification accuracy: k-NN (77.5%), NB (95%), RFT (97.5%), and LDA (98.75%). Moreover, the study also reveals that it’s easier to classify normal leaf and *greasy spot* leaf samples as compared to *scab* and *melanose* (least classification rate) leaf samples. A novel two-level feature descriptor is presented to detect and classify orange leaves as normal or infected with *canker*, *black spot*, *scab*, and *melanose* [17]. The work presented is meant for images collected in fields. The proposed descriptor employs enhanced AdaBoost (SceBoost) to separate the background and combines color-texture zone-based local features to get the required descriptor. Experimental comparisons with various feature descriptors prove effectiveness of the proposed descriptor. Also the descriptor is used to train several classifiers, namely, Adaboost, RBF-NN, k-NN, and SVM. SVM reports a minimum classification rate of 63% and Adaboost achieves a maximum accuracy, i.e., 88%. Further the study shows that the results obtained by the proposed approach are closer to those obtained by human experts, which confirms the feasibility of the system.

For a clear performance comparison, summary of all the studies related to fruit crops is given in Table 2.2. It can easily be observed that mildew in **grape** fruit is explored the most in the past 10 years. Accordingly, outstanding results (100% detection using feed forward BPNN [65] and statistical analysis of HSI [53]) are obtained for small image datasets (≤ 40). However, the superiority of multi-class SVM is discovered by looking at the number of training and testing images [5]. Multi-class SVM reports an

Table 2.2 Summary of the horticulture (fruits) crops.

[Reference]Year	**Disease or Healthy leaves or Deficiency**	**Number of images**		**Classifier**	**Performance measure (Accuracy)**
		Training images	**Testing images**		
Grape					
[53] 2006	Greasy spot	20	20	Statistical analysis	95% (H and S) – 100% (HSI)
	Melanose	20	20		
	Scab	20	13		
	Healthy leaves	20	20		
[5] 2008	Scab	497	39	Multi-class SVM	83.5%
	Rust	489	41		82.5%
	Healthy leaves	492	35		91.7%
[75] 2012	Downy mildew	30	20	GRNN or PNN	94.29%
	Powdery mildew	20	15		
[65] 2013	Downy mildew and Powdery mildew	29	2	Feed forward BPNN	100%
[108] 2014	Healthy leaves and Downy mildew	Total images = 31		Fuzzy set theory	87.09%
[107] 2016	Healthy leaves and Potassium deficiency	Total images = 50		k-NN based segmentation	Visual Examination
Orange					
[17] 2011	Canker, Black spot, Scab, and Melanose	N/A	N/A	SceBoost	88%
				RBN	73.25%
				k-NN	69.25%
				SVM	63%
Lemon					
[79] 2013	Greasy spot, Scab, Melanose and Healthy	80	80	k-NN	77.5%
				NB	95%
				LDA	98.75%
				RFT	97.5%

accuracy of nearly 83% on a large dataset of more than 1500 images. The results obtained for other fruits are not as good as those for **grapes**. Also *scab* is found to be the most studied disease followed by *spots*, *melanose*, and *mildew*. For this category of crops, although the best performance is reported with feed forward BPNN, but others like SVM, NB, LDA, random forest, and statistical analysis are also shown the potential of their applicability.

2.2.2 Vegetables

Chili, a high-risk horticultural good, generally gets affected by various diseases caused by bacteria, micro-organisms, and pests. An accurate and fast system for early detection of infected **chili** leaves is designed [27]. Instead of concentrating on some specific set of diseases, the system basically examines each plant on a healthiness scale. The criterion to measure healthiness is based on the percentages of a few colors in a leaf image. Tested on 107 samples, the system reports acceptable results. Although, the main focus is to reduce the usage of harmful chemicals by early recognition of potential problems in plants.

Next in this category are **cucumbers** which have valuable nutritional benefits, especially the hydrating properties. The production of this crop is commonly affected by *powdery mildew*, *downy mildew*, *brown spot*, *angular leaf spot*, *blight*, and *anthracnose* leaf infections. A study compares linear, polynomial, radial basis, and sigmoid kernel based SVM with artificial NN to identify *powdery mildew* and *downy mildew* [76]. Each of the four kernels is individually trained with color, texture and shape features. The highest recognition performance is reported by linear kernel based SVM, utilizing least number of vectors, in all the considered cases and the color features are shown to have the lowest running time. Similarly, the best result is achieved with color features followed by a combination of texture and shape. The experimental results show that SVM is more appropriate than ANN for efficient recognition of *powdery* and *downy mildew* in **cucumber** leaves. Another similar work presented a system to detect *angular leaf spot* and *brown spot* along with *downy mildew* [109]. Radial basis function based SVM generates higher recognition rates than sigmoid and polynomial kernel based SVMs. The study suggests that using image of each *spot* in an infected leaf during training improves the system efficiency. The applicability of ANN to detect *fungal* infections (*downy* *mildew* and *powdery mildew*) is proved by presenting an autonomous device [110]. It is based on Levenberg–Marquardt back-propagation algorithm; and perceives leaf symptoms using normalized thermal and textural parameters. The developed device is able to detect an infection as well as an hour post inoculation using images. In another such attempt PNN is trained using a 38 dimensional vector (24 color, 4 shape, 5 texture, and 5 meteorological features) [111]. The study focuses to enhance the recognition accuracy of *downy mildew*, *blight* and *anthracnose* infected leaves for images which are acquired under different environmental conditions. Achieving a recognition rate of 91.08% using 300 image samples shows the ability of combined features to successfully train the PNN.

Tomato is another important commercial crop which frequently gets infected and leads to low production quality. Common **tomato** leaf diseases are bacterial (*canker*, *spec*, *spot*), *anthracnose*, *fungal blight*, *viral curl*, etc. A system to automatically detect nitrogen and potassium deficiencies in **tomato** culture is presented using a unique combination of GA for feature selection and fuzzy k-NN for classification [55]. An optimum set of color-texture features is used to train fuzzy k-NN. The results present a classification accuracy of 90% and 85% for nitrogen and potassium deficient leaves. The study shows that the chosen feature set using GA gives more accuracy in comparison to the whole feature set. Moreover, a classification framework is presented in the form of a binary tree to identify a nutrient deficient leaf.

The developed system is claimed to successfully identify disease 6–10 days before the actual disease symptoms become visible to an expert. Another work uses a unique hyperspectral imaging concept to automate yellow leaf *curl* detection without visible scars [62]. Efficiency of the system to distinguish nine texture features of healthy and infected leaf samples computed in different spectrums is shown using receiver operator characteristic (ROC) curve. The study indicates that leaf edges are more prone to diseases than its midrib area. Accuracy of the system varies with the employed texture feature and ranges from 87.2% to 92.3%. Another study uses simple color descriptors to train 1NN classifier [74]. The main focus is to compare color structure, scalable color, and color layout descriptors to detect a mycotic infection (*early blight*). Using a nested-leave-one-out cross validation method, the study summarizes that the classification accuracy of color structure is superior to either of the other two descriptors for mycotic infection detection. The work also suggests usage of texture features for improved accuracy. Working on the same domain, another work compares the performance of SVM with different kernel functions: linear, RBF, MLP, and polynomial [97]. Instead of focusing a disease type, the system attempts to differentiate healthy-unhealthy **tomato** leaves using texture features. The presented system is highly efficient as it uses only 400 images in training and testing is performed on 800 image samples. The highest classification accuracy of 99.83% is achieved when SVM utilizes a linear kernel function.

Another work uses decision tree for classification of five diseases such as *bacterial canker, bacterial leaf spot, fungal late blight, septoria leaf spot, leaf curl*, and healthy leaves [112]. Average recognition accuracy of 78% is achieved. Another attempt to detect same diseases with different number of training and testing images is performed using fuzzy and BPNN approach [113]. However, BPNN provides the best average

Table 2.3 Summary of the horticulture (vegetables) crops.

[Reference] Year	Disease or healthy leaves or Deficiency	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
Chili					
[27] 2012	Healthy and diseased	Total images = 107		Color based analysis	Visual Examination
Cucumber					
[76] 2008	Powdery mildew and Downy mildew	Total images = 40		SVM-Linear	100%
				SVM-Polynomial	95%
				SVM-Radial	90%
				SVM-Sigmoid	95%
				ANN	78%
[109] 2010	Brown Spot	20	12	SVM-RBF	91.7%
				SVM-Polynomial	83.3%
				SVM-Sigmoid	75%
	Downy mildew	20	12	SVM-RBF	83%
				SVM-Polynomial	75%
				SVM-Sigmoid	66.5%
	Angular leaf spot	20	12	SVM-RBF	75%
				SVM-Polynomial	75%
				SVM-Sigmoid	66.7%
[110] 2013	Downy mildew, Powdery mildew, and Healthy	250	30	ANN	Visual Examination
[111] 2015	Downy mildew	150	150	PNN	90%
	Blight				92%
	Anthracnose				92%

Table 2.3 (Contd.) Summary of the horticulture (vegetables) crops.

[Reference] Year	Disease or healthy leaves or Deficiency	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
Tomato					
[55] 2011	Healthy, Deficiency (Nitrogen and Potassium)	120	120	Fuzzy k-NN classifier	82.5%
[74] 2014	Early Blight and Healthy	Total images = 147		1-Nearest neighbor	100%
[62] 2013	Yellow leaf curl and Healthy	Total images = 116		ROC curve	87.2% - 92.3%
[97] 2015	Healthy and Unhealthy leaves	200	200	SVM- Linear	99%
				SVM-Quadratic	98%
				SVM-RBF	88.50%
				MLP	94%
		400	400	SVM- Linear	99.83%
				SVM-Quadratic	99.25%
				SVM-RBF	93.25%
				MLP	95.25%
[112] 2016	Healthy	78	20	Decision tree	70%
	Bacterial leaf spot	104	26		69.2%
	Septoria leaf spot	104	26		80.7%
	Fungal late blight	104	26		69.2%
	Bacterial canker	104	26		84.6%
	Leaf curl	104	26		92.3%
[113] 2016	Healthy	58	58	BPNN	96.5%
	Bacterial leaf spot	32	32		93.3%
	Septoria leaf spot	14	14		64.7%
	Fungal late blight	22	22		54.5%
	Bacterial canker	22	22		100%
	Leaf curl	32	32		87.5%

accuracy of 87.2%.

The vegetable category of horticulture crops is well explored for **chili**, **cucumber**, and **tomato**. The same is shown by the summary of results provided in Table 2.3. SVM-Linear as well as 1-Nearest neighbor classifiers are observed to perform the best by reporting 100% accuracy in detecting the cucumber and tomato leaf diseases respectively. Various version of SVM are popularly explored in this group of crops followed by neural networks and nearest neighbor classifiers in sequence. It is clearly visible that more than 90% accuracy is achieved in nearly 50% of the studies covered in this literature. For **tomato**, it’s hard to determine the most explored disease; however for **cucumber** *mildew* is the most explored as well as the most correctly detected and classified leaf disease.

2.3 Floriculture Crops

Oil palm is another valuable crop that easily gets infected by several leaf diseases like *wilt*, *rots*, *streaks*, *blast*, *colored spots*, and *blight*, etc. These diseases are mainly caused due to virus, bacteria or nutrient deficiencies. A system able to suggest an appropriate fertilizer to cure several infections caused due of macro- and micronutrients deficiencies is developed [78]. Deficiencies caused due to Nitrogen, Phosphorous, Potassium, Boron, Magnesium, Manganese, and Zinc are considered. A fuzzy classifier is developed using color and shape information, and the rules are designed after interviewing the domain experts. The presented method is a nondestructive way to identify deficiencies, improve productivity, and optimize fertilizers usage. On the contrary, another work focuses on identification of those leaf diseases showing visual symptoms like hawar leaf, *anthracnose*, and leaf *spot* [95]. Each pixel identified as a spot is used to extract a color and shape feature vector. NN with different number of hidden neurons (3, 6, and 12) are trained using the extracted features. The best classification accuracy of 87.75% is achieved using a NN with 6 hidden neurons.

The economy of any country is dependent not only on the commercial crops but on a range of decorative crops too. Popular in this category are maple and hydrangeas which are characterized by large and heavy flower heads. Leaf diseases like *anthracnose*, *wilt*, *leaf spot*, and *powdery mildew* are common for this category of crops, occasionally leaf scorch is also observed. Combining techniques of NN and fuzzy logic, a two phase system is presented that automates the detection of leaf spot and leaf scorch diseases [7]. The first phase utilizes NN to identify a leaf as either maple or hydrangeas and the second phase does the actual disease classification using fuzzy logic. The system uses five fuzzy rules to grade severity of a disease. The experimental results show the effective of the presented system over manual recognition methods. Further the study may assist agricultural experts to automate several decisions like identification of correct pesticide, their quantity, etc.

Next important decorative crop is **phalaenopsis** which comes from a popular orchid family. Among all the leaf diseases, **phalaenopsis** is infected the most from *bacterial soft rot*, *phytophthora black rot*, and *bacterial brown spot*. For automatic detection of these diseases in initial stages of crop, a system that analyzes seedlings is developed [58]. The system utilizes two classifiers, Bayes for differentiating leaves from a pot, and BPNN for disease classification. BPNN trained with color-texture features reports a classification accuracy of 89.6% and a really nice infected leaf detection accuracy of 97%. The system is not able to detect covered blade infections but can be of great help to make automatic observations in greenhouses.

The past 10 years studies included in the literature for floriculture crops are reviewed in Table 2.4. The maximum accuracy of only 90.9% is reported for **Phalaenopsis**, which is almost 9% lesser in comparison to previous crop categories. Leaf disease detection and classification is not explored much for floriculture crops. *Spots* are cleverly explored in all the studies using either the concepts of neural networks or fuzzy logic. Although, the fuzzy logic implementation is supported by visual examination, but the results presented in the corresponding works are quite acceptable. As far as neural networks are concerned, BPNN outperforms others.

2.4. Food Grains

A large variety of food grains is studied in literature, for their effective presentation in the

Table 2.4 Summary of the floriculture crops.

[Reference] Year	Disease or Deficiency	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
Maple and Hydrangea					
[7] 2015	Leaf spot and Leaf scorch	20	8	Fuzzy logic ANN	Visual Examination
Oil palm plant					
[78] 2011	Nitrogen, Phosphorous, Potassium, Boron, Magnesium, Manganese, and Zinc	N/A		Fuzzy logic	Visual Examination
[95] 2013	Leaf spot, Anthracnose, and Hawar	102	22	Neural network	87.75%
Phalaenopsis					
[58] 2007	Phytophthora black rot Bacterial brown spot Bacterial soft rot	145	144	BPNN	88.8% 90.9% 88.8%

study a few sub-groups are thus formed. They are 1. **Clover** and **corn**. 2. **Legumes**. 3. **Rice** and **wheat**. The details of works identified in each of the three sub-groups are included in the following sub-sections.

2.4.1. Clover and Corn

Kruse et al. (2014) applies pixel classification method to detect ozone-induced visible injuries on **clover** leaves [72]. The study aims to identify an efficient and robust approach for leaf surface injury detection. The system individually uses four classifiers trained on different color and texture features. The classification approaches used are fit to a pattern multivariate image analysis combined with T2 statistics, residual sum of square statistics, k-means clustering, and linear discriminant analysis (LDA). Evaluations made using manually segmented images as the ground-truth prove that LDA is superior to other approaches. Other observations show computational efficiency and robustness of LDA to variable backgrounds as well as degrees of injury. Also, LDA trained on color feature is observed to detect leaf surfaces injury rapidly, thus is suitable for real-time applications.

A few of the hazardous **corn** leaf diseases are *leaf blight* (*turcicum*, *maydis*, *sheath*), banded *leaf*, *mildew* (*powdery* and *downy*), and *bacterial stalk rot*. Among these leaf blight and mildew are focused the most. A system to recognize **corn** leaf *spot* diseases, i.e., *leaf blight*, *sheath blight*, and *southern leaf blight* is designed using YCbCr color space [73]. After recognizing the infected region on the leaf, texture features using spatial GLCM are extracted to train a back propagation neural network (BPNN). The developed system is shown to classify **corn** leaf diseases with a success rate of 98%. In another work, a histogram feature based system to identify and grade *turcicum leaf blight* disease is developed [114]. The proposed methodology is shown to have the capability to identify different **corn** diseases. For example, the classification accuracies reported to detect *downy mildew* and *powdery mildew* are 83.5% and 95.2%, respectively. Another work uses the concept of locality sensitive discriminant analysis (LSDA) to develop the supervised, robust, and orthogonal nonlinear dimensionality reduction algorithm, named as orthogonal locally discriminant projection (OLDP), for plant disease detection [8]. The system is based on 1- nearest-neighbor classifier and is trained for five types of **corn** leaf diseases. Results presented on real (infected) leaf images of maize reports 93.42

classification accuracy using 18 training images. Similarly, one more system is developed using k-nearest neighbor (k-NN) classifier and trained with color, texture, and shape features of infected leaf regions for five types of **corn** leaf diseases [67]. The results are compared with three more systems having different classifiers, i.e., color-texture features with neural network, principal component analysis with neural networks, and Bayesian. The reported classification rate of more than 90% with 18 training samples shows the applicability of k-NN classifier over other combinations. A histogram based system combining a SRG and curvelet modulus correlation (CMC) algorithms is also developed to detect six **maize** diseases; they are *leaf blight*, *rust spots*, *gray leaf spot*, *curvularia leaf spot*, *brown patch*, and *small spot* [115]. A histogram template of diseases is used to detect the disease category of the target image. Using n-fold cross-validation algorithm, the system reports a classification accuracy of 94.45%.

Summary of the **clover** and **corn** food grains is given in Table 2.5. Clover leaf diseases are not explored much in literature. Among the results from available classifiers,

Table2.5 Summary of the food grains (clover and corn) crops.

[Reference] Year	Disease or healthy leaves	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
Clover leaves					
[72] 2014	Leaf surface injury and Healthy	Total images = 36		LDA K-means FPM-T^2 FPM-RSS	95% 93% 86% 85%
Corn					
[73] 2011	Leaf spot (Leaf blight, Sheath blight, and Southern leaf blight)	10	30	BPNN	98%
[114] 2012	Downy mildew Powdery mildew Normal	210 210 210	43 48 47	Histogram feature based	83.5% 95.2% 96.7%
[8] 2013	5 types of diseases	50 90	50 10	1- nearest-neighbor	78.32% 93.42%
[115] 2015	Leaf blight, Rust spots, Gray leaf spot, Curvularia leaf spot, Brown patch, and Small spot	Total images = 744		Feature based (SRG and CMC)	94.45%
[67] 2015	5 kind of diseases	90	10	k-NN PCA and NN Bayesian	90% 86.95% 89.25%

LDA provides the highest accuracy of 95% followed by K-means and statistical methods. For **corn** leaf diseases a range of classifiers are studied and BPNN is observed to achieve the highest accuracy of 98% in leaf *spots* detection considering a small dataset of 40 images. Surprisingly, the results reported using feature based classifiers prove their applicability in corn leaf disease detection. A histogram based system reports an average accuracy of 91.8% and another similar feature based (SRG and CMC) system reports the second highest average accuracy, i.e., 94.45% for **corn** culture. One more clear observation is the larger size of the dataset considered in feature based systems (253 to 744 images) as compared to other available works. This further clarifies the superiority of these systems. Also *spots* are found to be the most commonly studied **corn** leaf disease. Considering only the reported accuracy values, the BPNN can be seen as the preferred classifier for group of food grains as well.

2.4.2. Legumes

Digital color imaging finds its usage in diagnosing nutrient deficiencies in plants by observing changes in the color of a leaf. This also provides a method to automate a system. One such work attempts to identify macronutrient (Nitrogen, Potassium, Phosphorus and Magnesium) deficiencies in three legume species, viz. **pea**, **yellow lupine**, and **faba bean** [71]. Leaf color variations of the three species are observed in L*a*b* and HSI color spaces by means of Euclidean distances. Observations presented show that crop species direct the phenomenon of changes in color of a leaf caused due to some deficiency. Study reveals that the potassium deficiency is greatly visible in **pea** and **faba bean** where as **yellow lupine** responded to phosphorus deficiency the most. Another popular crop in this category is **soybean**. In recent years, various **soybean** diseases, including fungal (*brown spot*, *frog eye*, *rust* etc.), bacterial (*pustule* and *blight*), and viral (*bean pod mottle virus*) are explored for automatic detection. A few systems that can work on images captured in the field with different conditions are developed [60, 94]. Using mobile phone for image acquisition, one method detects and classifies only two **soybean** diseases, i.e., *brown spot* and *frog eye* [60]. For 50 testing samples, the k-NN classifier trained with shape based feature vector is shown to identify *brown spot* and *frog eye* with 70% and 80% accuracy. In addition, two more diseases *rust* and *bacterial blight*, are also detected in another work [94]. LDA is trained using a combination of structural texture and normalized DCT based features. The system reports an average classification accuracy of 89.9%. The proposed hybrid feature is said to classify other infections like *downy mildew* and *sudden death syndrome* well. Further, the system can be used for **rice**, **beans**, **cotton**, **fruits**, **vegetables**, etc. A system based on the concepts of reference histogram, correlations, and likelihood function is developed for automatic detection of nine diseases, namely, *bacterial blight*, *rust*, *phytotoxicity*, *stem canker*, *corynespora leaf spot*, *myrothecium leaf blight*, *downy* & *powdery mildew*, and *septoria brown spot* [116]. Results are presented in the form of a confusion matrix. Diseases with least confusion are *myrothecium leaf blight* and *downy mildew*. It is also observed that for good classification accuracy in a system dealing with many diseases consideration of external parameters is keenly required. A disease independent and level estimation method is developed based on three parameters, i.e., ratio of infected area, lesion color index, and damage severity index [117]. The applicability of the system is shown by accurately identifying various leaf diseases; they are *rust*, *bacterial blight*, *brown spot*, *sudden death syndrome*, *frog eye*, and *downy mildew*. An effective and fast disease detection method based on the techniques of local descriptors and bag-of visual words is presented [50]. Five common local descriptors (SURF, HOG, DSIFT, SIFT, and PHOW) are compared on a real-world dataset containing 300 healthy leaf samples and 900 samples infected from *mildew*, *rust tan*, or *rust RB*. The system uses SVM and is evaluated on correct classification rate (CCR) metric. The results prove the dominance of PHOW over others and its applicability to color spaces. The study also reveals that local descriptors can classify *mildew* more accurately in comparison to *rust RB* and *rust tan*. The discussed method is very general and can easily be used for other crops. Another study trains SVM with SIFT features to develop an autonomous decision support system [118]. The presented approach uses leaf shape to identify its species and can also classify the sample as healthy or infected. The system reports an average classification accuracy of 93.79%. The main focus of this study is to effectively assist farmers' as much as possible using minimal amount of input information, which in this case is only a mobile captured image.

Besides detection, a system to grade severity of a disease is developed for reducing the usage of pesticides or other pest control measure [29, 44, 48]. One system is based on neural networks and works for *downy mildew*, *frog eye*, and *bacterial pustule* infections with a classification accuracy of 93.3% [44]. Another system presented a severity grading system using k-means clustering to automatically detect diseases

Table 2.6 Summary of the food grains (pea, yellow lupine, faba bean, and soybean) crops.

[Reference] Year	Disease or healthy leaves or Deficiency	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
Pea, Yellow lupine and Faba bean					
[71] 2009	Deficiencies (Nitrogen, Potassium, Phosphorus, and Magnesium)	N/A	N/A	Color based analysis	Visual Examination
Soybean					
[60] 2014	Frog eye	50	22	k-NN	80%
	Brown spot	50	22		70%
[94] 2014	Rust	Total images = 57		LDA	100%
	Brown spot				100%
	Bacterial blight				75%
	Frog eye				84.6%
[116] 2015	Bacterial blight	38	18	Feature based	37.5%
	Rust	45	20		71%
	Phytotoxicity	16	7		73.9%
	Stem canker	15	7		9.1%
	Corynespora leaf spot	43	19		19.4%
	Myrothecium leaf blight	1	1		100%
	Downy mildew	31	15		91.3%
	Powdery mildew	52	24		61.8%
	Septoria brown spot	14	6		20.0%
[117] 2015	Healthy, Rust, Bacterial blight, Brown spot, Sudden death syndrome, Frog eye, and Downy mildew	Total images = 1000		Parameter based level estimation	Visual Examination
[44] 2015	Downy mildew, Frog eye, and Bacterial pustule	25	3	NN	93.3%
[29] 2015	Bacterial leaf blight, Septoria brown spot, and Bean pod mottle virus	N/A	N/A	k-means clustering	Visual Examination
[118] 2015	Healthy and diseased	Total images = 120		Feature based	93.79
[48] 2016	Rust	N/A	N/A	Grade based	Visual Examination
[50] 2016	Healthy, Mildew, Rust tan and Rust RB	Total images = 1200		SVM	98%

(*bacterial leaf blight*, *septoria brown spot*, and *bean pod mottle virus*) [29]. Efficacy of the system is evaluated by comparing the obtained results with those measured using a manual technique. One more system aims to study color distribution and pixel relationship at every stage of disease growth [48]. Observations are made for 25 days using local as well as global features of rust infected leaf images. Again percentage disease index (PDI) based on severity levels is computed to categorize a *rust* disease. The minimum PDI of 0.2 and maximum of 95.5 are observed on 6th and 25th day respectively. The study reveals that higher PDI indicates a decrease in spatial relationship among color and gray pixels due to lesser contribution of green color region.

Studies to detect and classify a variety of leaf diseases in legumes species such as **pea**, **yellow lupine**, **faba bean**, and **soybean** are summarized in Table 2.6. Considering 57 leaf images infected from *rust*, *brown spot*, *bacterial blight*, and *frog eye*, LDA is observed to be 100% accurate in detecting *brown spot* as well as *rust*. Similar observation is obtained for detecting *myrothecium leaf blight* using the concepts of reference histogram, correlations, and likelihood function. But nothing can be said much in later case as the dataset consists of only 2 images. On the contrary, the same concept reports 9% accuracy for stem canker detection using a bit larger number of images, i.e., 22. In all, this concept needs some more supporting observations for its strong recommendation in future studies. However, the results obtained with SVM classifier is the highest, i.e. 98% followed by NN and feature based systems which have reported approximately the same accuracy of 93%. Also automatic detection of various diseases is attempted in case of **soybean** culture, particularly *blight, rust, brown spots*, and *frog eye* are explored the most.

2.4.3. Rice and Wheat

Majority of the diseases in one of the most common food grain **rice** can easily be recognized by observing the appearances of *spots* around the infected areas. Moreover unbalanced mineral compositions cause several deficiencies which lead to a disease. Both of these causes are explored by the researchers so as to automate their detection process. A prototype system using BPNN is developed to identify Nitrogen, Iron, Magnesium, Potassium, Boron, and Manganese deficiencies in **rice** crop [46]. The system combines the outcomes of two BPNNs trained individually with color and texture features. The segmentation mappings obtained at the output layers categorize 88.56% of pixels accurately. Another variant of NN, i.e., self organizing map neural network (SOM-NN) is employed to detect *brown spot* and rice *blast* diseases [85]. The network is trained using gray feature values of the spots and is tested on RGB as well as Fourier transform features. The system performed better in case of RGB features. Results are also generated after transforming an image in the frequency domain, but they are inferior to those obtained with the original image. A faster version of NN, probabilistic neural network (PNN) trained using fractal texture descriptors are also explored [64]. The system reports good classification accuracy for four diseases, namely, *tungro* (97.96%), *leaf blast* (83%), *bacterial leaf blight* (96.25%) and *brown spot* (92.31%). The observations show that the color variability in leaf blast leads to its higher rate of misclassification. In other way, the method presented fails to differentiate among diseases having similar color characteristics. To achieve efficiency in such cases, other features like shape are also required.

SVM is also explored to identify **rice** bacterial *leaf blight*, *rice blight*, and *sheath blight* diseases [56]. Radial basis function based three SVM models trained individually on various features (texture, shape, and their combination) are presented and compared for efficient identification of *bacterial leaf blight*, *rice blight*, and *sheath blight*. Maximum overall classification accuracy of 97.2% is observed when combination of features is employed. On the other hand minimum overall classification accuracy is achieved with shape features, due to instability in shape of rice *blight* and *sheath blight* spots. The study thus recommends the usage of shape and texture features for accurate disease detection not only for **rice**, but for other crops too. Another work developed a two stage system and presented a comparison of Bayes classifier with SVM [40]. At first stage, the system identifies if the sample is healthy or not. In infectious case, the second stage classifies the sample as *brown spot* and *blast*. The system accurately recognizes 92% of healthy samples and performs better for *brown spot* than *blast*. Besides, the classification accuracy of Bayes (79.5%) is found to be higher than that of SVM (68.1%). The study also reveals the time efficiency of Bayes over SVM. Another work compares the performance of k-NN with SVM using an automatic disease detection and classification system [119]. The identification phase utilizes Haar features to train AdaBoost classifier that reports a detection accuracy of 83.33%. The second phase does the required comparison of SIFT trained k-NN and SVM in classifying test leaf samples as *brown spot*, *leaf blast*, and *bacterial blight*. This study also reports that k-NN (93.33%) is better than SVM (91.10%). Moreover, the authors claimed that the proposed system is helpful in early identification of disease.

Contrary to this, a set of membership functions is designed to automatically identify rice *sheath blight*, *brown spot*, and rice *blast* diseases [52]. The system utilizes the nearest neighbor classification concept to put a test sample in the appropriate diseaseclass. It is fast and provides good results with reasonably high-quality images. The proposed system recognizes *brown spot* (85%) more successfully as compared to other two, although the reported average classification accuracy is 70%. Another study interviewed agricultural experts and obtained some production rules with forward chaining method to detect *brown spot*, *narrow brown spot* and *blast* diseases [92]. The main focus of the study is to reveal the importance of threshold in local entropy threshold and Otsu method segmentation techniques. The developed system achieves 94.7% classification accuracy using local entropy threshold as it deals effectively with different intensities and illumination issues. A novel idea of Fermi energy is introduced to segment an image and then a rule generation algorithm using classification rule mining techniques is presented [6]. The developed rule based classifier reports a classification accuracy of 92.29% for identifying *brown spot* and *blast*. Comparison done with various traditional classifiers, like C4.5, NB, Part, Kstar, SMO and bagging, further proves the efficacy of a rule based system in accurate detection of plant diseases. Capability of fuzzy c-mean clustering algorithm is also utilized to detect blast fungal disease and the associated production loss for **rice** crop [120]. This pixel based approach separates the leaf region into three classes: healthy pixels, medium and highly affected by blast pixels. The main focus of this study is to estimate the loss of production due to *blast* instead of its detection. The system reports 85% accuracy and is meant to assist farmers for precision farming using decision support systems.

Wheat, the third most produced food grain, is observed to get affected generally by any form of *rust* diseases. Common types of wheat rust are *powdery mildew*, *stripe*, *septoria leaf spot*, *tan spot*, and *snow mold*. Researchers have proposed several solutions for automatic and accurate classification of these *rust* categories at early stages. In one such attempt, a combination of various color, texture, and shape features are used to train different neural networks; they are BPNN, radial basis function NN (RBF-NN), generalized regression NN (GRNN), PNN [75]. All the NN are compared on accurately classifying the *stripe rust* and leaf *rust* fungal diseases. The observations show that the least accuracy is reported by RBF-NN and the optimum performance is achieved when BPNN is trained on features which are reduced with PCA. The fitting and prediction accuracy of the presented system is 100%. Usage of PCA is optional in this system, but mandatory in case disease recognition is performed via Internet. In addition, the study also suggests replacing PCA with some other dimension reduction method in case non-linear features are employed. Improved rotation kernel transformation based directional feature, termed as IRKT is developed to classify stripe *rust* and *powdery mildew* [66]. Experimental results show that IRKT is noise insensitive and provides better edge related information. Compared to edge orientation histograms (EOH), the proposed feature classifies stripe rust (97.5%) more correctly. However, both EOH and IRKT are shown to report same classification accuracy for *powdery mildew* (100%). Overall it can be summarized that the proposed directional feature, IRKT can successfully be used to recognize a range of **wheat** diseases. Two systems are developed to detect and recognize four types of rust, viz., *powdery mildew*, *septoria leaf spot*, *tan spot* and *snow mold* [70, 121]. The system presented using fuzzy c-means is simple, fast, and focuses on identifying a set of best suited features [121]. First phase of the system separates a set of diseased leaf images and second phase classifies a test sample in one of the four disease classes. This system reports a recognition accuracy of only 56% and also requires images of all types of infections during training. To resolve this issue along with and improved accuracy, another system based on BPNN is presented which is trained on a combination of color and texture features [70]. The system estimated 290 out of 342 test samples accurately. The improved system also attempted to rate the severity of *rust* infection in addition to diagnosis. The applicability of the system is proved through manual examination by two experienced doctors.

Table 2.7 summarizes all the studies considered in the literature for **rice** and **wheat** food grains. For both the food grains, neural network based classifiers are shown

Table 2.7 Summary of the food grains (rice and wheat) crops.

[Reference] Year	Disease or healthy leaves or Deficiency	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
Rice					
[46] 2007	Healthy and Deficiency of Boron, Iron, Magnesium, Manganese, Nitrogen, and Potassium	N/A	N/A	BPNN	88.56%
[85] 2008	Blast and Brown spots	N/A	N/A	SOM-NN	92%
[56] 2009	Rice spot (Blast, Sheath blight, and Bacterial leaf blight)	Total images = 72 36	36	SVM	97.2%
[52] 2009	Blast Sheath blight Brown spot	Total images = 50		Nearest neighbor classification	80% 60% 85%
[92] 2009	Blast, Narrow brown spot, and Brown-spot	Total images = 94		Production Rules with forward chaining method	94.7%
[120]2011	Healthy and Blast	N/A	N/A	Fuzzy c-mean clustering	85 %
[40] 2012	Brown spot and Blast	Total images = 1000		Bayes SVM	79.5% 68.1%
[64] 2013	Bacterial leaf blight Tungro Leaf blast Brown spot	Total images = 40		PNN	96.25% 97.76% 83% 92.31%
[6] 2013	Brown spot and Blast disease	N/A	N/A	Rule mining technique	92.29%
[119] 2016	Brown spot, Leaf blast and Bacterial blight	90	30	k-NN SVM	93.33% 91.10%
Wheat					
[75] 2012	Stripe rust and Leaf rust	60	40	BPNN	100%
[66] 2014	Stripe rust and Powdery mildew	Total images = 200		IRKT	97.5%
[121] 2014	Four types of Rust (Powdery mildew, Septoria leaf spot, Tan spot, and Snow mold)	Total images = 310		Fuzzy c-means	56%
[70] 2015	Four types of Rust (Powdery mildew, Septoria leaf spot, Tan spot and Snow mold) and Healthy	20	342	BPNN	84.8%

to achieve the best results. In particular, BPNN reports the highest accuracy of 100% in detecting **wheat** leaf diseases (*stripe rust* and *leaf rust*) for a 100 images dataset. Similarly, using much lesser images (only 40), PNN is observed to detect *tungro* rice leaf disease with 97.76% accuracy. Researchers have reported an accuracy > 60% in case of rice and > 50% in case of wheat to correctly identify as well as classify diseases. As far as the diseases are considered, *blast* and *brown spots* are equally studied for **rice** leaves; also *rust* is explored in all the works focused on wheat leaves. Considering the number of images, the best performing classifier is Bayes for **rice** and BPNN for **wheat** both have reported accuracies in the range of 80% – 85% with larger datasets. Here also BPNN can easily be observed to outperform other classifiers.

2.5. Assorted Cultures

There exist a range of works that concentrate on automatic detection of a general leaf disease affecting a group of cultures. All such works use dataset that contains images of leaves infected from some specific diseases irrespective of their culture. One such work utilizes the techniques of image processing to detect and classify five leaf diseases, namely, *late scorch*, *early scorch*, *ashen mold*, *small whiteness*, and *cottony mold* [69]. The system trains a 10 hidden layer based feed forward BPNN using optimized color-texture feature set obtained from infected leaf region. Five such models considering various color components (HSI, HS, H, S, and I) are compared. The model using HS components reports a maximum classification accuracy of 89.5%. The study also shows that the computational complexity improves if an intensity component (I) is not considered. Another work proposes a 2D Fourier transform based wilting index for early detection of temporary wilting caused due to drought stress [122]. Inspired from leaf morphology, the proposed index depends on the curvatures of the points on the surface of 3D laser scanned leaf images. The applicability of the wilting index is shown using **zucchini** leaves and may also be suitable for **Cucurbita pepo** leaves as both the species have same leaf shape. However, the generalization of index is questioned due to variability of plants, their leaf shapes, and wilting morphologies. The study motivates all the researchers in this domain to explore leaf morphology in 3D space to extract physiologic information using mathematical tools. An intelligent and specialized image sequence capture device is integrated to capture a series of images to automate spore detection [84]. The set of images thus obtained are processed and identified as *powdery mildew spores* using BPNN. The proposed approach counts the number of *spores* after detection. Using 155 training samples BPNN reports 95.5% accuracy, but the accuracy obtained for 89 testing samples is only 63.6%.

Infections may affect any part of a plant. Based on this fact, various fungal diseases are examined in different crop categories, viz. fruit, vegetable, commercial, and cereal using separate models [83]. The work considers infections in leaf, stem, and fruits of various cultures in the four categories; vegetable crops examined are **beans**, **bengal gram**, **soybean**, **sunflower**, and **tomato**; commercial crops examined are **chili**, **cotton**, and **sugarcane**; cereal crops examined are **jowar**, **wheat**, and **maize**; and fruits examined are **grape**, **mango**, and **pomegranate**. The model presented for vegetable crops uses local binary patterns of both sides of a segmented leaf to analyze several infections (*anthracnose*, *blight*, *rust*, and *mildew*). For this category neuro-k-NN classifier, a combination of BPNN and k-NN is introduced and compared with ANN. Neuro-k-NN (91.54%) is shown to report better average classification accuracy. PCA reduced discrete wavelet transform (DWT) features are utilized to train Mahalanobsis distance based- and PNN classifiers for commercial crops. PNN classifier reports an average classification accuracy of 86.48% to detect *anthracnose*, *rot*, *powdery mildew*, *alternaria leaf spot*, *smut*, *gray mildew*, and *wilt* infections. In contrast to these models, the one developed for cereal crops employs different combinations of color, texture, and shape features to train SVM. At the first stage radon transform differentiates between a healthy and diseased plant followed by SVM classification to identify an infection as *leaf blight*, *leaf spot*, *powdery mildew*, *leaf rust*, and *smut*. The results presented shows that the training done with color-texture features is the most appropriate for this category as it reports the maximum average classification accuracy of 85.33%. The study considers the infections affecting fruits as well. The corresponding model achieves an average classification accuracy of 94.085% and classifies the test sample as normal or infected (partial, moderate, and severe). The presented architecture can be used in remote monitoring of crops to detect diseases at early stages. Moreover, this work is effective yet complex and challenging due to varying outdoor conditions. Another work from the same author compares the performance of ANN with SVM on different features (color, texture, and their combination) [28]. Both the systems are trained using 900 images taken from plant pathology department, Dharwad. They can identify an agriculture crop test sample as fungal, bacterial, nematodes, viral, deficiency and normal. ANN models report an average classification accuracy ranging from 82% to 87%, but those based on SVM performs better for all the features. The maximum and minimum classification accuracies achieved by SVM models are 84% and 92%, respectively. Here as well, SVM performs the best when trained with a combination of color and texture features. The performance of SVM trained using texture features is also studied for other cultures to detect a range of diseases [68]. The approach is designed to detect *blight*, *sun burn*, *scorches*, *spots*, *bacterial*/*fungal* infections, and *mold* in leaves of different cultures (**banana**, **beans**, **jackfruit**, **lemon**, **guava**, **mango**, **potato**, **sapota**, and **tomato**). Trained on H image based texture features, the minimum distance and SVM classifiers report accuracies of 86.77% and 97.74%, respectively. The study again made the same observation of "SVM is better". In contrast with the trend, one study analyzes the effect of Salmonella Typhimurium (human pathogenic bacterium) on immune system of **Arabidopsis** leaves [123]. Although the study uses the popular as well as successful linear SVM which is trained using color features. The system classifies infected foreground region pixels with 95.8% accuracy and additionally refines the final image using neighborhood-check method. The presented system is shown to provide accurate results for the considered dataset and can also be extended to detect other diseases. An effective blend of color features (histograms and transformations) with pairwise-correlation based classification is also presented for disease recognition [124]. Dataset used to design the system contains 82 biotic and abiotic stresses of 12 plant species. The performance is evaluated by means of confusion matrix which is prepared for each plant species. The obtained results are not very impressive and can be improved by considering some measures related to capturing of images. All the required points of consideration are also discussed thoroughly by the author [13]. This work is very similar to the one discussed in Section 2.4.2 for **soybean**.

Focusing uneducated farmers, a human-mobile interface (HMI) is designed recently that can assist them in automatic examination of fields at any phase, just on a mobile click [9]. The initial steps (pre-processing and segmentation) are implemented at the client mobile device, and the remaining steps (feature extraction and classification) are performed on the pathology server. The final classification result is intimated to the user using short messaging service (SMS). In the present state, the interface runs only on Android operating system using concepts of Gabor, GLCM, and k-NN classifier. It is able to identify infections as leaf *spots* and leaf *blotch* with a classification accuracy of 93%.

Similar to this, a deep convolution NN model for crop disease diagnosis on large scale using a smart-phone is presented [42]. The study focuses on popular AlexNet and GoogLeNet architectures. The applicability of the designed system is proved using a subset of PlantVillage dataset containing 54,306 images of 14 crop species infected from 26 diseases. As per the expectations, the approach reported an accuracy of 99.35%. Moreover, on an online collected heterogeneous image dataset this system achieves 31.4% accuracy. The study suggests the usage of diverse training set to attain feasible results in case of general purpose large datasets. Working on the same domain of deep convolutional NN (CNN), a system to recognize 13 types of leaf diseases is developed using a popular Caffe framework [125]. The system is tested using a dataset of 30,000 images prepared after suitable transformations of more than 3,000 original images collected online. For all the considered diseases the system reports an overall classification accuracy of 96.3%, more specifically the obtained values lie in the range of 91.11% and 98.21%. The study reveals that augmentation process is more important than fine-tuning to obtain desired overall accuracy.

The current research trend is much focused in the development of a system that is capable of detecting and classifying a range of diseases over a variety of cultures. To highlight this observation, Table 2.8 presents the summary of all the studies covered for assorted cultures. A gradual increase in the number of diseases being identified by a system is clearly visible in Table 2.8. As far as classifiers are concerned, systems under this category have explored neural networks and SVM the most. SVM is observed to achieve more than 90% accuracy in any of the cases, but different types of neural networks have shown varied performances with accuracies in the range of 31% to 100%. Also neural networks further prove their applicability for larger datasets. It is evident from Table 2.8 that deep convolution neural networks can effectively be used in systems dealing with some thousands of images (as large as 55,000). Although BPNN has reported 100%, but in the absence of the size of the dataset, 96.3% accuracy reported by deep convolution neural networks for 33,469 images is observed as the best.

Table 2.8 Summary of the assorted cultures

[Reference] Year	Disease or healthy leaves	Number of images		Classifier	Performance measure (Accuracy)
		Training images	Testing images		
[69] 2010	Early scorch	N/A	N/A	BPNN	100%
	Tiny whiteness				88%
	Ashen mold				100%
	Cottony mold				96%
	Late scorch				80%
[84] 2013	Powdery mildew spore	155	89	BPNN	63.6%
[68] 2013	Early scorch, Brown spots, Late scorch, and Yellow spots	Total images=500		Minimum distance	86.77%
				SVM	94.74%
[123] 2014	Healthy and unhealthy leaves	Total images=1200		SVM	95.8%
[83] 2015	Anthracnose	N/A		BPNN	80.62%
	Anthracnose, Powdery mildew, and Downey mildew			Nearest Neighbor	94.08%
	Anthracnose, Powdery mildew, Downey mildew, Early blight, Late blight, and Rust			Neuro- k-NN	91.54%
	Anthracnose, Fruit rot, Powdery mildew, Alternaria leaf spot, Gray mildew, Smut, Fusarium wilt, and Red rot			PNN	86.48%
	Leaf blight, Powdery mildew, Leaf spot, Leaf rust, and Smut			SVM	90.83%
	Leaf blight, Powdery mildew, Leaf spot, Leaf rust, and Smut			SVM	91.16%
[9] 2016	Leaf spots and Leaf blotch	Total images=297		k-NN	93%
[28] 2016	Fungal, bacterial, Nematodes, Viral, and Deficiency	450	450	SVM	93%
				ANN	84%
[125] 2016	13 diseases	30880	2589	Deep convolution NN	96.3%.
[42] 2016	26 diseases	Total images=54306		Deep convolution NN	31.4% - 99.35%
[124] 2016	82 diseases	Total images=1335		Feature based	58%
				Feature based (manual)	63%

2.6. Overall Analysis

Image acquisition and size of the database: The leaf images database corresponding to a particular type of infection in a specific culture is very difficult to obtain. This fact is clearly visible by the limited size of the image databases used in the studies. Only a few works (less than 10%) have used large size databases which are ranging in thousands [5, 40, 42, 50, 117, 123-125]. Also, the ratio of training is to testing images, is varying a lot from one work to another. A common observation is to use a large proportion of the database images during training than that of the testing phase. However, some exceptions are always there [70, 73, 101]. It is observed that an efficient acquisition of a leaf image is the need of an hour. If taken in real world scenario (i.e., an uncontrolled environment) then its acceptance would automatically be increased. As per the current state-of-the-art works, images taken using a mobile are slowly gaining the popularity. Another issue of concern is the association of a few numbers of varying stages with some of the leaf infections, which complicates the process of image acquisition further. Although, some complicated yet effective image acquisition techniques as well as a single click image systems are presented, but much more is still supposed to be done. The concept of using a leaf back can also be considered in sensor based system for proper detection of an infection in early stages. Because of all these reasons, a transition from the consideration of a specific culture disease to a disease common to a set of cultures can easily be observed in the domain of plant leaf disease detection systems. This transition may help in overcoming several issues related to the database size.

Techniques employed in pre-processing, segmentation, and feature extraction modules: The nature of database images plays an important role in selecting appropriate techniques to perform the task of pre-processing and segmentation efficiently. Among various available ones, those which are suitable for a particular form of acquisition usually serve the purpose. A large variability span is observed in the algorithms available under different modules. Similar observations are made for a feature extraction module as well. In other words, the standardization of techniques is yet to be achieved. This observation is palpable in an automatic plant leaf disease detection system as it is a type of content based image retrieval system only. It is also observed that proposing a universal technique for an individual module is very difficult in case of these types of systems. As per the current trends, the dependency of techniques employed in any module of the system is very high on the database being used by that system.

Difficulties in the classification module: Automation of plant leaf disease detection and classification system is been focused by the researchers since a very long time. Highly acceptable results on a few numbers of images are reported in some of the studies. Also a range of classifiers are explored in this domain. As observed in the preparation of this literature, the classifiers back propagation neural network, support vector machine and linear discriminant analysis perform better across all the cultures followed by random forest tree, feature based, Naive Bayes, probabilistic neural network, k-nearest neighbor, multi layer perceptron, and rule mining. Talking specifically, then among all the considered systems 41% have used either SVM or a feature based classifier. Both these classifier types are used equally in the past 10 years. The next classifier that is used popularly in 17% of the articles is BPNN and k-NN is utilized in 14% of the studies. The remaining classifiers are used in the rest 28% of the articles and are excluded from this discussion due to their lesser number. Recently, the deep convolution neural networks are used in systems working with assorted cultures. They are yet to be explored in systems pertained to a single culture. Proper utilization of convolution networks may help in improving the effectiveness of a system on large databases.

Limitations of available systems: Bock et al. [126] provides literature which mainly focus on how leaf disease severity is calculated by using different methods (visual rating, image analysis, and hyperspectral imaging) in early 1900s to 2010. The paper provides review of how quantity of leaf disease is measured visually by raters using different scales. But due to its numerous disadvantages, image analysis and hyperspectral imaging techniques are used. However these techniques also have some drawbacks. The efficiency of any system depends on the quality of training data used in the training phase; indirectly it is actually the number of training images and their extracted features. So it can be said that a well trained system is highly efficient. But, all the existing systems have a well-defined set of requirements which are essential to be fulfilled for their accurate performance [127]. If one or the other constraint is not fulfilled, then the considered system may produce inaccurate results which lead to inappropriate disease detection. For example, the problem of over-training or over-fitting is commonly observed in the studied systems that employ the powerful techniques of NN, SVM, and GA improperly. In such a scenario, researchers must think of hybrid and adaptive systems which can work on various sets of requirements instead of a single one. Otherwise, some generalized techniques which can work on a group of heterogeneous environments must be developed. Also an in-depth knowledge of several techniques as well as proper usage of classy tools cannot be compromised for efficiency. All these issues of concern actually fall in the area of domain adaptation, which in itself is a very popular research problem these days.

Disease stage quantification: Another related area of research is quantification, i.e., detecting the infected proportion of a culture. This research objective is particularly important as it controls the amounts of pesticides or other chemicals to be used for disease prevention. In the present scenario, chemicals are applied periodically without any prior analysis of the infection or the quantification. This practice may have harmful effects on human health. Effective application of image processing methods would help in determining if the chemicals are required or not. In case of an accurate quantification the analysis would further be used to control the quantity of chemicals to be applied.

Development of new applications: Various solutions do exist in literature, but the corresponding system is not available for a public use. Only a few Web portals and mobile based applications are accessible to provide an online assistance for a specific disease set of a particular culture. To the best of our knowledge, Leaf Doctor and Assess software are available but they work on images with black background only [104]. So the development of an online system, for plant disease detection and then classification, may also form another research objective in this domain. The availability of any such software would help farmers to a great extent. In the near future, these systems may replace the requirement of specialist suggestions in the initial stages of infections. For helping farmers in the remote area, the system may also provide an option of “analysis report generation” which can further be send to an expert for getting proper suggestions.

Real world application: Most of the works presented in the literature considers images that are collected offline by picking leaves which is somewhat destructive. The research conducted till date identifies diseases under some specified conditions. To the best of our knowledge, none of the studies perform disease identification in real world scenario with acceptable accuracy. The real world implementation of these studies may be attempted to get a practical method. For example, one may think of a real time system which uses the concept of continuous remote monitoring of a field area for disease identification. This research objective is also related directly to several computational complexity and memory requirement issues.

As per the literature, it can be easily observed that dataset required for leaf disease identification and classification is very difficult to attain. Also systems are mainly dependent on the dataset acquisition and environmental conditions. So an attempt has been made to develop frameworks that can work across different datasets irrelevant of these conditions. Various studies are performed for different crops for disease detection and classification. As soybean crop is among one of the top ten staple foods as importance as discussed in Section 1.1. So the main focus in this research work is soybean crop and an online PlantVillage dataset of legume plant i.e. soybean is considered for the same. A rule-based system as well as a generic semi-automatic system is planned to be developed. As per analysis of various research conducted in literature, it has been observed that majority of studies developed offline systems that cannot be utilized by other users. So, web application is also developed to share the system online. Frameworks are not destructive in nature. Developed framework is flexible to work using best performing classifiers. The developed frameworks should be flexible enough and has capability to work with different algorithms or techniques. Another important concern is non-destructive nature of these frameworks i.e. achieved by making them workable in fields. Thus a web-application is developed and one can easily operate it using any smart phone.

Chapter 3

PROPOSED RULE-BASED SYSTEM

From the literature survey presented in Chapter 2, it is evident that the development of automatic/semi-automatic disease detection and classification system is significantly explored in precision agriculture. In the past few decades, researchers have studied several cultures exploiting different parts of a plant. Working on the same lines, this chapter presents a semi-automatic system for soybean using leaf images. A rule-based system using concepts of k-means is designed and implemented to distinguish healthy leaves from diseased leaves. In addition, a diseased leaf is classified into one of the three categories (*downy mildew*, *frog eye*, and *septoria leaf blight*). Experiments are performed by separately utilizing color features, texture features and their combinations to perform training in three different scenarios based on SVM classifier. Results are generated using thousands of images collected from PlantVillage dataset. Acceptable average accuracy values are reported for all the considered combinations which are also found to be better than existing ones. An attempt has also been made to discover the best performing feature set for leaf disease detection in soybean. The proposed system is shown to efficiently compute the disease severity as well. Visual examination of leaf samples further proves suitability of the proposed system in detection, classification, and severity calculation.

In this chapter system architecture is explained in detail. Section 3.1 provides the overall architecture of the proposed rule-based system and gives an insight of all the basic steps involved in various system phases. Section 3.2 provides a simple equation to calculate the percentage of infected leaf area. The performance of the system in terms of qualitative i.e. visually as well as quantitative i.e. accuracy measures discussed in Section 3.3. Finally, Section 3.4 presents the chapter summary.

3.1 Overall Architecture of System

Fig. 3.1 demonstrates complete process flow of the proposed rule-based system. The basic preprocessing module is followed by an important segmentation module. Color and texture features for each segmented cluster are extracted to train the system. The training phase is completed in two stages. The first stage classifier learns distinguishing features between a healthy and an unhealthy soybean leaf image samples. Classifiers in the second stage learn from features of infected clusters to classify a leaf image sample into one of the three disease categories. While training, observations are made and some rules are

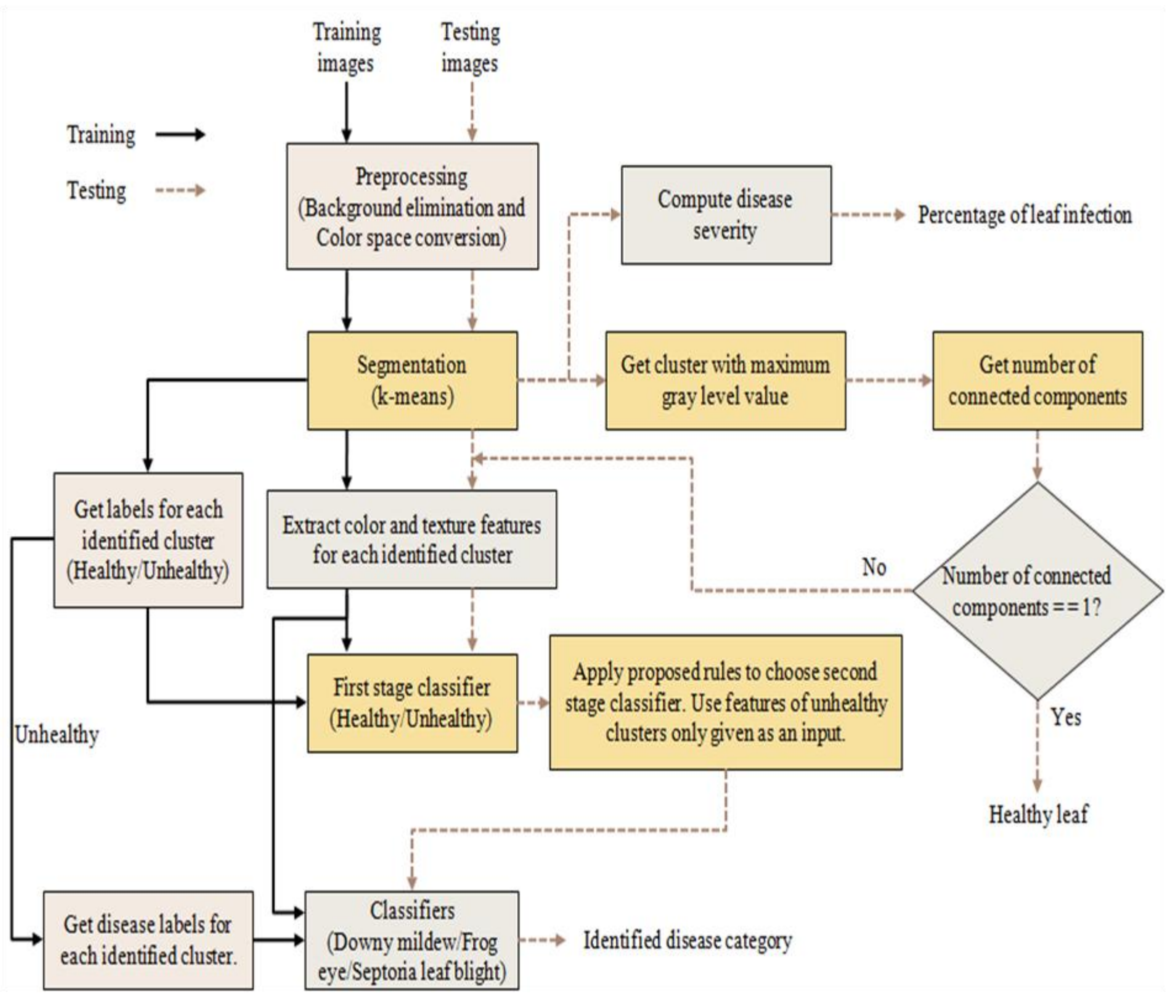


Fig.3.1 Process flow of the proposed leaf image sample recognition and disease classification system.

designed to ease classifier selection during testing. The testing phase starts with the same two initial modules. Among all the segmented clusters, if the one with maximum gray level value contains single connected component, then the process stops giving healthy leaf as an output. In case this condition holds false, then depending upon the output of first stage classifier a matching rule is referred to select one of the classifiers from the second stage. Output of the selected classifier gives the disease infecting the test leaf image sample. Also, the amount of infected leaf region is computed using another module and is provided as severity percentage. Algorithms employed in main modules of the proposed method are discussed in the following subsections. As the proposed rules are dependent on the type of images contained in the dataset, thus a description of the dataset used during experiments is already included in advance, i.e. Section 1.6 of Chapter 1.

3.1.1 Pre-processing

The complex leaf image background is first removed to get a clear leaf sample before further processing. A simple method used for this is given in Fig. 3.2. First the region of interest (ROI) is selected manually followed by the computation of binary mask and then its complement. Addition of the original image to the obtained complement removes irrelevant background. Finally, cropping eliminates extra whitespaces and completes background elimination process.

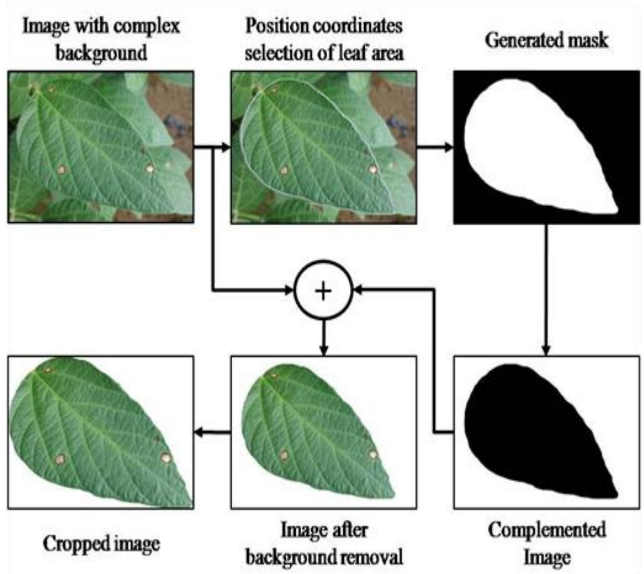
The next preprocessing step is color space conversion. The color space is converted from a device dependent RGB model into a device independent model. The proposed system uses $L^*a^*b^*$ (L^* signifies lightness, a^* and b^* are chromaticity layers) color space. It closely resembles human perception and also splits information about chrominance better than other models [128]. The background eliminated resized RGB image is converted into $L^*a^*b^*$ first then segmentation module is executed on "a^*b^*" channel. Using only two channels for color representation decreases the processing time as well. Moreover, $L^*a^*b^*$ has strong aptness towards good segmentation results as compared to other color models [129].

Input: An input image *I*.

Output: A cropped image *I* without complex background.

```
Manually select ROI in I and store it in R.
// Generate binary mask.
    Set J to I.
    for each pixel (m, n) in I
        if pixel (m, n) is in R
            Set J(m, n) to 1.
        else
            Set J(m, n) to 0.
  Compute complement J '.
  Set I to I + J '.
  Crop I to eliminate unwanted image
```

(a) Background elimination method.



(b) An illustrative example.

Fig.3.2 Background elimination method and an illustrative example.

3.1.2 Segmentation and Rule Formation

Segmentation is an important step in the proposed system, as its output is observed to design rules for classifier selection. A well-known k-means clustering algorithm is utilized to separate infected and healthy leaf regions. The iterative k-means reassign each pixel to nearest cluster so as to decrease the sum of distances and recalculate cluster centroids [130, 131]. This process results in an input image partitioned into three[1] clusters each containing different portions, also termed as regions here after, of a leaf image sample. Index values corresponding to the three clusters are used to label image pixels in the original image. Gray level values of the resultant three color clusters are then utilized for further processing. The concept of thresholding is successfully explored in various studies that deal with detection of diseases showing a particular color symptom [117]. However, diseases considered in this work demonstrate large color variation (from pale green to yellow or gray to brown) and such heterogeneity in colors may remain undetected by thresholding approach. Thus, the proposed system works using some rules as listed in Table 3.1.

[1] Experiments are also performed by partitioning leaf into two as well as four clusters. With two clusters the average classification accuracy is less than 40% and with four clusters the improvement is not too much. Thus, number of clusters is fixed to three.

Table 3.1 The proposed rules for disease identification and classification.

Category	Rule #	Cluster 1	Cluster 2	Cluster 3
Healthy	**Rule 1:**	ROI	Healthy region	Healthy region
Frog eye	**Rule 2:**	ROI with disease region	Healthy region	Healthy region
Septoria leaf blight	**Rule 3a:**	ROI with disease region	Healthy region	Healthy region
	Rule 3b:	ROI with disease region	Disease region	Healthy region
	Rule 3c:	ROI with disease region	Disease region	Disease region
Downy mildew	**Rule 4a:**	ROI with disease region	Disease region	Healthy region
	Rule 4b:	ROI with disease region	Healthy region	Disease region

These rules are framed after closely observing the formed clusters. In the current implementation, each leaf is divided into three clusters, namely **Cluster 1**, **Cluster 2**, and **Cluster 3** which may occur in any order. A particular rule is applicable depending upon cluster types (healthy or diseased) irrespective of their orders. However, in the study **Cluster 1** represents the highest, **Cluster 2** represents intermediate, and **Cluster 3** represents the lowest gray level value region.

Rule 1 signifies a healthy leaf sample, in which **Cluster 1** contains only a single connected component, i.e. ROI and the other two clusters are simply the healthier ones having variations of green color due to different chlorophyll level. Similar observations are made for *frog eye* infected leaf sample but in this case **Cluster 1** has more connected components denoting gray to brown spots which generally fall in same cluster (***Rule 2***). ***Rule 3a***, ***Rule 3b***, and ***Rule 3c*** are in correspondence to the symptoms of three stages of *septoria leaf blight* fungal attack. ***Rule 3a*** is for the first stage which contains only tiny spots lying in one cluster and other two clusters are again of type healthy. The second stage is identified by tiny spots surrounded with yellow color. In this case spots and yellow symptoms fall separately, thus there are two infected clusters and one healthy cluster containing the green color (***Rule 3b***). ***Rule 3c*** signifies the third stage in which leaf gets covered completely with tiny spots and yellowish brown color variations. The two infected clusters are similar to ***Rule 3b***, but instead of the healthy cluster there is an infected one containing yellow to brown color variations. Lastly, for *downy mildew* two out of three are infected clusters, one having tiny brown spot and other contains pale yellow to green lesion region. The said observations can easily be seen in Fig.3.3.

It is observed that **Cluster 1** should always be an infected one and either of the remaining two clusters be infected, thus two rules are considered here (***Rule 4a*** and ***Rule***

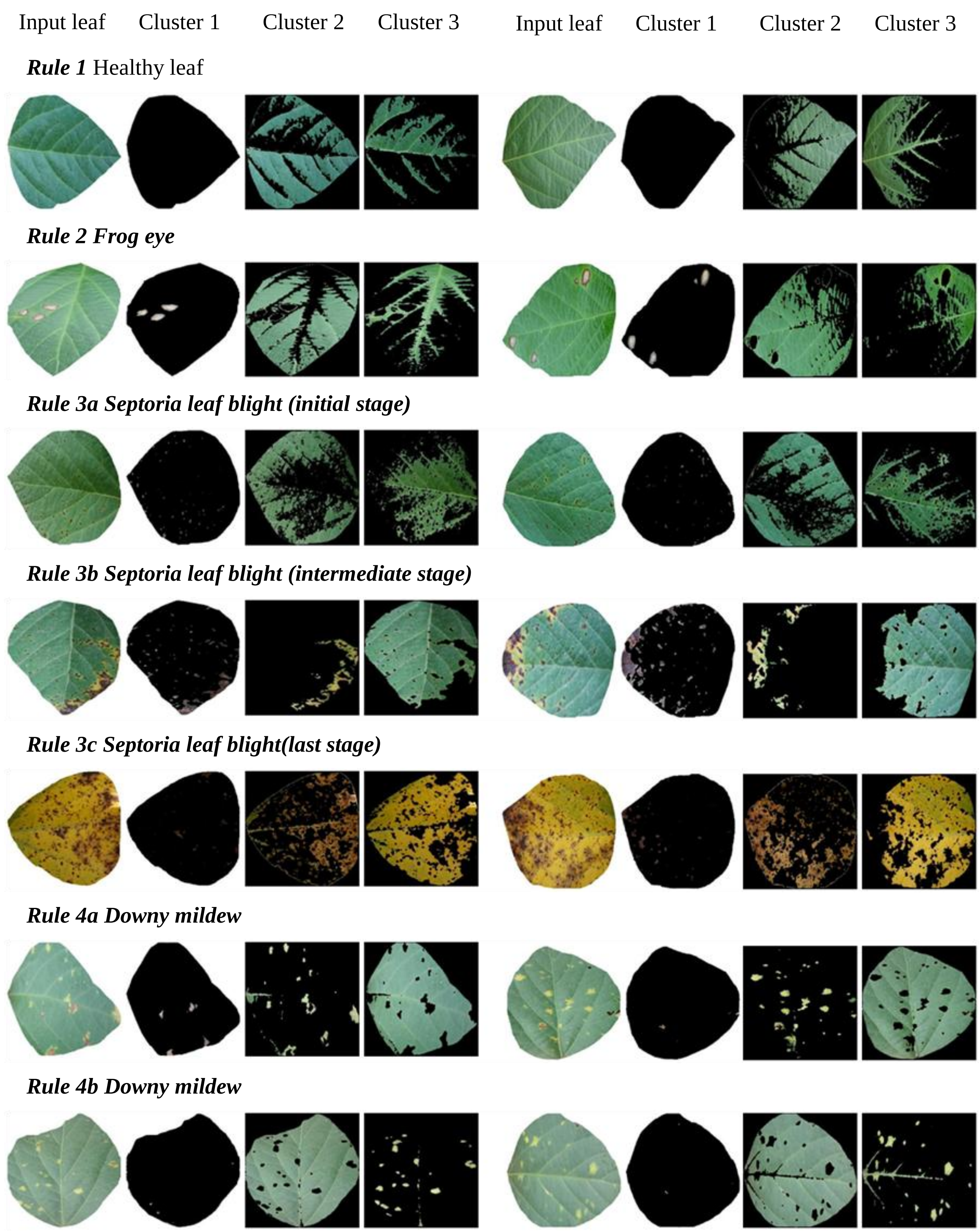


Fig.3.3 Clusters obtained after segmentation of various leaf samples

4b). Fig. 3.3 shows two examples for each of the proposed rules for clearer explanations. The confusion among rules listed in Table 3.1 can easily be observed. ***Rule 2*** and ***Rule 3a*** show similar scenarios but for different infection, i.e. *frog eye* and *septoria leaf blight*.

Similarly, there is a conflict between *septoria leaf blight* and *downy mildew* due to overlapping of ***Rule 3b*** with ***Rule 4a*** or ***Rule 4b***. These confusions are resolved by utilizing separate classifiers for the two cases which are trained accordingly and then chosen effectively to classify disease during testing.

3.1.3 Feature Extraction

Literature for plants disease detection conveys that color and texture play an important role in disease lesion classification [93]. Thus, several color features (Color moments, autocorrelogram, and HSV histogram), texture features (Haralick features, Gabor features, wavelet features), and their combinations are explored in this study to design a valuable system and to validate its performance. List of all the features examined is summarized in Appendix A. During initial observations different patterns are being observed for same disease classes considered in this study, thus shape features are completely eliminated in experimental evaluation of the proposed system.

3.1.4 Purpose and Training of Classifiers

The dataset used is imbalanced thus for improved classification accuracy the proposed system uses three classifiers to determine the following issues. One classifier is to identify a healthy or infected cluster and other two are utilized to resolve the clashing among rules as discussed in Section 3.1.2. $\mathbf{CLFR_{HD}}$ learns healthy and infected clusters, $\mathbf{CLFR_{FE\_SLB}}$ differentiates *frog eye* from *septoria leaf blight*, and $\mathbf{CLFR_{DM\_SLB}}$ classifies between *downy mildew* and *septoria leaf blight*. $\mathbf{CLFR_{HD}}$ uses features of all the clusters, $\mathbf{CLFR_{FE\_SLB}}$ and $\mathbf{CLFR_{DM\_SLB}}$ are trained from the features of infected clusters only.

Training process starts with labelling, i.e., healthy or diseased, of the three clusters obtained after segmentation. Methods used to automatically identify clusters as **Cluster 1**, **Cluster 2**, and **Cluster 3** are shown in Fig. 3.4. As mentioned earlier, **Cluster 1** contains region with the highest gray level value, **Cluster 2** has an intermediate gray level value region, and **Cluster 3** holds the lowest gray level value region. Based on the proposed rules following points are observed. In case a leaf sample is healthy, then **Cluster 1** contains only one connected component. However, based on initial analysis

Input: A cropped image I without complex background.

Output: **Cluster 1**, **Cluster 2**, and **Cluster 3** respectively stored in $I_{max_{g1}}$, $I_{max_{g2}}$, and $I_{max_{g3}}$.

Assumptions:

I_k and G_k be the color and gray images corresponding to the k^{th} cluster obtained after segmentation.

$Max_k = max(max(G_k))$, is the maximum gray level value in the k^{th} cluster.

// **Cluster 1**
if $(Max_1 > Max_2)$ && $(Max_1 > Max_3)$
    $I_{max_{g1}} = I_1$
elseif $(Max_2 > Max_1)$ && $(Max_2 > Max_3)$
    $I_{max_{g1}} = I_2$
else
    $I_{max_{g1}} = I_3$
// **Cluster 2**
if $(Max_1 < Max_2$ && $Max_3 > Max_2 \,\|\, Max_3 < Max_2$ && $Max_1 > Max_2)$
    $I_{max_{g2}} = I_2$
if $(Max_3 < Max_1$ && $Max_2 > Max_1 \,\|\, Max_2 < Max_1$ && $Max_3 > Max_1)$
    $I_{max_{g2}} = I_1$
if $(Max_2 < Max_3$ && $Max_1 > Max_3 \,\|\, Max_1 < Max_3$ && $Max_2 > Max_3)$
    $I_{max_{g2}} = I_3$
// **Cluster 3**
if $(Max_1 < Max_2)$ && $(Max_1 < Max_3)$
    $I_{max_{g3}} = I_1$
elseif $(Max_2 < Max_1)$ && $(Max_2 < Max_3)$
    $I_{max_{g3}} = I_2$
else
    $I_{max_{g3}} = I_3$

Fig.3.4 Method used to identify Cluster 1, Cluster 2, and Cluster 3.

Cluster 1 is found to always contain diseased region if a leaf sample belongs to any of the disease class. However, **Cluster 2** is healthy if leaf image is healthy or infected either from *frog eye* or first stage of *septoria leaf blight*. For second stage of *septoria leaf blight* this cluster is identified as diseased region. Lastly, **Cluster 3** is labelled as healthy for healthy leaves as well as for *frog eye* disease, but for third stage of *septoria leaf blight*, it’s found to be a diseased region. This automatic process does not work in case of *downy mildew* because infectious pale yellow region appears either in **Cluster 2** or in **Cluster 3**. Thus, semi-automatic labelling of clusters is followed. After appropriate tagging of clusters three classifiers, $\mathbf{CLFR_{HD}}$, $\mathbf{CLFR_{FE\_SLB}}$, and $\mathbf{CLFR_{DM\_SLB}}$ are trained using features of the required clusters. The particulars of the classifiers employed in this work are discussed in Section 3.3.

3.1.5 Testing

The proposed system is examined using a step by step process as is shown in Fig. 3.5. Test image samples are first gone through initial modules (pre-processing and segmentation) in a manner similar to that of training. The three clusters are then marked as **Cluster 1**, **Cluster 2**, and **Cluster 3**. If number of connected components in **Cluster 1** is one, i.e. a background, then the tested leaf sample is healthy. In case **Cluster 1** has more connected components then the sample is infected and need to be classified into one of the three disease types using rules listed in Table 3.1 and classifiers. After feature extraction first stage classifier, **CLFR**$_{\mathbf{HD}}$ identify each of three clusters as either healthy or diseased region. If two clusters are healthy, then disease type is obtained by classifying **Cluster 1** using **CLFR**$_{\mathbf{FE\_SLB}}$. For two diseased clusters, there are two possibilities. In addition to **Cluster 1**, if **Cluster 3** is diseased then the output is *downy mildew*. In other case, disease type is identified after classifying **Cluster 2** using **CLFR**$_{\mathbf{DM\_SLB}}$. If all the

Input: Cluster 1 ($I_{max_{g1}}$), **Cluster 2** ($I_{max_{g2}}$), and **Cluster 3** ($I_{max_{g3}}$) of a test image I.

Output: Classification of I as HEALTHY, DOWNY MILDEW, FROG EYE, or SEPTORIA LEAF BLIGHT.
Assumptions: h and d are counters for healthy and diseased clusters respectively.

Binarize image $I_{max_{g1}}$ and get *numCC*, i.e., number of connected components.
if (*numCC* = = 1)
    Leaf I is HEALTHY.
else
    Set $h = d = 0$
    for k = 1 to 3
        Compute feature vector fv_k of $I_{max_{gk}}$.
        Classify fv_k using **CLFR**$_{\mathbf{HD}}$ to mark $I_{max_{gk}}$ as HEALTHY or DISEASED.
        if ($I_{max_{gk}}$ is healthy)
            h++
        else
            d++
    if (h = = 2) // two HEALTHY clusters.
        Classify fv_1 using **CLFR**$_{\mathbf{FE\_SLB}}$ to detect leaf I as FROG EYE or SEPTORIA LEAF BLIGHT.
    elseif (d = = 2) // two DISEASED clusters.
        if ($\mathrm{I}_{\mathrm{max}_{g3}}$ is marked as DISEASED)
            Leaf I is infected from DOWNY MILDEW.
        else
            Classify fv_2 using **CLFR**$_{\mathbf{DM\_SLB}}$ to detect leaf I as DOWNY MILDEW or SEPTORIA LEAF BLIGHT.
    elseif (d = = 3) // all DISEASED clusters.
            Leaf I is infected from SEPTORIA LEAF BLIGHT.

Fig.3.5 Method used to identify the type of disease.

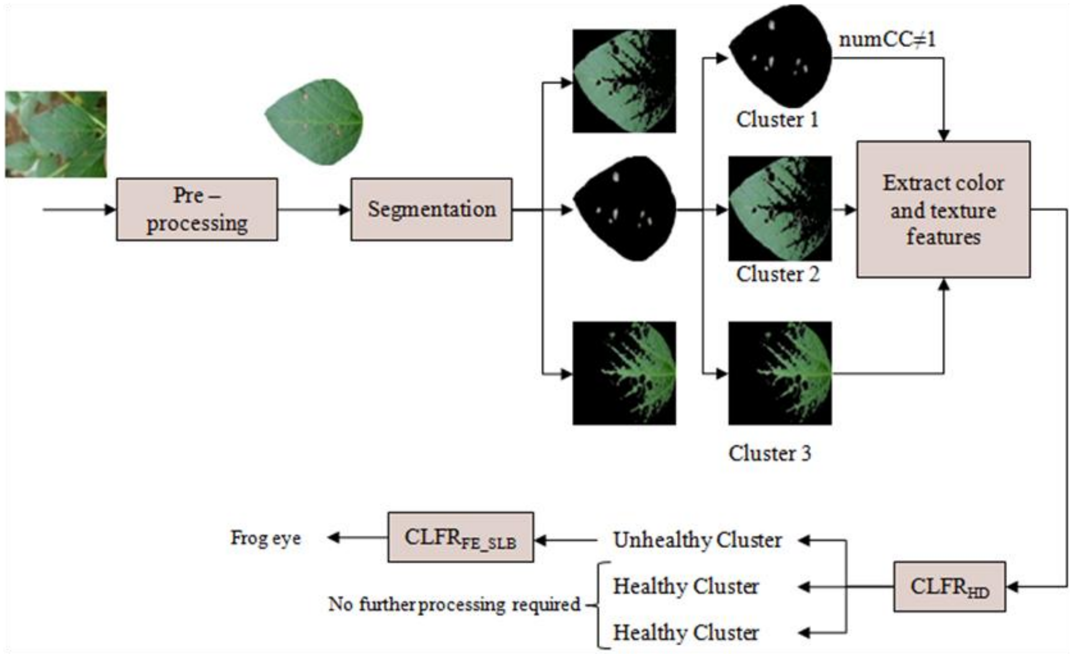


Fig.3.6 Flowchart to identify the type of disease.

clusters are diseased, then the test leaf sample is infected from *septoria leaf blight*. An illustrative example for a test leaf image infected from *frog eye* is shown in Fig. 3.6. Clearly, **Cluster 2** and **Cluster 3** are healthy regions; and **Cluster 1** contains more than one connected component. The proposed system identified this leaf correctly as *frog eye*.

3.2 Disease Severity Computation

The severity of a disease depends upon the region being covered by an infection. After identifying the test leaf image sample, percentage of leaf infection can easily be computed as a ratio of number of infected pixels is to number of healthy pixels in the leaf image [141]. In this study the complete leaf ROI is segmented into three clusters, thus disease severity is calculated using an expression given in Eq. (1). Here, A_i are the number of infected image pixels in **Cluster *i*** (where i = 1, 2, 3) and A_L represents total number of pixels lying in the extracted ROI of a leaf image.

$$A = \frac{\sum_{i=1}^{3} A_i}{A_L} \times 100 \qquad (1)$$

Table 3.2 Summary of the three training/testing scenarios used to perform experiments.

Class \ Scenario	# Images	Scenario 1		Scenario 2		Scenario 3	
		Training (50%)	Testing (50%)	Training (60%)	Testing (40%)	Training (70%)	Testing (30%)
Healthy	1079	539	540	647	432	755	324
Downy mildew	105	53	52	63	42	73	32
Frog eye	1662	831	831	997	665	1163	499
Septoria leaf blight	1929	964	965	1157	772	1350	579
Total	**4775**	**2387**	**2388**	**2864**	**1911**	**3341**	**1434**

3.3 Performance of the System

For rigorous examination, images collected from PlantVillage for the four classes' viz. healthy, *downy mildew*, *frog eye,* and *septoria leaf blight* are randomly divided between training and testing phases in three different ways as is given in Table 3.2. The three scenarios are individually trained and tested using color features, texture features, and their several combinations to detect and classify a leaf image into one out of four classes. Any classifier can be chosen, but in the current implementation SVM is used due to its numerous advantages in high dimensional space (Appendix A). Various binary classifiers are tested such as k-NN, SVM, RUSBoost, LogitBoost, and Decision tree. SVM performs better than other classifier which can be observed from results shown Table 3.3 as well as visualized in Fig. 3.7. Performance of the system is presented in terms of classification accuracy and the results obtained in each of the three SVM based scenarios for different features are shown in Table 3.4.

Literature of disease classification indicates texture as an important feature [93], but the proposed system for three soybean crop diseases is observed to give better classification accuracy using a set of color features in each of the three scenarios. However, the texture features selected in this study identify *frog eye* very efficiently but do not work for *septoria leaf blight* and *downy mildew*. Also, looking at the results for combined feature set, then again combinations considering all color features (**Color + Gabor (CG), Color + Haralick (CH)**, and **Color + DWT (CD)**) have reported better average classification accuracy ranging from 80.5% to 85% in all the scenarios. On the other hand, combining individual color features with a set of texture features (**Texture + Autocorrelogram (TA)**, **Texture + HSV Histogram (TH)**, and **Texture + Color**

Table 3.3 Classification results obtained using color and texture features for rule-based system (All values are in percentage)

Scenarios	**Downy mildew**	**Frog Eye**	**Healthy**	**Septoria leaf blight**	**Average Accuracy**
k-nearest neighbor					
Scenario 1	62.2	58.4	79	61.6	65.3
Scenario 2	66.7	59.1	80	62.5	67.07
Scenario 3	68.7	59.7	82	62.9	68.32
Decision tree classifier					
Scenario 1	73.5	83.9	79	81.4	79.45
Scenario 2	76.1	85.1	80	81.7	80.72
Scenario 3	75	85.3	82	83.9	81.55
Random undersampling boosting					
Scenario 1	78.1	83.8	79	83.9	81.28
Scenario 2	80.9	85.6	80	84.2	82.68
Scenario 3	81	86.9	82	84.4	83.57
Logitboost					
Scenario 1	77	85.2	79	86.2	81.85
Scenario 2	78.5	86.9	80	88.4	83.45
Scenario 3	81	87.7	82	89.1	84.95
Support vector machine					
Scenario 1	77	87	79	89	83
Scenario 2	77.5	90	80	89.5	84.25
Scenario 3	79.9	90.7	82	90	85.65

moments (TC)) seems to degenerate the performance in all the cases. For these three combinations, the reported average minimum accuracy is 55.25% (**Scenario 1**) and average maximum accuracy is 65.95% (**Scenario 3**) only. Uniting both the feature sets completely has provided the best average classification accuracy (**Scenario 1** - 83%, **Scenario 2** - 84.25%, and **Scenario 3** - 85.65%). These best results are found very much close to **Color + Gabor (CG)** combination and then to a mix of **Color + Haralick (CH)** features. This shows that one can compromise with lower dimensional feature set if accuracy is not that much important. Although in disease detection systems accuracy is of prime importance, thus it is suggested to work with the **Color + Texture (CT)** which results in a 203 dimensional feature vector.

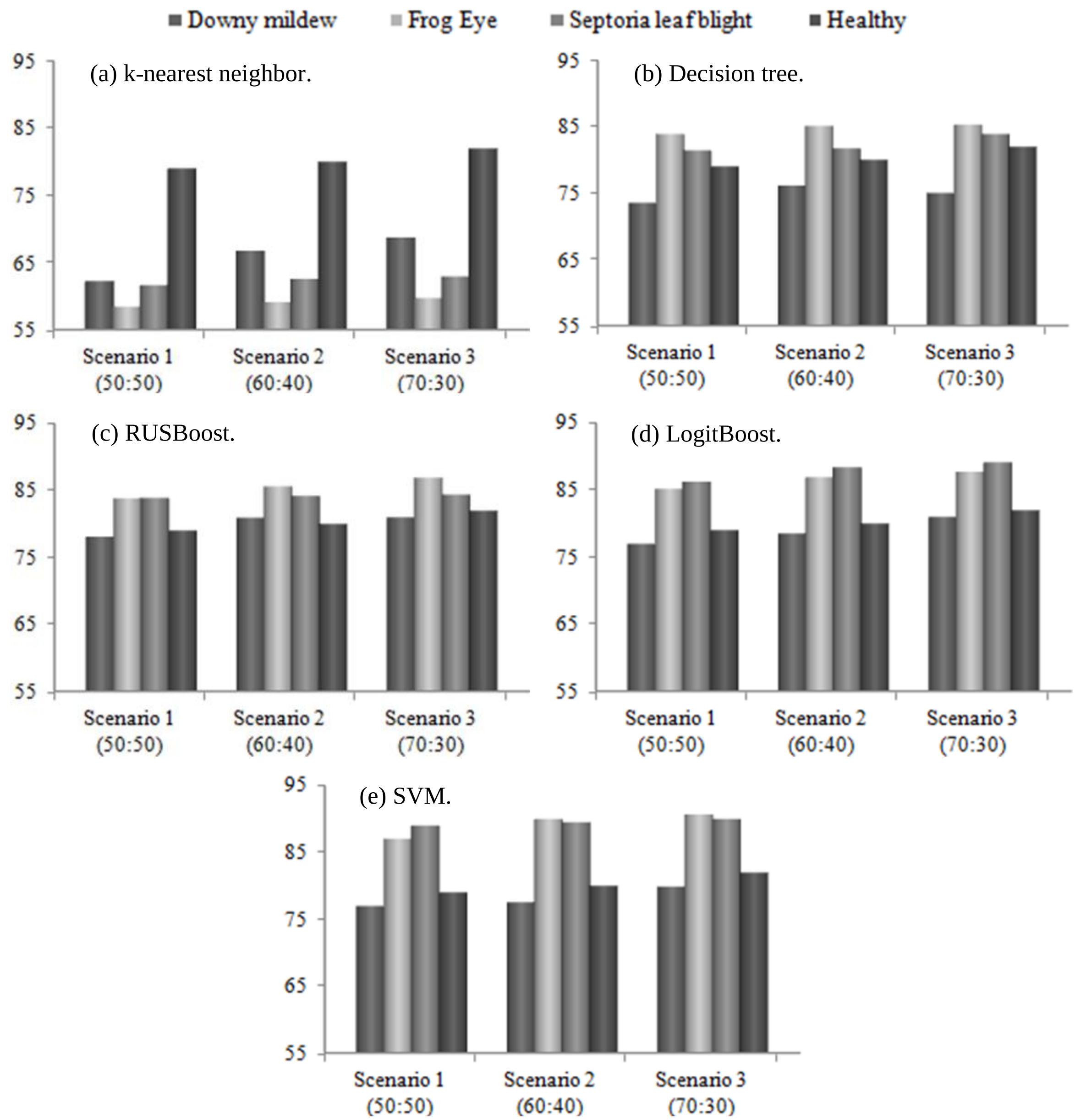


Fig.3.7 Classification results obtained using color and texture features for rule-based system with different classifiers.

Another obvious observation is increase in number of training images improves classification accuracy of the system (Fig. 3.7). The results obtained in **Scenario 1** are the least and those reported in **Scenario 3** are the best for all considered feature combinations. Next observation is related to healthy leaf sample, for this category the classification accuracy is independent of the feature set employed. Same results are obtained in all the scenarios because decision is being made on the number of connected components that is unrelated to features.

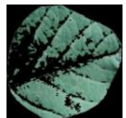
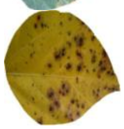
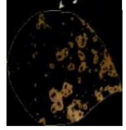
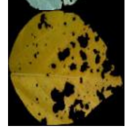
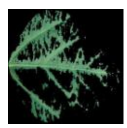
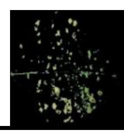
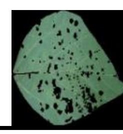
Table 3.4 Classification accuracy results obtained in different scenarios. All values are in percentage.

(C: Color, T: Texture, CT: Color + Texture, CH: Color + Haralick, CG: Color + Gabor, CD: Color + DWT, TA: Texture + Autocorrelogram, TH: Texture + Histogram, TC: Texture + Color moments)

	C	**T**	**CT**	**CH**	**CG**	**CD**	**TA**	**TH**	**TC**
Dimension	**101**	**102**	**203**	**115**	**149**	**141**	**166**	**134**	**107**
Scenario 1									
Healthy	79	79	79	79	79	79	79	79	79
Downy mildew	71	30	77	74	75.5	75	41	40	22
Frog Eye	82	90	87	85.8	87	83	90	81	75.5
Septoria leaf blight	87	42	89	88	88	85	40	21	50
Average	**79.75**	**60.25**	**83**	**81.7**	**82.38**	**80.5**	**62.5**	**55.25**	**56.63**
Scenario 2									
Healthy	80	80	80	80	80	80	80	80	80
Downy mildew	71.9	54	77.5	74.5	76	77.3	42.8	41	23
Frog Eye	84.9	92	90	86.7	89	84.5	93	95	88.8
Septoria leaf blight	87.7	43	89.5	88.2	88.5	86.7	40.7	22.6	50.7
Average	**81.13**	**67.25**	**84.25**	**82.35**	**83.38**	**82.13**	**64.13**	**59.65**	**60.63**
Scenario 3									
Healthy	82	82	82	82	82	82	82	82	82
Downy mildew	72	54.5	79.9	75	79	78	46.8	43	28
Frog Eye	85	95	90.7	88.5	90	85.5	94	95.9	89.1
Septoria leaf blight	87.9	43.9	90	88.5	89	87	41	24	51
Average	**81.73**	**68.85**	**85.65**	**83.5**	**85**	**83.13**	**65.95**	**61.23**	**62.53**

On the basis of results listed in Table 3.4, top three combinations of features that reported maximum classification accuracy are selected. The best three candidates are **Color + Texture (CT)**, **Color + Haralick (CH)**, and **Color + Gabor (CG)**. Intermediate clusters formed, classification, and disease severity for each of them are shown in Table 3.5 using leaves from each considered class (i.e., healthy, *downy mildew*, *frog eye,* and *septoria leaf blight*). Again, mixture of **Color + Texture (CT)** features provides accurate results as compared to other two feature sets. Observing the results shown in Table 3.4

Table 3.5 Results of the proposed system for a few images from PlantVillage dataset in Scenario 3.

Leaf	**Clusters**			**Results**			
	Cluster 1	**Cluster 2**	**Cluster 3**	**Color + Texture (CT)**	**Color + Gabor (CG)**	**Color + Haralick (CH)**	**Severity**
Healthy							
				Healthy	Healthy	Healthy	0
Septoria leaf blight							
				Septoria leaf blight	Septoria leaf blight	Frog eye	0.4
				Septoria leaf blight	Frog eye	Septoria leaf blight	8.3
				Septoria leaf blight	Septoria leaf blight	Septoria leaf blight	100
Frog eye							
				Frog eye	Frog eye	Frog eye	1.5
Downy mildew							
				Downy mildew	Downy mildew	Septoria leaf blight	15

and Table 3.5, it can correctly be said that the proposed system is capable to distinguish a healthy leaf image from a diseased leaf image. In addition it can accurately classify a diseased leaf image into one of the three classes most of the time. It can efficiently compute the severity too as can be verified manually by looking at the leaf image. For example, consider Table 3.5, image in Row 1 is healthy and that at Row 4 is infected from third stage of *septoria leaf blight*. As per the study percentage of infected leaf area should respectively be 0% and 100%; the same results are registered by the proposed rule-based system.

3.4 Summary

This chapter describes how image processing techniques are effectively utilized to find leaf class and severity. It is important to detect whether a leaf is healthy or diseased. Once detected, the disease needs to be classifed. Three different soybean disease classes are detected in this study, i.e., *downy mildew*, *frog eye*, and *septoria leaf blight*. Currently, three scenarios with different ratios of training is to testing images are explored. The proposed system utilizes three SVM classifiers, although it is flexible to work with different classifiers as well. Based on several combinations of color and texture features, classification is performed using the proposed rules. It has been observed that 203 dimensional vector formed by combining all the considered color and texture features provides accurate results than individual features. The proposed rule-based system is found to be better on many criteria as compared to existing studies. The maximum average classification accuracy reported is approximately 90% using a dataset of 4,775 images. Also, visual examination of the test images confirms the acceptance of the proposed system. However, the system is designed only for soybean and based on some rules. Next chapter thus aims to design a system that can work across different plants for detection of several disease classes.

Chapter 4

PROPOSED SEMI-AUTOMATIC SYSTEM

The rule-based system proposed in Chapter 3 is database dependent due to the formation of the specific rules. Though, automatic leaf disease detection systems capable to work across variety of cultures are being developed since long to gain significant results. Thus the main focus of this chapter is to design a generalized framework that can detect a leaf image as healthy or unhealthy and, in case of diseased sample classify its type. The basic idea to develop this system is to eliminate rules completely as discussed in Chapter 3. A two-stage framework based on the concepts of two-/multi-class classifiers and k-means clustering is proposed to reduce complexity which increases with the number of diseases being identified. Classifiers are trained on numerous features (texture and color). The framework may employ fusion in case of more than one classifier and is flexible to work with the best performing classifiers. The chapter also discusses a fusion method to combine the results logically. The performance of the proposed system is tested on six different datasets formed with leaf images of legumes, vegetables, fruits, and commercial crops which are collected using Google image search engine. The results for the base dataset i.e. PlantVillage are presented in this chapter and for the remaining datasets are included in Chapter 5.

Section 4.1 provides the overall architecture of the proposed semi-automatic system along with all the details involved in classifier selection. It also includes basics of other system modules as well as the process followed in the development of a generic framework i.e. capable to detect disease classes irrespective of the culture type.

4.1 Overall Architecture of Framework

The complete process flow of the proposed semi-automatic framework is shown in Fig. 4.1. It's a two stage framework, first-stage differentiates between healthy and diseased leaf sample followed by second-stage which is solely meant to detect and classify the unhealthy leaf sample. The initial stages of this framework are same as in Chapter 3 starts with pre-processing step which is basically meant for background elimination, if a leaf image has complex background. Next is segmentation to obtain different clusters of an input leaf which are to be examined separately in later steps of the proposed system. As observed earlier, color and texture features are found to work well, thus this framework is also designed using a combination of multiple color and texture features. Color and

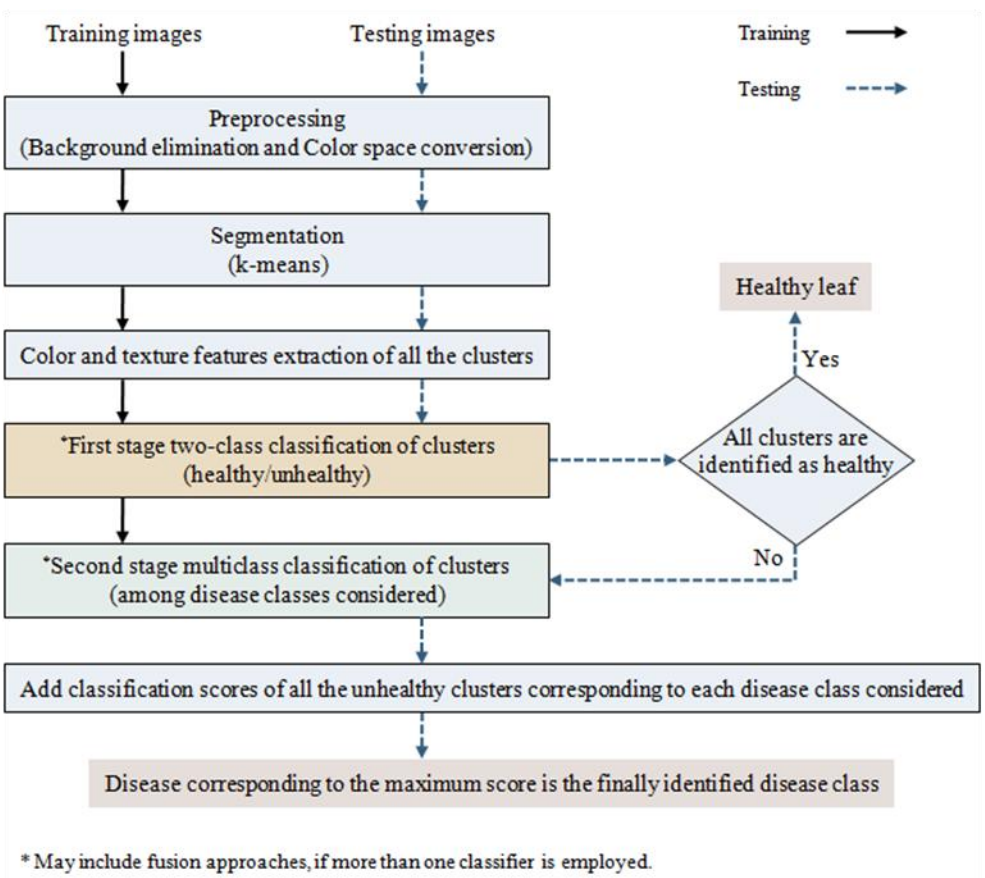


Fig.4.1 Process flow of the proposed semi-automatic generalized framework for plant leaf disease detection

texture features for each segmented cluster are extracted to train the system.

Training is split into two stages. The first-stage two-class classifier(s) is meant to separate a healthy cluster from an unhealthy one. However, the second-stage multi-class classifier(s) are trained from features of infected clusters only for each of the disease category considered. In case any stage has a set of classifiers then fusion approaches can be employed to get a single output label which is then used to mark a cluster as healthy/unhealthy (first-stage) or with a particular disease label (second-stage). Testing phase starts with the same three initial modules, i.e. preprocessing, segmentation, and features extraction. If first-stage classification identifies all the clusters as healthy, then leaf sample is considered as a healthy leaf. Otherwise, features of unhealthy clusters are passed to second-stage and a classification score is computed for each of the considered disease category corresponding to all unhealthy clusters of a test leaf image. Summation of classification scores gives the final score for a disease category and one with maximum score is the final identified disease class for the test leaf image.

As mentioned here an attempt has been made to develop a generalized framework for disease identification using plant leaf images. This framework is shown in Fig. 4.1. It is tested for four groups of crop categories (legume, vegetable, fruit, and commercial). The results are given in Chapter 5. While developing the model for any category all the phases may implement either same or different techniques. The work presented here considers same techniques for all the categories in various phases of the proposed framework. However classifiers to be used in each of the two stages of the proposed framework have to be decided and the process adopted for the same is discussed in Section 4.1.2.

4.1.1 Pre-processing, Segmentation, and Feature extraction

The same pre-processing, segmentation, and features (color and texture) extraction modules as used in rule-based system are employed. For each type of leaf disease 10% of the training samples are observed to determine the rank of cluster(s) which are diseased or healthy. Once a particular rank is marked as diseased or healthy, the same is followed for clusters obtained with remaining 90% of the images from that disease category. In this way, three clusters of an image are labeled as either healthy or diseased. This information is utilized by the first-stage classification only as for the second-stage classification disease type is needed and it is known in advance.

4.1.2 Classification Stages of the Proposed Framework

The proposed framework does classification in two stages. The first-stage differentiates between healthy or diseased thus requires a binary classifier. While the classifiers at the second-stage may either be binary or multi-class depending upon the number of diseases a system is classifying. In the proposal presented here, the assumption is that first-stage is purely binary and second-stage classifies among various diseases. Thus, selection of classifiers is to be done for both the stages. Base dataset considered to finalize the classifiers is a subset of images collected from PlantVillage [26]. It contains 4,775 images of soybean leaf (healthy – 1,079, *downy mildew* – 105, *frog eye* – 1,662, and *septoria leaf blight* – 1,929). The performance of several two- as well as multi-class classifiers is explored in this study to pick the best performing one. Discussion related to classifiers considered and their selection is included in Section 4.1.5.

4.1.3 Training

The classifiers are examined for three commonly used ratios, i.e. 50:50, 60:40, and 70:30, of training:testing images as per the literature survey. As mentioned, a leaf image is pre-processed and segmented into clusters. The obtained clusters are labelled and their features are extracted individually. The first-stage two-class classifier(s) is trained with features of healthy as well as diseased clusters. Looking at the second-stage, then system has to classify among more than two disease classes, so a multi-class scenario is there. One can opt for binary classifier(s) too in the second-stage and their count will be given by a combination, $C(d, 2)$, where d is the number of disease classes considered. Thus, for three disease classes, $C(3,2) = 3$ binary classifiers are needed at second stage, similarly for four disease classes $C(4,2) = 6$ classifiers have to be used. In other words, the system's complexity increases with the number of diseases being identified. Thus, multi-class classifier(s) are explored. Features of the unhealthy clusters are only utilized to train multi-class classifier(s) of the second-stage. Since the disease type of a leaf image is known at the time of training there is no need of any manual labeling here. If more than one classifier is selected in any of the two stages then fusion methods have to be used.

4.1.4 Testing

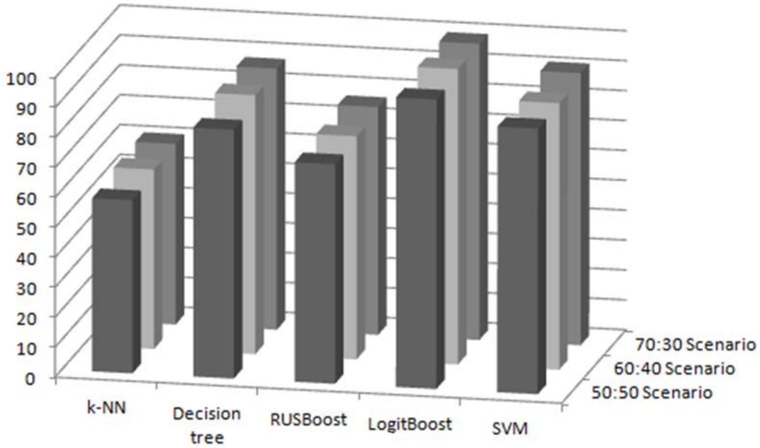
Test image samples are pre-processed and segmented using the same algorithms as are employed during training. Features of all the clusters are extracted and given as an input to the first-stage classifier one by one. The test leaf sample is identified as healthy only if all the three clusters are recognized as healthy by the first-stage two-class classifier, otherwise it is a diseased leaf sample. Features of all the clusters classified as unhealthy at the first-stage are passed to each of the second-stage classifiers individually. The classification scores are used to recognize the disease. Summation of classification scores from all the classifiers gives final score for a disease category. This procedure is repeated for all the unhealthy clusters and the obtained score values for each cluster are added corresponding to each disease. Finally, category with highest score is the identified disease class for the test leaf sample.

4.1.5 Classifier Selection

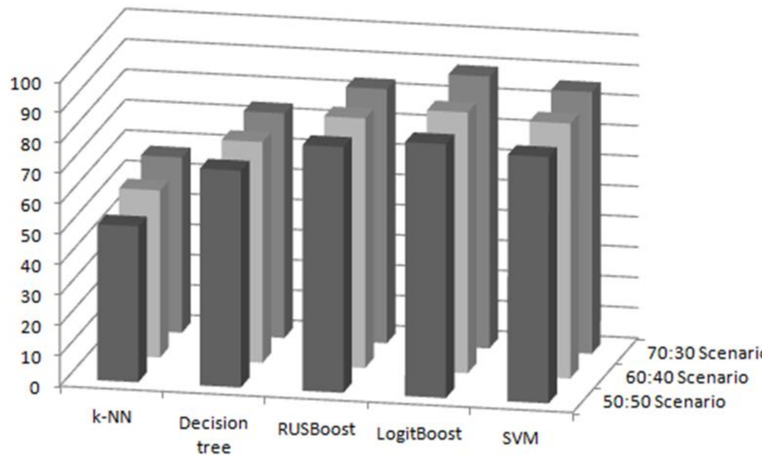
Performances of five classifiers (k-NN, Decision tree, RUSBoost, LogitBoost, and SVM) are examined over base dataset consisting of 1,079 healthy and 3,696 unhealthy images. Accuracy obtained for the three considered training:testing ratios for each of the five classifiers is listed in Table 4.1. The same has been visualized in Fig. 4.2. It is evident that LogitBoost performs the best in all scenarios, thus it is chosen in the proposed framework. In case of healthy leaf sample accuracies achieved by LogitBoost is above 96% and for diseased samples all values are above 83%. The results obtained by

Table 4.1 Classification accuracy values obtained using various classifiers for the base dataset, i.e. PlantVillage (all values are in percentage).

	50:50 Scenario		60:40 Scenario		70:30 Scenario	
Classifier	**Healthy**	**Disease**	**Healthy**	**Disease**	**Healthy**	**Disease**
k-NN	57.9	51.4	60	55.1	60.4	57.8
Decision tree	83.2	71.5	86.6	72.8	87.3	74.1
RUSBoost	73.4	81	74.6	82.3	76.2	83.8
Logitboost	96.7	83.7	98.8	86	99.1	89.7
SVM	88.7	81.2	89.1	84.1	91	86.3



(a) Healthy



(b) Disease

Fig.4.2 Classification accuracy values obtained for the base dataset, i.e. PlantVillage using various classifiers.

classifiers at the first-stage are utilized to finalize second-stage classifiers as well. This stage deals solely with the disease classification so accuracy obtained for the diseased leaf identification is only focused. As already noted LogitBoost is performing the best, however classifiers SVM and RUSBoost are also performing very close to LogitBoost. Specifically, the respective accuracies in scenarios 50:50, 60:40, and 70:30 are ranging from 81.0% – 83.7%, 82.3% – 86.0%, and 83.8% – 89.7% for the three classifiers. Therefore second-stage combines results from classifiers LogitBoost, SVM, and RUSBoost to identify the disease class of a leaf. LogitBoost is replaced by its multi-class version, AdaBoost M2, during implementation.

The proposed system has segmented a leaf into three clusters, thus instead of classifying each cluster independently the scores generated by the three classifiers for all the three clusters are summed up and the disease class with maximum score is the final

Table 4.2 Confusion matrix obtained by second stage of the proposed framework for the base dataset, i.e. PlantVillage. (DM –Downy mildew, FE – Frog eye, SLB – Septoria leaf blight).

(a) 50:50 Scenario

	DM	FE	SLB
DM	**0.50**	0.00	0.50
FE	0.00	**0.83**	0.17
SLB	0.00	0.07	**0.93**

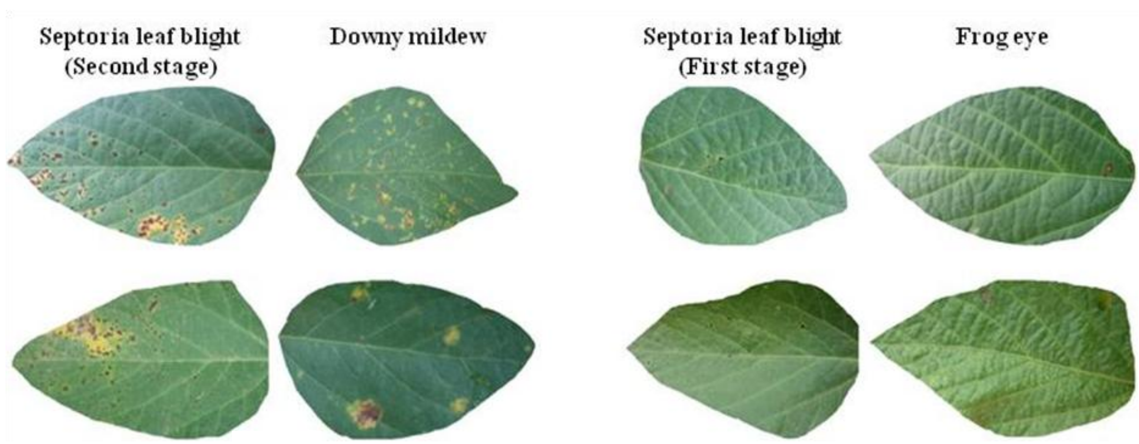
(b) 60:40 Scenario

	DM	FE	SLB
DM	**0.55**	0.00	0.45
FE	0.00	**0.87**	0.13
SLB	0.00	0.08	**0.92**

(c) 70:30 Scenario

	DM	FE	SLB
DM	**0.59**	0.00	0.41
FE	0.00	**0.94**	0.06
SLB	0.00	0.07	**0.93**

result. Following the said process, confusion matrix obtained for *downy mildew* (105 images), *frog eye* (1,662 images), and *septoria leaf blight* (1,929 images) is given in Table 4.2. It is evident that the combination has reported the poor classification accuracy for *downy mildew* specifically which is being confused with *septoria leaf blight*. The base dataset used for classifier selection contains leaf images for three stages of *septoria leaf blight* and the symptoms of *downy mildew* are very close to that of the second stage of *septoria leaf blight*. Fig 4.3 (a) shows images that help is visualizing the similarity between the two, both have yellow color spots. Similarly, frog eye and first stage of *septoria leaf blight* have common tiny spots, thus a few images from both the categories get mis-classified. Fig. 4.3 (b) further clarifies the observation made. It's difficult to separate symptoms (of *frog eye* and *septoria leaf blight*) as even in 70:30 training:testing scenario 6% to 7% of images are not identified correctly by the proposed system. Another possible reason may be the number of images for *downy mildew* category which is relatively lesser. Thus performance of the opted combination is observed on a randomly selected dataset of 300 images (100 per disease category) as well. The results



(a) (b)

Fig.4.3 Similarity between symptoms of various diseases (a) Septoria leaf blight (second stage) and downy mildew, (b) Septoria leaf blight (first stage) and frog eye.

Table 4.3 Confusion matrix obtained by second stage of the proposed framework for the randomly selected 300 images of base dataset, i.e. PlantVillage. (DM –Downy mildew, FE – Frog eye, SLB – Septoria leaf blight).

(a) 50:50 Scenario

	DM	FE	SLB
DM	**0.88**	0.00	0.12
FE	0.00	**0.87**	0.13
SLB	0.00	0.06	**0.94**

(b) 60:40 Scenario

	DM	FE	SLB
DM	**0.90**	0.00	0.10
FE	0.00	**0.88**	0.12
SLB	0.00	0.05	**0.95**

(c) 70:30 Scenario

	DM	FE	SLB
DM	**0.94**	0.00	0.06
FE	0.00	**0.91**	0.09
SLB	0.00	0.03	**0.97**

are tabulated in Table 4.3 and clearly the improvement obtained is acceptable for *downy mildew* leaf images. For this category the misclassification rate drops from the range of 41% – 50% to 6% – 12%, however no significant improvement is observed in the confusion between *frog eye* and *septoria leaf blight* but is bit lesser now. Earlier *septoria leaf blight* is mis-classified as *frog eye* for 7% – 8% of images and now this range is varying from 3% – 6%. Similarly, 6% – 17% of *frog eye* images are identified as *septoria leaf blight* formerly and with the balanced dataset its 9% – 13% now. Thus, it can be said that a balanced dataset leads to better performance of the proposed system as sometimes it's difficult to completely distinguish the symptoms of two or more plant leaf diseases.

Selected Classifiers. For the first phase, a two-class classifier, LogitBoost, reporting maximum accuracy value is chosen, but for second phase three classifiers (LogitBoost, SVM, and RUSBoost) achieving the best classification accuracy are picked. The final framework thus has one two-class classifier at first-stage and three multi-class classifiers at second-stage. As three classifiers are opted for the second stage, fusion logic has to be implemented. The complete process for the proposed framework is illustrated in Fig 4.4. Here *d* signifies the number of disease classes considered during the implementation.

4.2 Comparison Between Rule-based System and Semi-automatic System

Considering the overall system performance as given in Table 4.4, then the proposed rule-based system performs only a bit better than semi-automatic system for Scenario 1 and Scenario 2; and for Scenario 3 comparable performance is achieved. The respective accuracy values (in percentage) for both the systems are Scenario 1: 83 and 80.07;

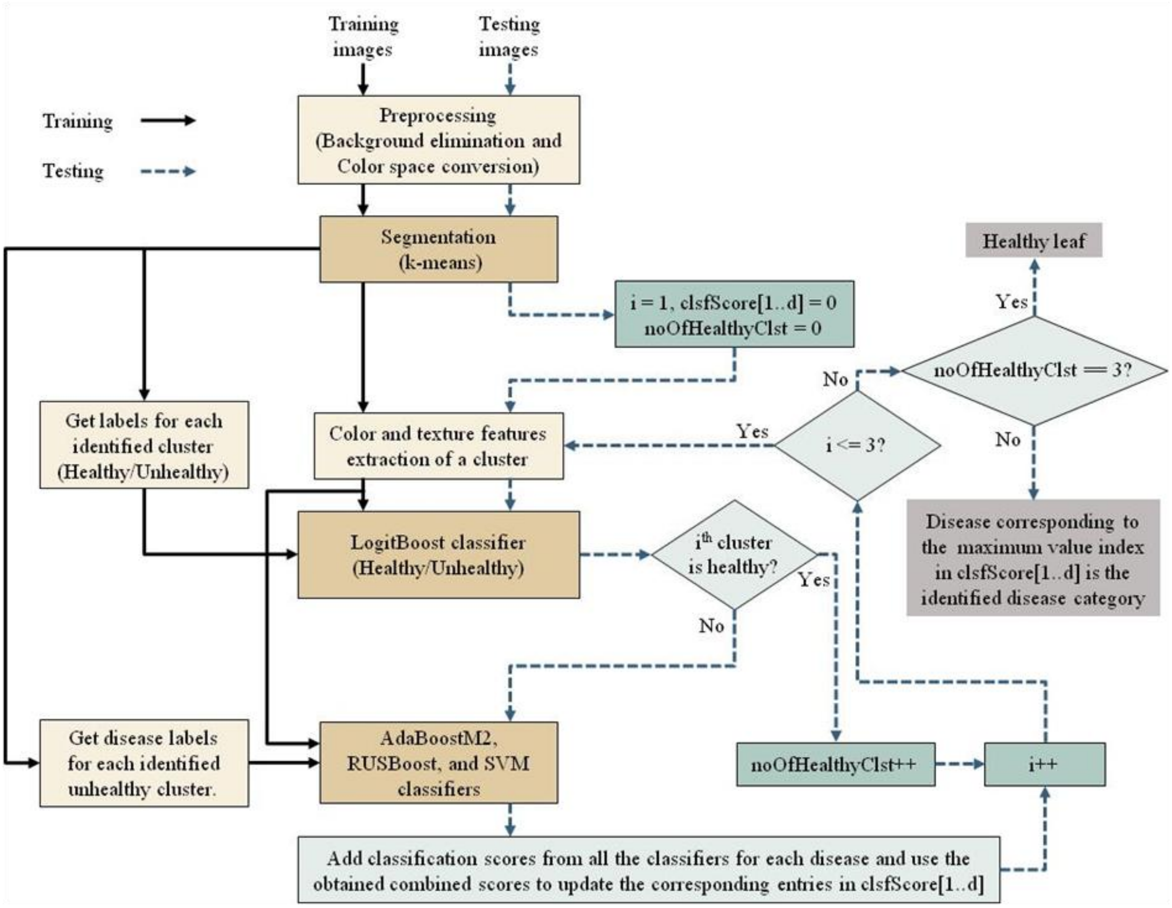


Fig.4.4 Detailed process flow of the proposed framework.

Table 4.4 Comparative evaluation of the rule-based system and semi-automatic system

	Rule-based	Semi-automatic
Scenario 1		
Healthy	79%	96.7%
Frog Eye	87%	82.2%
Septoria leaf blight	89%	91.4%
Downy mildew	77%	50%
Scenario 2		
Healthy	80%	98.8%
Frog Eye	90%	86.3%
Septoria leaf blight	89.5%	91.6%
Downy mildew	77.5%	54.7%
Scenario 3		
Healthy	82%	99.1%
Frog Eye	90.7%	92.9%
Septoria leaf blight	90%	92.7%
Downy mildew	79.9%	59%

Scenario 2: 84.25 and 82.85; and Scenario 3: 85.65 and 85.92. When Scenario 3 is utilized the accuracy to detect frog eye using semi-automatic system is more as compared to rule-based system. However in case of other scenarios rule-based system performs better than semi-automatic system. For septoria leaf blight disease detection, accuracy increases for all the scenarios respectively by 2.4%, 2.1%, 2.7%. As already discussed for downy mildew semi-automatic system does not works much efficiently as it reports decrease in accuracy values. In case of healthy class, there is large increase in accuracy values ranging from 17.1% to 18.8%, in all the scenarios. Overall analysis indicates that the classification accuracy for all the classes increases except for downy mildew, when semi-automatic system is employed.

4.3 Summary

Highlighting the significance of a generalized leaf disease detection system, the chapter has proposed a framework utilizing the concepts of image processing, segmentation, and machine learning to identify a leaf image as healthy or its disease class if unhealthy. The system is simple as well as flexible to employ any algorithm at any phase. The current implementation is developed with three classifiers LogitBoost (AdaBoost M2 for multi-class), RUSBoost, and SVM along with several color/texture features and k-means algorithm. The general applicability of the system is proved by performing experiments on six datasets collected online and results for the same are presented in Chapter 5. The classification accuracies reported in all the cases are comparable to the existing systems and this observation supports the desired generalization notion focused in this study. However, it’s still difficult to efficiently categorize disease classes having similar symptoms. The low accuracy values may be due to various reasons, like initially disease symptoms are less visible and hence presents very less information, dataset imbalance, multiple infections, etc. In future, focus is to make the presented system more robust against these observations.

Chapter 5

RESULTS AND DISCUSSIONS

This chapter presents the complete results of both the proposed system discussed in Chapter 3 and Chapter 4. Performance of rule-based system and semi-automatic system is analysed and compared with existing ones as well as in cross-domain scenario.

Section 5.1 compares proposed systems with other systems existing in literature. It also presents comparative analysis after implementing a few of the existing works. Section 5.2 discussed system performance in provides the cross-domain scenario testing phase of the system. Section 5.3 describes the results of generic framework i.e. semi-automatic system in various scenarios. Section 5.4 presents the evaluation of semi-automatic framework with other existing systems. Section 5.5 describes developed the web application in this study.

5.1 Comparative Analysis

In the domain of automatic leaf disease detection there is still lack of standardized dataset, thus it’s difficult to make direct comparisons with any of the existing works. Though an attempt has been made to analyse the scenario in as clearer way as possible. Table 5.1 summarizes several studies that focused on soybean culture. It also includes results obtained for the best three candidates, i.e. Color + Texture (CT), Color + Haralick (CH), and Color + Gabor (CG) by the proposed rule-based system.

Accuracy is specified in the form of a range so as to comprise the results in all the

Table 5.1 Comparative evaluation of the proposed system with other existing works.

[Reference] Year	Disease or healthy	Number of images		Classifier	Performance measure	
		Training	Testing		Accuracy	Computes Severity
[118] 2015	Healthy and diseased	Total images = 120		Feature based	93.79%	No
[60] 2014	Frog eye Brown spot	Total images = 100		k-NN	80% 70%	No
[94] 2014	Rust Brown spot Bacterial blight Frog eye	Total images = 57		LDA	100% 100% 75% 84.6%	No
[44] 2015	Downy mildew, Frog eye, and Bacterial pustule	25	3	NN	93.3%	No
[116] 2015	Bacterial blight	38	18	Feature based	37.5%	No
	Rust	45	20		71%	
	Phytotoxicity	16	7		73.9%	
	Stem canker	15	7		9.1%	
	Corynespora leaf spot	43	19		19.4%	
	Myrothecium leaf blight	1	1		100%	
	Downy mildew	31	15		91.3%	
	Powdery mildew	52	24		61.8%	
	Septoria brown spot	14	6		20.0%	
[117] 2015	Healthy, Rust, Bacterial blight, Brown spot, Sudden death syndrome, Frog eye, and Downy mildew	Total images = 1000		N/A	No	Yes
[29] 2015	Bacterial leaf blight, Septoria brown spot, and Bean pod mottle virus	N/A	N/A	k-means clustering	No	Yes
[48] 2016	Rust	N/A	N/A	N/A	No	Yes
Proposed Rule-based (Color+ Texture: CT)	Healthy Downy mildew Frog eye Septoria leaf blight	Healthy-1079 Downy mildew-105 Frog eye-1662 Septoria leaf blight-1929		SVM	79-82% 77-79.9% 87-90.7% 89-90%	Yes
Proposed Rule-based (Color+ Gabor:CG)	Healthy Downy mildew Frog eye Septoria leaf blight			SVM	79-82% 75.5-79% 87-90% 88-89%	Yes
Proposed Rule-based (Color+ Haralick: CH)	Healthy Downy mildew Frog eye Septoria leaf blight			SVM	79-82% 74-75% 85.8-88.5% 88-88.5%	Yes
Proposed Semi-automatic (Color+	Healthy Frog eye			LogitBoost, RUSBoost, SVM, AdaBoostM2	96.7-99.1% 82.2-92.9%	Yes

Texture:CH)	Downy mildew Septoria leaf blight	50-59% 91.4-92.7%

three scenarios. Undoubtedly, the number of training and testing images utilized in this work is far more than any of the existing works. The proposed system achieves an average accuracy of 85.65% (Table 3.3, Column: Color + Texture (CT), **Scenario 3**) to distinguish healthy and unhealthy leaf samples in a dataset of 4,775 images, which is approximately 8% less than [118] that uses only 120 images.

Studies in [44], [60], [94], and [116] classify diseased leaves into one of the various categories. For *frog eye*, the achieved accuracy in [94] and [60] is less than 85%; and the proposed system has reported a minimum accuracy of 85.8% (Table 3.3, Column: Color + Haralick (CH), **Scenario 1**) and a maximum accuracy of 90.7% (Table 3.3, Column: Color + Texture (CT), **Scenario 3**) for this category.

On contrary, for *septoria leaf blight* [94] reported 100% results for small dataset of 57 images and for this category the proposed rule-based system achieves accuracy ranging from 88% (Table 3.3, Column: Color + Haralick (CH), **Scenario 1**) and a maximum accuracy of 90% (Table 3.3, Column: Color + Texture (CT), **Scenario 3**). Studies in [44] and [116] considered only unhealthy samples. In summary, [44] achieves 93.3% average accuracy with 28 images and [116] reports an average accuracy of 53.78% for 372. In comparison the proposed rule-based system reports an average maximum accuracy of 86.87% (Table 3.3, Column: Color + Texture (CT), **Scenario 3**) and average minimum accuracy of 82.6% (Table 3.3, Column: Color + Haralick (CH), **Scenario 1**). Again number of images used is very large thus the achieved accuracy range can be considered as acceptable. Lastly, [29], [48], and [117] have worked only on severity and by means of visual comparison it can rightly be said that the proposed rule-based system is computing disease severity in all cases efficiently.

Results obtained using proposed semi-automatic system is more accurate as compared to existing work. For healthy sample detection accuracy improves to 99%. Also for *septoria leaf blight* and *frog eye* accuracy values are higher than respective existing works [116, 94].

Table 5.2 Comparative evaluation of the proposed system with approaches in [60, 94, and 118].
[*] Detects only healthy and unhealthy leaf samples. [+] Detects only disease class of an unhealthy leaf samples.
(All values are in percentage) CT: Color+Texture, CH: Color+Haralick, CG: Color+Gabor

	Dandawate [*] [118]	**Shrivastava [+][94]**	**Shrivastava [+][60]**	**Rule-based method**			**Semi-automatic method**
				CT	**CH**	**CG**	**CT**
Scenario 1							
Healthy	49	-	-	79	79	79	96.7
Frog Eye	58 Unhealthy	62	59	87	85.8	87	82.2
Septoria leaf blight		69	58	89	88	88	91.4
Downy mildew		43	-	77	74	75.5	50
Overall, disease	53.5	58	58.5	83, 84	81.7, 82	82.3, 83	80.1, 74.5
Scenario 2							
Healthy	51	-	-	80	80	80	98.8
Frog Eye	58.9 Unhealthy	65.5	60	90	86.7	89	86.3
Septoria leaf blight		70.1	59.9	89.5	88.2	88.5	91.6
Downy mildew		44	-	77.5	74.5	76	54.7
Overall, disease	54.9	59.8	59.9	84.3,85	82.4, 83	83.4, 84	82.8, 77.5
Scenario 3							
Healthy	53.3	-	-	82	82	82	99.1
Frog Eye	60 Unhealthy	67	63	90.7	88.5	79	92.9
Septoria leaf blight		71	62	90	88.5	89	92.7
Downy mildew		48	-	79.9	75	79	59
Overall, disease	56.7	62	62.5	85.6, 86	83.5, 84	85, 86	85.9, 81.5

5.1.1 Comparative Analysis After Implementing few of the Existing Works

For better evaluation, three approaches [94, 60, and 118] are implemented and compared in Table 5.2 with the best three candidates of the proposed rule-based system as well as semi-automatic system on PlantVillage dataset. Clearly, in comparison to [94] and [60] both the proposed systems have reported higher classification accuracies for all the classes i.e. healthy as well as diseased.

5.2 Performance Evaluation in Cross-domain Scenario

An attempt has also been made to validate the proposed rule-based system in a cross-domain scenario and the results for the same are shown in Table 5.3. Three scenarios trained using PlantVillage dataset are tested using diseased images from three different datasets, namely IPM images [41], a few image from [60], and a Self-collected dataset. Images in the Self-collected dataset are personally collected from the fields of Punjab Agricultural University, India. IPM images contain leaf samples for the three disease categories considered in this study; however the other two has leaves only for *septoria leaf blight*. Results reported with IPM images and self-collected dataset are not very encouraging, but for images from [60] the system has reported fairly good results. One probable reason behind the obtained results can be different environmental conditions used for capturing leaf images. As capturing conditions play a major role and the importance of capturing conditions has already been established in literature [13].

However, these results can be improved by updating the intermediate algorithms utilized in the current implementation of the proposed systems. Cross-domain results obtained using proposed semi-automatic system is described in Table 5.4. The result obtained from self collected image samples and IPM images (*frog eye* and *downy*

Table 5.3 Results for cross-domain setting using rule-based system with Color + Texture (CT) feature set.

Dataset	**Training dataset**	**Testing dataset**				
		IPM Images [41]			**Shrivastva [60]**	**Self-collected**
		Frog eye	**Downy mildew**	**Septoria leaf blight**	**Septoria leaf blight**	**Septoria leaf blight**
Scenario 1	PlantVillage	5/11 (45.4%)	6/22 (27.2%)	4/12 (33.3%)	14/27 (51.8%)	16/71 (22.5%)
Scenario 2	PlantVillage	5/11 (45.4%)	6/22 (27.2%)	4/12 (33.3%)	16/27 (59.2%)	20/71 (28.2%)
Scenario 3	PlantVillage	5/11 (45.4%)	6/22 (27.2%)	4/12 (33.3%)	17/27 (62.9%)	20/71 (28.2%)

Table 5.4 Results for cross-domain setting using semi-automatic system with Color + Texture (CT) feature set.

Dataset	Training dataset	Testing dataset				
		IPM Images [41]			Shrivastva [60]	Self-collected
		Frog eye	Downy mildew	Septoria leaf blight	Septoria leaf blight	Septoria leaf blight
Scenario 1	PlantVillage	3/11 (27.3%)	4/22 (18.18%)	6/12 (50%)	20/27 (74%)	32/71 (45%)
Scenario 2	PlantVillage	4/11 (36.4%)	4/22 (18.18%)	9/12 (75%)	20/27 (74%)	49/71 (69%)
Scenario 3	PlantVillage	4/11 (36.4%)	8/22 (36.4%)	10/12 (83.3%)	21/27 (77.8%)	56/71 (78.8%)

mildew) are not much efficient, however the results obtained from [60] and IPM images (*septoria leaf blight*) are better. However results obtained with semi-automatic system are better than the rule-based system for cross domain testing as shown in Table 5.4. However, the accuracy value is low in some cases, but this may be due to different environmental conditions in which images are acquired.

5.3 Performance Evaluation of Semi-automatic System on Different Datasets

Literature shows the unavailability of public dataset as well as lack of dataset standardization in this research area, thus self-collected datasets are observed in majority of the studies. In such a scenario, efficacy of the proposed semiautomatic system with selected classifiers is proved via different datasets collected using Google image search engine. Search is performed using several combinations of diseases and plants name. All the datasets used in this study are described in Table 5.5.

As stated in Section 1.6, a dataset consisting of 4,775 images is considered as a base dataset to finalize classifiers in the proposed semi-automatic framework. This is considered as Dataset_1 and is collected from PlantVillage [100]. It contains 1,079 healthy soybean leaf images as well as leaf images infected from three diseases, namely

Table 5.5 Description of the datasets considered to examine the proposed generalized framework.

Dataset	Crops	Classes considered	#Images
Dataset_1	Soybean	Healthy, Septoria leaf blight, Frog eye, Downy mildew	4,775
Dataset_2	Groundnut	Healthy, Early blight, and Late blight	180
Dataset_3	Soybean and Groundnut	Healthy, Septoria leaf blight, Frog eye, Downy mildew, Early blight, and Late blight	360
Dataset_4	Groundnut, Tomato, and Potato	Healthy, Early blight, and Late blight	540
Dataset_5	Apple and Mango	Healthy, Scab, Rust, and anthracnose	240
Dataset_6	Sugarcane, Cotton, Chilli, and Jute	Healthy, Rust, and Spots (caused by Bipolaris spicifera and Cercospora)	300

downy mildew (105), *frog eye* (1,662), and *septoria leaf blight* (1,929). Dataset_1 contains leaf images for three stages of *septoria leaf blight* disease. Dataset_2 is similar to Dataset_1 as it contains leaf images for another legume species groundnut. Dataset_2 has three types of leaf images healthy (60) and infected from either *early blight* (60) or *late blight* (60). Dataset_3 is formed by merging leaf images from Dataset_1 and Dataset_2. Thus, it has five diseases and one healthy category. Dataset_4 is formed using leaf images from three cultures, namely groundnut, tomato, and potato for two diseases (early blight and late blight) along with a healthy category. 60 images per category per culture are collected to have a total of 540 images. Dataset_5 is formed using fruit culture, particularly apple and mango. It has 60 healthy images and 60 images from each of the three disease classes' *apple scab*, *apple rust*, and *mango anthracnose*. Lastly, Dataset_6 is formed from commercial crops considering sugarcane, cotton, chilli, and jute cultures. Similar to others, this dataset also has leaf images from various categories healthy (60), *bipolaris spicifera leaf spot* (60), rust (60), and *cercospora leaf spot* (60 each from cotton and chilli).

The proposed framework is validated against the notion of generalization by observing its performance on six datasets for three commonly used training:testing images ratio, i.e. 50:50, 60:40, and 70:30, in this domain. Experimental results for Dataset_1 are presented and discussed in Chapter 4 as this is the base dataset in this study. The same framework i.e. techniques, algorithms, and classifiers, as is utilized in the experimental setup for Dataset_1 is considered for all the remaining five datasets. However, the system has to be trained in each case individually.

Classification results obtained using the proposed framework for all the datasets is shown in Table 5.6. Clearly the performance of first-stage in the proposed semi-automatic system is acceptable ranging from ≈ 75% (Dataset_4 50:50 scenario) to ≈ 90% (Dataset_2 70:30 scenario) accuracy. Another observation is that the first-stage is classifying better than second-stage for all the considered cases, the probable reason behind this is the two-class versus multi-class scenario. Results for Dataset_1 are already discussed in Chapter 4 (Section 4.1.7). Dataset imbalance is one of the reasons identified

Table 5.6 Classification accuracy for all the six datasets. (Accuracy values are in percentage).

Datasets	**50:50 Scenario**	**60:40 Scenario**	**70:30 Scenario**
Dataset_1			
Healthy	96.7	98.8	99.1
Frog eye	82.2	86.3	92.9
Downy mildew	50	54.7	59
Septoria leaf blight	91.4	91.6	92.7
Average classification accuracy (first stage, second stage)	**(85.6,74.5)**	**(88.2, 77.5)**	**(90.3, 81.5)**
Dataset_2			
Healthy	100	100	100
Early blight	63.3	70.8	83.3
Late blight	73.3	75	77.8
Average classification accuracy (first stage, second stage)	**(84.2, 68.3)**	**(86.5, 72.9)**	**(90.3, 80.6)**
Dataset_3			
Healthy	85	89.5	94.4
Downy mildew	73.3	79.1	83.3
Frog Eye	83.3	87	88.9
Septoria leaf blight	76.7	79.1	83.3
Early blight	73.3	75	77.8
Late blight	80	87.5	94.4
Average classification accuracy (first stage, second stage)	**(81.2, 77.3)**	**(85.5, 81.5)**	**(89.9, 85.5)**
Dataset_4: Tomato, Potato, and Groundnut			
Healthy	78.8	80.5	83.3
Early blight	70	72.2	77.8
Late blight	71.1	73.6	75.9
Average classification accuracy (first stage, second stage)	**(74.7, 70.6)**	**(76.7, 72.9)**	**(80.1, 76.9)**
Dataset_5: Apple and Mango			
Healthy	86.7	91.6	94.4
Scab	73.3	75	77.8
Rust	76.7	79.1	83.3
Anthracnose	83.3	87.9	88.9
Average classification accuracy (first stage, second stage)	**(82.2, 77.8)**	**(86.1, 80.7)**	**(88.9, 83.3)**
Dataset_6: Cotton/Chilli, Sugarcane, and Jute			
Healthy	83.3	87.5	88.9
Cercospora spot	78.3	79.1	83.3
Rust	70	70.8	77.8
Bipolaris spicifera leaf spot	73.3	75	83.3
Average classification accuracy (first stage, second stage)	**(78.5, 73.9)**	**(81.2, 74.9)**	**(85.2, 81.5)**

earlier for Dataset_1 and therefore rests of the datasets are collected in a balanced way. This minor precautionary point helped in achieving better accuracy values. Looking at the 70:30 scenario (leaving Dataset_1), then classification is more than 80% for all the classes except four (*early blight*, *late blight*, *scab*, and *rust*). Overall it's worthy to state that suitable accuracy values are reported for a variety of datasets and certainly 70:30 scenario outperforms than 50:50. At disease category level, maximum gain of 20% is reported for early blight (Dataset_2), also increment of >5% is achieved in more than half of the cases. Similarly, for first-stage classification accuracy increases in ranges 4.7% (Dataset_1) to 8.7% (Dataset_3) and for second-range the respective growth is from 5.5% (Dataset_5) to 12.3% (Dataset_2).

5.4 Comparative Analysis of Semi-automatic System With Existing Works

Comparative analysis with the exiting studies is covered under following head, namely image acquisition, segmentation, and number of training/testing images. Overall average classification accuracies of proposed semi-automatic system are shown in Table 5.7. The respective minimum and maximum reported values in each of the three scenarios across all the datasets are 73.3% and 80.07% (50:50), 75.43% and 83.4% (60:40), and 79% and 87.03% (80:20). However, individual class detection accuracy is quite high (Table 5.6), for example, classification accuracy for healthy varies from 78.8% to 100% and similarly for *septoria leaf blight* the reported range is 91.4% to 92.7%.

Image acquisition is important in leaf disease detection systems. Several studies have utilized controlled conditions to capture leaf images [9, 28, 69]. The observed accuracy values for these studies are 89.5% [69], 92% [28], and 93% [9]. In contrast, the proposed semi-automatic system is developed aiming its independence over the capturing conditions. This factor is clearly supported by the obtained results for different datasets having varying capturing procedures (different angles, lighting/reflection controls, etc.). Though the maximum average reported value is only 87.03%, but looking at other factors then it is considerable.

Table 5.7 Comparative evaluation with the existing systems. (Accuracy values are in percentage).

[Reference]	**Disease or healthy leaves**	**Number of images**		**Classifier**	**Accuracy**		
		Training	**Testing**				
[69]	Early scorch, Tiny whiteness, Ashen mold, Cottony mold and Late scorch	N/A		BPNN	89.5		
[68]	Mold, Scorches, Sun burn, Blight, Bacterial/Fungal infections, And Spots	500		SVM	94.74		
[28]	Fungal, bacterial, Nematodes, Viral, and Deficiency	450	450	SVM	92		
				ANN	87		
[9]	Leaf spots and Leaf blotch	297		k-NN	93		
[83]	Anthracnose, Powdery mildew, Downey mildew, Early blight, Late blight, and Rust	N/A		Neuro- k-NN	91.54		
	Anthracnose, Fruit rot, Powdery mildew, Alternaria leaf spot, Gray mildew, Smut, Fusarium wilt, and Red rot			PNN	86.48		
	Leaf blight, Powdery mildew, Leaf spot, Leaf rust, and Smut			SVM	85.33		
[125]	13 diseases	30880	2589	Deep CNN	96.3		
[42]	26 diseases	54306		Deep CNN	31.4 -99.35		
[124]	82 diseases	1335		Feature based	58		
				Feature based (manual)	63		
Proposed framework					**50:50 Scenario**	**60:40 Scenario**	**70:30 Scenario**
Dataset_1	Healthy. Frog eye, Downy mildew, and Septoria leaf blight	4775		LogitBoost, RUSBoost, SVM,	80.07	82.85	85.92
Dataset_2	Healthy, Early blight, and Late blight	180		AdaBoostM2	78.86	81.93	87.03
Dataset_3	Healthy. Frog eye, Downy mildew, Septoria leaf blight, Early blight, and Late blight	360			78.6	82.86	87.01
Dataset_4	Healthy, Early blight, and Late blight	540			73.3	75.43	79
Dataset_5	Healthy, Scab, Rust, and Anthracnose	240			80	83.4	86.1
Dataset_6	Healthy, Cercospora spot, Rust, and Bipolaris spicifera leaf spot	300			76.22	78.1	83.32

A few of the systems have collected precise leaf image data, like patch containing 50% of the disease region [68] and front and back images of a leaf [83], to achieve better detection. Indirectly, this makes these systems bit complex in comparison to the proposed one as it needs only single side of a leaf. Moreover obtaining diseased patch effectively is difficult in few scenarios, like initial stage of *septoria leaf blight* considered in this study. It is already stated that initial stage has very small symptoms in comparison to the final stage which has good amount of disease symptoms. Still the proposed system has successfully identified this category with a bit of mis-classification as discussed in Section 3.3.

Systems utilizing deep learning have achieved 96.3% accuracy by utilizing approximately 93% images for training and < 8% images for testing [125]. Another system has also reported unexceptional results, but when tested on an online dataset reports a low accuracy of 31.4% only [42]. In comparison to these systems, the proposed semi-automatic system is performing better considering training:testing images ratio. Also, Dataset_2 to Dataset_6 are collected through different online web-sources, so achieving accuracy values in the range of 73.3% to 87.03% is quite challenging and thus are acceptable.

5.5 Web Application

The proposed semi-automatic system is made available to users like pathologists and farmers by means of developing a web application. The web application is required in ASP.NET C# language in Microsoft Visual Studio. Users only require web browser to access the developed application. However Matlab runtime compiler is also required for execution and is freely available by Mathworks.

Overall architecture of web application and layers involved in development is described in Appendix B. Also, it describes all the general steps involved in developing the online application. A few of the screenshots of the web and mobile application are shown in Fig. 5.1-5.5. The application can easily be used on any smart phone. The screenshots for the same are shown in Fig. 5.6.

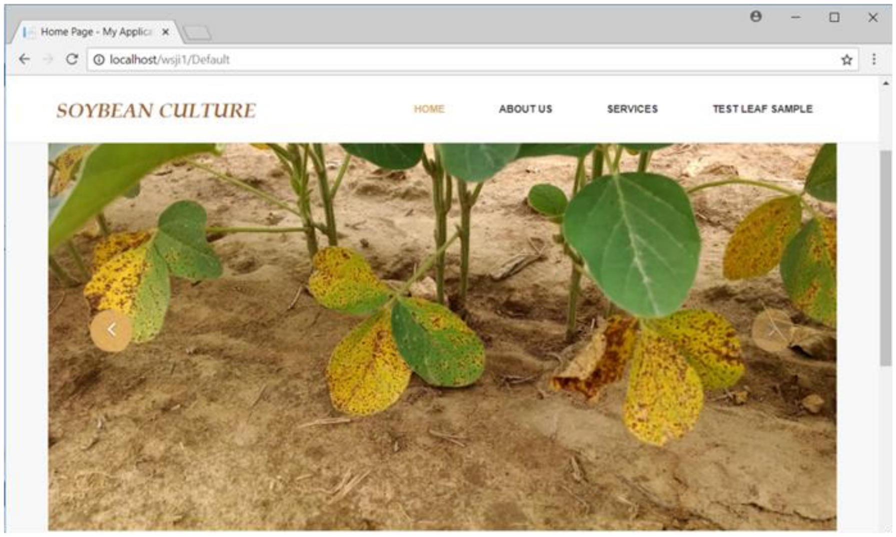


Fig 5.1 Home page of web application

In chapter overall results obtained from both the proposed systems are described. The proposed systems are compared with several existing systems in Table 5.1 and Table 5.7. For effective comparison purpose three papers [94, 60 and 118] are implemented on same dataset and the results are tabulated in Table 5.2. Considering cross-domain scenario systems are tested on three other datasets namely IPM images, Shrivastava[60], and self-collected. However, the results are not much efficient, due to variations of environmental and geographical conditions in which images are captured. Finally the applicability of semi-automatic system is verified on several datasets. Lastly the developed web application is discussed.

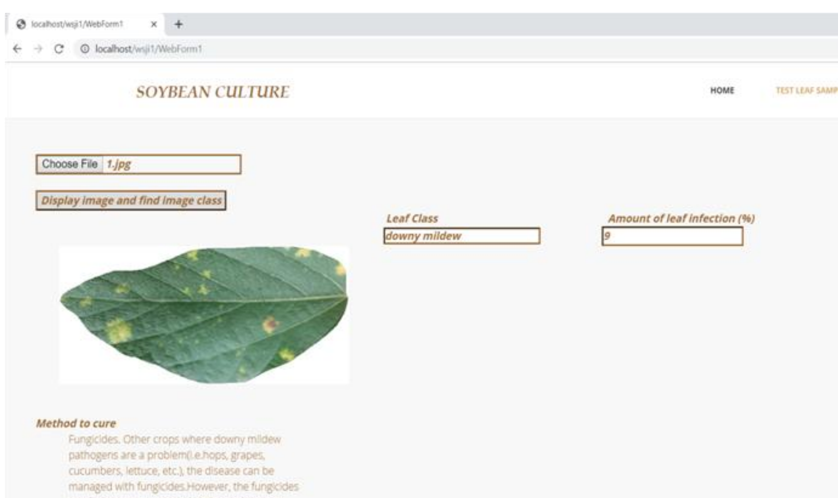


Fig 5.2 Results obtained using web application for a leaf sample infected from downy mildew disease class.

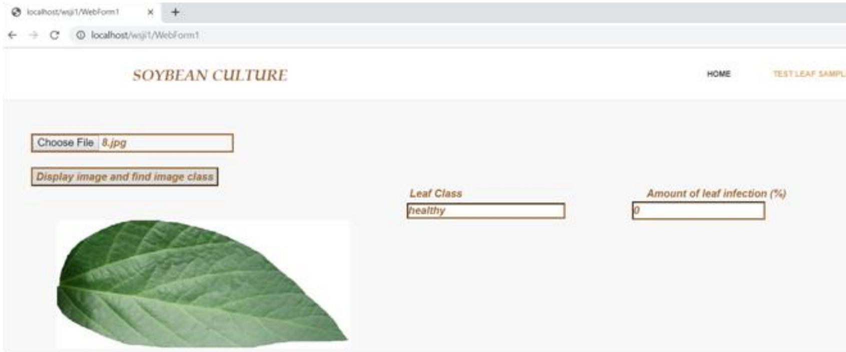


Fig 5.3 Results obtained using web application for a healthy leaf class sample.

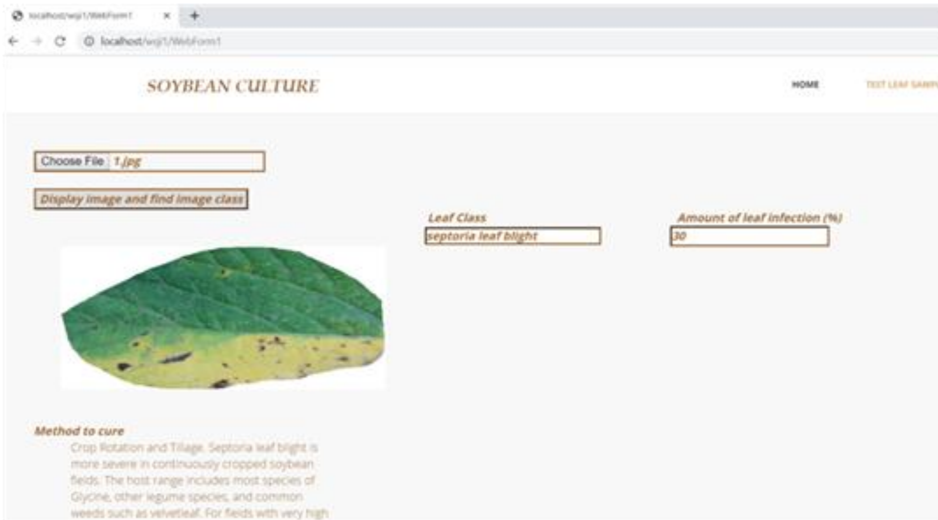


Fig 5.4 Results obtained using web application for a leaf sample infected from septoria leaf blight disease class.

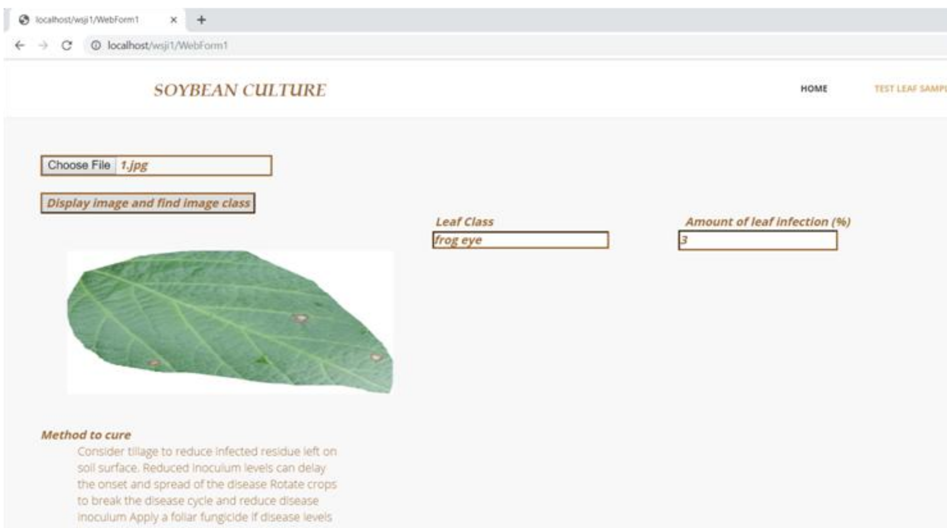
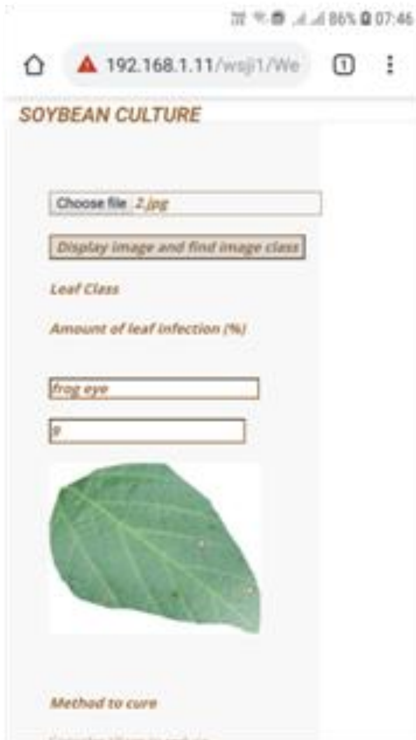
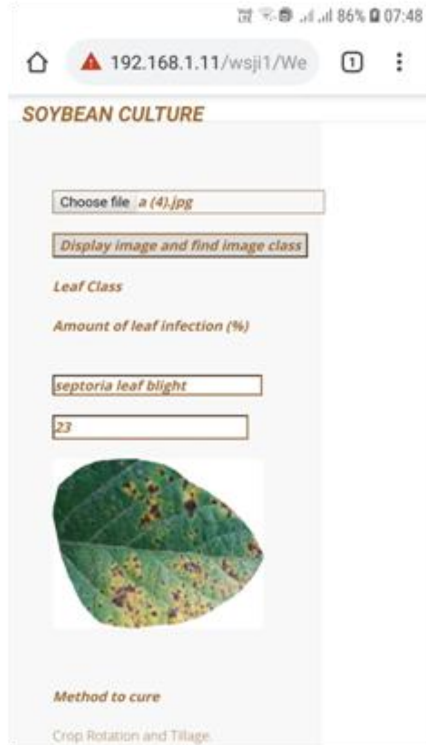


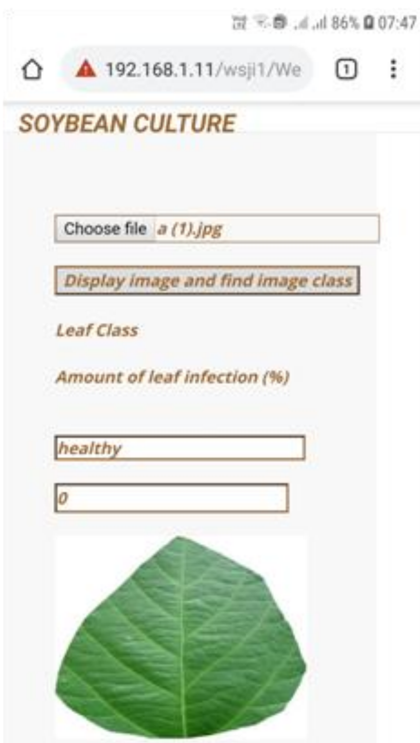
Fig 5.5 Results obtained using web application for a leaf sample infected from frog eye disease class.



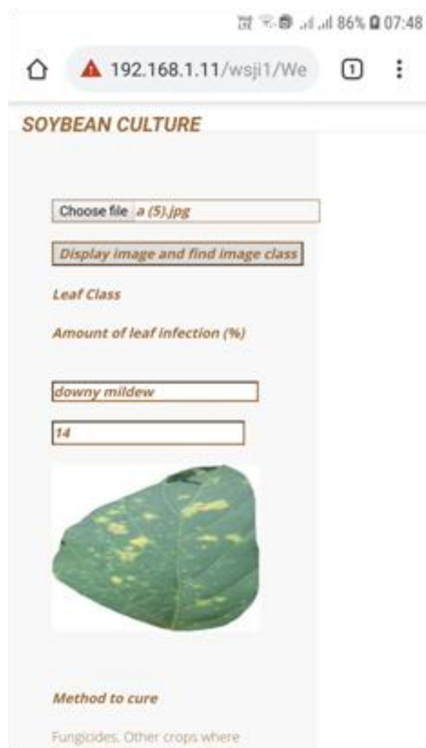
(a) Frog eye.



(b) Septoria leaf blight.



(c) Healthy.



(d) Downy mildew.

Fig 5.6 Results obtained using web application on a Smartphone for various test leaf samples.

Chapter 6

CONCLUSIONS AND FUTURE SCOPE

In precision agriculture, disease detection system is a significant area of research. Various studies are carried out in this area on different cultures. This chapter concludes the work presented in thesis and also lists the several future scope possible. Following are the main conclusions drawn from the thesis.

6.1 Conclusions

The following main conclusions are drawn from the thesis:

The first objective involves study and analysis of image processing techniques for identification of fungal disease in plants. This objective is achieved by doing review of various plant disease detection systems which are illustrated in Chapter 2. As per literature there is uncertainty whether combination of technology, particularly image processing with machine learning will evolve and solve the problems related to disease and replace the pathologists and specialist. Because there are many aspects that are needed to be taken care such as atmospheric conditions, soil variability, seeds variation and also the image capturing system environment conditions. The most difficult part is dataset availability to train the system about disease classes. Symptoms variations of disease also lead to difficulty in disease detection and classification. Various diseases have very much similar type of symptoms too. So basically there are many aspects which demand care in successful detection of disease.

The second objective is to develop an eAgriculture application to detect disease a class infecting the leaf. This goal is achieved by providing frameworks using various image processing techniques in Matlab language and then deploying this system to web server using Microsoft visual studio which is described in Chapter 3, Chapter 4, and Chapter 5. It is important to detect whether a leaf is healthy or diseased. The study presented here attempts to detect whether a soybean leaf is healthy or diseased. A diseased leaf is further classified into three disease categories namely *septoria leaf blight*, *frog eye*, and *downy mildew*. PlantVillage dataset is taken as base dataset for study. First a rule-based system is developed. The semi-automatic system is proposed which eliminates the existing rules and provides a generic framework that works across different datasets (such as legumes, vegetables, fruits, and commercial crops). Average accuracy reported after execution of proposed framework in MATLAB R2018a is better than existing ones. The combination and subset of the features generates an effective result for classification rather than individual features. The system also computes the disease severity of leaf as well.

Currently, three scenarios with different ratios of training is to testing images are explored. The proposed system utilizes three best performing classifiers, although it is flexible to work with different classifiers as well. Based on several combinations of color and texture features, classification is performed using the proposed rule-based system. Later the system is automated and generalized. It has been observed that 203 dimensional vector formed by combining all the considered color and texture features provides accurate results than individual features. The proposed systems are found to be better on many criteria as compared to existing studies. Moreover, the proposed system is designed and tested using a dataset collected from PlantVillage which contains images with complex backgrounds. The maximum average classification accuracy reported is approximately 85% using a dataset of 4,775 images. Also, visual examination of the test images confirms the acceptance of the proposed system.

Third objective is to test and validate the system. This system implementation is performed on PlantVillage dataset. The results obtained using proposed systems are given in Chapter 3, Chapter 4 and Chapter 5. Also, systems are tested on other datasets such as self-collected, Shrivastava dataset, and IPM images. A few of the existing systems are implemented and results are compared on PlantVillage dataset. However, the results are not as much accurate due to variation in atmospheric conditions in which images are taken, also due to existence of different soil and weather conditions. Studies have used several parameters which have to be fixed properly for better performance. It's good to note that no such parameter setting is needed in the proposed system in this thesis. For comparison purpose, another system is implemented but is unable to detect the disease class because they have used the fix thresholding value.

Robustness of the proposed systems is measured by evaluating its performance against various datasets. In the research four different datasets collected under different environmental conditions are used. PlantVillage acquired leaf samples in fields under full light using digital camera (Sony DSC - Rx100/13 20.2 megapixels) [99]. Similarly, IPM images are collected in fields by various photographers using a digital camera [41]. On contrary, images used by Shrivastava are not captured in fields [60]. A mobile phone camera with following specifications: Samsung GT-S3770, 2MP, resolution 1600x1200 pixels, exposure time 1/1,756 seconds is used to click a leaf image without flash under the sunlight with white/light background. Finally, a set of self-collected leaf images is created by picking leaves from fields under the College of Agriculture, Punjab Agricultural University, Ludhiana, Punjab. The leaves are then scanned with high resolution scanner.

Results presented in Chapter 3 are obtained by training and testing the system using PlantVillage dataset only. However, those in Chapter 5 results are generated by training the system on PlantVillage and testing is done using images from other three datasets. In all the cases PlantVillage is chosen for training as it contains sufficient number of images for all the four categories. In comparison, IPM has 45 images in all for the three disease classes. The proposed rule-based system is working on some rules which are formed after some observations. And these observations are difficult to be attained using such less number. On the other hand, Shrivastava [60] (27 images) and Self-collected (71 images) contain only *septoria leaf blight* infection. Thus, they cannot be used for training.

6.2 Future Scope

The literature survey presented in the Chapter 2 summarizes various studies that automate the identification and classification of plant leaf diseases using machine learning and image processing techniques. The survey shows the well-acceptance of a huge range of computer vision techniques in this domain and thus makes it a wide area of research in the near future. Here are some research points which may be explored to enhance the current state-of-the-art in near future.

Disease stage identification and quantification: Usually, a disease has certain stages, but the proposed work mainly focuses on disease detection and its classification. Thus, the design and implementation of a system that can detect a particular stage of a disease would be of great interest. In addition, these systems should possess the capability to suggest a suitable measure depending on the identified disease stage. Detection of a disease in an early stage, also known as disease forecasting, may help agriculturists to take proper precautions and reduce the damage percentage.

Background elimination: In the proposed study, background elimination is implemented manually. Selecting the Region of Interest (ROI) is just an intermediate module. The process of automatically removing background from leaf image is different research application. It is very difficult to remove different objects from the background such as other leaves, soil, hands/feet, etc. However, the system is trained using leaf images with complex background but these images are taken from an online dataset. To the best of author knowledge, background elimination is an essential step in systems focusing on leaf disease detection and classification. Researchers do capture images in complex background but they align the required ROI accurately for robustness. Considering the datasets used in the research the ROI is not that precisely focused. Also there exists more than one leaf (sometimes infected from different diseases) in a single image. As a result, background elimination is used in the current implementation of the system and is thus not robust against ROI identification step. Thus this point can be focused in future.

Accurate classification: Disease identification is bit simpler than its proper classification. Sometimes it becomes difficult for an expert to classify a particular infection with 100% confidence. Development of a system that can correctly categorize various fungal, viral, and bacterial diseases may also be focused. Literature considers minerals or nutrients deficiencies as another form of plant disease. The development of a system that can effectively differentiate between an infection and a deficiency is another interesting topic of research in this domain. This can be considered as a very difficult objective because from the experts perspective separating an infected leaf from a deficient leaf is a complex task.

Reliability of fully automatic systems: Another issue of concern is the semi-automatic nature of the proposed system. Attempts have been made for complete automation of these systems to mimic the judgmental ability of an expert. But at some stage or other an agriculture engineer (or a plant pathologist) intervention is often required to keep a check on the accuracy. One possible solution to this is an intelligent blend of the expert system concept into the computer vision and machine learning techniques. An attempt to develop such a system may also be of great interest to researchers in this domain. The research can also be extended to three dimensional leaf images [142]. The rules are framed after observing some segmentation characteristics for the training images. Thus, an attempt to develop a fully automatic system for identification as well as classification can be targeted in the near future. Finally, several diseases categories are not a part of the current study, this range can also be broadened.

Appendix A

FEATURES AND CLASSIFIER

Color and texture features used for classification are described here.

A.1. Color Features

Color plays an important role in disease lesion classification. Color moments, autocorrelogram, and HSV histogram are the main features used in this system. Color moments consist of 5-features such as mean (μ), standard deviation (σ), variance (V), kurtosis (K) and skewness (S). Mean is average value of intensities at pixel. Standard deviation indicated the gray level spread in the region of the mean. Around the mean, the third moment skewness calculates the gray levels asymmetry and symmetry. Kurtosis measures the intensity distribution closeness to the shape of normal Gaussian [134]. Moments are described in Eq. 1 – Eq. 5.

$$\mu = \frac{1}{mn} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} P(i, j) \quad \text{(A.1)}$$

$$\sigma = \sqrt{\frac{1}{mn} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} (P(i, j) - \mu)^2} \quad \text{(A.2)}$$

$$V = \sqrt{\sigma} \quad \text{(A.3)}$$

$$S = \frac{1}{mn} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} \left(\frac{P(i,j) - \mu}{\sigma}\right)^3 \quad \text{(A.4)}$$

$$K = \left\{\frac{1}{mn} \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} \left(\frac{P(i,j) - \mu}{\sigma}\right)^4\right\} - 3 \quad \text{(A.5)}$$

$P(i, j)$ represents the intensity value of image with size $m \times n$.

Color autocorrelogram integrates the spatial information along with color histogram. Let in image I, pixel A_1 with color C_k, j distance from b_1 choose another b_2 pixel, then find what would be probability b_2 is having color C_k. The image auto-correlogram is then defined as Eq. 6.

$$\gamma_{C_k}^{j}(I) = A_r\left[|b_1 - b_2| = j, b_2 \in I_{C_k} \middle| b_1 \in I_{C_k}\right] \text{.......} \quad (A.6)$$

From this equation space and color information is integrated. $O(j \times n^2)$ where j indicates the count of neighborhood pixels which depends upon the distance selection [135]. 64-dimensional feature vector is extracted. HSV histogram involves color space conversion, quantization, and histogram computation [136]. Computational complexity is minimized by HSV color space quantized and 32-dimensional feature vector is extracted.

A.2. Texture Features

Literature for plants disease detection indicates that texture is the most suitable feature for disease detection [93]. Haralick features, Gabor features, wavelet features are popular features which are utilized in this study. 14 texture features are calculated using the gray-level co-occurrence matrix (GLCM) to provide a measure of the intensity variation at a particular pixel proposed by the Haralick et al. [137]. Gabor functions give the optimal resolution in together the spatial domain and frequency domain [138]. Gabor features are extracted with 4 scale and 6 orientations. Fast fourier transform (FFT) is used to perform convolutions. Gabor function is described as Eq. 7.

$$\mathrm{g(x,y)} = \left(\frac{1}{2\pi\sigma_x\sigma_y}\right) exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx\right] \text{.......} \quad (A.7)$$

and Fourier transform is described as Eq. 8.

$$\mathrm{G(u,v)} = exp\left[-\frac{1}{2}\left(\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2}\right)\right] \text{.......} \quad (A.8)$$

$where, \sigma_u = 1/2\pi\sigma_x, \sigma_v = 1/2\pi\sigma_y, g_{mn}(x, y) = a^{-m} g(x', y'), x' = a^{-m}(xcos\theta + ysin\theta)$,

$y' = a^{-m}(-xsin\theta + ycos\theta)$, $\theta = n\pi/K$, and m, n are integers, K represents orientations .

Mean-squared energy and mean amplitude is calculated for each scale and orientation. 48-dimensional feature vector is utilized from the Gabor features.

For expressing the features efficiently and easily by signal, wavelet transform gives a basis. The main advantage of this method over the Fourier transform is localization within the both domains (frequency and spatial). The image can be decomposed into 4 sub images. Filters are applied vertically then horizontally [139]. 2D discrete wavelet transform (DWT) contains four elements such as three wavelet functions ($g^H(x,y)$, $g^V(x,y)$, and $g^D(x,y)$) and scaling function $s(x,y)$. The wavelets measure the variations of intensity along with dissimilar images orientations, as g^H, g^V, and g^D represents variations computed in horizontal, vertical and diagonal orientation respectively. DWT function of image $i(x,y)$ with size $m \times n$ is described in Eq. 9 and Eq. 10.

$$\mathrm{W_g}(\mathrm{j_0}, \mathrm{r}, \mathrm{c}) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} i(x,y) g_{j_0,r,c}(x,y) \quad \text{(A.9)}$$

$$W_g^i(j,r,c) = \frac{1}{\sqrt{mn}} \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} i(x,y) g_{j,r,c}^i(x,y), i = \{H,V,D\} \quad \text{(A.10)}$$

where j_0 indicates the arbitrary initial scale, image approximation is indicated by coefficients of $W_g(j_0,r,c)$. The vertical, horizontal, and diagonal aspects are given by coefficients of $W_g^i(j,r,c)$ for $j \geq j_0$ [139, 140]. Two coefficients mean and standard deviation of discrete wavelet transform is calculated which provides 40-dimensional feature set.

A.3. Classifier

SVM is used due to their numerous advantages in high dimensional space. SVM deals beautifully with noisy data and also prevents over fitting [132]. A two-class SVM learns an optimal hyperplane $\boldsymbol{g(x)}$, as given in Eq. (11), to properly split data points of the classes. Here, $\boldsymbol{a_n}$ is a non-zero valued Lagrange multiplier, $\boldsymbol{z_n}$ is its corresponding support vector, $\boldsymbol{t}$ is the size of input data, $\boldsymbol{c}$ is the threshold, and $\boldsymbol{x_n}$ represents an input data point [133].

$$\mathrm{g(x)} = \sum_{1 \leq n \leq t} \mathrm{z_n a_n (x_n + x) + c} \quad \text{(A.11)}$$

Appendix B

WEB APPLICATION ARCHITECTURE

B.1. Overall Architecture of the System

In this system front end is designed using C# and backend contains MATLAB programming. MATLAB Compiler SDK provides .NET assemblies which can be used to enable MATLAB application conversion into enterprise application. Initially, functions are created in MATLAB, then new project is created by adding main function and supporting files using library compiler. Then the generated .dll files are used as reference

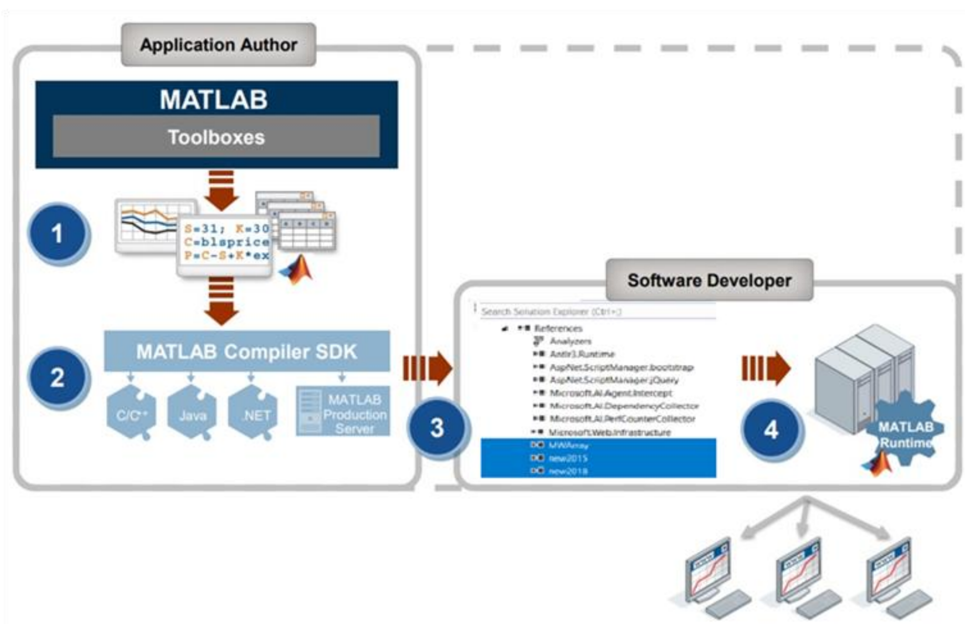


Fig.B.1 General flow of complete system

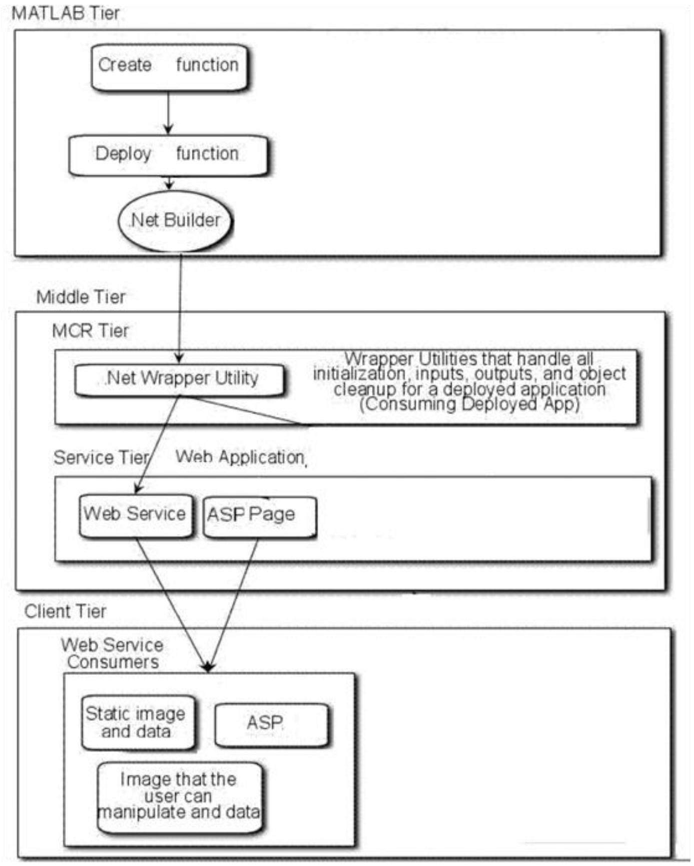


Fig.B.2 Layers involved in overall web application development

in C# web application. Internet Information Services (IIS) is used as web server to run ASP.NET web application.

1 The first stage in a deployment application is writing a code in MATLAB 2018a using MATLAB toolboxes. If MATLAB programmer wants to share the code with others, this objective is accomplished by using MATLAB Compiler.

2 MATLAB compiler* enables MATLAB code to be used by other people in various environments who do not have MATLAB knowledge. Then MATLAB programmer provides the deployable components to the front end developer.

3 MATLAB Compiler SDK provides .NET assemblies which can be used to enable MATLAB application conversion into enterprise application. MATLAB compiler SDK broadens the functionality of MATLAB Compiler for building Microsoft .NET assemblies; C/C++ shared libraries, Python packages, and Java classes. These components then can be incorporated with the custom applications and further used in

* www.mathworks.com

deployment to web, desktop, and enterprise systems. Applications build using software components from the MATLAB Compiler SDK is shared freely with other users which do not require MATLAB.

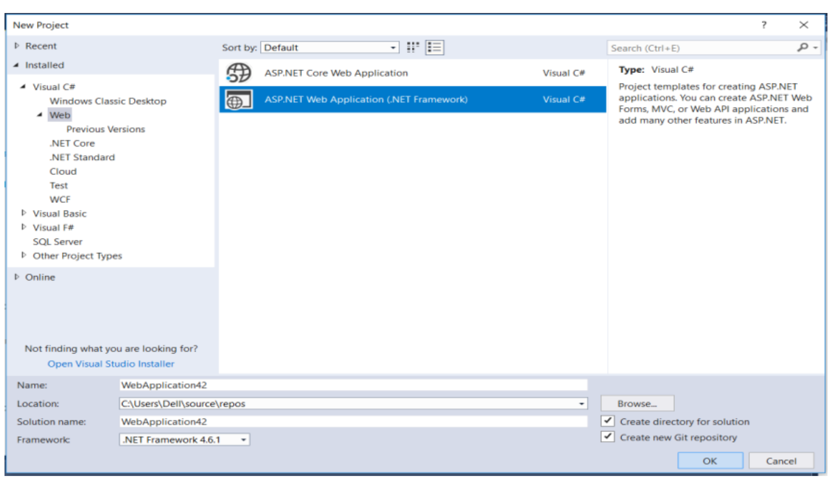
4 This application cannot be shared directly to users which work on other network. For providing the access to other users, some other programming languages which support web application is required as front end. To share this application to other users such as pathologists, farmers, this system is further utilized to create web application in ASP.NET C# language in Microsoft Visual Studio. Users only required web browser to access the application.

B.2. Layers Involved in the System

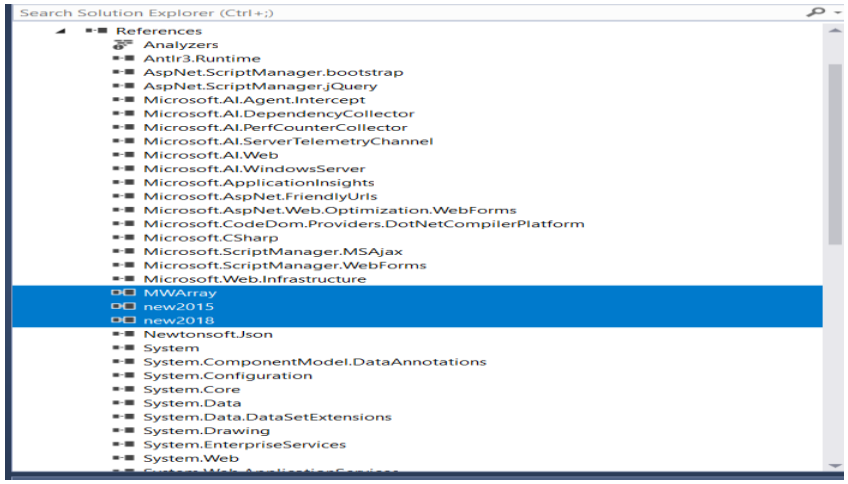
First tier is MATLAB tier, in which application author make use of existing toolbox functions to implement some new useful system also described in Fig. B.1. Then to share the function with other users, deployment function is used. MATLAB compiler SDK is utilized in this tier and is shown in Fig. B.2 to create web application. Then MCR wrapper utility is used to handle all initialization inputs, outputs and object cleanup for deployed application. Web application is developed by software developer which acts as front end. Then application is utilized by the users in client tier.

B.3. Steps Involved in Integrating MATLAB Components in Web Application

1. In Visual Studio Create a new project



2. Create a reference to .NET assembly generated from Library compiler in MATLAB and define in web application.



Incorporating compiled MATLAB functions to the .NET application needs to use various APIs combination. MATLAB Compiler SDK uses these APIs for the initialization of the MATLAB Runtime, and loads the compiled MATLAB functions to the MATLAB Runtime, and then manages data which is passed between the .NET code and the MATLAB Runtime. The compiler generates a few of the APIs depending upon the signatures of compiled functions. MATLAB Runtime gives other APIs which are consistent for all the applications.

3. Add UI objects to web application
4. Add reference to the API in the C# code by ‘using’ command.
5. Call the specific class by passing input arguments.

REFERENCES

[1] Savary S, Ficke A, Aubertot JN, Hollier C (2012) Crop losses due to diseases and their implications for global food production losses and food security. Food Security 4 (4):519-537.

[2] ‘Diagnosing plant problems: plant diseases and disorders’, https://firstdetector.org/static/pdf/NPDNDiagnosingPlantProblemsPlantDiseaseforreview2.pdf, Last accessed: February 2019.

[3] Bagde S, Patil S, Patil S, Patil P (2015) Artificial Neural Network Based Plant Leaf Disease Detection. International Journal of Computer Science and Mobile Computing 4(4):900-905.

[4] Weizheng S, Yachun W, Zhanliang C, Hongda W (2008) Grading method of leaf spot disease based on image processing. In: IEEE Int. Conf. on Computer Science and Software Engineering, Wuhan, Hubei, December 8, pp. 491-494.

[5] Meunkaewjinda A, Kumsawat P, Attakitmongcol K, Srikaew A (2008) Grape leaf disease detection from color imagery using hybrid intelligent system. In: IEEE 5th Int. Conf. on Electrical Engineering / Electronics, Computer, Telecommunications and Information Technology ECTI-CON., Krabi, pp. 513-516.

[6] Phadikar S, Sil J, Das AK (2013) Rice diseases classification using feature selection and rule generation techniques. Comput Electron Agr 90:76-85.

[7] Rastogi A, Arora R, Sharma S (2015) Leaf disease detection and grading using computer vision technology and fuzzy logic. In: IEEE 2nd Int. Conf. on Signal Processing and Integrated Networks SPIN, pp. 500-505.

[8] Zhang S, Zhang C (2013) Orthogonal locally discriminant projection for classification of plant leaf diseases. In: IEEE Int. Conf. on Computational Intelligence and Security CIS, Leshan, pp. 241-245.

[9] Prasad S, Peddoju SK, Ghosh D (2016) Multi-resolution mobile vision system for plant leaf disease diagnosis. Signal Image Video P 10(2):379-388.

[10] Patil JK, Kumar R (2011) Advances in image processing for detection of plant diseases. Journal of Advanced Bioinformatics Applications and Research 2(2):135-141.

[11] Khirade SD, Patil AB (2015) Plant Disease Detection Using Image Processing. In: IEEE Int. Conf. on Computing Communication Control and Automation (ICCUBEA), pp. 768-771.

[12] Barbedo JGA (2013) Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus 2(1):1-12.

[13] Barbedo JGA (2016) A review on the main challenges in automatic plant disease identification based on visible range images. Biosyst Eng 144:52-60.

[14] Prashar K, Talwar R, Kant C (2015) A Review on Efficient Identification of American Cotton Leaf Diseases through Training Set. Int J Comput Appl 132(7):32-39.

[15] Martinelli F, Scalenghe R, Davino S et al. (2015) Advanced methods of plant disease detection. A review. Agron Sustain Dev 35(1):1-25.

[16] Sankaran S, Mishra A, Ehsani R, Davis C (2010) A review of advanced techniques for detecting plant diseases. Comput Electron Agr 72(1):1-13.

[17] Zhang M, Meng Q (2011) Automatic citrus canker detection from leaf images captured in field. Pattern Recognit Lett 32(15):2036-2046.

[18] 'Staple food', https://en.wikipedia.org/wiki/Staple_food#cite_note-12, Last accessed February 2019.

[19] 'Nitrogen fixation', https://en.wikipedia.org/wiki/Nitrogen_fixation, Last accessed February 2019.

[20] 'Soybean', https://en.wikipedia.org/wiki/Soybean#cite_note-gsp-5, Last accessed February 2019.

[21] 'Signs and symptoms of plant disease: Is it fungal, viral or bacterial?', http://msue.anr.msu.edu/news/signs_and_symptoms_of_plant_disease_is_it_fungal_viral_or_bacterial, Last accessed January 2019.

[22] Li X, Yang XB (2009) Similarity, pattern, and grouping of soybean fungal diseases in the United States: Implications for the risk of soybean rust. Plant Disease 93(2):162-169.

[23] 'Soybean growth and development', http://corn.agronomy.wisc.edu/Crops/Soybean/pdfs/L004.pdf, Last accessed February, 2019.

[24] Mian MA, Missaoui AM, Walker DR et al. (2008) Frogeye Leaf Spot of Soybean: A Review and Proposed Race Designations for Isolates of Cercospora sojina Hara. Crop Science 48(1):14-24.

[25] 'Characteristics of soybean foliar diseases from bacterial blight to rust', http://extension.cropsciences.illinois.edu/fieldcrops/classics/2005/characteristicsofsoybeandiseases.php, Last accessed January 2019.

[26] 'Soybean diseases', https://alliedcooperative.files.wordpress.com/2014/07/soybeandiseases.pdf, Last accessed January, 2019.

[27] Husin ZB, Aziz AHBA, Shakaff AYBM, Farook RBSM (2012) Feasibility study on plant chili disease detection using image processing techniques. In: IEEE 3rd Int. Conf. on Intelligent System Modeling and Simulation ISMS, Kota Kinabalu, pp. 291-296.

[28] Pujari JD, Yakkundimath R, Byadgi AS (2016) SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique. International Journal of Interactive Multimedia and Artificial Intelligence 3:6-14.

[29] Jadhav SB, Patil SB (2015) Grading of Soybean Leaf Disease Based on Segmented Image Using K-means Clustering. International Journal of Advanced Research in Electronics and Communication Engineering 4(6):1816-1822.

[30] Gonzalez RC, Woods RE: Digital Image Processing. 3rd edn. Prentice Hall (2008).

[31] Sharma S, Singh D, Chauhan S (2017) An interactive computer vision system for tree ring analysis. Current Science (00113891), 112(6): 1262-1265.

[32] Shweta, Singh D (2016) Computer aided leaf morphometric approach for the identification of regional plant species. Environment and Ecology 34(3C):1556-1561.

[33] Shivling VD, Singla A, Ghanshyam C, Kapur P, Gupta S (2011) Plant leaf imaging technique for agronomy. In: IEEE Int. Conf. in Image Information Processing ICIIP, Shimla, India, pp. 1-5.

[34] Grewal GK, Sawhney BK, Sirari A (2016) A Mobile Application in Regional Language (PUNJABI) for Identification and Management of Diseases in Mungbean. International Journal of Research and Scientific Innovation 3(8): 176-180.

[35] 'Machine Learning', https://en.wikipedia.org/wiki/Machine_learning, Last accessed February 2019.

[36] 'Types of Machine Learning Algorithms You Should Know', https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861, Last accessed February 2019.

[37] Dewangan SK (2016) Importance & Applications of Digital Image Processing. International Journal of Computer Science & Engineering Technology (IJCSET) 7(7):316-320.

[38] 'Applications of image processing in biology and agriculture', http://www.barc.gov.in/publications/nl/1999/199905-01.pdf, Last accessed February 2019.

[39] Saxena L, Armstrong L (2014) A survey of image processing techniques for agriculture. In: Proceedings of Asian Federation for Information Technology in Agriculture, Perth WA Australian Society of Information and Communication Technologies in Agriculture, pp. 401-413.

[40] Phadikar S, Sil J, Das AK (2012) Classification of Rice Leaf Diseases Based on Morphological Changes. International Journal of Information and Electronics Engineering 2(3):460-463.

[41] 'IPM images', https://www.ipmimages.org/about/. Last accessed 15 February 2019.

[42] Mohanty SP, Hughes D, Salathé M (2016) Using Deep Learning for Image-Based Plant Disease Detection. Front Plant Sci 7:1-10.

[43] 'APS Image database', https://imagedatabase.apsnet.org/search.aspx. Last accessed 16 February 2019.

[44] Gharge S, Singh P (2016) Image Processing for Soybean Disease Classification and Severity Estimation, In: Shetty N, Prasad N, Nalini N (eds) Emerging Research in Computing, Information, Communication and Applications. Springer, New Delhi, India, pp. 493-500.

[45] 'University of Minnesota Extension', https://www.extension.umn.edu/. Last accessed 16 February 2019.

[46] Sanyal P, Bhattacharya U, Parui SK, Bandyopadhyay SK, Patel S (2007) Color texture analysis of rice leaves diagnosing deficiency in the balance of mineral levels towards improvement of crop productivity In: IEEE 10th Int. Conf. on Information Technology ICIT, Orissa, pp. 85-90.

[47] Camargo A, Smith JS (2009) An image-processing based algorithm to automatically identify plant disease visual symptoms. Biosyst Eng 102(1):9-21.

[48] Pujari JD, Yakkundimath RS, Jahagirdar S, Byadgi AM (2015) Quantitative detection of soybean rust using image processing techniques. J Crop Prot 5(1):75-87.

[49] Rumpf T, Mahlein AK, Steiner U, Oerke EC, Dehne HW, Plümer L (2010) Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Comput Electron Agr 74(1):91-99.

[50] Pires RDL, Gonçalves DN, Oruê JPM, Kanashiro WES, Rodrigues JF, Machado BB, Gonçalves WN (2016) Local descriptors for soybean disease recognition. Comput Electron Agr 125:48-55.

[51] Barbedo JGA (2014) An automatic method to detect and measure leaf disease symptoms using digital image processing. Plant Dis 98(12):1709-1716.

[52] Anthonys G, Wickramarachchi N (2009) An image recognition system for crop disease identification of paddy fields in Sri Lanka. In: IEEE Int. Conf. on Industrial and Information Systems ICIIS, Sri Lanka, pp. 403–407.

[53] Pydipati R, Burks TF, Lee WS (2006) Identification of citrus disease using color texture features and discriminant analysis. Comput Electron Agr 52(1):49-59.

[54] Bauer SD, Korč F, Förstner W (2011) The potential of automatic methods of classification to identify leaf diseases from multispectral images. Precis Agric 12(3):361-377.

[55] Xu G, Zhang F, Shah SG, Ye Y, Mao H (2011) Use of leaf color images to identify nitrogen and potassium deficient tomatoes. Pattern Recognit Lett 32(11):1584-1590.

[56] Yao Q, Guan Z, Zhou Y, Tang J, Hu Y, Yang B (2009) Application of support vector machine for detecting rice diseases using shape and color texture features. In: IEEE Int. Conf. on Engineering Computation ICEC, Hong Kong, pp.79-83.

[57] Mirik M, Michels GJ, Kassymzhanova-Mirik S, Elliott NC, Catana V, Jones DB, Bowling R (2006) Using digital image analysis and spectral reflectance data to quantify damage by greenbug (Hemitera: Aphididae) in winter wheat. Comput Electron Agr 51(1):86-98.

[58] Huang KY (2007) Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. Comput Electron Agr 57(1):3-11.

[59] Oberti R, Marchi M, Tirelli P, Calcante A, Iriti M, Borghese AN (2014) Automatic detection of powdery mildew on grapevine leaves by image analysis: Optimal view-angle range to increase the sensitivity. Comput Electron Agr 104:1-8.

[60] Shrivastava S, Hooda DS (2014) Automatic brown spot and frog eye detection from the image captured in the field. American Journal of Intelligent Systems 4(4):131-134.

[61] Cui D, Zhang Q, Li M, Zhao Y, Hartman GL (2009) Detection of soybean rust using a multispectral image sensor. Sens Instrum Food Qual Saf 3(1):49-56.

[62] Lu J, Cui D, Jiang H (2013) Discrimination of Tomato Yellow Leaf Curl Disease using Hyperspectral Imaging. American Society of Agricultural and Biological Engineers, Kansas City, Missouri, July 21-July 24, (p.1).

[63] Kumbhar NP, Dhaygude SB (2013) Crop Disease Detection Using CBIR. International Journal of Emerging Technologies in Computational and Applied Sciences 3(2):113-118.

[64] Asfarian A, Herdiyeni Y, Rauf A, Mutaqin KM (2013) Paddy diseases identification with texture analysis using fractal descriptors based on Fourier spectrum. In: IEEE Int. Conf. on Computer, Control, Informatics and Its Applications IC3INA, Jakarta, pp. 77-81.

[65] Sannakki SS, Rajpurohit VS, Nargund VB, Kulkarni P (2013) Diagnosis and classification of grape leaf diseases using neural networks. In: IEEE 4th Int. Conf. on Computing, Communications and Networking Technologies ICCCNT, Tiruchengode, pp. 1-5.

[66] Wang L, Dong F, Guo Q, Nie C, Sun S(2014) Improved rotational kernel transformation directional feature for recognition of wheat stripe rust and powdery mildew. In: IEEE 7th Int. Conf. on Image and Signal Processing CISP, Dalian, pp. 286-291.

[67] Zhang SW, Shang YJ, Wang L (2015) Plant Disease Recognition Based On Plant Leaf Image. J Anim Plant Sci 25(Suppl. 1):42-45.

[68] Arivazhagan S, Shebiah RN, Ananthi S, Varthini SV (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agricultural Engineering International: CIGR Journal 15(1):211-217.

[69] Bashish AD, Braik M, Bani-Ahmad S (2010) A framework for detection and classification of plant leaf and stem diseases. In: IEEE Int. Conf. on signal and image processing ICSIP, Chennai, pp. 113-118.

[70] Majumdar D, Kole DK, Chakraborty A, Majumdar DD (2015) An integrated digital image analysis system for detection, recognition and diagnosis of disease in wheat leaves. In: 3rd Int. ACM Internat. Symposium on Women in Computing and Informatics., New York, pp. 400-405.

[71] Wiwart M, Fordoński G, Żuk-Gołaszewska K, Suchowilska E (2009) Early diagnostics of macronutrient deficiencies in three legume species by color image analysis. Comput Electron Agr 65(1):125-132.

[72] Kruse OMO, Prats-Montalbán JM, Indahl UG, Kvaal K, Ferrer A, Futsaether CM (2014) Pixel classification methods for identifying and quantifying leaf surface injury from digital images. Comput Electron Agr 108:155-165.

[73] Kai S, Zhikun L, Hang S, Chunhong G (2011) A research of maize disease image recognition of corn based on BP networks. In: Proceedings of 3rd Int. Conf. on Measuring Technology and Mechatronics Automation ICMTMA. IEEE, Shanghai, pp. 246-249.

[74] Molina JF, Gil R, Bojacá C, Gómez F, Franco H (2014) Automatic detection of early blight infection on tomato crops using a color based classification strategy. In IEEE XIX Symposium on Image, Signal Processing and Artificial Vision, September, pp. 1-5.

[75] Wang H, Li G, Ma Z, Li X (2012) Application of neural networks to image recognition of plant diseases. In: IEEE Int. Conf on Systems and Informatics ICSAI, Yantai, pp. 2159-2164.

[76] Youwen T, Tianlai L, Yan N (2008) The recognition of cucumber disease based on image processing and support vector machine. In IEEE Congress on Image and Signal Processing CISP'08, May, Vol. 2, pp.262-267.

[77] Shinde RC, Mathew CJ, Patil CY (2015) Segmentation Technique for Soybean Leaves Disease Detection. International Journal of Advanced Research 3(5):522-528.

[78] Hairuddin MA, Tahir NM, Baki SRS (2011) Overview of image processing approach for nutrient deficiencies detection in elaeis guineensis. In: IEEE Int. Conf. on System Engineering and Technology ICSET, Shah Alam, pp.116-120.

[79] Bandi SR, Varadharajan A, Chinnasamy A (2013) Performance evaluation of various statistical classifiers in detecting the diseased citrus leaves. International Journal of Engineering Science and Technology 5(2):298-307.

[80] Xia C, Lee JM, Li Y, Song YH, Chung BK, Chon TS (2013) Plant leaf detection using modified active shape models. Biosyst Eng 116(1):23-35.

[81] Bankar S, Dube A, Kadam P, Deokule S (2014) Plant Disease Detection Techniques Using Canny Edge Detection & Color Histogram in Image Processing. International Journal of Computer Science and Information Technologies 52(2):1165-1168 62.

[82] Singh V, Misra AK (2015) Detection of unhealthy region of plant leaves using image processing and genetic algorithm. In: IEEE Int. Conf. on Advances in Computer Engineering and Applications ICACEA, pp.1028-1032.

[83] Pujari JD, Yakkundimath R, Byadgi AS (2015) Image processing based detection of fungal diseases in plants. Procedia Comput Sci 46:1802-1808.

[84] Danping W, Botao W, Yue Y (2013) The identification of powdery mildew spores image based on the integration of intelligent spore image sequence capture device. In: IEEE 9th Int. Conf on Intelligent Information Hiding and Multimedia Signal Processing IIH-MSP, Beijing, pp. 170-180.

[85] Phadikar S, Sil J (2008) Rice disease identification using pattern recognition techniques. In: IEEE 11th Int. Conf. on Computer and Information Technology ICCIT, Khulna, pp. 420–423.

[86] Masood R, Khan SA, Khan MNA (2016) Plants Disease Segmentation using Image Processing. International Journal of Modern Education and Computer Science 8(1):24-32.

[87] Cui D, Zhang Q, Li M, Hartman GL, Zhao Y (2010) Image processing methods for quantitatively detecting soybean rust from multispectral images. Biosyst Eng 107(3):186-193.

[88] Pang J, Bai ZY, Lai JC, Li SK (2011) Automatic segmentation of crop leaf spot disease images by integrating local threshold and seeded region growing. In: IEEE Int. Conf. on image analysis and signal processing, October, pp. 590-594.

[89] Shire A, Jawarkar U, Manmode MM (2015) A Review Paper On: Agricultural Plant Leaf Disease Detection Using Image Processing. International Journal of Innovative Science, Engineering & Technology 2(1): 282-285.

[90] Gui J, Hao L, Zhang Q, Bao X (2015) A new method for soybean leaf disease detection based on modified salient regions. Int. J. Multimedia Ubiquitous Eng 10:45-52.

[91] Krishnan M, Sumithra MG (2013) A novel algorithm for detecting bacterial leaf scorch of shade trees using image processing. In: IEEE Malaysia Int. Conf. on Communications MICC, Kuala Lumpur, pp. 474-478.

[92] Kurniawati NN, Abdullah SNHS, Abdullah S, Abdullah S (2009) Investigation on image processing techniques for diagnosing paddy diseases. In: IEEE Int. Conf. on Soft Computing and Pattern Recognition SOCPAR, Malacca, pp. 272-277.

[93] Gavhale KR, Gawande U (2014) An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques. IOSR Journal of Computer Engineering 16(1):10-16.

[94] Shrivastava S, Singh SK, Hooda DS (2014) Statistical Texture and Normalized Discrete Cosine Transform-Based Automatic Soya Plant Foliar Infection Cataloguing. British Journal of Mathematics & Computer Science 4(20):2901-2916.

[95] Aji AF, Munajat Q, Pratama AP, Kalamullah H, Setiyawan J, Arymurthy AM (2013) Detection of Palm Oil Leaf Disease with Image Processing and Neural Network Classification on Mobile Device. International Journal of Computer Theory and Engineering 5(3):528-532.

[96] Gurjar AA, Gulhane VA (2012) Disease detection on cotton leaves by eigenfeature regularization and extraction technique. International Journal of Electronics, Communication and Soft Computing Science & Engineering 1(1):1-4.

[97] Mokhtar U, El Bendary N, Hassenian AE, Emary E, Mahmoud MA, Hefny H, Tolba MF (2015) SVM-based detection of tomato leaves diseases. In Intelligent Systems', Springer International Publishing, pp. 641-652.

[98] ‘A Brief List of Plant Diseases Caused by Fungi’, https://gardenerdy.com/plant-diseases-caused-by-fungi, Last accessed February 2019.

[99] Hughes DP, Salathé M (2015) An open access repository of images on plant health to enable the development of mobile disease diagnostics. CoRR abs/1511.08060.

[100] Giesler LJ ‘Frog eye leaf spot of soybean’, http://extensionpublications.unl.edu/assets/pdf/g2213.pdf, Last accessed February 2019.

[101] Ghaiwat SN, Arora P (2015) Cotton Leaf Disease Detection by Feature Extraction, in Research Advances in the Integration of Big Data and Smart Computing, Oct 13:89.

[102] Camargo A, Smith JS (2009) Image pattern classification for the identification of disease causing agents in plants. Comput Electron Agr 66(2):121-125.

[103] Rothe PR, Kshirsagar RV (2015) Cotton leaf disease identification using pattern recognition techniques. In: IEEE Int. Conf. on Pervasive Computing (ICPC), January, pp. 1-6.

[104] Ramakrishnan M (2015) Groundnut leaf disease detection and classification by using back probagation algorithm. In: IEEE Int. Conf. on Communications and Signal Processing (ICCSP), April, pp. 0964-0968.

[105] Barbedo JGA (2013) Automatically measuring early and late leaf spot lesions in peanut plants using digital image processing. In Congresso brasileiro de agroinformática Vol. 9.

[106] Nisale SS, Bharambe CJ, More VN (2011) Detection and Analysis of Deficiencies in Groundnut Plant using Geometric Moments. World Acad Sci Eng Technol 5:608-612.

[107] Rangel BM, Fernández MA, Murillo JC, Ortega JC, Arreguín JM (2016) KNN-based image segmentation for grapevine potassium deficiency diagnosis. In: IEEE Int. Conf. on Electronics, Communications and Computers (CONIELECOMP) Feb 24, pp. 48-53.

[108] Kole DK, Ghosh A, Mitra S (2014) Detection of Downy Mildew Disease present in the Grape Leaves based on Fuzzy Set theory, In Advanced Computing, Networking and Informatics, Volume 1, pp. 377-384, Springer International Publishing.

[109] Jian Z, Wei Z (2010) Support vector machine for recognition of cucumber leaf diseases. In: IEEE Int. Conf. on Advanced Computer Control, ICACC, March, Vol. 5, pp. 264-266.

[110] Asefpour Vakilian K, Massah J (2013) An artificial neural network approach to identify fungal diseases of cucumber (Cucumis sativus L.) plants using digital image processing. Archives of Phytopathology and Plant Protection 46(13):1580-1588.

[111] Yun S, Xianfeng W, Shanwen Z, Chuanlei Z (2015) PNN based crop disease recognition with leaf image features and meteorological data. International Journal of Agricultural and Biological Engineering 8(4):60-68.

[112] Sabrol H, Kumar S (2016) Intensity based feature extraction for tomato plant disease recognition by classification using decision tree. International Journal of Computer Science and Information Security 14(9):622-626.

[113] Sabrol H, Kumar S (2016) Fuzzy and Neural Network based Tomato Plant Disease Classification using Natural Outdoor Images. Indian J Sci Technol 9(44):1-8.

[114] Samanta D, Ghosh A (2012) Histogram Approach for Detection of Maize Leaf Damage. International Journal of Computer Science and Telecommunications 3(2):26-28.

[115] Luo J, Geng S, Xiu C, Song D, Dong T (2015) A curvelet-SC recognition method for maize disease. Journal of Electrical and Computer Engineering 2015(25):1-8.

[116] Barbedo JGA, Godoy CV (2015) Automatic Classification of Soybean Diseases Based on Digital Images of Leaf Symptoms. SBI AGRO.

[117] Shrivastava S, Singh SK, Hooda DS (2015) Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation. Multimed Tools Appl 74(24):11467-11484.

[118] Dandawate Y, Kokare R (2015) An automated approach for classification of plant diseases towards development of futuristic Decision Support System in Indian perspective. In: IEEE Int. Conf. on Advances in Computing, Communications and Informatics (ICACCI), August, pp.794-799.

[119] Mohan KJ, Balasubramanian M, Palanivel S (2016) Detection and Recognition of Diseases from Paddy Plant Leaf Images. Int J Comput Appl 144(12):34-41.

[120] Verma T, Satpathy SK, Sharma LK (2011) A Step towards Precision Farming of Rice Crop by Estimating Loss Caused by Leaf Blast Disease Using Digital Image Processing and Fuzzy Clustering, Int. Journal of Computer Trends and Technology pp.1-6.

[121] Majumdar D, Ghosh A, Kole DK, Chakraborty A, Majumder DD (2014) Application of fuzzy c-means clustering method to classify wheat leaf images based on the presence of rust disease. In: Int. Conf. on Frontiers of Intelligent Computing: Theory and Applications (FICTA), Springer International Publishing, pp. 277-284.

[122] Cai X, Sun Y, Zhao Y, Damerow L, Lammers PS, Sun W, Tang Y (2013) Smart detection of leaf wilting by 3D image processing and 2D Fourier transform, Comput Electron Agr 90:68-75.

[123] Schikora M, Neupane B, Madhogaria S, Koch W, Cremers D, Hirt H, Kogel KH, Schikora A (2012) An image classification approach to analyze the suppression of plant immunity by the human pathogen Salmonella Typhimurium. BMC bioinformatics 13(1):171.

[124] Barbedo JGA, Koenigkan LV, Santos TT (2016) Identifying multiple plant diseases using digital image processing. Biosyst Eng 147:104-116.

[125] Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D (2016) Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification. Comput Intell Neurosci 2016:1-11.

[126] Bock CH, Poole GH, Parker PE, Gottwald TR (2010) Plant disease severity estimated visually, by digital photography and image analysis, and by hyperspectral imaging. CRC Crit Rev Plant Sci 29(2):59-107.

[127] Pethybridge SJ, Nelson SC (2015) Leaf Doctor: A new portable application for quantifying plant disease severity. Plant Dis 99(10):1310-1316.

[128] ‘Hunter lab color scale’. Insight on color, Hunter Labs Reston, Aug 1996, 8(9).

[129] Wang X, Hänsch R, Ma L, Hellwich O (2014) Comparison of different color spaces for image segmentation using graph-cut. In: IEEE Int. Conf. on Computer Vision Theory and Applications (VISAPP), 2014, January, pp. 301-308.

[130] Seber GAF ‘Multivariate Observations’, 1984, Wiley, New York.

[131] Spath H ‘Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples’, translated by J. Goldschmidt, Halsted Press, New York, 1985.

[132] Amsaveni V, Singh NA, Dheeba J (2014) Application of Support Vector Machine Classifier for Computer Aided Diagnosis of Brain Tumor from MRI, in Panigrahi B, Suganthan P, Das S (eds) Swarm, Evolutionary, and Memetic Computing. SEMCCO. Lecture Notes in Computer Science, vol 8947. Springer, Cham, pp.514-522.

[133] Yu H, Kim S (2012) SVM tutorial-Classification, Regression and Ranking, in Rozenberg, G., et al. (eds.): Handbook of Natural Computing, (Springer Berlin Heidelberg, pp. 479-506.

[134] Sachdeva J, Kumar V, Gupta I et al. (2016) A package-SFERCB-Segmentation, feature extraction, reduction and classification analysis by both SVM and ANN for brain tumors. Applied Soft Computing 47:151-167.

[135] Huang J ‘Color-spatial image indexing and applications’. Ph.D. thesis, Cornell University, 1998.

[136] Mistry Y, Ingole DT, Ingole MD (2017) Content based image retrieval using hybrid features and various distance metric. Journal of Electrical Systems and Information Technology article in press doi:http://dx.doi.org/10.1016/j.jesit.2016.12.009.

[137] Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. Transactions on Systems, Man, and Cybernetics 3(6):610–621.

[138] Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence 18(8):837-842.

[139] Cheng FH, Chen YL (2006) Real time multiple objects tracking and identification based on discrete wavelet transform. Pattern Recognition 39:1126-1139.

[140] Yadav R, Anand RS, Dewal ML et al. (2015) Performance analysis of discrete wavelet transform based first-order statistical texture features for hardwood species classification. Procedia Computer Science 57:214-221.

[141] Kaur S, Pandey S, Goel S (2017) An automatic leaf disease detection system for legume species. Journal of Biology and Today's World 6:115-122.

[142] Cai J, Miklavcic S (2012) Automated extraction of three-dimensional cereal plant structures from two-dimensional orthographic images. IET Image Processing 6(6):687-696.