Quality Assessment and Classification of Basil Using Computer Vision

A thesis

Submitted in fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

in

Electronics and Communication Engineering

By

Gittaly

Regn. No: 901306011

Under the supervision of

Dr. Vinay Kumar | **Dr. Hem Dutt Joshi**

Associate Professor | **Associate Professor**

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

Department of Electronics and Communication Engineering

Thapar Institute of Engineering and Technology

(Deemed to be University),

Patiala-147004, India

September 2020

Declaration

I, Gittaly Dhingra, hereby certify that the work, which is being presented in the thesis, entitled "Quality Assessment and Classification of Basil Using Computer Vision" submitted in Thapar Institute of Engineering and Technology, Patiala in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Electronics and Communication Engineering, is an authentic record of my own work carried out during the period Jan 2014 to June 2019 under the supervision of Dr. Vinay Kumar and Dr. Hem Dutt Joshi.

I have also cited the reference about the text(s)/figure(s)/table(s) from where they have been taken. The matter presented in this thesis has not been submitted elsewhere for the award of any other degree or diploma from any institution.

Date: 15/Sep/2020

Gittaly

Gittaly

Candidate

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

Dr. Vinay Kumar

Associate Professor

Dr. Hem Dutt Joshi

Associate Professor

........dedicated to

my family

Abstract

The natural products are inexpensive, non-toxic, and have fewer side effects. Thus, their demand especially herbs based medical products, health products, nutritional supplements, cosmetics etc. are increasing worldwide. Majority of medicines are ready from medicinal plants. But, due to diseases medicinal plants growth diminish severely. The diseases possess threats to economic, and production status in the medicinal industry worldwide. So, it is mandatory to continuous measuring quality of plants to predict the disease extremity. Earlier, manual observation is used to analyze quality but somehow it is tedious, inconsistent and costly. By now, studies show that digital image processing methods work as effective tools for the identification and classification of plants diseases.

Medicinal plants as basil, neem, aloe, pepper, and turmeric are widely used for preparation of Ayurvedic and allopathic medicines. Particularly, basil has an intense significance in medicine prospective. So, basil disease detection and classification using computer vision is the motivation of presented work. Pathologists focus on diseases in different parts of the plant like roots, kernel, stem and leave. The presented thesis concentrate, particularly on leaves. The work present in this thesis is focus on to design a new framework for segmentation, feature extraction and classification. After, data set preparation a new segmentation technique with neutrosophic logic is used to detect and identify region of disease. Features are extracted from segmented regions using amalgamation of texture and color features. New texture feature is also introduced named as bin binary pattern. Then, we used different classification models for diseases predication. As comparison to existing segmentation techniques, proposed method gives promising results.

A classification algorithm using survival of fittest approach is proposed in other work. Best solution is obtained through the fitness function with minimum distance and maximum similarity values using maximum aggregation analysis. As comparison to existing machine learning methods; proposed classification method provides best results. Moreover, present thesis contributes in the area of healthcare and medicines, which plays significant role in daily life.

Acknowledgement

Pursuing for PhD was an idea to me planted by the enthusiasm of dreams. Persevering for 5 years and 150 yards to the Electronics & Communication Engineering Department, Thapar Institute of Engineering & Technology; has made it possible to accomplish my work and producing this piece of thesis after conducting several trials and finding the new perspective in form the research work. A keen interest in image processing and computer vision led the drive in me to push hard for this critical work. Guiding my path was encouragement from my family and constant faith of my mentors Dr. Vinay Kumar, Associate Professor and Dr. Hem Dutt Joshi, Associate Professor. It wasn't been without the guidance of my supervisors who portrayed my dream into the substantial research and paved the pathway in a sequential phased manner, this work would have left to be completed. This piece of thesis is encrypted by the colossal efforts of my supervisors. It's very difficult to thank him in words for being the source of my pillar strength and in every other aspect. I would like to express my gratitude towards Dr. Alpana Aggarwal, Professor and Head of Department, she made sure that the infrastructure surely should suffice with our requirements.

I also want to thanks my PhD committee members Dr. Kulbir Singh, Professor, Dr. Amit Kumar Kohli Associate Professor and Dr. Neeraj Kumar, Associate Professor (CSED), they played a pivotal role for being a source of inspiration during all the years. I would like to thank fellow researchers & faculty members Dr. Amanpreet Kaur, Dr. Sakshi, Dr. Manish Deshwal, Madhu, Harishankar, Amneet, Kanwarpreet Kaur, Navneet Kaur, Vishal and Vivek Singh Bawa for their support. My parents Mr. Raghu Nath Dhingra, Mrs. Nirmal Dhingra, Sr. Jagjit Singh Sandhu, Mrs. Bhupinder Kaur, brothers Gaurav Dhingra and Saurav Dhingra for whom I am dedicating my research work. I also want to thanks family members Mrs. Pritpal Kaur, Ms. Darshan Kaur, Aman Sohi, Mrs. Harjit Kaur for giving me shelter and unconditional support during my research work.

My husband Sandeep Sandhu and son Abeer Sandhu who provide, me an encouragement and valuable support to carry out my dreams. Their love has been unconditional for all my life. It was their dream and love who gave me strength to overcome all the hurdles and accomplish this.

List of Publications

Published:

- ✓ ***Springer Nature:***

Multimedia Tools and Applications

Title: Study of digital image processing techniques for leaf disease detection and classification.

G. Dhingra, V. Kumar and H.D. Joshi, “Study of digital image processing techniques for leaf disease detection and classification,” Multimedia Tools and Applications, vol. 77, no. 15, pp. 19951-20000, 2017 (IF= 2.3).

- ✓ ***Journal of the International Measurement Confederation:***

Measurement

Title: A novel computer vision based neutrosophic approach for leaf disease identification and classification.

G. Dhingra, V. Kumar and H.D. Joshi, “A novel computer vision based neutrosophic approach for leaf disease identification and classification,” Measurement, vol.13, pp. 782-794, 2018 (IF= 3.34).

- ✓ ***An International Journal Sponsored by the Chemometrics Society:***

Chemometrics and Intelligent Laboratory Systems

Title: Basil leaves disease classification and identification by incorporating survival of fittest approach

G. Dhingra, V. Kumar and H.D. Joshi, “Basil leaves disease classification and identification by incorporating survival of fittest approach,” Chemometrics and Intelligent Laboratory Systems, vol. 186, pp. 1- 11, 2019 (IF= 2.789).

Table of Contents

List of Figures

List of Tables

Abbreviations

ANN	Artificial Neural Networks
AUC	Area under the Curve
ANFIS	Adaptive Neuro Fuzzy Inference System
AYUSH	Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy
BBP	Bin Binary Pattern
CNN	Convolution Neural Networks
CIGDFNN	Cross Information Gains Depth Forward Neural Network
CBIR	Content Base Image Retrieval
DI	Damage Index
DSR	Diseases Sequence Region
CCM	Color Co-occurrence Method
FLSVM	Fuzzy Least Square Vector Machine
FP	False Positive
FN	False Negatives
GLCM	Gray Level Co-occurrence Matrix
HIS	Hue, Saturation and Intensity
HIC	Histogram Information Content
LUT	Look Up Table
LDA	Linear Discriminant Analysis
LM	Linear Model
LBP	Local Binary Pattern Features
KNN	K-Nearest Neighbor
MRF	Markov Random Field
OLDPA	Orthogonal Locally Discriminant Projection Algorithm
PNN	Probabilistic Neural Network
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
PPV	Positive Predictive Value
RBF	Radial Basis Function
RF	Random Forest
ROC	Receiver Operating Characteristics
SGLDM	Spatial Gray Level Dependence Matrices
STEPDISC	Step Wise Discriminant Analysis
SVM	Support Vector Machines
SVD	Singular Value Decomposition
TN	True Negative
TP	True Positive

Chapter 1

Introduction

This section discusses the importance of medicinal plants, their benefits and traditional quality assessment measures. Benefits of basil leave and their diseases are discussed in section 1.2 and 1.3. Applications of computer vision and digital image processing techniques in agricultural and plants science sector are briefly described in section 1.4 and 1.5. The later section discusses research gaps, objectives and contributions of the proposed thesis. Later the concise overview of thesis is highlighted.

1.1 Introduction

Indian continent is huge storehouse of medicinal plants. The medicinal attributes make them very useful to mankind [1]. Though rural masses use them for their prime health care issues in developing countries, it is also promising for modern medicines dominated developed countries. From the last 4000 years, Indian Vaids, Mediterranean cultures and Europeans are using medicinal plants for healing purposes. Approximately 25,000 plant-based formulations are used in traditional medicine in India [2]. These medicinal plants are recognized worldwide to harmonize health maintenance, treatment and diseases prevention. The health treatment using medicinal plants is measured very safe over allopathic medicines.

AYUSH (Ayurveda, Yoga and Naturopathy, Unani, Siddha and Homoeopathy) department in India, have summarized around 25,000 herbal treatments [3]. The utilization of medicinal plants is enhancing rapidly worldwide, due to increasing demands of natural healthy goods and herbal drugs. In countries like India and China, involvement of medicinal plants in medicine is around 80% [4]. From Figure 1.1, it can be easily realized that India and China are utilizing around 7,500 and 11,146 species, respectively. It is comparatively a massive amount as compare to the other countries such as Columbia, Southern Africa, etc [5].

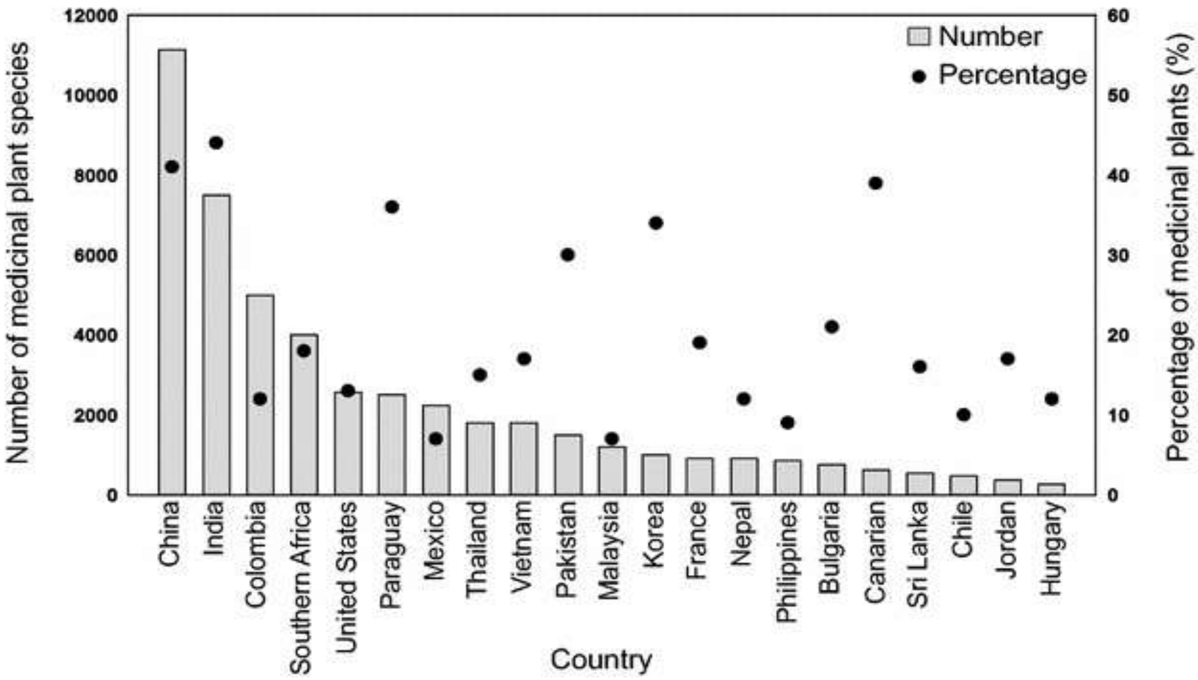


Figure 1.1: Number and percentages of medicinal plants species in all over countries [5]

Medicinal plants-based medicines have numerous advantages over modern allopathic medicines [6]:

- Renewable sources
- Lesser side effects
- Environment friendly processing
- Lower cost
- Efficient in chronic health conditions
- Easily available

So, due to above mentioned benefits, populations rely on traditional medicinal plants-based medicines. Medicinal plants such as basil, neem, aloe, pepper, and turmeric are well known medicinal plants used in drug development. Specially, basil has a profound significance in medicine and religious prospective. It is used fresh, dried, and processed for flavoring, fragrances, and in traditional medicines. Using unique amalgamation of pharmacological actions, basil can handle chemical, physical, psychological and metabolic stress [7]. It has been used for around 300 different herbal treatments to support healthy response to stress, energy booster, increase stamina, healing properties, cardiovascular health, cancer, heart diseases, cardio-protective, anti-hypercholesterolemia, anti-ulcer, anti-emetic, anti-spasmodic, anti-hypertensive, anti-diabetic, anti-carcinogenic, anti-asthmatic, anti-tussive, neuro-protective, etc. [8-9].

Swiss Federal Institute of Technology observed the existence of high quantities of (E)-beta-caryophyllene (BCP) in basil, which is believed to be helpful in the treatment of arthritis and inflammatory bowel diseases [10]. Figure 1.2 illustrates the benefits of medicinal plant basil. Basil is indigenous to the countries of Iran and India as well as other tropical regions of Asia. It is cultivated extensively in India, Greece, Egypt, Indonesia, United States, France, Hungary, Morocco, and Israel. It contains the essential oil and oleoresin required for manufacturing perfumes, food flavors, and aromatherapy products [11]. The compound composition of basil oil has been the subject to a considerable number of studies. All the basil medications are accessible in the form of decoction, oil, powder and paste. Basil based medicines does not contain any kind of side effects.

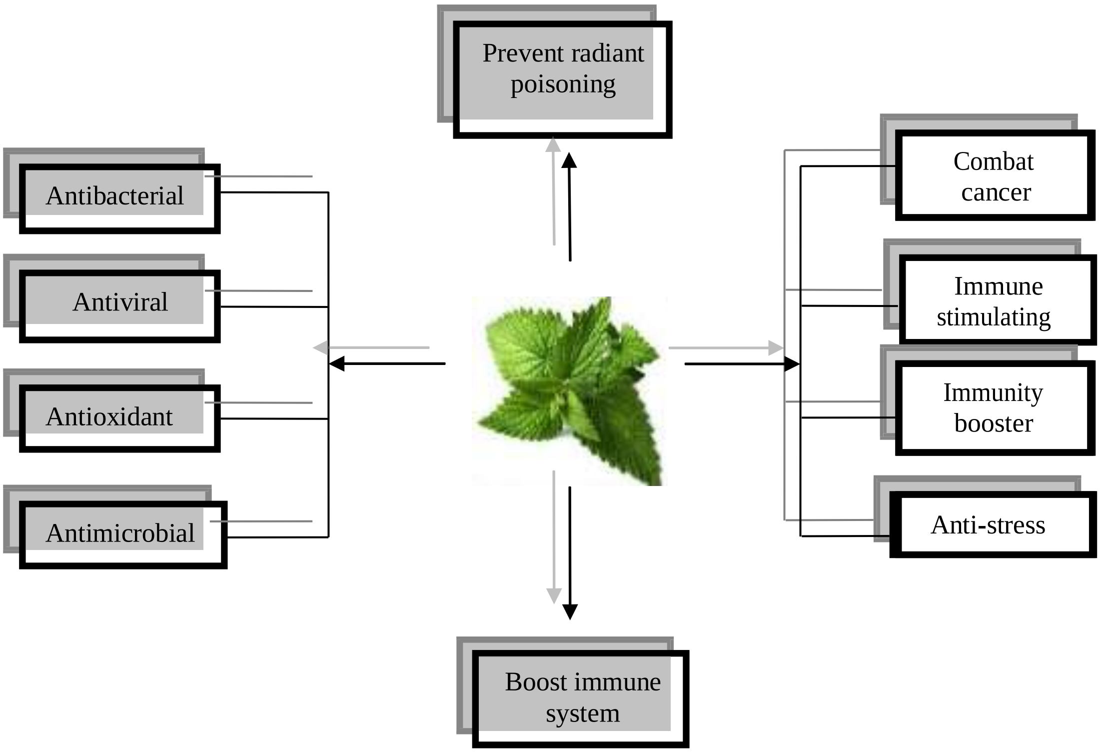


Figure 1.2: Basil leaves medicinal benefits

1.2 Problem statement

According to study of world wildlife and International Union for Conservation of Nature approximate 15,000 species of medicinal plants are dying out due to habitat destruction, diseases existence etc. [5-12]. Foundation of quality assessment is dependent on features of product such as appearance, cracks, texture and infection where human alertness could be easily fooled. Human may be tired and loose concentration, thus decreasing their accuracy. There can be substantial inter and intra-rater variability (subjectivity). Re-training required to maintain quality and human are prone to various illusions (for example, lesion number/size and area infected). It is assessed that 30 to 40% of basil plants are lost each year through the production chain. Losses from diseases in basil plants also have an important economic impact, causing a drop in income for producers, higher cost for consumers and distributors. So**,** production of diseased free good quality leave is of prime concern.

The traditional manual visual quality inspection of leaves for disease identification cannot be defined systematically since it is unpredictable and inconsistent [13]. Moreover, it involves a remarkable amount of expertise in the field of plant disease diagnostics (phytopathology) in addition to disproportionate processing times. So, in this digital era, it is mandatory that farmers use modernize techniques for efficient supervision of their plants. In last few years, internet has made information readily available for farmers to get updates on plants diseases. But they are not able to identify each disease properly. Thus, there is a need to evaluate the quality of leave that can be a potential danger to the life of plants.

Although researchers have worked intensively to identify the diseases using various techniques such as thermography, immunofluorescence techniques, fluorescence imaging, gas chromatography techniques, chain reactions and Deoxyribonucleic Acid (DNA) and Ribonucleic Acid (RNA) based affinity biosensor [14] etc. for quality evaluation of basil leaves. The problem of above-mentioned techniques are inefficiency, inconsistency and time consuming. Therefore, consideration for an automatic and precise along with inexpensive and efficient technique to identify and classify basil plant disease is of great realistic significance. By now, studies show that digital image processing methods work as effective tools for the identification and classification of plants diseases. Major progresses have been made to enhance reliability and accuracy of image analysis techniques for identifying and classifying plant diseases [15-16].

1.3 Disease classification

Quality of leaves characterizes the measure of excellence or a position of being free from deficits, substantial variations and defects. Plant leave diseases characterized based on their key causal agent (i.e. infectious and non-infectious). Infectious leave diseases are originated by a pathogenic organism such as a fungus, virus, nematode, bacterium, mycoplasma and viroid etc. Non-infectious plant diseases owe their origin to critical growing conditions, disadvantageous relationships between moisture and oxygen, excesses of temperature, toxic constituents in the soil or atmosphere, and deficiency of an essential mineral.

The intensive use of fungicides, depositions due by higher RH favour, unfavorable temperature, fungi, virus, high humidity, water and soil pollution etc. lead to severe epidemics in basil plants. Diseased basil leaves can adversely influence health as well as yield. The most common diseases of basil plants are downy mildew, cercospora leave spot, Fusarium wilt, aphid and gray mold. Table 1.1 briefly represents basil leaves diseases, their symptoms and descriptions. In the thesis, mainly focused on downy mildew, cercospora leave spot and fusarium wilt diseases.

Table 1.1: Basil leave diseases and their symptoms

Disease symptoms	Agriculture information	Plant pathogen group
Downy mildew	Fuzzy growth, brown color lesion	Fungus
Gray mold	Leave dying	Fungus
Bacterial leave spot	Circular to irregular dark spots with light center	Bacteria
Fusarium wilt	Dropping of leave	Fungus
Aphids	Small sized insects usually yellow or green in color residing on underneath of leave or stem	Insect

1.3.1 Downy mildew

In the past decagon, downy mildew became a severe disease in basil production worldwide. It occurs by fungal pathogen named as Peronospora belbahrii. It mainly affects leaves, stems and branches of basil. It primarily emerges underneath the leave as greyish/black little spores and yellowish/brownish upper surface of leave [17]. Due to the high humidity level, green area of leave twirl to dark brown color.

When spores are produced, a characteristic fuzzy dark gray to purple growth on the underside of leaves is evident, which looks like a bright layer of dirt [18]. Severe impression of diseases can lead to leave death. As, per study young plants are more vulnerable than old plants.

1.3.2 Gray mold

Gray mold is fungal disease that occurs due to excessive moisture or humid conditions. Diseases symptoms are fuzzy growth and brown to gray fungal expansion on both leaves and stems [19]. The appearance of disease is denser and fuzzy than the downy mildew. This disease spreads very rapidly and severe lesion on stem, can damage the entire healthy plant. Figure 1.3 and 1.4 represents diseased basil leave samples.

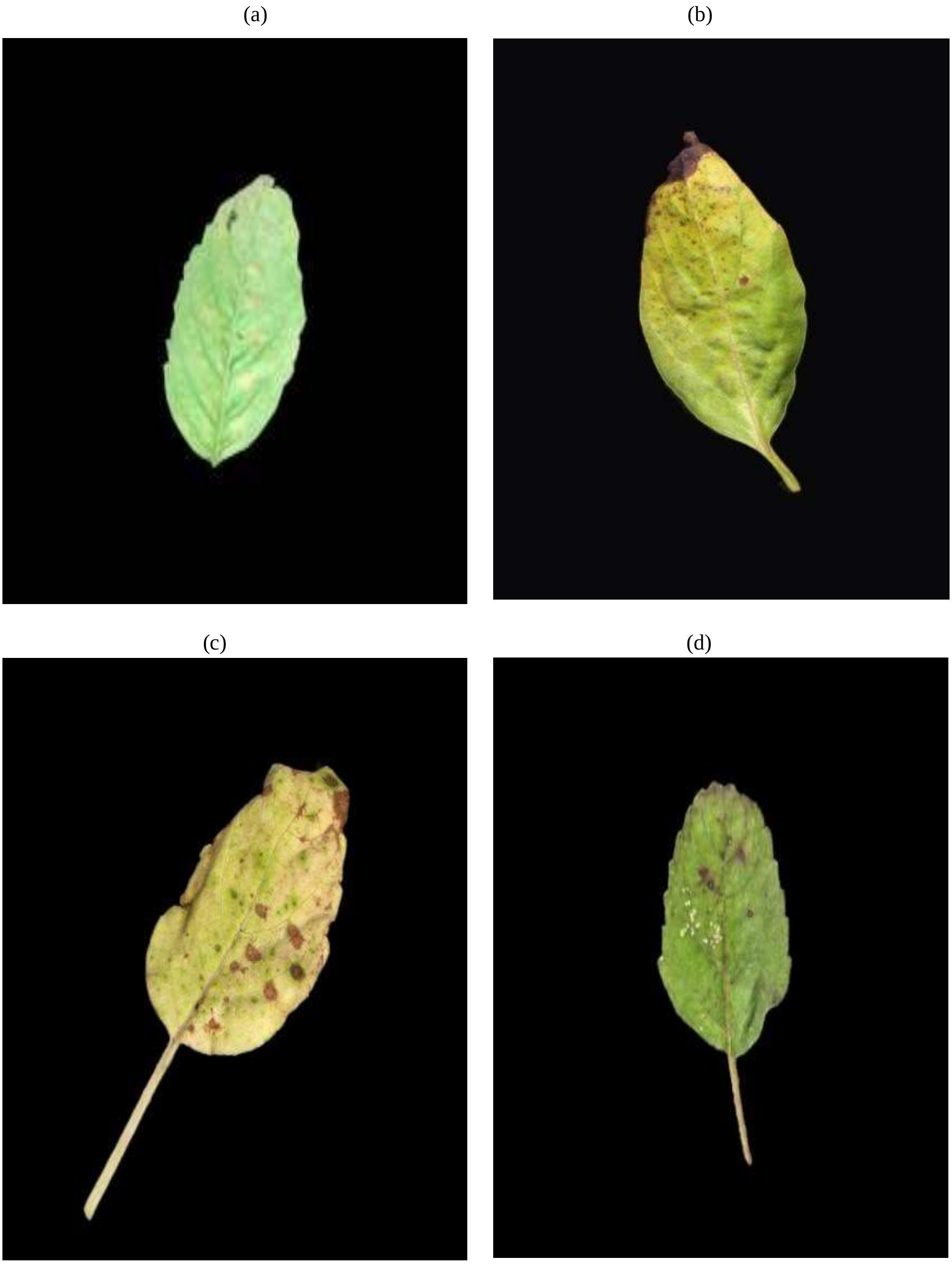


Figure 1.3: Diseased basil leave samples (a) and (c) Images depicts downy mildew, (b) and (d) Represents bacterial leave spot

(a)

(b)

(c)

(d)

Figure 1.4: Diseased basil leave samples (a) Image depict fusarium wilt disease (b) Depicts downy mildew, (c) Represents bacterial leave spot (d) Depicts aphid diseased leave

1.3.3 Bacterial leave spot

Lesion or black spots on leave are often indication, of disease named as shoot blight. It occurs by bacterial pathogen named as Pseudomonas cichorii. Theses lesions are circular to an irregular shape, dark brown; stem lesions, delineated by the small veins and hugrophanous. They appear and spread mainly due to overhead watering and high humidity levels [20-21]. Bacterial leave spot is mainly prospered in a moistened soil. Severe disease symptoms can cause leave to defoliate effortlessly, tip dieback and sometimes complete loss of plant.

1.3.4 Fusarium wilt

Fusarium wilt also known as Fusarium oxysporum basilicum occurs due to contaminated seeds, equipment, air and soil-borne fungus. The symptoms are wilted and stunted plant with asymmetric growth, reddish brown discoloration of tissues, brown steaks on stem tissue, black lines along the petioles or stem, severely twisted stem and rotted leave [22]. As disease symptoms grow, leaves fall off from plant and eventually sudden death of plant. Fusarium wilt is mainly prospered in a warm, moist condition.

1.3.5 Aphids

Aphids are small sized insects usually yellow or green in color residing on underneath of leave or stem [23]. It may curl, distort and change the color of leave.

1.4 Application of computer vision in plant science

Plant science is experiencing computerized revolution. Computer vision techniques have immense perspective in the plant safety field [24-25]. They characterize as a proficient set of tools for plant diseases recognition and classification. It concerned with automated processing of images from real world to take out characteristics from computers and understand information on a real time scenario [26]. These techniques are generating greater effectiveness and profitability through improved harvest and lower operational cost.

In the first stage images of basil leave are acquired using color camera and to avoid effect of noise and lighting conditions, images are pre-processed using filters. To avoid lighting conditions and variations, a shadow detection algorithm is used. CLAHE algorithm is also used to enhancing contrast of an image. In the subsequent stage, after segmentation, features are extracted based on detailed properties of image. Later on, to minimize redundancy best features are selected using feature selection technique. Redundant features did not define any relevant information in relation to other important features. They also increase the classification and training time. We used Random Forest for feature selection, which efficiently defines the rank of important features by measuring Ginni index and mean decrease value. Finally, images are classified using various machine learning algorithms. Accuracy of the machine learning models is predicted using various evaluation parameters. Figure 1.5 represents the basic structure of computer vision system.

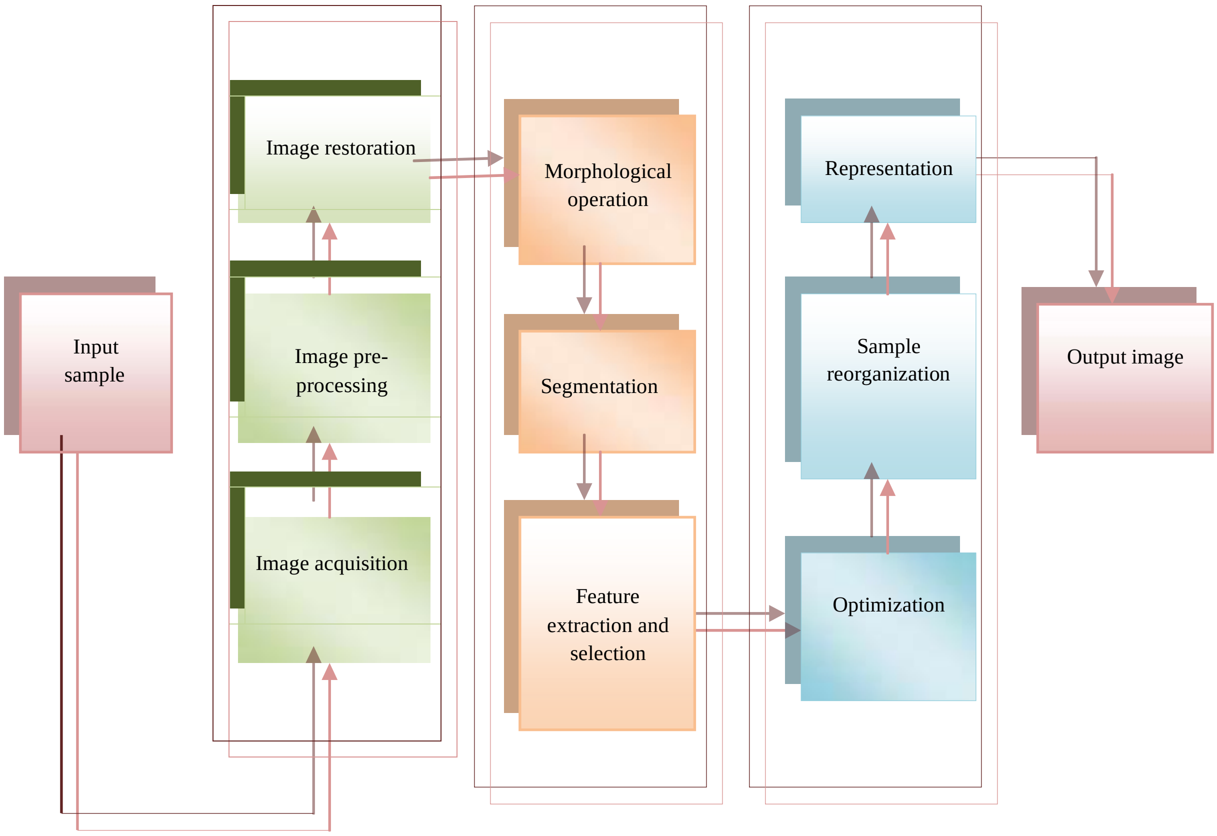


Figure 1.5: Computer vision system building steps

1.5 Machine learning applications

Machine learning, a subdivision of artificial intelligence, is to recognize organization of data and fit that data into different models that can be utilized and understood. Machine learning models have appeared simultaneously among big data techniques and recognize data intensive methods in various agricultural working environments. The main goal of the machine learning is to permit the computers to be trained without any human interference and regulate actions accordingly. It analyzes data faster and delivers accurate results [27-28]. Figure 1.6 shows the basic block diagram of machine learning, where supervised and unsupervised data is classified with respect to some tuning parameters. In this figure, structure of machine learning methodology is described.

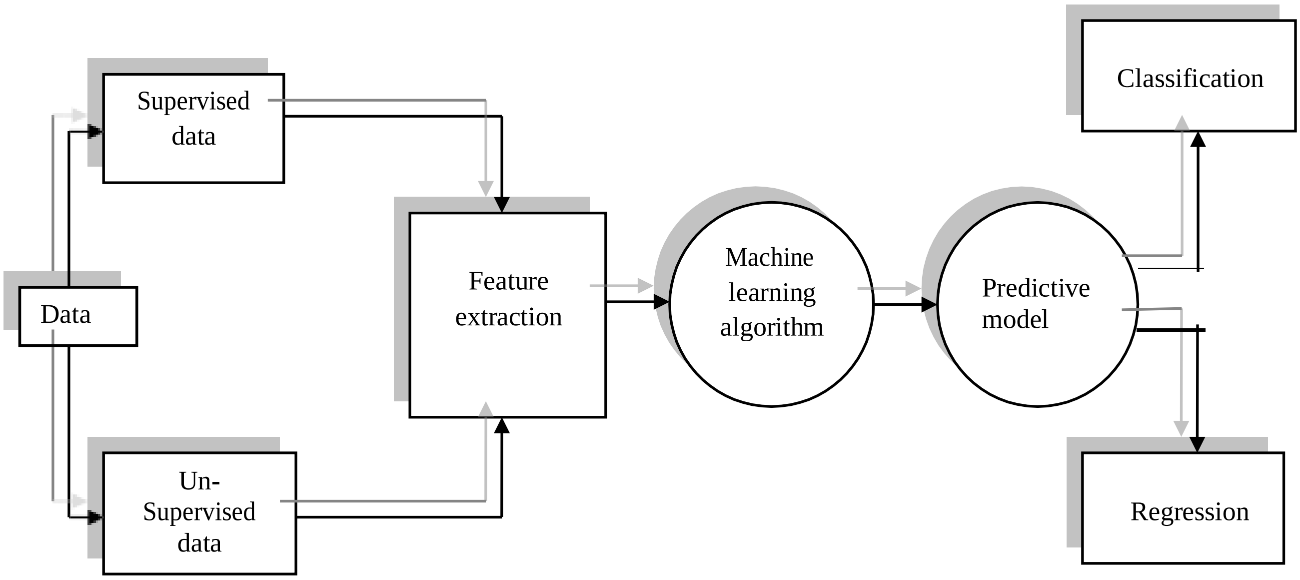


Figure 1.6: Machine learning methodology

Machine learning models are divided into two groups as supervised and un-supervised learning. The main motive is to characterize an image using distinctive characteristic that differentiate the interpretation from one category to other. We used machine learning models for diseases classification of basil leaves.

1.6 Gaps in Study

From literature survey, it is revealed that quality analysis for basil using computer vision is an undeveloped area. But based on existing techniques of diseases detection and classifications some key points to be considered are:

(i) Setup conditions

Presence of noise and shadow in captured images makes the quality analysis process very difficult. Another constraint is environment conditions for capturing images in controlled lighting conditions. Existing techniques exhibit high accuracy while examined under controlled laboratory lighting conditions. However, their use in outdoor condition is problematic, since the accuracy in such conditions drops dramatically. Robust calibration is required to reduce the effect of illumination, angle of camera and distance between object and camera. The other significant challenge arises due to inherent inconsistency of color under natural lighting conditions.

(ii) Complexity of images

It is challenging to separate desired leave region from the complex image background.

(iii) Reliability of data

The additional concern is related to the unavailability of complete and reliable information of data.

(iv) Lack of technical knowledge or selection of image processing schemes

The choice of image-processing techniques, segmentation techniques as well as classification strategies is crucial for the effective performance of any computer vision system. Assessment of segmentation and classification techniques cause problems like oversized sample set, under-fitting etc.

(v) Image quality

For better diseases detection and recognition, high quality images dataset is recommended. The lighting setup must be in proper position because it also affects, the quality of captured image.

(vi) Evaluation of new features extraction and classification measurement

Various existing feature extraction and classifications techniques are being used for disease detection and classification. But, due to complex leave structure they are not able to detect diseases having same color or structure. So, there is a scope to enhance performance of diseases detection methods considering new features and classification.

(vii) Processing speed

Processing speed of the diseases identification and classification techniques is an important issue as we need to process huge amount of data.

1.7 Research Objectives

The main aim of present thesis is to develop calibration model based on computer vision to estimate basil leave quality. This outmost goal achieved by following specific objectives.

(i) To develop an algorithm for efficient pre-processing and segmentation of herbs (basil).

(ii) To develop a model for the analysis of herb (basil) parameters, such as color and texture.

(iii) To develop an automated model to classify basil based on quality.

(iv) Test and validate the experimental accuracy of the proposed methods.

1.8 Contribution

This research integrates existing image segmentation, feature extraction, selection and classification techniques to motivate the development of new segmentation and classification methods to identify disease. The subsequent research contributions are investigated based on mentioned objectives. The first objective plans to design a segmentation technique using neutrosophic logic (fuzzy extended form). The primary function of proposed segmentation technique is accurately identifying region of disease. The first objective is achieved as follows:

(i) We first created basil leave dataset. Leave dataset consists of four types of healthy and diseased basil leave images; Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (holy basil) and Ocimum gratissimum (Vana-holy basil). The data is collected from various locations of Punjab during summer and winter seasons but they were captured under field-controlled conditions. These samples are collected from herb gardens at Punjab Agriculture University Ludhiana (30.9°N 75.85°E), National Institute of Pharmaceutical Education and Research (NIPER), Mohali (30.78°N 76.69°E), Chandigarh Botanical garden and Punjabi University Patiala (30.32°N 76.40°E) for reflective study. We took 400 image samples.

(ii) To improve contrast, Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm is used.

(iii) A novel neutrosophic logic is developed to segment a leave image and extract region of diseases. Based on segmentation, whole image is converted into three neutrosophic segments as true, false and intermediate. As compare to existing techniques, proposed segmentation technique introduced new segment as, intermediate segment which is neither defined as healthy nor diseased.

(iv) Experimental result shows effectiveness of proposed segmentation technique.

The second objective aim is to extract different features using color and texture for further classification of leaves. Extracted features are used to train classifier to categorize leave as healthy or diseased. Features are mainly responsible for accurate image classification or pattern recognition applications. The second objective is achieved as follows:

(i) Complex background, shadow, illumination factors and presence of other occlusion must be considered, while selecting efficient features. Usually color and texture are of most important concern in order to detect and identify diseases. So, in proposed diseases detection framework, features are extracted using texture and color analysis.

(ii) Using histogram information content, damage index and disease sequence region features are extracted. A new "bin binary pattern" texture feature is introduced. In this work, combinations of color and texture features are used. Moreover, correlation between diverse features may design an efficient feature vector and produced efficient results.

(iii) These extracted features are fed to nine different classifiers as Random Forest (RF), Naive Bayes, Artificial Neural Networks (ANN), Decision tree (DT), Linear Discriminant Analysis (LDA), K Nearest Neighbors (KNN), Support Vector Machine (SVM), AdaBoost and Linear model. These machine learning methods are trained with different tuning parameters and finally classify leave as healthy or diseased. Finally, Random forest outperforms and achieves highest accuracy as 98.4% in distinguishing diseases without requiring any human intervention.

The third objective aims to introduce new optimization model for diseases classification. The third objective is achieved as follows:

(i) A new classification model using survival of fittest approach is introduced, where leave diseases are further classified as downy mildew and bacterial leave spot.

(ii) In the first stage, after enhancement and segmentation, using gray level co-occurrence matrix features are extracted.

(iii) Irrelevant features make difficult classifier task, computationally complex and also reduces classifier accuracy. To tackle with this problem, feature selection technique RF is used. RF selects best features, and redundant features are ignored.

(iv) Finally using different distance measures survival of fittest approach is introduced. Where, best solution is obtained through the fitness function with minimum distance and maximum similarity values by calculating maximum aggregation.

(v) Moreover, classification algorithm has the ability to discriminate leave diseases with almost same color values with accuracy 95.73%.

The four objective aims to test and validate the experimental accuracy of the proposed methods. The fourth objective is achieved as follows:

(i) We test and validate our proposed methods with existing segmentation, feature extraction and classification techniques.

(ii) Moreover, the proposed technique definitely reduces the computational complexity and achieves better results than existing classification and segmentation techniques. For better perceptive, Figure 1.7 illustrates the thesis work structure.

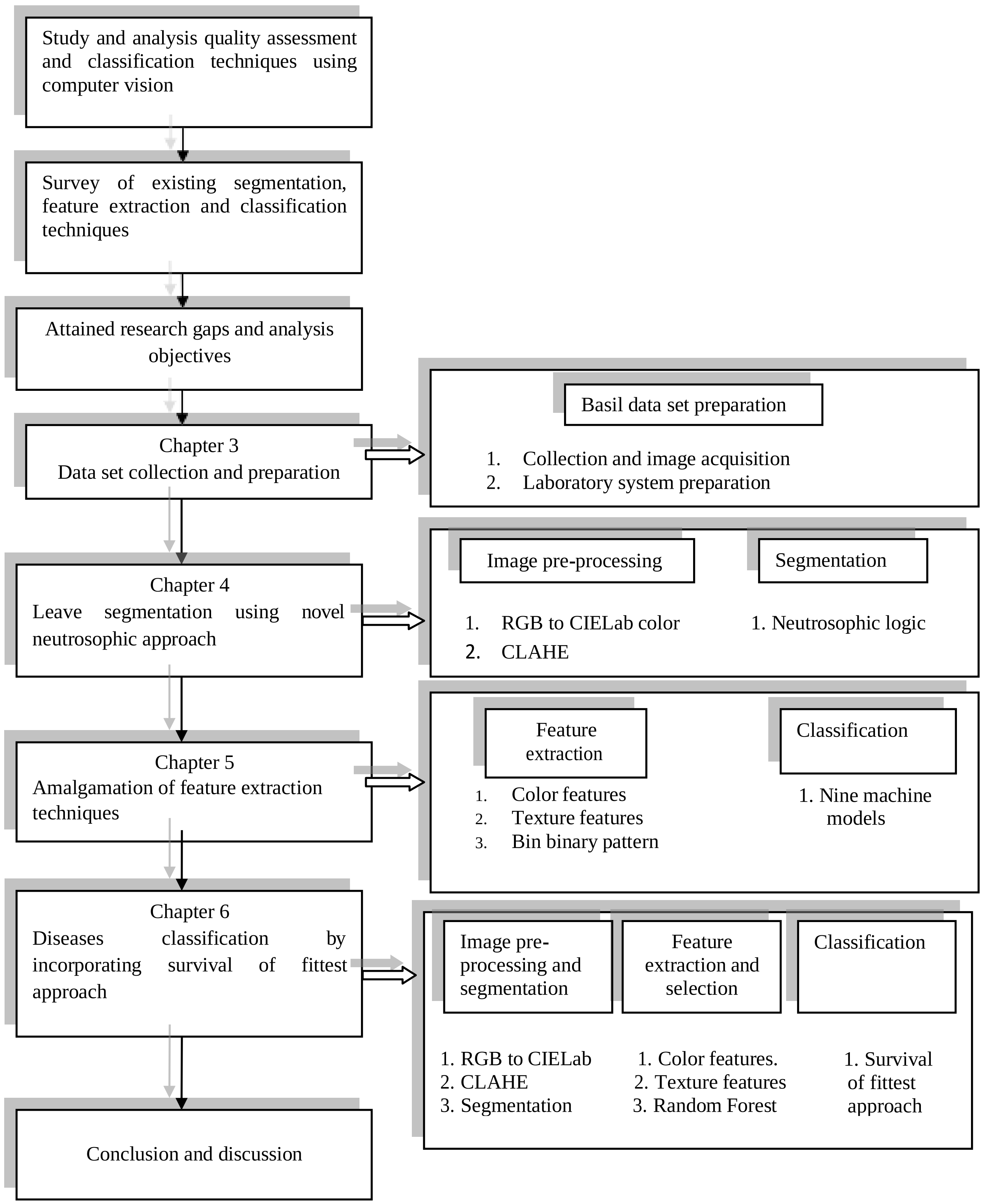


Figure 1.7: Flow diagram of thesis structure

1.9. Organization of Thesis

The thesis is structured into six chapters including conclusion and a concise summary of thesis is as follows:

Chapter 1: This chapter starts with introduction of basil plant, its diseases and causes. It covered applications of image processing technique, computer vision and machine learning techniques. It also discusses advantages of above-mentioned techniques in precision agricultural sciences.

Chapter 2: Chapter describes detailed study of existing segmentation, feature extraction, feature selection and classification techniques. It also describes research gaps, research objectives, methodology, and motivation of the work.

Chapter 3: The chapter presents an image acquisition system, description of various factors while preparing data set. It discusses details of equipment, lightening conditions, location selection and time selection for image capturing.

Chapter 4: The chapter represents overview of existing segmentation techniques and designed a new segmentation technique. The uniqueness of presented approach is to define new region known as intermediate region with new segmentation technique using neutrosophic logic. It includes various texture and color properties to classify image as healthy and diseased using new set of features.

Chapter 5: Represents various color and texture features extraction methods. It defined new amalgamation technique of color and texture features. New texture feature as bin binary pattern is also described. In the experimental section, proposed technique is compared with existing techniques and provides satisfactory results.

Chapter 6: Various existing classification techniques have been discussed and a new classification algorithm using survival of fittest is proposed to classify basil leave diseases. After segmentation and feature extraction, using Random Forest most important features is selected and trained using new optimization algorithm for further classification.

Chapter 7: Concludes the presented thesis, limitations and future scope.

Chapter 2

Literature review

In this chapter, we address a comprehensive study on disease recognition and classification of leaves using existing image processing methods. The work is divided into two sections. First sub- section emphasizes on contribution of researchers with respect to technical explanation and a summarizing table which contains information regarding scientific culture. In subsequent sub-section, a concluding section is presented. Section 2.1 describes existing diseases detection and classification techniques. Section 2.2 explains different evaluation metrics used for measure the effectiveness of proposed system model.

2.1 Traditional approaches

Recent developments in diseases identification and classification arrangement are considerably explored in this chapter. A review of discussion and opinion of experts is also discussed. Various conventional techniques of disease detection and classification are categorized by utilizing image processing approaches and machine learning models for classification etc. as shown in Figure 2.1. This figure complete gives overview of complete literature.

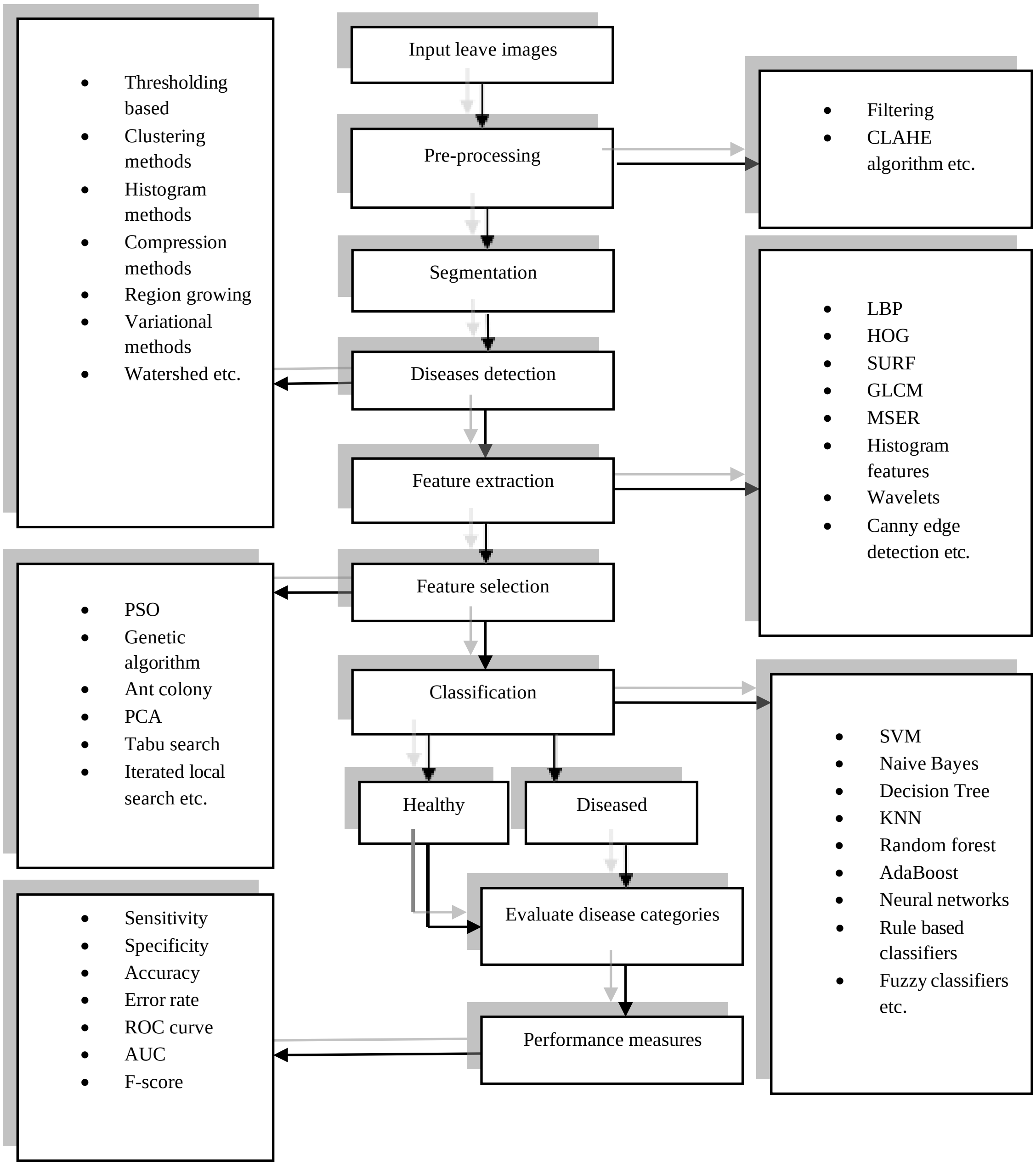


Figure 2.1: General Structure for diseases identification and classification of plant leaves

2.1.1 Color and texture characteristics analysis

Pydipati et al. [29] described Color Co-occurrence Method (CCM) to determine texture based on Hue, Saturation and Intensity (HSI) in combination with statistical classification algorithms. In first stage, features are extracted using CCM by considering four different categories of leaves. Further, Statistical Analysis System (SAS) is used to evaluate potential classification accuracy with reducing variable sets and color features individually. Four models are used for classification in referred paper, model 1 consisted of a reduced set of hue and saturation texture, model 2 consisted of a reduced set of intensity features, model 3 consisted of a reduced set of hue, saturation, and intensity features, and the fourth model consisted of all combined 33 HSI texture features. However, best results are attained using combined HSI features. The authors explicated that hue and saturation features are influenced by low lighting conditions. These methods prove to be best under controlled laboratory lightning conditions. The most significant challenge is to use this method under natural lightning conditions.

A. Camargo et al. [30] presented image processing method to identify plant diseases visual symptoms from exploration of color images. They transformed RGB color space into H, I3a and I3b space. The I3a and I3b conversions are generally developed from an adjustment of original I1 I2 I3 color transformations to meet requirements. By analyzing scattering intensities in a histogram, transformed images are segmented. After post processing, accuracy of manually segmented images are compared with automatically segmented images. The authors concluded that this technique is suitable to identify a diseased segment, even when segment is represented by an extensive range of various intensities.

Haiguag Wang et al. [31] described a method for identification of diseases in grape and wheat leaves. Captured images are transformed from RGB color space to XYZ color space and then again transformed from XYZ color space to L*a*b color spaces. Squared Euclidean distance (D) as a similarity distance is used to evaluate color difference in a*b, two dimensional data space. Total 4 shape features, 21 color features and 25 texture features are extracted to distinguish various diseases. Furthermore, for reducing dimensions, Principal Component Analysis (PCA) is performed. Later on, neural network techniques with different parameters are used to classify and identify wheat and grape leave diseases respectively. The author confirmed that these techniques provide a fast and precise recognition of plant diseases.

Qinghai He et al. [32] proposed diseases identification method for cotton leave using three color models. After image acquisition, spatial non-linear and frequency domain filters are used to eliminate influence of noise. Histogram equalization is used to enhance contrast of captured images. The enhanced images are converted into RGB, HIS, and YCbCr colormodel. The percentage of damage (γ) is selected as a feature to evaluate amount of damage due to leave diseases. After applying different color models, comparative results are obtained. The YCbCr color model dominates other models. The author validated that projected algorithm fails to handle random noise interference and shadow in case of outdoor

Auzi Asfarian et al. [33] presented work on paddy diseases identification using fractal descriptors based on Fourier transform with texture analysis. Injuries (lesions) are measured manually and each of these lesion images is transformed to HSV color space. After selecting only saturation components, histogram equalization is performed for reducing lightning effects. For each lesion, fractal descriptors are extracted and given to Probabilistic Neural Networks (PNN) classifier. For verification and validation of data, 5-fold cross validation is used. The author concluded, it is difficult to recognize disease precisely, when two diseases comparatively have same color.

Pradnya Ravindra Narvekar et al. [34] discussed an effective way of disease detection of grape leaves using SGDM (Spatial Gray Level Dependence Matrices) method. They used four classes of diseased leaves i.e. black rot, downy mildew, powdery mildew and normal. After color transformation of RGB to HSI, H component is considered for further analysis. The S and I components are not considered since they do not provide any useful information. Based on specified and varying threshold, green pixels are masked using Otsu thresholding method. Later on, features based on color and texture are extracted and classified. The authors concluded that proposed work defines valuable approach with little computational efforts.

Vinita Tajane et al. [35] suggested Content Base Image Retrieval (CBIR) method for identifying medical plants leave diseases using canny edge detection algorithm and histogram analysis. In first stage, CBIR finds image gradients to highlight regions with high spatial derivatives. Canny edge detection method is used to evaluate strong edges of leave and draw histogram. Color features are evaluated using histogram of each red, green and blue color plane individually to measure intensity of each color pixels. The author concluded that canny edge detection algorithm is an effective method for distinguishing edges of healthy and infected plants and preserve significant structural properties in an image.

Juan F. Molina et al. [36] presented a color-based strategy to detect an early blight disease or any kind of infection on tomato leave lets. Color characterization is evaluated using Hue-Max-Min method, scalable color descriptors and color layout descriptors with YCbCr color space. After calculation of all these descriptor values, a novel strategy based on nested leave one out cross validation method is used to achieve better classification ratio. The individual descriptor configuration performed using an inner loop permit. Outer loop measures, performance assessment between different descriptors. The author concluded that color structure descriptor provided better accuracy than other methods. Table 2.1 represents summarization of color and texture technique for diseases detection.

Table 2.1: Summarization of color and texture technique proposals for diseases identification

Proposal	Culture	Diseases	No. of images	Accuracy	Feature	Image format	Environ-ment condition	Acquisition device/ dataset
Pydipati et al.[2006]	Citrus leave	Melanose, greasy spot , scab	Total 40 images	96%	Intensity Texture	Un-compressed Jpeg format	Controlled lighting condition	3 CCD camera (JAI, MV90) captured from central Florida
Camargo et al.[2008]	Banana, corn, soya, alfalfa	Fungal	Total images not defined 20 images [testing]	N/A	N/A	Jpeg format		Images collected from Department of Entomology, at the University of IOWA, USA University of Georgia, supplied images of cotton and soya. The International Network for Improvement of Banana and Plantain field
Wang et al. [2012]	Grapes, wheat leave	Rust, downy, powdery mildew	Total 185 images 110[training] 75 [testing]	94.29%	Color, shape & texture feature	Jpeg format	Controlled lighting conditions	Common digital camera
He et al.[2013]	Cotton	Fungal & bacterial	N/A	N/A	Ratio of damage	N/A	Controlled lighting conditions	Common digital camera

Proposal	Culture	Diseases	No. of images	Accuracy	Feature	Image format	Environ-ment condition	Acquisition device/ dataset
Asfarian et al. [2013]	Paddy crop leave	Leave blast, brown spot, bacterial leave blight & tungro	Total 40 images 5 fold cross valida-tion	83%	Fourier descriptor	Jpeg format	N/A	Images are taken in paddy field in Laladon and Cipanas, West Java, Indonesia
Narvekar et al. [2014]	Grapes leave	Black rot, powdery, downy mildew	100 images	N/A	Color & Texture	Jpeg format	Controlled lighting conditions	Common digital camera
Tajane et al. [2014]	Medicinal plants leave	Fungal, virus & bacterial	N/A	N/A	Color Histogram	Jpeg format	Controlled lighting conditions	Common digital camera
Molina et al. [2014]	Tomato Leave	Early Blight	147 images	100% [CSD64]	Colors descriptor	Jpeg format	Controlled lighting conditions	Captured at tomato green-house crop Boyacá Region, Colombia

2.1.2 Thresholding and clustering parameters analysis

Tucker and Chakraborty [37] focused on identifying and detecting diseases of oat and sunflower leave. Images are segmented with varying threshold w.r.t the disease being measured (blight or rust). Using clustering, subsequent pixels are clustered into signifying diseased regions. Based on features of lesions, they are segregated into proper categories. The authors observed good results but due to inappropriate illumination during the capturing of the images, some errors exist.

D.G. Sena Jr. et al. [38] proposed work to develop a procedure for identifying damaged maize plant leave diseases. Damaged and healthy maize plant leave images are captured in three different light intensities and in eight different locations. The proposed algorithm is divided into two stages: pre-processing and image analysis. In, first stage original RGB images are transformed into 256 grey level images and rescale pixel values to create binary images. Then, monochrome images are threshold by iterative method. In second stage, images are sectioned into blocks. Finally, images are classified as healthy or damaged based on number of objects originated in each block. The author concluded proposed algorithm performs well and gives good accuracy.

Shen Weizheng et al. [39] have implemented an image processing-based technique for analyzing leave spot disease. They investigated all influencing issues that existed in process of segmentation. Leave regions are segmented using Otsu's thresholding method. After color transformation from RGB to HSI color system, H element is selected for segmentation to reduce disruption of lightning changes. Further, Sobel operator is considered in order to observe edges of leave disease spots. Finally, grading is achieved by assessing quotient of diseased region and leave areas. The author indicates that this method is fast and accurate.

Sekulska-Nalewajko et al. [40] presented a method to detect disease signs in cucumber and pumpkin leaves. First, images are converted from RGB to HSV color space format. After thresholding leave disease area, brightness section is discarded. Further, Fuzzy C Means (FCM) clustering algorithm is applied in hue-saturation space to group pixels into clusters. Ultimately, authors justified that present approach gives good classification results. However, this approach requires too many processes to achieve the desired results.

Zulkifli Bin Husin et al. [41] discussed effective method used for evaluating primary detection of chilli disease using leave features inspection. Now days, from time to time chemicals are applied to plants without seeing the prerequisite of each plant life cycle. This technique confirms that chemicals are applied only when plants are noticed to be instigated with the diseases. Leave images are captured under controlled field conditions and pre-processing is done to achieve improved image information. Features are extracted using color information for differentiating healthy and non-healthy leaves. In the end, histogram graphs are evaluated to measure the healthiness of leaves. The author concluded that proposed method is fast and efficient in recognition of plant chilli disease.

Mokhled S. Al-Tarawneh [42] suggested a method to detect diseases of olive leave. The proposed method combines auto cropping technique with fuzzy c means clustering to identify disease. After, color transformation of RGB to CIELab color space, region of interest (diseased region) is cropped using automatic polygon. The polygon cropping follows the edging contour of entire image to define the masked polygon points of diseased segments. At the end, classification is done using fuzzy c-means clustering for statistical usage to define the defects and severity. The severity percentage is calculated based on the classification of detected diseased and total leave area. The author actualized that experimental result of FCM algorithm with polygon auto cropping segmentation shows encouraging accuracy prospects.

D. Zhihua et al. [43] discussed cotton mite diseases recognition approach using color features and thresholding methods. Proposed technique is distributed into three steps: 1) Diseases spots and stems extraction of the green plants. 2) Special features detection in gray histogram and transform the segmented image into 8-bit gray scale image using thresholding. The black pixels signified the corresponding disease spotted part taking mite disease spots and stems. White pixels representing non-disease spotted areas. 3) Segmentation of binary images using area thresholding and comparing the areas with spots to the stems. The author concluded that mite and stem disease spotted portions are almost similar in color therefore it is a difficult task to compare the areas with spots on the stems.

Krishnan et al. [44] defined approach to analyze shade leaves disease (bacterial scotch). Firstly, shade tree leaves images are captured under controlled lightning conditions. K-means clustering approach segregate image into distinct clusters and attain cluster center to evaluate region of interest. Finally, subtract reference image from base image to evaluate only diseased region. The authors concluded that proposed algorithm gives high accuracy with less operational time. But, classification of algorithm can be improved by updating clusters through repetition procedure. The clustering algorithms built on limitations and certain relaxation algorithms could also be improved to provide accuracy to proposed scheme.

Eric Hitiman et al. [45] explained technique for leave injury detection and severity assessment of coffee leave. Gaussian kernel is used to suppress effect of noise and RGB color space is transformed into CIELAB color space. Look Up Table (LUT) based gamma correction is used to increase contrast of image. Later, boundary features of threshold image are detected using Canny edge detector and joined to modify overall structure of object. Using luminance and color, only leave area (foreground) is extracted and processed with YUV color space (i.e. V channel) to maximize leave injury detection. Finally, leave damage severity is assessed with respect to percentage of pixel distribution of healthy leave. The author consummated that proposed method is fast and avoids defoliation. Moreover, proposed method can handle all images and provides good accuracy rate.

Jiangsheng Gui et al. [46] proposed a novel method for identification of diseases of soybean leave by means of salient regions with complex background. Based on idea of Itti method, low level features of luminance and color, salient regions are detected. In last, segmentation using threshold is performed on R component of RGB space. Morphology algorithm is used to fill small holes and eliminate insignificant small sized regions to precisely calculate diseased segment. The author finally conferred that proposed technique can precisely detect segment of diseases in complicated farmland backgrounds.

K. W. V Sanjaya et al. [47] proposed system to measure and predict orchid leave healthiness and diseases. Gaussian operator and histogram equalization method is used for removing noise and enhance contrast of image. First transform the RGB color space into HSV color model and applied threshold values for hue, saturation and value bands separately. After segmentation, shape and color features are extracted to detect orchid leave diseases. The extracted geometric features are used to predict orchid species and color features are used to predict healthiness of orchid plant. The author concluded that proposed system accuracy is reliable but system can identify only three orchid species and three orchid diseases.

Sourabh Shrivastva et al. [48] presented a technique to automatically identify and estimate level of disease severity in soybean plant foliar. The concerned method converts RGB image into Y, Cb, Cr channels and then segments through simple thresholding method. The authors used various novel parameters such as disease level parameter, severity index, diseases severity index and infected area to automatically measure disease level severity. The author highlighted that it is a low-cost method and has potential for the extensive usability in field conditions. Moreover, performance of this method might be improved by using unconventional background separation approach to distinct leave object from a complex background.

Amar Kumar Dey et al. [49] considered image processing procedure to identify leave rot disease by classifying color feature of rotted leave region of betel vine. After analyzing different color spaces, HVS gives promising results where hue component provides strong remark of rotted leave area. Afterwards, using Otsu thresholding method rotted area is segmented using various threshold values and area of rotted portion is considered as feature set. A leave disease severity scale can be prepared by calculating proportion of diseased area. Using frequency of precise amounts of pesticide application, disease severity volume can be regulated, which moderates pesticide cost used for management. The author concluded that technique presented promising results of this automatic vision-based system with easy validation. Table 2.2 represents the summarization of thresholding and clustering technique for diseases classification and detection.

Table 2.2: Summarization of thresholding and clustering proposals for diseases identification

Proposal	Culture	Diseases	No. of images	Accuracy	Feature	Image format	Environment conditions	Acquisition device/ dataset
Tucker and Chakraborty [1997]	Sunflower & oat leave	Rust & blight	Upto 20 for each disease	N/A	Area, size & diameter	N/A	Controlled lighting condition	TMC-76 color CCD
SenaJr et al. [2003]	Maize leave	Fall armyworm	720 images	94·72%	N/A	N/A	Controlled different lighting conditions	Digital camera, MS3100/ captured at greenhouse at Embrapa Corn & Sorghum, SeteLagoas, Brazil
Weizheng et al. [2008]	Soybean	Leave spot	N/A	N/A	Diseased spot & area	N/A	Controlled lighting conditions	Common digital camera
Nalewajko et al. [2011]	Pumpkin & cucumber	Bacterial	N/A	N/A	N/A	Jpeg format	Controlled lighting conditions	Scanner
Husin et al. [2012]	Chili leave	Fungal & bacterial	107 images	N/A	Color	Jpeg format	Controlled lighting conditions	Common digital camera
Al-Tarawneh et al. [2013]	Olive leave	Leave spots	100 images	86%[FCM] 66% [KMC]	Color segment	Jpeg format	Uncontrolled lighting conditions	N/A
Zhihua et al. [2013]	Cotton leave	Mite diseases	30 images	94.79%	Diseases spot area	Jpeg format	Controlled lighting conditions	Nikon D90 SLR digital camera
Krishnan et al. [2013]	Shade tree leave	Bacterial	N/A	N/A	Area	Jpeg PNG	N/A	N/A
Hitiman et al. [2014]	Coffee leave	Bacterial	Training data not defined 27 images testing	N/A	Area	N/A	Controlled lighting conditions	Common digital camera

Proposal	Culture	Diseases	No. of images	Accuracy	Feature	Image format	Environment conditions	Acquisition device/ dataset
Jiangsheng Gui et al. [2015]	Soybean leave	Fungal & bacterial	N/A	N/A	Salient region	N/A	Controlled lighting conditions	Hefei academy
Sanjaya et al. [2015]	Orchid leave	Phyllostict , yellowing & black rot	250 image set	80.7365% [J48] 79.0368% [Naives Bayes]	Color & geometric	N/A	Uncontrolled lighting conditions	Smart phone camera
Shrivastva et al. [2015]	Soybean leave	Frog eye, rust bacterial blight, downy mildew & sudden death syndrome	1000 images	N/A	Color & area	Jpeg	Images captured in field	Samsung GT-S3770 phone/ captured at Guna, Madhya Pradesh, India.
Dey et al. [2016]	Betel vine	Rotten	12 images	N/A	Color	Jpeg	Images captured in field	CanoScanLiDE 110 [Scanner] / captured at Dhara, Rajnandgaon district of Chhattisgarh

2.1.3 Membership function analysis

Anthonys et al. [50] presented framework to recognize paddy diseases using membership functions. Color images captured by digital camera in laboratory with dark background are used to avoid effect of environmental factors. The images are then transformed into CIE XYZ color space. Using appropriate threshold values, images are segmented. Texture and shape features are extracted by considering parameters like area, roundness, shape complexity etc. Using these features, membership functions are calculated and defined for each class of disease using nearest neighbor. The author concluded that proposed work gives satisfactory results with less recognition time but noise affects performance. Table 2.3 represents summarization of membership technique used for disease classification.

Table 2.3: Summarization of membership technique used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition/ Standard dataset
Anthonys et al. [2009]	Rice	Rice blast, rice sheath blight and brown spot	50 samples	70%	Color, texture and shape	Jpeg format	Controlled lighting conditions	Digital color camera

2.1.4 Minimum path evaluation theory

Pixia et al. [51] suggested a novel scheme based on minimum distance to identify cucumber leave diseases. Leave images are pre-processed using median filtering and image smoothing. Images are segmented using color range of different diseases to attain lesion segment. Texture and color features are extracted and normalized. Finally, calculate shortest distance path and average center distance among diseases. As the recognized diseases are more similar with standard diseases, the distance is assumed to be less. The authors concluded that the characteristics based on the minimum path evaluation are very effective. Table 2.4 represents summarization of minimum path evaluation theory used for disease classification.

Table 2.4: Summarization of minimum path evaluation theory used for disease classification

Proposal	Culture	Diseases	No. of Images	Accuracy	Features	Image format	Environment conditions	Acquisition device / Standard dataset
Pixia et al. [2013]	Cucumber leave	Downy mildew, powdery mildew & anthracnose	N/A	More than 96%	Color, shape &texture	N/A	Controlled lighting conditions	Digital color camera

2.1.5 Artificial Neural Networks (ANN)

Kuo-YI Huang et al. [52] explained an approach using artificial neural networks for recognizing Phalaenopsis seedling diseases. Lesion areas are segmented using underlining principle to differentiate background and object (leaves). An exponential transform is used to enhance the image. Grey Level Co-occurrence Matrix (GLCM) is used to evaluate texture features. After that, tri-color mean values of lesion areas are employed to classify diseases using ANN. The authors are able to distinguish and categorize visible lesion areas but they are incapable to examine infected area on the enclosed blades. The recognition and classification system can be also useful for the development of flowers and plant leaves in greenhouse.

Noor Ezan Abdullah et al. [53] classified rubber tree leaves diseases using multilayer perception neural networks. Color features are extracted using RGB distribution indices to calculate region of interest (lesion spots). PCA is employed on progression values of each image. The improved artificial neural networks designed based on optimized dominant pixel (mean) of RGB and normalized PCA data. Finally, optimized models are evaluated and validated through analysis of various performance indexes. According to authors, proposed method is found to be the best model for identify bird eye spot and otrichum diseases. However, effectiveness of this classification can be increased providing a better process and high-resolution camera.

Santanu Phadikar et al. [54] defined framework to identify diseases of rice plant leaves using zooming algorithm and neural networks. The input RGB images are converted into HIS color space and infected regions are extracted using entropy-based thresholding method. Zooming algorithm is used to extract features and classified using Self Organizing Map (SOM) neural network. The authors explained that it is a simple and computationally efficient technique, provides satisfactory classification results. However, it has been observed that transformation of image in frequency domain does not produce better classification results.

Dheeb Al Bashish et al. [55] suggested a structure for classification and identification of diseases such as early scorch, ashen mold, cottony mold, late scorch and tiny whiteness etc. Subsequently, after color transformations from RGB to HIS, images are clustered with K-means clustering approach using Squared Euclidean distance. This method segregates leave image into four clusters and evaluates the color and texture features. SGDM matrices are created for each pixel map of H and S component only. Statistical examination tasks are accomplished to reducing feature redundancy. Then for final classification, neural network is used. It is found that proposed method can expressively support precise and automatic recognition of diseases in leaves. However, due to invariability in lighting conditions some samples of diseased leaves are misclassified.

H. Al. Hiary et al. [56] shed light on an automatic disease detection method. In proposed method, pixels are categorized into K number of clusters on the basis of green pixel values obtained using the Otsu's thresholding technique. Features are extracted using color co-occurrence matrix to analyze affected leaves texture. The leave spots are considered as an indicator of crop diseases. In the end, classification of diseases is evaluated using artificial neural networks. The author concluded that this method requires small computation of time and computational effectiveness.

Wanrat Abdullakasim et al. [57] elucidated an image analysis method for recognition of brown leave spot disease. Various color descriptors i.e. RGB and HIS descriptors are used to recognize different regions of leave based on its color. The range of color indices of RGB and HSI varies from 0 to 1. Then, ANN is used to differentiate between an infected region and a healthy region. Brier score is employed to estimate the recognition capability of the ANN with more hidden layers. It is found that algorithm properly identifies 79.23% diseased leaves and 89.92% healthy leaves. This method can be improved by combining lightning conditions, effects of infection phases and proper segmentation.

Kai et al. [58] suggested a technique to recognize diseases in maize leaves based on neural networks. After color transformation of RGB to YCbCr, evaluate Cb and Cr components because they are less affected by brightness. In this two-dimensional plane, lesion area is relatively concentrated in accordance with Gaussian distribution. Texture features are evaluated using SGLCM matrix. Furthermore, back propagation based neural networks with sigmoid function using 3 hidden layers are used to classify various diseases. The exploratory results proclaimed that accuracy is very high. The author concluded that neural network error value is very low, weight design is reasonably correct and overall performance is good.

D.S. Guru et al. [59] discussed innovative idea for extracting lesion areas from tobacco leave using neural networks. Contrast stretching transformation with an adjustable parameter and morphological operations are used to segment lesion areas. R channel based statistical texture features are extracted from lesion area to identify and analyze diseased category.

Based on extracted texture features, Probabilistic Neural Network (PNN) is used to categorize tobacco seedling leave diseases as anthracnose and frog-eye spots. The author concluded that first order statistical texture features perform better than gray level co-occurrence matrix features. Classification accuracy can be improved with combination of different color and texture features.

Zhang et al. [60] proposed new method to identify six varieties of jujube leave diseases i.e. jujube fruit rust disease, jujube rust, jujube white rot, jujube anthracnose, ascochyta spot of jujube, jujube witch broom. After image acquisition, nine color, eleven texture and five morphological features are extracted corresponding to red, green and blue plane of image. Best twelve features using Step Wise Discriminant Analysis (STEPDISC) and four principal components using PCA are selected from original 24 features space. Furthermore, cross validation is performed for each feature set space. In last stage, a two-layer tan sigmoid model using 12 parameters diagnose jujube diseases. The authors concluded that detection precision of jujube white rot disease is highest, due to big disease spots. The high similarity structure index between jujube diseases decreases the classification accuracy.

Pranjali Vinayak Keskar et al. [61] suggested leave disease analysis and recognition method for assessment of injured leaves and classifying category of diseases. To enhance quality of acquired images, image enhancement methods are applied. The proposed model is involved with four stages: first stage includes transformation of RGB to the HIS color space, analyzing histogram and intensity adjustment. Further next stage, contains adaption of fuzzy feature algorithm to segment image with adjustable parameter. In third stage, features are extracted using component labelling based on color, size and shape of diseased spots. The fourth stage is classification, which includes artificial neural networks to categorize disease category. The author concluded that proposed model clearly differentiates between healthy and diseased leaves with maximum accuracy rate.

S. Sannakki et al. [62] proposed organization model to classify grape leave diseases using neural networks. Initially, grape leave image with complex background is captured and background is removed using green color pixel masking. Noise is removed using anisotropic diffusion up to 5 iterations to preserve infected part information. K-means clustering is used for segmentation. Textural information is calculated from GLCM matrix. Extracted nine features are used by feed forward BPNN for its classification. The author concluded that the maximum accuracy is attained using the hue features only.

Kholis Majid et al. [63] presented technique for paddy malady diseases identification system, using Probabilistic Neural Networks (PNN) and fuzzy entropy classifier that keep running on versatile android working system. Initially images are captured and significant features extracted using fuzzy entropy with membership functions to extracting brightness levels. After feature extraction by, probabilistic neural networks method is used for further classification. The authors concluded that PNN approach provides optimal result, but the tungro disease has lowermost accuracy of recognition, because certain features of tungro diseases are alike to bacterial leave blight diseases. But proposed method accuracy can be increased by increasing the dataset.

Revathi et al. [64] demonstrated a new method for detection of cotton diseases using improved Particle Swarm Optimization (PSO) feature selection technique. It implements skew divergence method with parameters like variances, texture, color and edge to excerpt required features. Color variance feature is evaluated using color histogram and color descriptor. Sobel and Canny edge detection method is used to extract shape. Finally, Cross Information Gains Depth Forward Neural Network (CIGDFNN) classify cotton leave spot diseases such as root rot, leave blight, micro nutrient, verticilium wilt, bacterial blight and Fusarium wilt precisely and diminishing the error rate.

Sachin B. Jagtap et al. [65] presented an integrated image analyzer with a diagnostic expert system model. Histogram equalization is used to increase contrast of image by adjusting its intensity. After color transformation from RGB to HIS color space, fuzzy c means clustering is used for identifying region of interest. Finally, diseases are classified using neural networks. The author effectuated that due to integration diagnosis, accuracy will increase. Additionally, proposed system focuses on specific identification disorder which can be further extended to include more disorders. Such extension of system is carried out in such a way that it will be capable of detecting and identifying abnormalities on the other parts of plants also e.g. fruit, stem and root.

J.W. Orillo et al. [66] projected a technique to detect rice plant diseases such as bacterial leave blight causing brown spot using Otsu thresholding method and neural networks. In the first stage, after transformation of RGB to HSV color space, diseased area of leave is evaluated using Otsu thresholding method. After segmentation, there are four features namely, (a) segment enclosed by the disease on leave (b) R, G, and B means value of disease (c) R, G, and B standard deviation values (d) H, S and V mean values of disease. In last stage, back propagation neural network is used to evaluate the accuracy and performance of proposed method. The author finally verified that representative lesions of the diseased leave are recognized accurately. Table 2.5 represents the summarization of neural networks techniques used for disease classification.

Table 2.5: Disease classification using Artificial Neural Networks

Proposal	**Culture**	**Diseases**	**No. of images**	**Accuracy**	**Features**	**Image format**	**Environment conditions**	**Acquisition device / dataset**
Huang et al. [2007]	Phalaenopsis seedling	Bacterial [bacterial soft rot, bacterial brown spot, Phytophthora & black rot]	289 Images	89.6 % With classifying diseases 97.25 % without classify diseases	Texture & color	TIFF format	Controlled lighting conditions	XC-711, Sony, / Dataset provided by Taiwan Sugar Research Institute, Tainan
Abdullah et al. [2008]	Rubber tree leave	Corynespora, frog eye & Collectotrich Um	800 Images	96.5%	Color	Jpeg format	Controlled lighting conditions	FinePix 6900 Zoom (Fuji Film)/ Samples collected from LembagaGetah Malaysia,
Phadikar et al. [2008]	Rice leave	Leave brown spot, rice blast, Sheath rot& brown spot	300 [training g]	94.21%	Features extracted w.r.t zooming algo.	Jpeg format	Controlled lighting conditions	Nikon COOLPIX P4 digital/ captured at East, Midnapur, India.
Bashish et al. [2010]	Leave & stem diseases	Early scorch, cottony mild, ashen mold, tiny whiteness & late scorch	192 Images	93% (approx.)	Texture	Jpeg format	Controlled lighting conditions	Common digital camera. Images taken from Al-Ghor area in Jordan
Hiary et al. [2011]	Various plants and fruits leave	Early scorch, cottony mold, ashen mold, late scorch & whiteness	192 Images	94.67%	Color &texture	N/A	Controlled lighting conditions	Common digital camera. Images taken from Al-Ghor area in Jordan

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device / dataset
Gulhane et al.[2011]	Cotton leave	Bacterial, fungal, viral and aphid	45 Images	85 to 91%	Color	N/A	N/A	Common digital camera
Abdullakasim et al. [2011]	Cassava leave	Brown spot	160 images	79.23% (diseased) 89.92% (healthy)	Color indices	N/A	Natural lighting conditions	Common digital camera/ Kamphaengsaen Camus at Kasetsart University, Thailand.
Kai et al. [2011]	Maize leave	Disease spot	40 Images	98% or more	Texture	N/A	N/A	N/A
Guru et al. [2011]	Tobacco leave	Anthracnose & frog eye	100 Images	88.59%.	Texture	Jpeg format	Captured in field under un-controlled lighting conditions	Sony digital camera. Images captured at Central Tobacco Research Institute, Hunsur, Karnataka, India
Zhang et al. [2013]	Jujube leave	Rust, anthracnose, white rot, fruit rust disease, ascochyta spot and witch broom	300 Images	rust [91%] anthracnose[89%] white rot [94%] fruit rust [84%] ascochyta spot of [73%] witch broom [81%]	Color, morphological &shape	Jpeg format	Controlled Lighting Conditions	Canon camera (Model 60D)
Keskar et al. [2013]	Various plant leave	Citrus canker	N/A	N/A	Color	Jpeg format	Controlled lighting conditions	Common digital camera
Sannakki et al. [2013]	Grape leaves	Downy mildew, powdery mildew & anthracnose	33 images	100% (using only hue features)	Texture	Jpeg format	Controlled lighting conditions	Nikon Coolpix P510, 16.1 Megapixel
Kutty et al. [2013]	Watermelon leave	Downy mildew & anthracnose	200 Images	75 % & 76.9%	Color	Jpeg format	Controlled lighting conditions	Nikon D80 digital camera
Majid et al. [2013]	Cassava leaves	Cassava Mosaic Disease	N/A	91.46%.	Fuzzy entropy	Jpeg format	Images are captured in fields	Mobile phone. Laladon,/Bogor, Indonesia

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environ-ment conditions	Acquisition device / dataset
Revathi et al. [2014]	Cotton leaves	Grey mildew & bacterial blight	390 Images 270 [train] 120 [test]	94% [EPSO]	Color, texture & feature Vector and Using Skew divergen Ce method.	N/A	Controlled Lighting Conditions	Nokia mobile camera/ south zone of Tamil Nadu at Andhiyaur district, India
Jagtap et al. [2014]	Crop leaves	Leave batches	N/A	N/A	Morphol Ogical	Jpeg format	Controlled Lighting Conditions	Common digital camera
Orillo et al. [2014]	Rice leave	Bacterial blight	134 Images	100%	Mean And standard deviation	Jpeg format	Controlled Lighting Conditions	Common digital camera. Images captured at Greenhouse of the International Rice Research Institute located at Los Banos, Laguna, Philippines.

2.1.6 Naive Bayes classifier

Aduwo et al. [67] presented an automated vision-based analysis system to detect cassava mosaic disease. The procedure begins with capturing leave images with a standard digital camera. Three different feature extraction techniques are used to attain representative data of leave images. For first dataset, color features of hue pixels of HSV color space using histogram are extracted. For next two dataset shape features using speeded up robust features and scale-invariant feature transform methods are extracted. Finally, images are classified using K-Nearest Neighbor (KNN), Naive Bayes and SVM. A comparison analysis of these three classifiers is generated. Naive Bayes provided best results. The author concluded that study of color histograms by use of divergence procedures has the ability to give good classification results. The extra information in the augmented feature sets lead to better generalization of the classifiers using them.

Dhiman Mondal et al. [68] proposed a technique to detect yellow vain mosaic virus disease in okra leaves with combination of K means and Naive Bayes classifier. Initially 23 features are evaluated on gray images and converted into invariant features by representing different features range. Using gray level co-occurrence matrix texture features are extracted in second phase. After feature extraction, Pearson correlation coefficient method is used to identify dominant features. Feature values analogous to dominant feature set are designated to create feature metrics and clustered. The final classification is done using Naive Bayes classifier. The author concluded that the proposed method identifies 87% diseases correctly and classification rate can be increased using more appropriate features. Table 2.6 represents the summarization of Naive Bayes used for disease classification.

Table 2.6: Disease classification using Naive Bayes

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device / dataset
Aduwo et al. [2010]	Cassava leave	Cassava Mosaic Disease	193 Images	N/A	Color and shape	Jpeg format	Controlled lighting conditions	Camera phone, Images captured at Namulonge Crops Resources Research Institute, Uganda
Mondal et al. [2015]	Okra leave	Yellow Vein Mosaic Virus diseases	79 images 40 [training] 39[testing]	87%	Texture	N/A	Controlled lighting conditions	Digital color camera

2.1.7 Fuzzy logic analysis

Azmi et al. [69] presents an orchid diseases detection system using combination of image processing and fuzzy logic. The proposed scheme comprises of two phases; in first phase, leaves are segmented using Otsu thresholding method. After segmentation, numbers of the diseased spot are calculated. In second phase, extracted diseased spots and diseased area are used as input of fuzzy logic system. The number of diseased spots for each leave is stored in a matrix form during batch processing of images. Based on these two features, fuzzy system determines disease category using nine "if-else" rules based on Mamdani method. The de-fuzzification is evaluated using center of maximum method, mean of maximum method and center of area method. The author concluded that this method did not provide actual number of diseases spots. They detect a greater number of diseased spots as compared to actual one.

Billah et al. [70] suggested a model for recognizing tea leave diseases, which uses color wavelet features and adaptive neuro fuzzy inference system. After pre-processing, color wavelet features are extracted with second order statistical representation of the wavelets transform. The extracted features classified by Adaptive Neuro Fuzzy Inference System (ANFIS) along with three types of disease categories. The author concluded that proposed techniques gives promising results and can recognize diseases accurately. The color wavelet analysis performs better than other feature extraction methodologies.

K. Muthukannan et al. [71] defined a fuzzy rule-based technique using color features for disease classification of tomato leave. Firstly, gradient operator is used to suppress noise or small fluctuation in image. Two important features, mean and standard deviation are extracted from cropped images. In the end, fuzzy inference system using fuzzy rules with selected color features classified image regions as healthy, slightly healthy and highly affected disease portion of plant leave. Where, orange color indicates healthy portion of image, yellow and red color indicated diseased region of leave. The author concluded that performance of fuzzy rule-based classification is satisfactory. The experimental results show proposed method can detect leave diseases with little computational effort. Table 2.7 represents the summarization of fuzzy logic analysis used for disease classification.

Table 2.7: Summarization of fuzzy classification techniques for diseases classification

Proposal	**Culture**	**Diseases**	**No. of images**	**Accuracy**	**Features**	**Image format**	**Environment conditions**	**Acquisition device / Standard dataset**
Azmi et al. [2013]	Orchid leave	Black spots, virus & yellow Spots	80 images	N/A	Centroid, area, diseased spots	Jpeg format	Fixed lighting conditions	Digital color camera
Billah et al. [2015]	Tea leave	Virus &fungal	75 images 45[Training] 30[Testing]	95.7%	Color wavelet feature extraction	N/A	N/A	Digital color camera
Muthukannan et al. [2015]	Tomato leave	Early blight, late blight &septoria	100 images	95%	Mean, standard deviation	N/A	N/A	N/A

2.1.8 Particle Swarm Optimization

Zhang et al. [72] proposed an improved Particle Swarm Optimization (PSO) algorithm and neural networks to recognize and diagnose maize leave diseases. Image is enhanced using histogram equalization method and de-noising is realized using image filter. Color, texture and shape features based on HSI color components are extracted. An improved PSO algorithm based on opposition learning method is proposed to decrease the prospect that exploration for particle swarm falls into the local optima so as to achieve a more optimal solution. To achieve high convergence speed, more compatible particle obtained to continuously optimize global area. The authors concluded that traditional neural networks have slow convergent speed, easy getting into local minimum and low rate of correct motion pattern recognition.

Muthukannan et al. [73] stated the importance of PSO based segmentation of images. After data set collection, Gaussian filter is used to remove unwanted information of speckles. The results of pre-processing are analyzed using peak signal to noise ratio and max error parameters. The segmentation is achieved using binary PSO to select best number of clusters. The centers of selected clusters are refined by K-means clustering. PSO is employed for allocating each pixel to a cluster. After segmentation, hybrid feature parameters based on texture, color and shape using GLCM matrices are extracted and classified. According to authors, hybrid feature extraction approach is useful for disease classification It also increases the correct prediction classification accuracy. Table 2.8 represents the summarization of PSO techniques for disease classification.

Table 2.8: Summarization of PSO techniques used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device / Standard dataset
Zhang et al. [2014]	Okra leave	Yellow vein mosaic virus disease	400 images	93.3%	Color, shape & texture	Jpeg format	N/A	Images collected from Hebei agricultural university
Muthukannan et al. [2015]	Tomato leave	Late blight, septoria leave spot,	N/A	N/A	Texture	N/A	N/A	N/A

2.1.9 Diseases classification using combination of classifiers

Tian et al. [74] proposed a system that utilizes stacked generalization structure to combine classification decisions attained from three kinds of Support Vector Machines (SVM) classifiers rather than using single SVM classifiers to identify wheat leave disease. The leave region is segmented using simple thresholding method. Color, texture and shape features are extracted and used as training set for three corresponding SVM classifiers. Later on, different extracted feature sets are classified by classifiers in low level and mid-level categories, which are partly described by the symptom of crop diseases according to knowledge of plant pathology. Then mid-level features are extracted from mid-categories produced from low-level classifiers. Finally using high-level SVM trained and correct errors made using different features. The author concluded that compared with other classifiers for wheat leave diseases recognition, the proposed approach can obtain better success rate of recognition.

El Massi et al. [75] proposed an approach using serial arrangement of two neural networks classifiers. Images are transformed from RGB to LAB color space and segmented using K means clustering to extract lesion region. Based on color, first classifier evaluates difference between classes. The damages having a same or an adjacent color are considered in same class. According to shape and texture features, second classifier is used to find difference between classes. The method is verified on four categories of diseases, including two categories of pest insects (i.e. Leave miners and the caterpillar Tuta absoluta), and symptoms of two fungal diseases (i.e. internal powdery mildew and downy mildew). The author concluded that proposed approach using serial classifiers arrangement method is stimulating and can resolve the difficulties of the individual classifiers.

Es-saady et al. [76] presented a system using serial grouping of two SVM classifiers. Firstly, images are transformed from RGB to LAB color space and segmented using K means clustering to extract the lesion region. Based on color, first classifier evaluates difference between classes. Damages having same or an adjacent color are considered in same class. According to shape and texture features, second classifier is used to evaluate difference between classes. The author concluded that proposed system using serial classifiers combination is interesting, and can resolve difficulties of individual classifiers. The performance can be improved using relevant features. Table 2.9 represents summarization of combinations of classifiers used for disease classification.

Table 2.9: Summarization of combinations of classifiers used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device/ Standard dataset
Tian et al. [2012]	Wheat leave	Powdery mildew, leave rust, leave blight & puccinia striiformis	800 samples	95.16%	Color, texture & Shape	N/A	Natural lighting conditions	Nikon D80 camera
Massi et al. [2015]	Vegetables crops leave	Leave miners caterpillar Tuta absoluta, downy mildew & powdery mildew	200 samples	Class of Whites 94% Class of Yellow 93%	Color, Texture &shape	N/A	Controlled lighting conditions	Digital color camera
Es-saady et al. [2016]	Vegetables crops leave	Leave miners, Thrips Tuta absoluta, Early blight, & powdery mildew	284 samples	87.80%	Color, texture & Shape	N/A	Controlled lighting conditions	Digital color camera

2.1.10 Orthogonal Locally Discriminant Projection Algorithm

Zhang et al. [77] presented Orthogonal Locally Discriminant Projection Algorithm (OPDPA) to classify maize plant diseases. The original leave images are pre-processed and converted to RGB to HSI color space. After that, green pixels are masked and removed to obtain the non-green component. Finally, every processed image is symbolized as a point in vector space. Segmentation is done using gray level selection thresholding method. Afterwards, training data is constructed from transformation matrix and assign to low-dimensional feature subspace which is orthogonal locally discriminant projected to test data points. The nearest neighbor graph weights are adjusted according to their reliability between two nodes. The final results prove that proposed method is achievable and effective. The effectiveness of kernel space in the OLDPA need to be more examined. Table 2.10 represents the summarization of technique used for diseases classification using orthogonal locally discriminant projection algorithm.

Table 2.10: Summarization of technique used for diseases classification using OPDPA algorithm

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition Device / Standard dataset
Zhang et al. [2013]	Maize leave	Fungus and leave spots	100 samples	94% (approx.)	Orthogonal locally discriminant feature Space	Jpeg format	N/A	N/A

2.1.11 Fractal dimensions

D. K. Wu et al. [78] proposed technique to recognize infected leave based on fractal dimensions. Two different kernels i.e. polynomial based kernel function and radial based kernel functions are considered for classify diseases. Fractal dimension of damaged cucumber leaves are used as input to threshold method and provide a quantitative index of the roughness of diseased leave image. Finally, radial kernel based SVM excels to polynomial–based kernel. The author concluded the recognition and precision are depends on input vectors. The threshold method is good only for linear problems.

V. Surendrababu et al. [79] defined a method for distinguishing rice leave disease with fractal dimension and chaos theory. The study of an unhealthy leave is carried out with respect fractal dimension, image pattern, especially box-counting ratio calculation, and chaos. The fractal pattern for unhealthy leave at concluding phase will have similar pattern during the preliminary phase for each category of rice leave disease. The author concluded that proposed method evaluating initial information for the progress of an early detection system. Table 2.11 represents the summarization of the fractal dimension, and chaos theory used for disease classification.

Table 2.11: Summarization of fractal dimension, and chaos theory used for disease classification

Proposal	Culture	Diseases	No. of Images	Accuracy	Features	Image format	Environment conditions	Acquisition device / Standard dataset
Ke et al. [2008]	Cucumber	Leave miner	95 Images 61 [training] 34 [testing]	90%	Fractal dimensions	N/A	Controlled lighting conditions	Casio EX-Z3 camera. lmages collected at, China
Surendrababu et al.[2014]	Rice leave	Bacterial	N/A	More than 96%	Color, shape & texture	N/A	N/A	N/A

2.1.12 Support Vector Machine (SVM)

Meunkaewjinda et al. [80] proposed a self-organizing feature map to distinguish diseases of grape leave. Before color extraction, anisotropic diffusion method is used to evaluate information of affected pixels. The resulting color pixels are assembled by modified unsupervised self-organizing feature map and genetic algorithm to optimize group of different colors. In next phase, Gabor filter is applied on segmented images to analyze diseases more efficiently. Subsequently, SVM is applied to categorize grape leave diseases. The authors concluded that appearance of grape leave disease features using proposed method can achieve a very efficient quality of classification, but there are certain restrictions with respect to background of image, regarding extraction of indistinct color pixels.

Youwen et al. [81] suggested a technique to recognize two diseases namely powdery mildew and downy mildew that can apparent in cucumber leaves. Statistic pattern recognition approach is used to segment healthy and diseased regions of leaves. To differentiate diseased varieties, color and texture features are extracted. Using all features, SVM performs final classification using multiple kernel functions such as linear, polynomial, radial basis and sigmoid function. The authors concluded that using radial basis kernel function with only one shape feature, SVM can recognize cucumber diseases accurately. While with linear kernel function, combination of two feature sets (texture and shape) provides good performance. SVM has more diseases recognition accuracy and high speed as compared to BP artificial neural network in same environment.

Jian et al. [82] explained method to identify diseases of cucumber leave using SVM with a polynomial function, radial basis function and sigmoid kernel function. The classifier is trained using above mentioned kernel functions and results are compared to decide best kernel function for identification of diseases. The author summarized that radial basis function is best method to classify leave diseases and promises good accuracy. But this method is preferable in case of single disease only. This method cannot diagnose accurately when two or more diseases occur simultaneously.

Asraf et al. [83] introduced method to detect the symptoms of oil palm leaves nutrient disease using support vector machine along with polynomial kernel with hard margin, polynomial kernel with soft margin, and linear kernel. Color and histogram-based texture features are extracted and separated into numerous sub-features. SVM classifier is then used for classification using all 27 extracted features using different kernel functions. After, classification, a polynomial kernel with soft margin is more proficient of accurately classifying nutrient disease according to its class as compared to other methods. The author finally deduced that more accurate results can be attained using proper kernel trick function.

L. Cunlou et al. [84] proposed a novel technique to recognize maize leave diseases, based on Fuzzy Least Square Vector Machine (FLSVM) Algorithm. Using YCbCr color space, spatial GLCM is used to calculate texture characteristics of maize leave. After feature extraction, fuzzy least square vector machine is used for categorization. Where, sample mean is calculated with respect to center of each class. Then, according to distance between center (sample class) and sample, it calculates initial membership function using fuzzy K nearest neighbor method. Membership functions reflected the distribution characteristics of samples. The author concluded that proposed model provides encouraging results. It gives smaller errors in identification process when the samples number is too small. The developed FLSVM also gives prediction uncertainty.

Arivazhagan et al. [85] discussed a system for automatic plant disease detection, which comprised of four stages. In first stage, using specific threshold value green pixels are masked and removed. Using, co-occurrence matrix texture features are calculated from diseases segments. Finally, useful features are handed over to SVM classifier where classifier gain is obtained by minimum distance criterion. The results specify that proposed technique can distinguish and classify leave diseases with a minimalistic computational effort. The author envisioned to improving disease identification rate along with color and texture features for any given input conditions.

Zhou et al. [86] explained a approach which focused on cercospora leave spot identification in sugar beet using SVM with hybrid algorithms of template matching. The technique divided in three stages. To discriminate leave from soiled background, automatic plant segmentation index of G-R is introduced. Secondly, a robust template matching technique is implemented for constant monitoring of foliar translation, dynamic object examining and disease development. At last, SVM used for disease classification using color features. Ultimately the author analyzed that segmentation process is not suitable for other Dicot family plant.

Ratnasari et al. [87] explained a model to identify sugarcane leave diseases and recognize severity of spot disease. Severity measurement is accomplished only on sample data consisting of regular sized leaves. Disease identification is achieved through segmentation of diseases spot using thresholding of (a*) component from L*a*b* color space. After segmentation, texture and color features are extracted. In the end, SVM classifies various sugarcane leave diseases (i.e. ring, rust and yellow spot) using different kernels as polynomial with third order, quadratic, Radial Basis Function (RBF) and linear function. The author effectuated that proposed model displays more accuracy with small average error severity calculation. Linear kernel function gives better results than others. On other hand, due to limitations of segmentation techniques, lesion cannot be identified accurately.

Mokhtar et al. [88] applied Gabor wavelet transform technique to extract various diseases in tomato leave. Various relevant features along with SVM having different kernel functions are used to identify and recognize different kind of disease that infects tomato plant. Initially, a wavelet-based feature technique is used to identify an optimum feature subset. Finally, SVM classifiers with different kernel functions such as the Cauchy kernel, Invmult kernel and Laplacian kernel are employed to identify and detect tomato leave infected with powdery mildew or early blight. Finally, it is concluded that Cauchy and Laplacian kernel functions provide good accuracy but choice of optimum factors of kernel functions is still one of the critical issues.

Zhang et al. [89] explained a new methodology to evaluate recognition rate of diseases in cucumber leave using Singular Value Decomposition (SVD). Watershed algorithm is used for segment disease spot from leave images in first phase. In second phase, every spot is distributed into little blocks. Combining features from each block is extracted using SVD. In third phase, main point vectors are assembled and their dimensionalities are adjusted. Finally, SVM classifier is used to identify category of unknown disease leave image. The author concluded that presented method achieve good results. The main limitation of the proposed method is that it needs more computation efforts to evaluate the singular values. Advanced color features can improve a better recognition result.

Qin et al. [90] investigated a technique to analyze and classify four categories of alfalfa leave diseases. Using artificial cutting from every captured disease image, a sub image with single or multiple typical diseases are achieved. Every sub image is transformed into HSV color space and L*a*b color space. Each pixel of sub-image, (a*) component value and the (b*) component value are regarded as color features. All pixels in image are clustered into ten classes. The mean value of H components of all pixels in each class is calculated. Then using twelve lesion segmentation methods combined with clustering techniques (i.e. fuzzy C-means clustering K-means clustering and K-median clustering) and supervised classification algorithms (i.e. Naive Bayes algorithm logistic regression analysis, linear discriminant analysis and regression tree) used to segment the sub images. The pixels in the class with the minimum mean are treated as typical lesion pixels, and the pixels in the seven classes with the largest means are treated as typical healthy pixels. From the lesion images based on color, shape and texture total 129 feature are extracted. Using ReliefF, IR and correlation-based feature selection methods useful features are selected. Finally, diseases recognized and classified using three supervised learning models using support vector machine, random forest and K-nearest neighbor. Evaluations of the classification results of the learning models are evaluated. The author concluded that the SVM model using most significant 45 features is the optimal model. Table 2.12 represents the summarization of support vector machine techniques used for disease classification.

Table 2.12: Summarization of support vector machine techniques used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device / Standard dataset
Meunkaewjinda et al. [2008]	Grape leave	Scab &rust	1478 [training] 115 [testing] Further divided into sub-images	97.8%	Color	N/A	N/A	N/A
Youwen et al. [2008]	Cucumber leave	Powdery mildew & downy mildew	40 images	100%	Texture, shape & color	N/A	Images captured in field	Images collected from field of cucumber research base of Shenyang Agricul-ture University
Jian et al. [2010]	Cucumber leave	Downy mildew, brown spot & angular leave spot	60 images	RBF provides best results	Texture, shape & color	N/A	N/A	N/A
Asraf et al. [2012]	Oil palm leave	Nutrition disease namely nitrogen, potassiu magnes-ium	420 images	95%	Color, texture & histogram based	N/A	Controlled lighting conditions	Digital color camera
Lu et al. [2013]	Maize leave	Blight, sheath blight & southern blight	Set of images	98% or more	Texture	N/A	N/A	N/A
Arivazhagan et al. [2013]	Rose, beans, lemon and banana leave	Bacterial disease, sun burn, early scorch, late scorch & fungal	500 images	94%	Texture	N/A	N/A	Digital color camera/ Images collected from Tamil Nadu
Zhou et al. [2013]	Sugar beet	Cercospora leave spot	N/A	99.07% [testing] 99.44% [training]	Histogram	N/A	Controlled lighting conditions	CMOS Camera (CMOS130-USB2, Fortissimo Co., Japan)

Proposal	**Culture**	**Diseases**	**No. of images**	**Accuracy**	**Features**	**Image format**	**Environment conditions**	**Acquisition device / Standard dataset**
Ratnasari et al.[2014]	Sugarcane leave	Sugarcane ring, rust & yellow spots	A set of images [training] 30 images [testing]	80%	Texture	N/A	Controlled lighting conditions	Digital color camera/ Images are collected from sugarcane fields Indonesia
Mokhtar et al. [2015]	Tomato leave	Powdery mildew & early blight	200 images	99.5%	Texture	Jpeg format	Controlled lighting conditions	Digital color camera
Waghmare et al. [2016]	Grape leave	Downy mildew & black rot	450 images	96.6%	Texture, Color & shape	Jpeg format	Controlled lighting conditions	Mobile camera
Zhang et al. [2016]	Cucumber leave	Leave spot	100 images	92% approx.	Global and local singular values	N/A	Controlled lighting conditions	Images collected from the agricul-tural demons-tration zone of Northwest A&F University
Qin et al. [2016]	Alfalfa leaves	Leave spot, rust, leave spot & Cercospora leave spot	899 images	80% approx.	Texture, shape & color	N/A	Controlled lighting conditions	Digital color camera. Images taken from Langtang Forage Experimental Base, Institute of Animal Science, Chinese Academy of Agricultural Sciences &alfalfa fields Hebei Province, China

2.1.13 Discriminant analysis

Bandi et al. [91] suggested a technique to classify four classes of citrus leave diseases i.e. normal, greasy spot, melanose and scab. Based on color co-occurrence method texture features are evaluated for each citrus leave sample. After extracting features, they are classified with leave age condition using Naive Bayes classifier (NBC), Linear Discriminate Analysis (LDA) classifier¸ k-Nearest Neighbor (KNN) and Random Forest Tree (RFT) algorithm classifier Eventually using earphone operative characteristics contour, all classifiers are compared. The author concluded that normal and greasy spot leaves can be classified easily from other classes of leaves. Whereas, melanose diseased leaves classification rate is less. The LDA outperforms than other classifiers.

Kruse et al. [92] explained technique to classify each pixel as injured or healthy of clove leaves through extraction of color and texture-based information. Four classification approaches are evaluated as Fit to Pattern Model Approach (FPM) combined with T statistics, Linear Discriminant Analysis (LDA), K means clustering and residual sum of squares for classifying and calculating leave surface injury. Predicted leave pixel classification is compared with manually segmented images. The ground truth is used to evaluate pixel classification accuracy. It is concluded that an LDA classifier performed well as compare to mentioned three methodologies in pixel identification. It is determined that simple feature vector with only color information is sufficient for leave pixel injury classification. The spatial information does not increase computation time significantly. Table 2.13 represents the summarization of discriminant analysis techniques used for disease classification.

Table 2.13: Summarization of discriminant analysis techniques used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device/ Standard dataset
Bandi et al. [2013]	Citrus leave	Greasy spot, melanose and scab	40 images 20 [Training] 20[Testing]	98.75%	Texture	Uncompressed Jpeg format	N/A	Digital color camera
Kruse et al. [2014]	Clove leave	Leave injury	72images	95%	Color & Texture intensity	TIFF format	Controlled lighting conditions	Canon EOS 20D SLR

2.1.14 AdaBoost algorithm

Min Zhang et al. [93] explained a new method using zone based local and global features to identify citrus canker from citrus leave images. In first phase, an improved AdaBoost algorithm is used to evaluate most imperative features (citrus lesions) after extracting lesions from their background. In second phase, for combining texture and color information about lesion, a special citrus canker feature descriptor is proposed. To disclose spatial properties of citrus canker, Local binary pattern descriptors are used. In last phase, to identify canker lesion, two-level hierarchical recognition structure is established. The author compared performance of the suggested method with human experts and concluded that accuracy of the presented approach is almost similar.

K. Jagan Mohan et al. [94] proposed a system to identify and classify paddy leave diseases i.e. brown spot disease, leave blast disease and bacterial blight disease. The proposed work is divided into two parts. Diseases are identified using HAAR features and classified using AdaBoost classifier to locate disease affected portion and provides identification rate 83.33% in the first phase. Secondly different diseases are recognized using SIFT feature extraction method to extract local features of image and classified using k-NN and SVM with recognition rate 91.10% and 93.33%. The author concluded that this approach can detect the disease at an early stage and thus can minimize the loss of production. Table 2.14 represents the summarization of AdaBoost algorithm used for disease classification.

Table 2.14: Summarization of summarization of AdaBoost algorithm used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device/ Standard dataset
Zhang et al. [2011]	Citrus leave	Citrus Canker	500 Images	88%	Color & Texture	N/A	Captured in the field	Digital Sony DSCP92 and Canon EOS350D/ Images are collected in winter 2005 and 2006 from Guangdong province, spring in 2007 at Guangxi province, China
Mohan et al. [2016]	Paddy leaves	Brown spot leave blast and bacterial blight	60 images	83.3% identify 91.1 %, recognize	HAAR & SIFT	N/A	N/A	N/A

2.1.15 Rule set theory

Phadikar et al. [95] proposed a method based of infected region features to classify different diseases categories of rice. To separate infected segment, Fermi energy segmentation method is used. Change in shape, position and color of diseases spot defined as features to classify diseases with respect to boundary of leave. Genetic algorithm is used to detect the shape of infected segment which estimates the structure of leave region. The disease spot is separated into different blocks and then settled as a quad tree at different labels. Consequently, position of infection is determined. Binary illustration of each block decreases computational complication reasonably. To minimize loss of information and complexity, significant features are selected using rough set theory. Finally, a rule base classifier using selected features has been assembled that cover all rice leave diseases. The author concluded that proposed algorithm gives superior outcomes as compare to traditional classifiers. It involves lesser computational complexity. Table 2.15 represents summarization of rules-based theory used for disease classification.

Table 2.15: Summarization of rules-based theory used for disease classification

Proposal	Culture	Diseases	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device / standard dataset
Phadikar et al.[2013]	Rice leave	Brown spot, bacterial blight, rice blast and sheath rot	500 images	91.19% [SMO]	Color & shape & position	N/A	N/A	UCI dataset

2.1.16 Deep learning methods

Sladojevic et al. [96] developed plants diseases identification model using Convolution Neural Networks (CNN) on 13 different kinds of datasets. To increase dataset, augmentation process is applied. New framework of training and fine tuning is proposed for diseases identification. Caffe framework is used to execute CNN training. It also contains multi-layers which progressively calculate features from images. Layer parameters includes different learnable for optimization. The author concluded that proposed method provides good results and avoid training over fitting.

B. Liu et al. [97] proposed apple leave disease detection system using deep neural networks. In the first phase, to decrease over fitting of images are generated using direction light disturbance, PCA jittering and direction disturbance. In the next phase, a new CNN model using AlexNet model is implementing by adding up pooling layers, initiating GoogleNet setting, eliminate partial full linked layers. In the end Nesterov's Accelerated Gradient (NAG) algorithm is used to evaluate system parameter to precisely identify disease. The author concluded that proposed CNN model gives good accuracy and quick convergence rate. This model also improves robustness of CNN model.

Toda et al. [98] compared and analyze four different visualization techniques for CNN to identify plant diseases. Features visualization, semantic dictionary, hidden layer visualization and attention map methods interpret the representation of CNN model. The evaluation of all visualization techniques highlighted the largely visible lesions of each image. The author concluded that semantic dictionary and feature visualization technique are useful to detect visual features for classify particular disease. Conforming to visualization results, 75 % network parameters are not significant and do not play an important role for classification. Table 2.16 represents summarization of deep learning methods used for disease classification.

Table 2.16: Summarization of CNN based methods used for disease classification

Proposal	Culture	Disease	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device/ standard dataset
Sladojevic et al [2016]	Apple, grapevine	N/A	30,000 images (approximate)	96.3 %	CNN neurons	NA	NA	Internet
B.Liu et al. [2017]	Apple	Rust, Brown spot, Mosaic and Alternaria leave spot	13,689	97.62%	NA	NA	NA	BM-500GE/BB-500GE digital camera/ Gansu Province Qingyang country and Shanxi province, China

Proposal	Culture	Disease	No. of images	Accuracy	Features	Image format	Environment conditions	Acquisition device / standard dataset
Toda et al. [2019]	Various plant diseases	NA	54306 images	99.9	N/A	N/A	N/A	Plant village data set

2.1.17 Work on basil

Diseases detection and classification of basil leaves using image processing is still an active topic for research.

M.B.AL-Otaibi et al. [99] proposed technique to recognize and classify fresh and infected basil and parsley leaves using neural networks. In first phase, RGB images are converted into binary and median filter is used remove the effect of noise. Using appropriate threshold values, images are segmented. Statistical features are extracted by considering parameters; area, median, centroid, Feret’s diameter, standard deviation, mean gay value and circularity. Finally, Euclidean distance and neural networks are used to recognize and classify fresh and infected parsley and basil leaves. The author concluded that proposed work gives satisfactory results. But there is need to evaluate proposed system accuracy on large dataset. Table 2.17 represents summarization of rules-based theory used for disease classification.

Table 2.17: Summarization of rules-based theory used for disease classification

Proposal	Culture	Disease	No. of images	Accuracy	Features	Image Format	Environment conditions	Acquisition device/ Standard Dataset
AL-Otaibi et al [2017]	Basil and parsley leaves	N/A	15 images of each category	80% (classification) 100% (recognition)	Statistical features	JPEG	N/A	Digital camera

2.2 Evaluation metrics

Various evaluation measures are used to calculate the effectiveness of proposed model; namely, specificity, sensitivity, error rate, accuracy, precision, Kappa coefficient, receiver operating characteristics and area under curve [100-101]. These all measures are depending on four parameters as true negatives, true positives, false positives and false negatives as shown in Table 2.18.

Table 2.18: Confusion matrix

	Predicted Class 1	**Predicted Class 2**
Class 1	TN Correct true negative prediction	FP Incorrect false positive prediction
Class 2	FN Incorrect False negative prediction.	TP Correct true positive prediction

(i) Specificity

It is measured as ratio of correct negative interpretation to total number of negative interpretations. It is also known as true negative rate.

$$\text{Specificity (True negative rate)} = \frac{TN}{TP+FP} \quad \textbf{(2.1)}$$

(ii) Precision

Precision is defined as ratio of accurately predicted positive interpretations to total predicted positive interpretations.

$$\text{Precision} = \frac{TP}{TP + FP} \quad \textbf{(2.2)}$$

It is also known as positive predictive value (PPV). The best case of precision is 1 and the worst case is 0.

(iii) Recall

Recall is defined as ratio of predicted positive observations to the total predicted positive and negative interpretation. It is also known as true positive rate.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

(2.3)

High value of recall signifies class or category of object is correctly recognized.

(iv) Accuracy

Accuracy is the proportion of accurately predicted interpretation to the total interpretation with acceptable errors.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

(2.4)

(iv) Error rate

Error rate is measured as the number of all wrong interpretations divided by the all correct and incorrect interpretation or dataset.

$$\text{Error rate} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

(2.5)

(v) False positive rate

FPR defined as number of false positive interpretations divided by the total number of negative interpretations. It is also known as negative predicted value.

$$\text{False positive rate} = \frac{\text{FP}}{\text{TN} + \text{FP}}$$

(2.6)

(vi) Receiver Operating Characteristics (ROC) and Area under the Curve (AUC)

ROC is an efficient method for evaluating discrimination power of statistical model [102]. It plots sensitivity on the y axis versus specificity on x axis across the different possible threshold values. It is a graphical representation that illustrate the trade-offs among the errors. Where, AUC process the whole two-dimensional area under the entire ROC curve. AUC portray the probability that an indiscriminately selected positive example is accurately rated with greater suspicion than a randomly chosen negative example. AUC ranges in value from 0 to 1. High value of AUC typically reflects good discrimination competence of a classifier. An area of 1 represents a perfect test and an area of .5 represents a worthless test. The model performs better if a ROC curve is lifted up and away from the diagonal. It provides the capability to access the performance of classifier.

(vii) Cost curve

Cost curve defined the classifiers performance based on the cost of misclassification. The x-axis represents probability cost function and y-axis represents normalized expected misclassification cost [103].

$$\mathrm{x\ axis} = \mathrm{pC}(+) = \frac{\mathrm{p} + \; * \, \mathrm{C} - | +}{\mathrm{p} + * \, \mathrm{C} - | + + \, \mathrm{p} - * \, \mathrm{C} + | -}$$

(2.7)

$$\mathrm{y\ axis} = 1 - \mathrm{PD} - \mathrm{PF} * \mathrm{PC} + + \mathrm{PF}$$

(2.8)

Where, (+) signifies the false values and (-) sign represents the true predicted values. $\mathrm{C} - +$ defined the cost of misclassifying false values as true values. C (+|−) represents the cost of misclassifying true values as false values. p $(-)$ and p $(-)$ explains the probability of software model being represents true and false value when classification model is deployed.

(viii) Kappa coefficient

The kappa coefficient compares an observed accuracy with expected accuracy (random accuracy) [104].

$$\mathrm{Kappa} = \frac{\mathrm{Observed\ accuracy} - \mathrm{expected\ accuracy}}{1 - \mathrm{expected\ accuracy}}$$

(2.9)

A kappa value of 1 represents perfect classification, while a value of 0 represents imperfect classification. Kappa range varies between −1 to +1, where 0 represents the amount of agreement that can be expected from random chance, and 1 represents perfect agreement between the raters. Values ≤ 0 as indicating no agreement and 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

Chapter 3

Image Acquisition System

This chapter describes structure of data set preparation. Chapter discusses various measures; image acquisition technique, location selection, leaves sample selection, camera selection involved in leaves analysis briefly. Section 3.1 explained role of image acquisition system, leaves sample selection, location selection and camera selection. Chapter summary is representing in section 3.2. Figure 3.1 defines the identification and classification framework. The colored box highlights the main proposed work of concerned chapter.

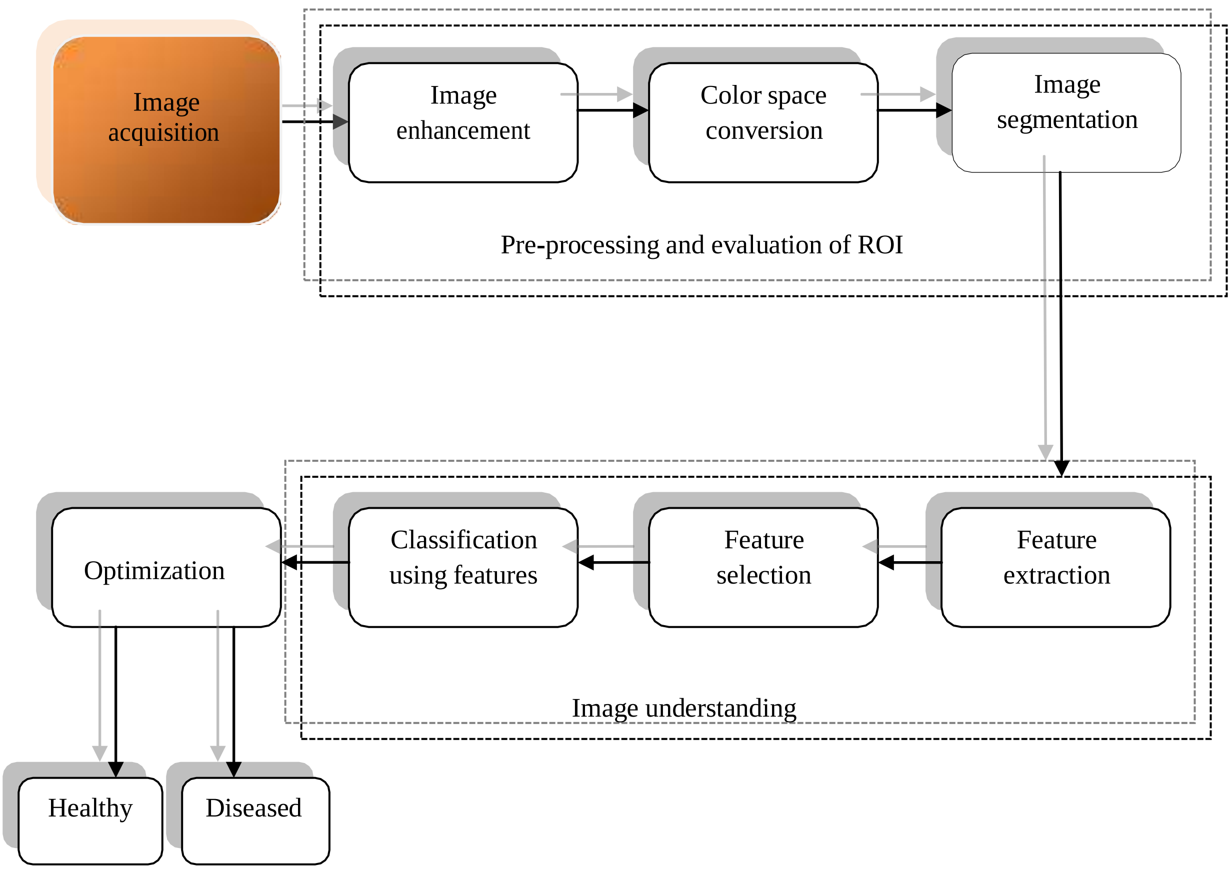


Figure 3.1: Highlighting steps of acquisition system in proposed system model

3.1 Introduction

In this chapter, data acquisition system of proposed work is outlined. Image acquisition is a task to capture and make record of different objects in realistic environment under different lighting conditions. In our thesis, we have endeavored to conquer limitations of previous works.

There are various challenges in accurately captured leaves images and identifying diseases [27-105]. We collect and acquired basil leaves from different basil gardens. Basil cultivate almost in every province with different environmental and geographical conditions. Even chemical composition of basil diverges with respect to season and cultivation location [106-107]. In order to obtain basil leaves, healthy as well as diseased, four different locations of Punjab are selected. We attempt to generate variations while collecting data from different herb gardens. During image acquisition, few things need to be considered carefully such as time, plant leaves sample and location. Diseases occurrence on basil depends on various features such as humidity, rainfall, nutrition, temperature, seasons etc.

We have worked with four types of major diseases of basil leaves. For image acquisition task, four main interpretations considered as (i) quality of leaves sample (ii) lighting conditions (iii) only visible diseases area to be detected (iv) camera selection. The chapter discusses method of leaves selection, location selection, method for acquiring image etc. Chapter also present measures concerned in image acquisition system used in research.

3.1.1 Location selection

Selection of location is an important issue as leave characteristics changes with respect to regions. In present work, leaves dataset consists of four types of healthy and diseased basil leaves images; these are Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (holy basil) and Ocimum gratissimum (Vana-holy basil). These samples are collected from herb garden at Punjab Agriculture University Ludhiana (30.9°N 75.85°E), National Institute of Pharmaceutical Education and Research (NIPER) Mohali (30.78°N 76.69°E), Botanical garden, Chandigarh and Punjabi University Patiala (30.32°N 76.40°E) India for reflective study during winter and spring seasons. These particular four places contain large variety of basil leaves. Figure 3.2 – Figure 3.4 represents various study sites of basil gardens.

Figure 3.2: Study sites of basil gardens of Chandigarh Botanical garden

Figure 3.3: Study sites of basil gardens of Punjab Agricultural University

Figure 3.4: Study sites of basil gardens of National Institute of Pharmaceutical Education and Research

3.1.2 Leaves sample selection

Both healthy and diseased leaves are selected as leaves samples. Diseases found as downy mildew, bacterial leave spots, gray mold and Fusarium wilt. Diseases can be occurred due to unfavorable environment conditions, soil, water contamination, excessive use of pesticides etc. Approximately 100 to 200 leaves are composed from selected locations.

3.1.3 Leaves sample preparation

Leaves sample are collected from different herb gardens require to be organized for further processing. Firstly, leaves are taken to the research laboratory. To attain similar surface condition, leaves are cleaned with water. While collecting leave samples, suggestions of various plant pathologists were considered.

3.1.4 Image acquisition

Leave images are captured with appropriate assortment of camera. The particulars are as follows.

3.1.4.1 Camera selection

Leaves are digitally captured in color using EOS 5D Mark III, 22.3-megapixel CMOS sensor with resolution of 5760 x 3840 pixels, 14-bit A/D Conversion, wide range ISO Setting 100 25600, which can shoot up to 6 frames per second (fps) from a constant height (45cm) over center of imaging station. Resolution of camera is set in normal mode. To diminish effect of reflection, flash is kept off.

3.1.4.2 Imaging station

Collected leave samples are taken to research laboratory. Figure 3.5 represents the experimental set up of proposed system. After cleansing of the leave samples, leaves are then taken to an imaging station. Images of leaves samples are acquired indoor to minimize the noxious effects of variants in ambient lighting conditions. To simulate outdoor environments and to avoid factors such as illumination and orientation four fluorescent bulbs with natural light filters and reflectors are used. A camera positioned vertically from samples to contain all components, with best possible resolution. The camera was stipulated on a camera stand which reduces movement of hand and capturing uniform images of basil leaves.

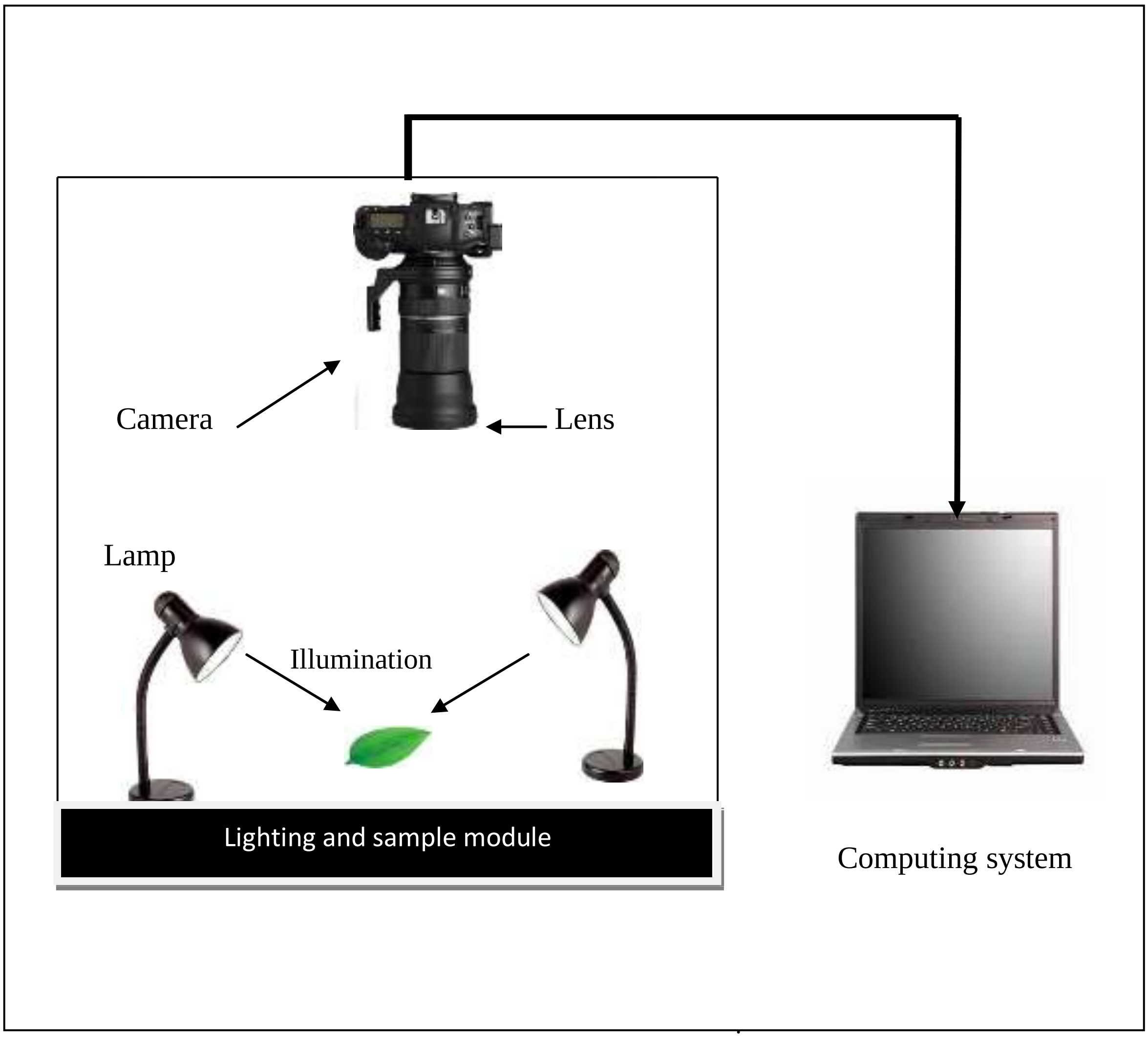


Figure 3.5: Imaging station

Lamps (bulbs) with natural light reflectors located at 45-degree angle to ensure proper lighting. The degree of damage varied between leaves samples. Images are captured under controlled field conditions to reduce unfavorable effects of deviation in surrounding lighting conditions.

To obtain uniform illumination four 16 W cool white fluorescent bulbs (4500 K color temperature) placed at 30 cm above imaging station surface. To sustain uniformity, images are acquired from same height. The focus and height of camera is adjusted to capture the image of whole leaves. To stay away from effect of reflection flash is kept off. Images are captured on black background as shown in Figure 3.6.

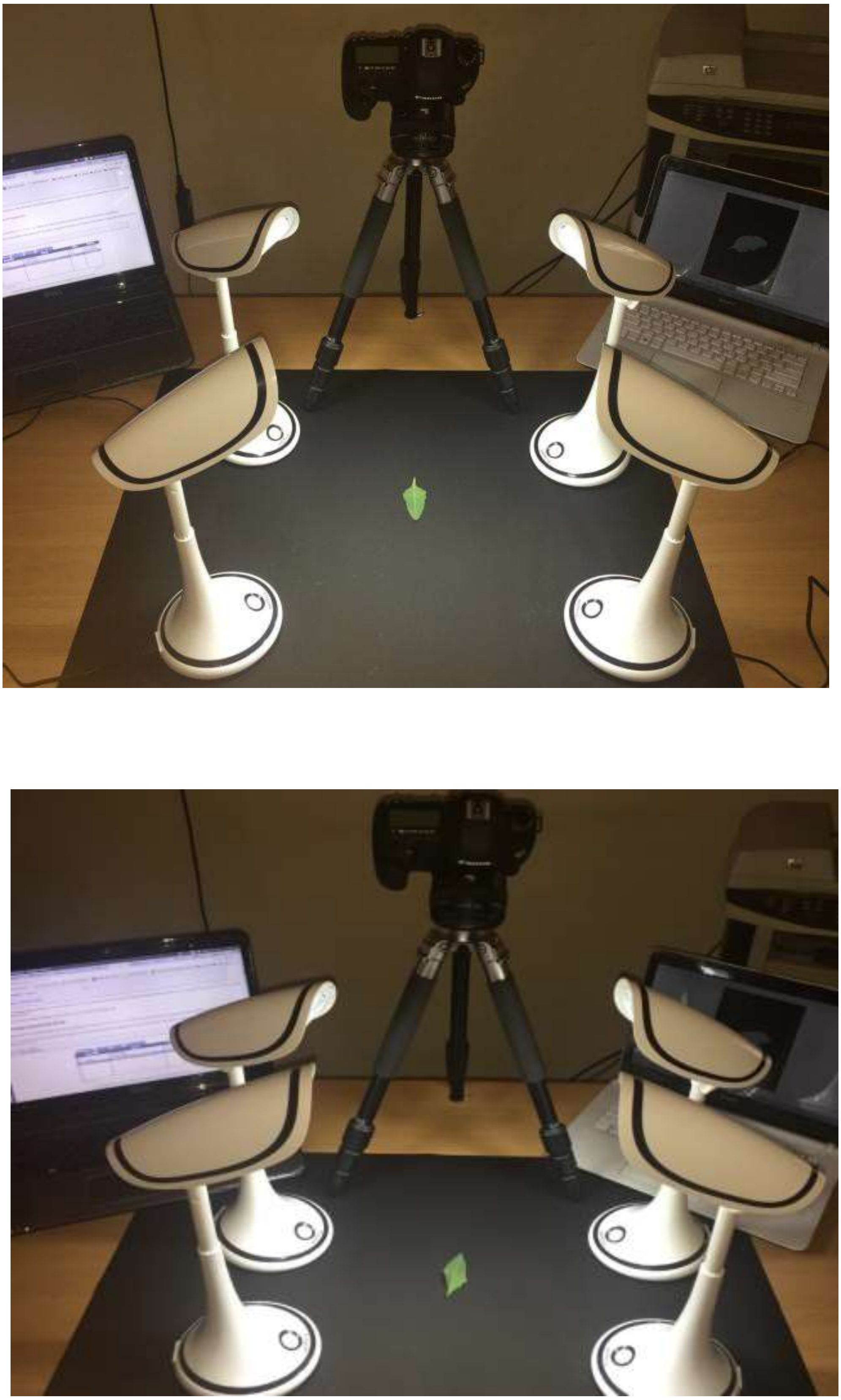


Figure 3.6: Live module of imaging station

Figure 3.7 represents captured basil leave images (samples) at imaging station.

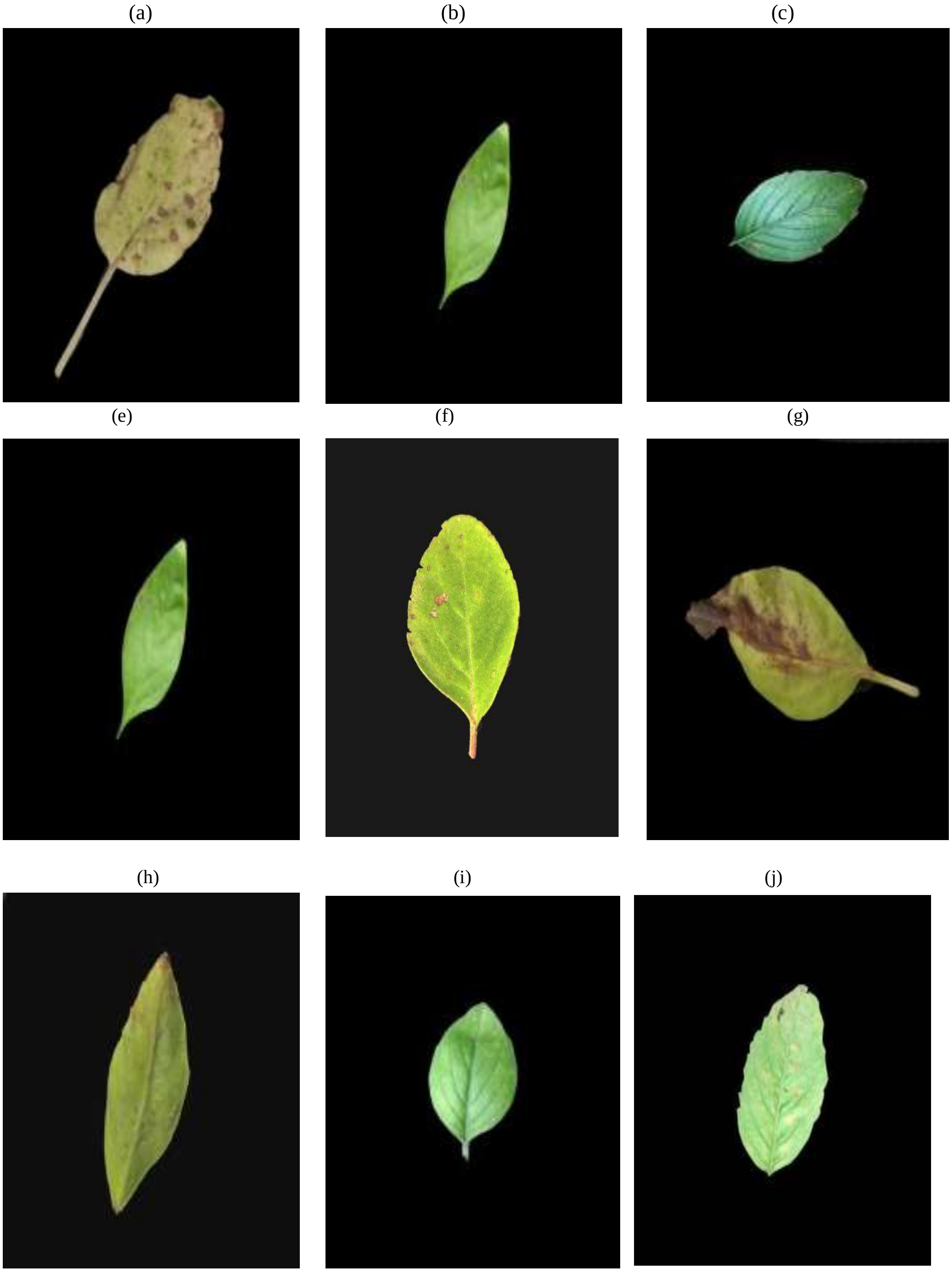


Figure 3.7: (a) and (g) depicts bacterial leave spot disease, (b), (c), (d), (h) defines downy mildew, (e) and (j) depicts cercospora leave spot

3.2 Chapter summary

Leave samples are composed as per guiding principle given by experts across India. Samples are collected from various fields (Ludhiana, Mohali and Patiala) of Punjab and Chandigarh. Images are captured under controlled lighting conditions. Images are further used for data analysis purpose using different image processing and machine learning algorithms.

Chapter 4

Segmentation using novel neutrosophic approach

In this chapter, novel fuzzy set extended form neutrosophic logic-based segmentation technique is outlined. The segmented neutrosophic image is distinguished by three membership elements: true, false and intermediate region. Using these three regions, region of interest is evaluated. Section 4.1 explained role of segmentation, existing segmentation models and their characteristics. In Section 4.2, color space conversion and pre-processing model are defined. New segmentation model is introduced in Section 4.3. Implementation details and results are specified in Section 4.4. Figure 4.1 represents flow chart highlighting steps of proposed segmentation model.

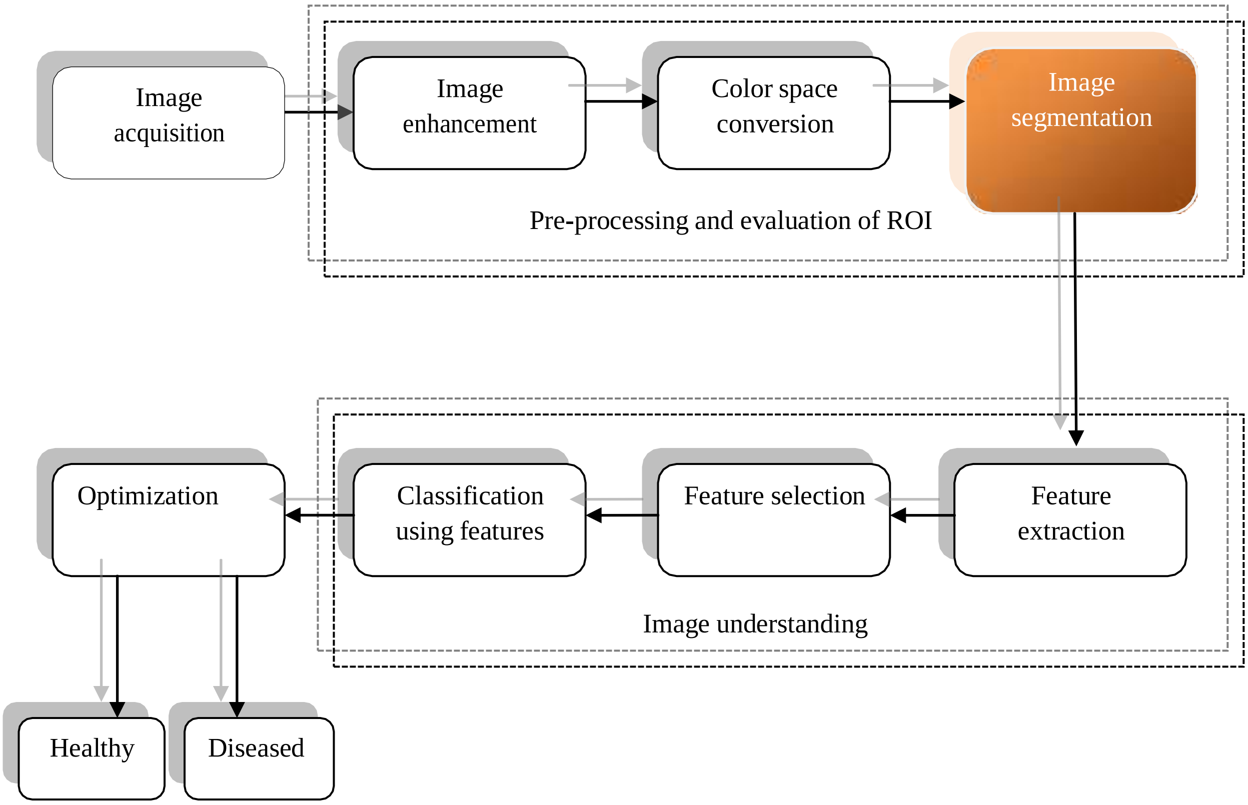


Figure 4.1: Highlighting steps of proposed segmentation system of system model

4.1 Introduction

Segmentation is a technique to divide an image into numerous segments according to various characteristics such as texture, color and intensity [108-109]. It helps to understand image representation in an easy way [110]. More precisely, segmentation allocates a label to each pixel of an image that shares visual characteristics. The accomplishment of image processing tasks depends on the strength of segmentation process.

However, the presence of distinct regions of texture, color, diversity, illumination and varying background makes process of segmentation quite challenging [111-113]. Mathematically, segmentation of an image (I) explained as spatial region (S) which is divided into z different sub-regions, S_1, S_2, S_3, ... S_z such that every region defines an image outline different from other regions.

- $\cup_i^z S_i = S$.
- S_i is a set of sub regions where $i = 1,2\ ,3$ ... .z.
- $S_i \cap S_j = \emptyset$, $\emptyset$ is a null set and for all i and j, where $i \neq j$.
- $S_i \cup S_j = false$, for adjacent regions of R_i and R_j.

For accurate segmentation, some points must be carefully considered as [114]

- Uniformity and homogeneity of regions with respect to image characteristics.
- The pixel values for the same segment should have same multivariate values and for different segments each pixel have dissimilar value.
- Boundaries of every segmented region should be clear.

Segmentation is an essential step in object representation, image analysis, image compression; image editing and others image processing applications [115-116]. The accomplishment of image analysis applications depends on consistency of segmentation. This chapter discusses the various existing segmentation techniques and proposed a new segmentation method. Various segmentation techniques are available in the literature for image processing applications. These approaches can be categorized as contextual or non-contextual [117]. Contextual approaches use homogeneity principle to segment an image e.g. region merging and splitting etc. This technique considers association among image features. Whereas, non-contextual ignore image features and group pixels based on some attributes e.g. thresholding techniques etc. On the concept of some global attribute pixels are simply grouped together. Various segmentation approaches are illustrated in Figure 4.2.

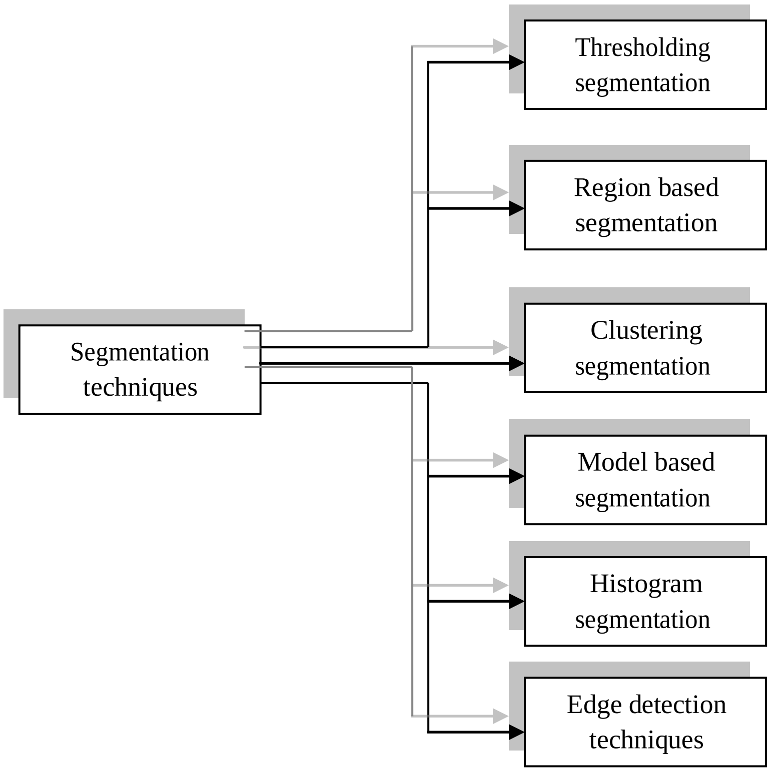


Figure 4.2: Segmentation techniques

4.1.1 Thresholding methods

It is one of the simplest and important segmentation techniques. The principle of thresholding is to separate pixels based on their intensity assessment [118]. Thresholding method is basically divided into two techniques as global thresholding and local thresholding.

(i) Global thresholding

This technique clearly recognizes intensity division among object and background. Using appropriate T_H , segmented thresholded image t (x, y) is expressed as

$$t(x,y) = \begin{cases} 1 \; if \; f(x,y) > T_H \\ 0 \; if \; f(x,y) \leq \; T_H \end{cases} \tag{4.1}$$

$f\ (x, y)$ represents grey level of pixels at co-ordinates (x, y) and (T_H) is threshold value For whole image, value of (T_H) will be same. In terms of pixel value, for the condition, $f\ (x, y) > T_H$, it is represents as object, whereas for $f\ (x, y) \leq T_H$, is considered as background.

(ii) Local thresholding

In local thresholding technique, the value of threshold is not same for whole image. For every pixel a threshold value is calculated using different parameters like variance and range etc. Mathematically, local thresholding technique defined as

$$b(x, y) = \begin{cases} 0, if\ f(x, y) \leq T(x, y) \\ 1\ otherwise \end{cases} \tag{4.2}$$

$f\ (x, y) \in [0,1]$ be the intensity of a pixel at location (x, y) of the image (I) and $b\ (x, y)$ defines binary image. $T(x, y)$ represents the thresholded values. For every pixel a threshold value is calculated using different parameters like variance and range etc. But noise affects performance of thresholding techniques. Thresholding using edge detection and image smoothing are two methods used to quash the consequence of noise.

4.1.2 Region based methods

Similarity or uniformity of regions is the main decisive factor for region growing technique [119]. It divides image into different regions based on continuity of grey level. It groups similar region of image to achieve a compact representation. Region based segmentation technique mainly divided into two categories as region growing method and region splitting and merges method.

(i) Region growing method

It is also known as pixel-based image segmentation technique, based on selection of primary seed points [120]. It also observes primary seed points of neighboring pixels. It examined relationship of neighboring pixels with region. For region growing, segment image R into small sub regions, R_e, e = 1,2,3 ... z, and satisfy subsequent situation.

$$R = \bigcup_{e=1}^{z} R_e\ ;\ R_1\ \cap\ R_j \tag{4.3}$$

$$U\ (R_e) = TRUE;\ e = 1,2,3, \ldots \ldots z \tag{4.4}$$

$$(R_e \cup R_j) = FALSE;\ e \neq j; \text{ Where, } R_e \text{ and } R_j \text{ are adjoining} \tag{4.5}$$

(ii) Region split and merges

Region split and merges efforts to segregate an image into homogeneous regions. Both techniques executed based on principle of quad tree data representation. Splitting defines iterative division of an image using same characteristics and merging combines neighboring similar regions [121]. Splitting is based on top down approach. Each region formed in splitting phase is coherent and varying in size. During merging, all coherent regions are grouped together to form a large region. However, main drawback of this technique is over-segmentation due to deviation of intensity.

4.1.3 Clustering methods

Process of segmenting image into different clusters based on their similar characteristics is known as clustering-based segmentation [122]. It groups cluster of similar objects in one place and dissimilar objects in another place. Clustering method is basically divided into two techniques as hard clustering and soft clustering.

(i) Hard clustering

It is a simple clustering method, where one pixel belongs to one cluster at a time. Membership function of hard clustering is either 0 or 1. K-means clustering is belonging to this type of clustering [123].

(ii) Soft clustering

Soft clustering segmentation technique is mainly helpful for segmentation in which belongings of pixel is not rigid [124]. In this type of clustering one pixel can belongs to one or more than one cluster at a time. Fuzzy c means clustering is falls into this category.

4.1.4 Model based segmentation

Model based segmentation requires specific knowledge about geometrical structure of object. It is helpful to segment object, which includes repetitive structure or geometry. It is mainly used for medical imaging applications. Segmentation using Markov Random Field (MRF) is an example of model-based segmentation and faster than existing segmentation methods [125].

4.1.5 Histogram based segmentation

Histogram segmentation technique computed histograms to choose grey level for combining pixels into regions. Histogram valleys and peaks are used to find regions in image. Entire image is divided into two parts as object and background. Background mainly contains most of the area of image and exhibit large peak of histogram. Object holds another gray level of image and exhibit smaller peak of the histogram. But main disadvantage of this method is difficult to evaluate important valleys and peaks of the image [126].

4.1.6 Edge detection segmentation

Edges are symbolized in an image as rapid inconsistency. Edge based segmentation technique rely on discontinuity of intensity value, which are supposed to signify object borders. Edge detection technique does not need any kind of prior knowledge of image. In this technique, first and second order derivatives implementation are used to find edges. After distinguishing edges, they all grouped together to evaluate object boundaries. Edge segmentation techniques are categorized into two groups as [127-128].

(i) Gradient edge detection

Gradient is used to estimate first order derivate of an image. Magnitude and direction of gradient is calculated as

$$|G| = \sqrt{[G_m^2 + G_n^2]} = \sqrt{\left[\left(\frac{\partial f}{\partial m}\right)^2 + \left(\frac{\partial f}{\partial n}\right)^2\right]} \tag{4.6}$$

$$D = \tan^{-1} \frac{G_y}{G_x} \tag{4.7}$$

Where, (G_x) and (G_y) are gradients respectively. Three different operators as Robert, Sobel and Prewitt are used to identify edges using first order derivative.

(ii) Laplacian edge detection

Edge points can be distinguished by evaluating zero crossings of second derivate of an image. But second derivate is very susceptible to noise. To filter out noise, Laplacian of Gaussian is used. $N * N$ mask is convoluting image with Gaussian filter to reduce noise.

$$\Delta^2 f(m,n) = \frac{\partial^2 f(m,n)}{\partial m^2} + \frac{\partial^2 f(m,n)}{\partial n^2}$$

(4.8)

These all approaches have their own advantages and disadvantages. Results of segmentation techniques are dependent on various aspects such as texture, image content and intensity. So, considering these all challenges, a novel leave segmentation technique is developed.

4.2 Image pre-processing and color space transformation

Before segmentation, one of the important tasks is image pre-processing to detect reflections and noise. It mainly comprises of changing a source image into an improved image which is different in certain aspects like contrast, illumination etc. Many factors such as environment, an acquisition equipment and storage device etc. ultimately affect the image eminence. Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm is used for image enhancement. CLAHE surmount the precincts of existing standard histogram equalization techniques such as over enhancement etc.

It works on small sections of the image, known as tiles, instead of whole image. CLAHE algorithm applies the histogram equalization after partition an image into contextual regions [129]. First, divide the image I into small non over-lapping areas of the same size and evaluates each region histogram [130]. Then, using preferred limit for contrast expansion, a clip limit for histogram is obtained. β is clip limit and is achieved by

$$\beta = \frac{MN}{L}\left(1 + \frac{\alpha}{100}(s_{max} - 1)\right)$$

(4.9)

Where, (α) defined as clip factor. If clip factor is equivalent to zero, clip limit turn into exactly equal to $\left(\frac{MN}{L}\right)$, if clip limit is equivalent to 100 the maximum allowable slope is (s_{max}). The resultant contrast limited histograms (I_C) are determined using greyscale mapping and clipping factor. To enhance image quality, two key parameters known as clip limit and block size are required. Higher value of clip limit results more contrast of processed or output image. Tile size of [8 8] is considered for proposed work.

After pre-processing, next step is to color transformation. We used color RGB and CIELab for segmentation. The characteristics of CIELab color space make it suitable for extracting color features, from a digital image. The Euclidian distance between two colors in CIELab color space is strongly correlated with human visual perception [131]. This space dramatically alters contrast of an image without affecting its color balance or saturation by adjusting a & b channels. L*a*b* color space is formulated from tristimulus values of XYZ color system.

$$L^* = 116\sqrt[3]{\left(\frac{Y}{Y_n}\right)} - 16 \tag{4.10}$$

$$a^* = 500\left(\sqrt[3]{\frac{X}{X_n}} - \sqrt[3]{\frac{Y}{Y_n}}\right) \tag{4.11}$$

$$b^* = 200\left(\sqrt[3]{\frac{Y}{Y_n}} - \sqrt[3]{\frac{Z}{Z_n}}\right) \tag{4.12}$$

Where, (w) signifies values of white points. X, Y and Z are the Tristimulus values of CIELab color space.

4.3 Proposed methodology

In this part, proposed segmentation framework for leave disease identification has been presented.

4.3.1 Proposed segmentation framework using neutrosophic logic

The goal of segmentation is to locate suspicious areas to diagnose disease. We have proposed new Neutrosophic logic approach as a segmentation technique. A neutrosophic set is an extended form of the fuzzy set, tautological set, dialetheist set, paradoxist set, intuitionistic set and paraconsistent set [132]. A pixel in the Neutrosophic logic domain is characterized as $P\ t, i, f,$ in the way as it is (t%) true, (i%) indeterminate and (f%) as false values. Where, these entire three elements T, I, F are independent to each other and real standard subset of $0^{-}1^{+}$. There is no restriction on the sum of T, F and I. (T), (I) and (F) are neutrosophic components respectively referring to neutrosophic set, neutrosophic logic, neutrosophic statistics and neutrosophic probability.

In literature, various fusion structure such as single valued neutrosophic rough set, rough neutrosophic set, bipolar neutrosophic set, single valued neutrosophic hesitant fuzzy set, etc. are proposed [133]. Neutrosophic proves as a more advantage technique than traditional theory in handling uncertainty [134]. Neutrosophic set has become an important tool in different areas as data mining, decision making, e-learning, engineering, medicine, social science etc. So, we are describing new neutrosophic theory for segmentation for identification of diseases.

For diseases analysis using new neutrosophic logic, diseased area of leave employed as the true part (T), healthy element represented as false part (F) and intermediate element (I) is defined as neither healthy (F) nor diseased (T). The neutrosophic domain provides extra element as "I" which provides a more efficient way to handle the degree of uncertainty. After enhancement and color transformation, T, I, F are mapped as follows:

Case (i): Initially, to evaluate diseased segment, consider input image as I_i (m, n). After contrast enhancement, represented as Ic (m, n) and diseased segment T_{IS}(m, n) is formularized as

$$T_{IS}\ (\mathrm{m}, \mathrm{n}) = \mathrm{I_c}\ (\mathrm{m}, \mathrm{n}) \times \mathrm{F_a}\ (\mathrm{m}, \mathrm{n}) \quad \textbf{(4.14)}$$

Where, F_a (m, n) is the binary mask obtained from (a*) chromaticity layer of CIELab color space and the color falls along the red-green axis. It performs bitwise multiplication. It separates the disease patches from different color populations. Diseases patches can be of dark brown, black and purple color. True section can eliminate healthy i.e. green section from the image

Case (ii): Healthy segment of leave is evaluated as

$$F_{IS}(m, n) = 1 - T_{IS}(m, n) \tag{4.15}$$

F_{IS} (m, n) represents a healthy segment of leaves. Healthy section represents the green color or section of leave image. Its case of affected leaves it considers green as well as yellow color.

Case (iii): Intermediate segment is considered as onset disease. In case of plants, starting of discoloration of leave indicates as on set diseases. Pale green, yellow and white yellow are the symptoms of onset disease.

These symptoms probably explain the plant or leave is not healthy exactly but not diseased as well. It is indication of disease stating that can damage plant if symptoms continuously growing. The main reasons of onset disease are chlorophyll breakdown, over watering, under watering, nitrogen, calcium and sulphur etc. [135]. To evaluate intermediate segment, symptoms of onset disease is to be evaluated. Firstly, create a mask by evaluating distance between colors M_y (m, n) and then subtract this mask from F_{IS} (m, n) that represents remaining portion of the leave where only remaining color portion of the image is considered.

$$I_{IS}(m, n) = F_{IS}(m, n) - M_y(m, n) \tag{4.16}$$

Finally, T_{IS} (m, n) represents the degree of being a diseased segment, F_{IS} (m, n) is the degree of being a healthy segment and I_{IS} (m, n) is a degree of being not healthy not diseased as well [136-137]. Figure 4.3 explains proposed segmentation model.

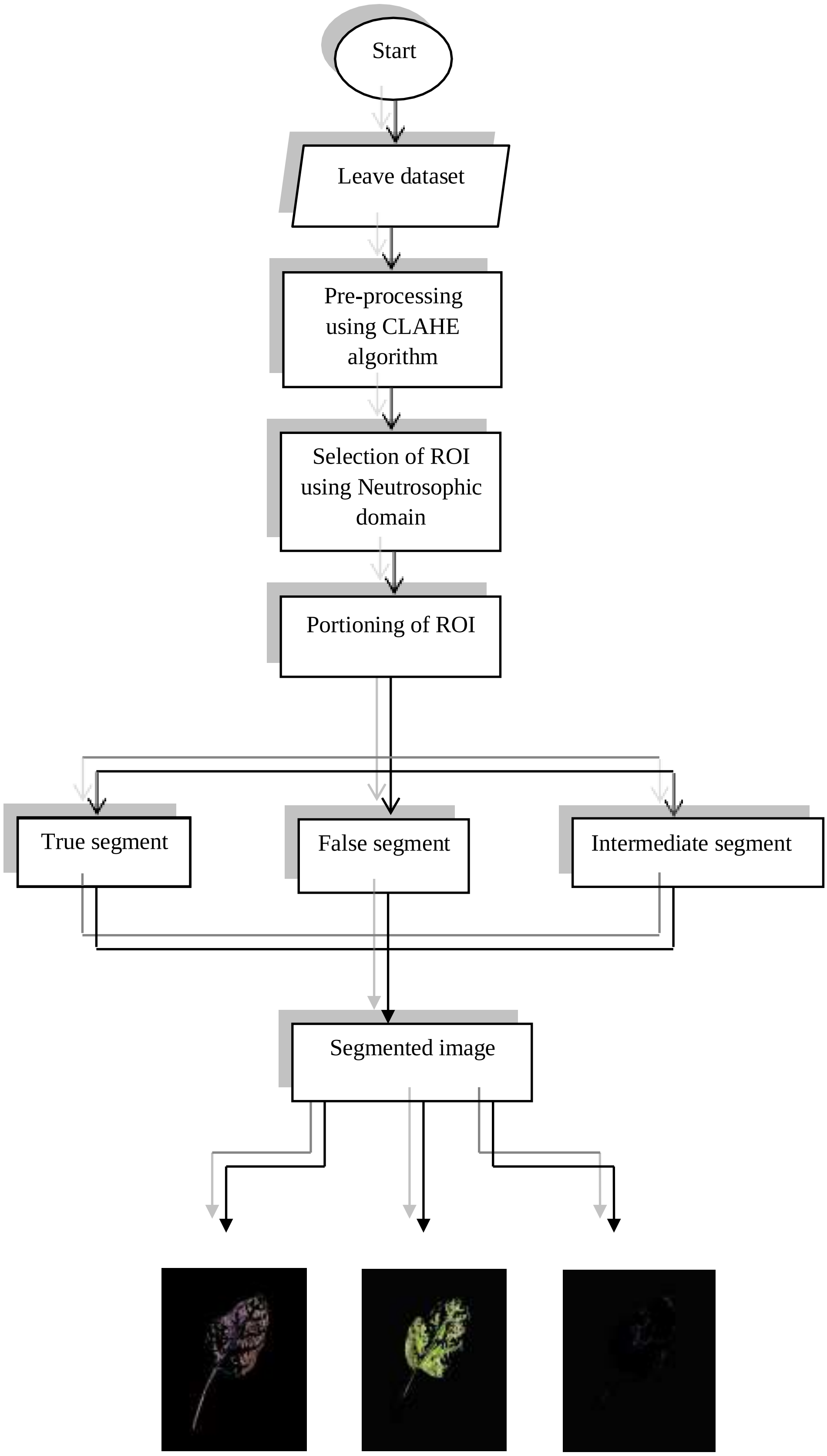


Figure 4.3: Proposed segmentation model

Color of leave plays a most dominating role to identify leave as healthy and diseases. Figure 4.4 displays the structure of color extraction during segmentation.

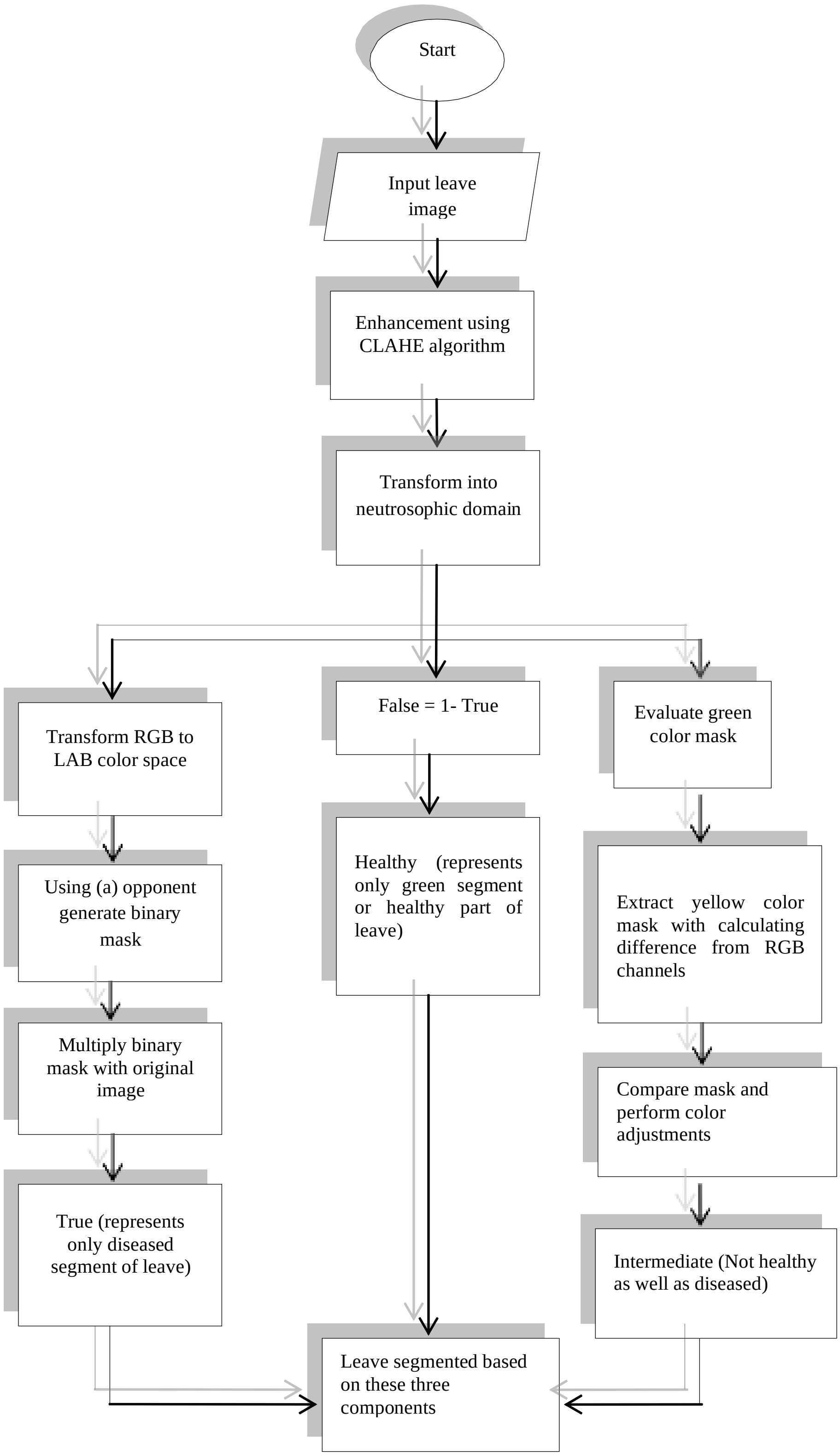


Figure 4.4: Color extraction method

Figure 4.5-4.7 represents the pictorial representation of extracted true, false and intermediate region of leave.

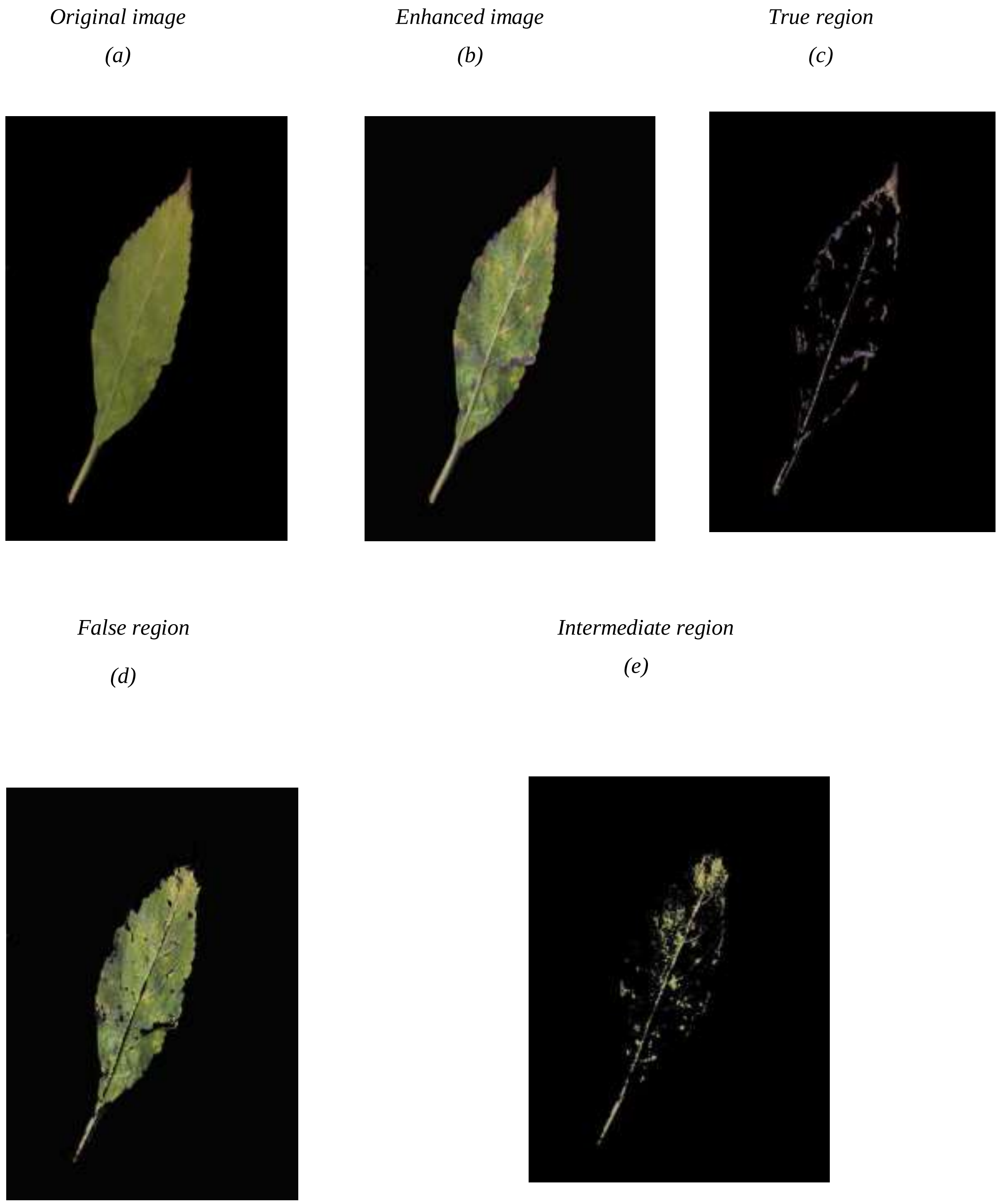


Figure 4.5: Segmented regions a) Represents original captured image, b) Pre-processing using CLAHE algorithm, c) True region which represents the diseases region, d) False region, where healthy region of leave is presented, e) Intermediate region of the leave

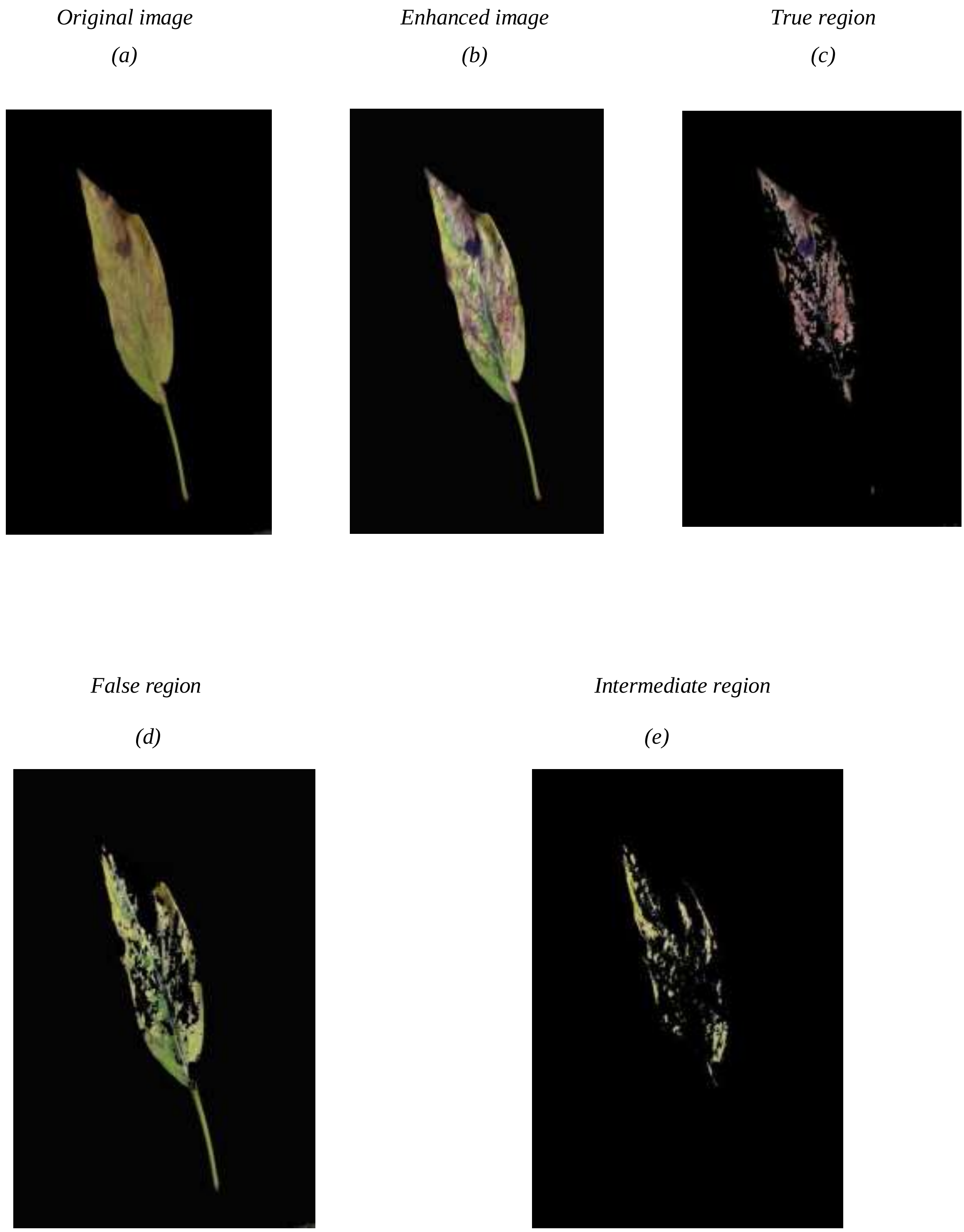


Figure 4.6 Segmented regions a) Represents original captured image, b) Pre-processing using CLAHE algorithm, c) True region which represents the diseases region, d) False region, where healthy region of leave is presented, e) Intermediate region of the leave

Original image

(a)

Enhanced image

(b)

True region

(c)

False region

(d)

Intermediate region

(e)

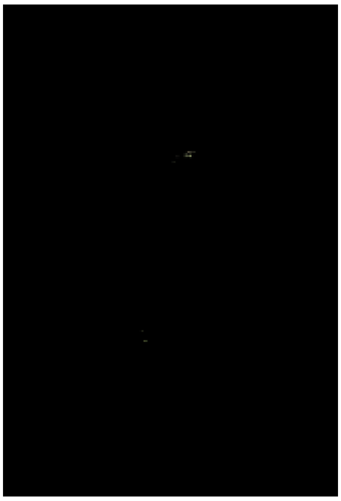


Figure 4.7 Segmented regions a) Represents original captured image, b) Pre-processing using CLAHE algorithm, c) True region which represents the diseases region, d) False region, where healthy region of leave is presented, e) Intermediate region of the leave

4. 4. Chapter Summary

In this section, we propose a new segmentation technique based on neutrosophic logic. The segmentation technique incorporates three different segments as true, false and intermediate. True segment defines the diseased region of image, false segment explains diseased element and intermediate defines the segment of leave that is not healthy as well as diseased. Based on these segments, further leave is classified as healthy or diseased. This algorithm is compared with existing segmentation algorithm, where proposed algorithm provides best results. This chapter contributes in the area of medicines and healthcare.

Chapter 5

Feature extraction analysis

In this chapter, various existing and new feature extraction techniques are presented. Features are extracted using color, texture and intensity relationship of segmented images. Using all extracted features, leaves are further classified as healthy or diseased. Section 5.1 explained role of features, existing feature extraction models and their characteristics. In Section 5.2, Amalgamation of color and texture feature extraction models is introduced. Implementation details and results are specified in Section 5.3. Figure 5.1 represents flow chart highlighting steps of proposed diseases detection and identification model.

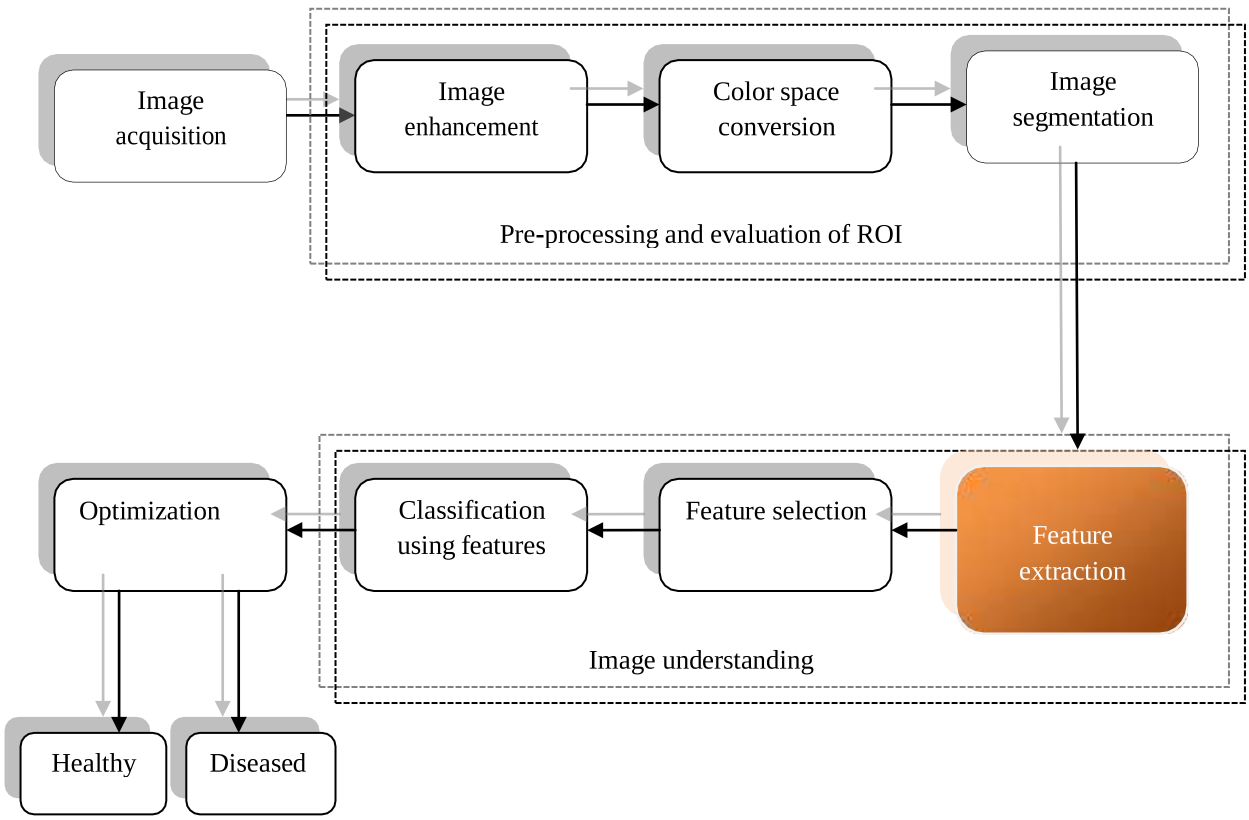


Figure 5.1: Highlighting steps of proposed feature extraction method of system model

5.1 Introduction

Features extraction is used to extract distinct characteristics or properties of an image [138-139]. They measure relevant and non-redundant information of image. It is also known as dimension reduction technique, where raw input data or information is abridged to more informative groups for efficient processing. The key responsibility of features extraction method is to convert image information into vector space [140]. While extracting features, some concerns must be carefully considered as [141].

- Features should be simple to compute.
- They should carry adequate information regarding the image for discrimination among the classes.
- They should have minimized the within class sequence variability and enhance variability between class sequence.
- Features should relate to human perceptual distinctiveness.
- Most desirable property for feature extraction is its repeatability.
- Abandon redundant information.
- No need to involve any kind of domain specific information.
- A local feature is an image pattern which differs from its immediate neighborhood. It is usually associated with a change of an image property or several properties, though it is not localized exactly on this change. The image properties considered are intensity, color, and texture.

In image processing and computer vision applications, features extraction techniques used for generalization of image information and further used this information for classification purposes. Feature extraction techniques further classified into two techniques as general features and domain specific features [142].

(i) General features

General features are defined as independent features such as color, texture and shape. They are further classified as pixel level features, global features and local features. In pixel level, all features are evaluated from each pixel location. Where, global features are measured from entire image. However, local features are evaluated from segment or subpart of an image.

(ii) Domain or application specific features

They evaluated specific application dependent features such as fingerprints and human faces etc.

5.1.1 Color features

Color is a dominant visual clue for information visualization [143]. For various image processing purposes like feature extraction, classification, segmentation, image matching, compression and object recognition color information plays an important role.

Color is a most important feature that human can distinguish when looking at an image. Color features define fundamental uniqueness content of an image. Color features are independent of orientation and size of an image and comparatively robust to complicated image background [144]. Color feature extraction methods mainly classified into five categories.

(i) Color histograms
(ii) Color moments
(iii) Color coherence vector
(iv) Color Correlogram
(v) Color descriptors

These methods extract mean, standard deviation and skewness of intensity of an image. For categorization and image retrieval, color histogram provides powerful information [145]. Histograms with more numbers of bins include additional information regarding image contents and reducing possibility of different colors being allocate to the similar bins.

So, change of color of leave section from original color is a crucial feature for disease recognition. Color of diseased segment varies for every different disease. Such color features are obtained by evaluating diseased spot color, non-diseased spot color corresponding to each plane individually.

5.1.2 Texture features

Texture is a repetitive outline of information of structure with usual intervals. It refers to properties and appearance of surface of the object [146]. Image texture provides information regarding spatial array of intensities of an image. Texture refers to characteristics of surface as coarse, smooth, fine, size, density and irregular.

It includes significant information regarding structural arrangement of structure and association with their surrounding neighborhood [147-148]. Hence, texture measure depends on dimension of neighborhood. Texture features are mainly categorized into four categories: i) Structural methods ii) Transform based methods iii) Model based methods iv) Statistical methods.

(i) Structural methods

Structure based texture analysis represents texture as composition of well-defined texture primitives such as regular parallel lines.

(ii) Transform based methods

Transform based methods represents texture using Fourier, wavelets and Gabor transforms and using spatial frequency characteristics of intensity variations.

(iii) Model based Methods

Model based methods such as Markov model and fractal models produce an empirical model based on weighted average of neighborhood pixel intensities of image.

(iv) Statistical methods

Statistical methods distinguish image texture according to non-deterministic characteristics of image that supervise associations between gray levels of an image such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Local Binary Pattern (LBP) features etc.

5.1.3 Shape features

Shape is a one of the significant and primitive features used to define image content description [149]. Shape descriptors mainly divided into two categories:

(i) Region based

Region based method utilize whole area of an object to analyze shape description.

(ii) Contour based

Contour based methods utilize boundary points of object to analyze shape.

Various shape parameters are: eccentricity, hole area ratio, elliptic variance, digital bending energy, solidity, centroid, circularity ratios, convexity, profiles, least inertia, Euler number and rectangularity [150-151]. But shape content depiction cannot be explained clearly because computing similarity among shapes is very difficult.

5.2 Analysis of combination of feature extraction techniques for proposed system model

After preprocessing and segmentation next step is to evaluate features for further classification of leaves. In this part, proposed feature extraction framework has been presented.

5.2.1 Need for combination of features

Under different experiment conditions, performance of features extraction techniques will be considerably different. Complex background, shadow, illumination factors and presence of other occlusion must be considered, while selecting efficient features. Usually color and texture are of most important concern in order to detect and identify diseases.

Two different kinds of information are extracted and analyzed from the leave images (i) Texture information (ii) Information regarding intensity. These features contain the considerable information about image. Figure 5.2 show flow diagram for combinational feature extraction module.

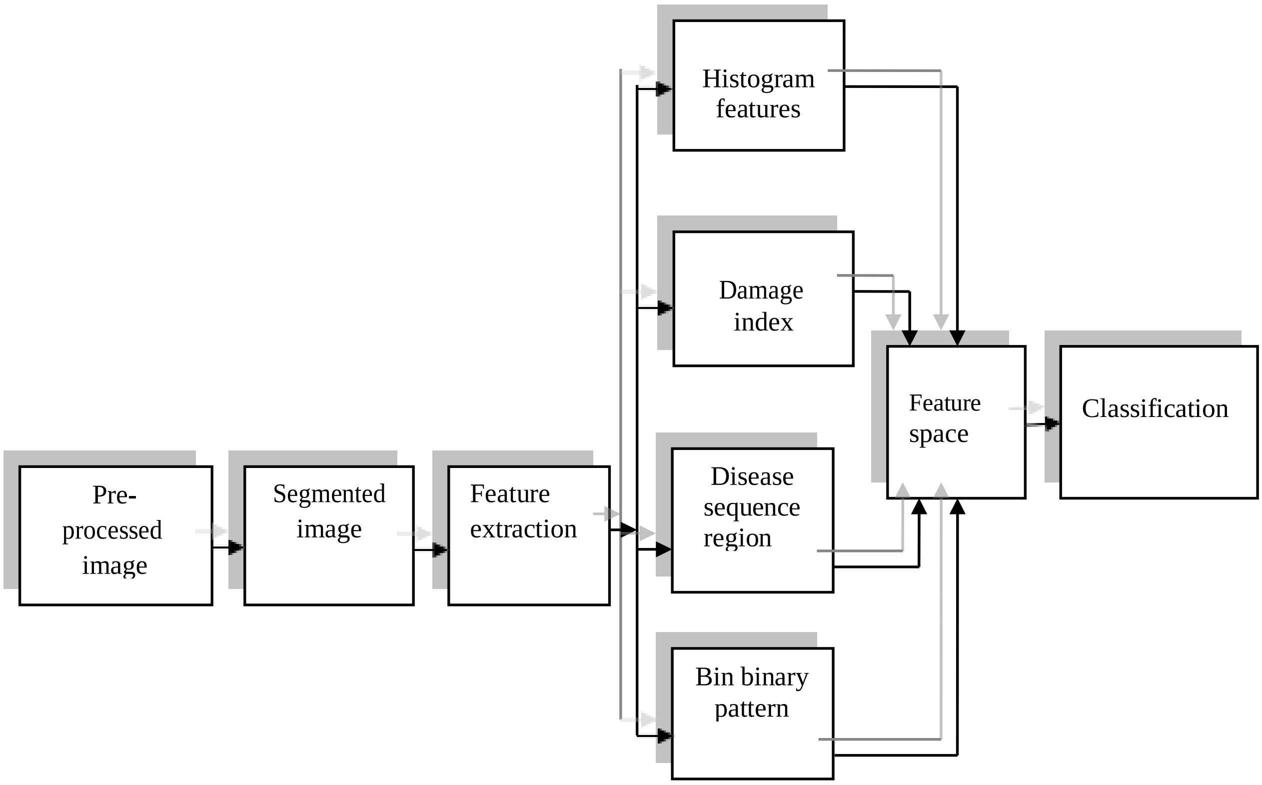


Figure 5.2: Proposed feature extraction model

Although using feature amalgamation, data dimensionality increases as well as amount of requisite training and testing data set increases exponentially. Moreover, correlation between diverse features may design an efficient feature vector and produced efficient results. The features extracted from image are histogram information content, damage index, disease sequence region and bin binary pattern.

5.2.2 Histogram Information Content (HIC)

Histogram is easy to calculate and effectual in describing both local and global allocation of colors in an image [152]. Histogram information content defines relative information content by finding the probability of occurrence of each plane of an image. Information will vary for every leave for each plane.

For true region:

$$\text{HIC (true)} = log\left(\frac{1}{histogram\ information\ contents\ of\ true\ region}\right) \tag{5.1}$$

For false region:

$$\text{HIC(false)} = log\left(\frac{1}{histogram\ information\ contents\ of\ false\ region}\right)$$

(5.2)

For intermediate region:

$$\text{HIC(intermediate)} = log\left(\frac{1}{histogram\ information\ contents\ of\ intermediate\ region}\right)$$

(5.3)

Histogram information content is evaluated for all three segments; true, false and intermediate for each red, green and blue plane.

5.2.3 Damage Index (DI)

The damage index defined as the amount of spaces taken by diseased segment of leave given as by

$$\text{DI} = \frac{\sum_{i=1}^{m}\sum_{i=1}^{n}(T_i+I_i)}{\sum_{i=1}^{m}\sum_{i=1}^{n}F_i}$$

(5.4)

Higher value of DI indicates more diseased region. DI represents possible presence of damage (disease) at leave structure.

5.2.4 Diseases Sequence Region (DSR)

Disease sequence region defines correlation of individual neighboring pixels with perceived pixel difference of an image, that is, pixel deviations between neighboring pixels. Theses pixels map then transformed into a single character value explaining difference in pixel interaction. We have calculated DSR for every extracted region (red, green and blue) of image vertically and horizontally. The DSR defined for vertical and horizontal orientation are given as:

Vertical deviation of intensity

$$\mathrm{DSR}_{(\mathrm{V})} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \left(P_{(x,y+\Delta y)} - P_{(x,y)} \right)$$

(5.5)

Horizontal deviation of intensity

$$\mathrm{DSR}_{(\mathrm{H})} = \sum_{\mathrm{x}=0}^{\mathrm{m}-1} \sum_{\mathrm{y}=0}^{\mathrm{n}-1} \left(\mathrm{P}_{(\mathrm{x}+\Delta\mathrm{x},\mathrm{y})} - \mathrm{P}_{(\mathrm{x},\mathrm{y})} \right)$$

(5.6)

$$\mathrm{DSR} = \left[\mathrm{DSR}_{(\mathrm{V})}\ \mathrm{DSR}_{(\mathrm{H})} \right]$$

(5.7)

(x) and (y) defines pixel location and (m) and (n) defines size of segmented image. Depending on horizontal and vertical deviations, we measure the deviation difference between healthy and non-healthy leaves.

5.2.5 Bin Binary Pattern (BBP)

A new texture descriptor as BBP is introduced to describe local structure information of leave. The BBP linearly interpolates pixel value of neighborhood to form an operator to define structure for distinguishing all individual patterns (healthy or non-healthy leaves). BBP determines local structure information around every pixel. To make it computationally simple, three separate planes (red, green, blue) are considered and the histogram is created for the same. The histograms are then split into 9 bins and mapped to (3 × 3) matrix to evaluate its mean intensity values and calculate the difference between the center pixel and neighboring pixels defined as

$$\mathrm{BBP}_{(\mathrm{k,R})}(\mathrm{x}_c, \mathrm{y}_c) = \overset{\mathrm{k}}{\sum\nolimits_{\mathrm{k}=0}^{\mathrm{k}}} (\mathrm{G}_\mathrm{k} - \mathrm{G}_c) 2^{\mathrm{k}}$$

(5.8)

Where, (G_K) represents center value, (G_C) defines neighborhood pixels, (x_C) and (y_C) represents pixel value and (k), defines number of pixels in the neighborhood. Finally, derived decimal number is referred as BBP value. Figure 5.3 depicts BBP structure step by step.

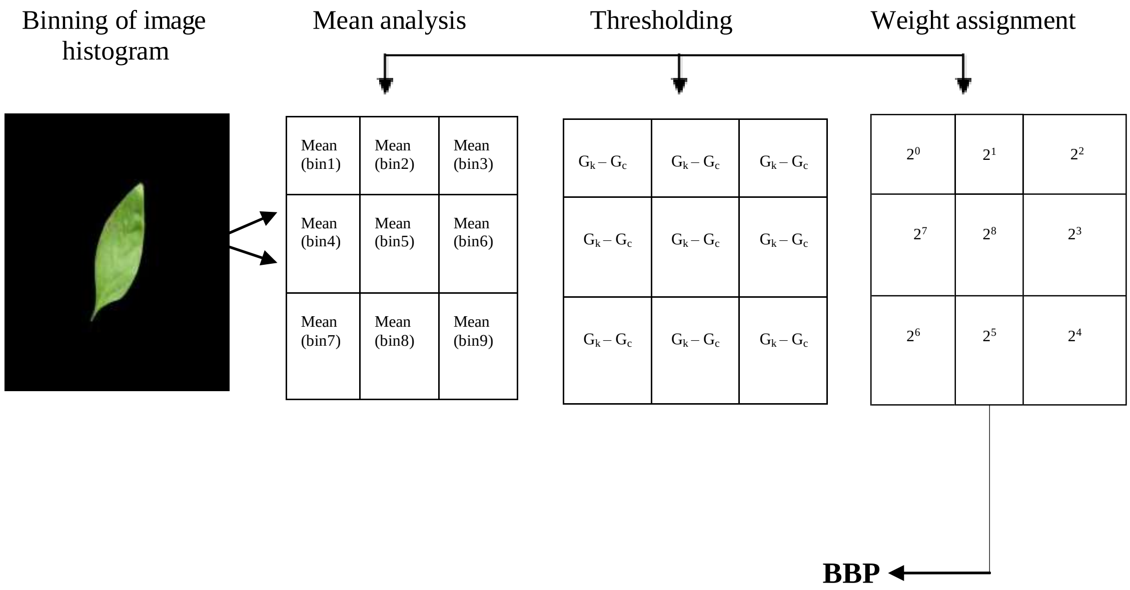


Figure 5.3: Bin binary pattern

The weights obtained for different binary pattern are given in a clockwise direction starting from top-left and its corresponding values. BBP represents as a local structure descriptor to analysis texture of an image. Algorithm 1 specifies working procedure of BBP.

Algorithm 1 Bin binary pattern

Input:

Input leave image data set

Set of true, false and intermediate regions Cluster

range of red, green and blue region

Output:

Set of unique decimal values represents the local structure information.

Step 1*: Calculate true, false and intermediate using neutrosophic segmentation*

Step 2: *Divide image into 9 different bins using histogram w.r.t to red, green and blue plane*

Step 3: *Calculate mean of all cluster bins.*

Step 4*: Evaluate the difference of all neighborhood bins w.r.t to center value using equation* ***(5.8).***

Step 5*: Assign weights*

Step 6: *Obtain unique decimal values*

Step 7: *end;*

5.3 Illustration of experimental results

The results are evaluated for all said criteria. The database consists of 400 images which includes 200 healthy and 200 diseased leaves of different categories of leaves (i.e. Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (Holy basil) and Ocimum gratissimum (Vana-holy basil). A hundred samples of four classes of leaves are collected. The disease of leave samples investigated is downy mildew, aphids, gray mold, bacterial leave spot and Fusarium wilt.

5.3.1 Classification of healthy and diseased leave using machine learning models

An efficient technique to categorize objects based on their significant difference and similarity is known as classification. We used nine classifiers; ANN [60], Naive Bayes [68], SVM [89], discriminant analysis [91], Adaboost [93], decision tree [153], RF [154], linear model [155] and K-NN [156] to evaluate effectiveness of proposed framework of features and segmentation. The summary of classification models is discussed in chapter 6. On the basis of parameters, classifiers will categorize leave as healthy or disease. Machine learning methods as are implemented using R open source software. R software is licensed under GNU GPL. The tuning parameters of machine learning methods are tabulated in Table 5.1.

Table 5.1: Tuning parameters of machine learning algorithm

Platform	Model	Method	Package	Tuning Parameter(s)
R (version 3.2.2)	Artificial Neural Networks	Neural net	nnet	Hlayers=10, MaxNWts=10000, maxit=100
	Naive Bayes	NBModel	e1071	No. of observation=150
	Support Vector Machine	SVM	e1071, kernlab	Kernel Radial Basis
	AdaBoost	Ada	Ada	Max Depth=30, Min Split=20, xval=10,
	Discriminant Analysis	LDA	mass	No. of observation=150, DiscrimType: “linear”
	Decision Tree	Rpart	rpart	Min Split=20, Max Depth=30, Min Bucket
	Random forest	RF	randomForest	Number of trees=500
	LM	Lm	Glm	Multinomial
	KNN	knn_model	caret	NumObservations=150, Distance= “euclidean”, NumNeighbours:5

5.3.2 Feature discriminant test (Training-testing)

The effectiveness of new segmentation technique and proposed features model is observed in this section. A set of color and texture features are used as an input to classification model. To compute the performance, various evaluation parameters; sensitivity, specificity, accuracy, error, cost curve and region of curve are calculated as discussed in chapter 2. The distribution of data in the training-testing experiment is set to 70% and 30% respectively for all models.

5.3.3 Classification results

An experiment result illustrates comparison analysis of various classifiers with respect to evaluation parameters as represent in Figures 5.4-5.7.

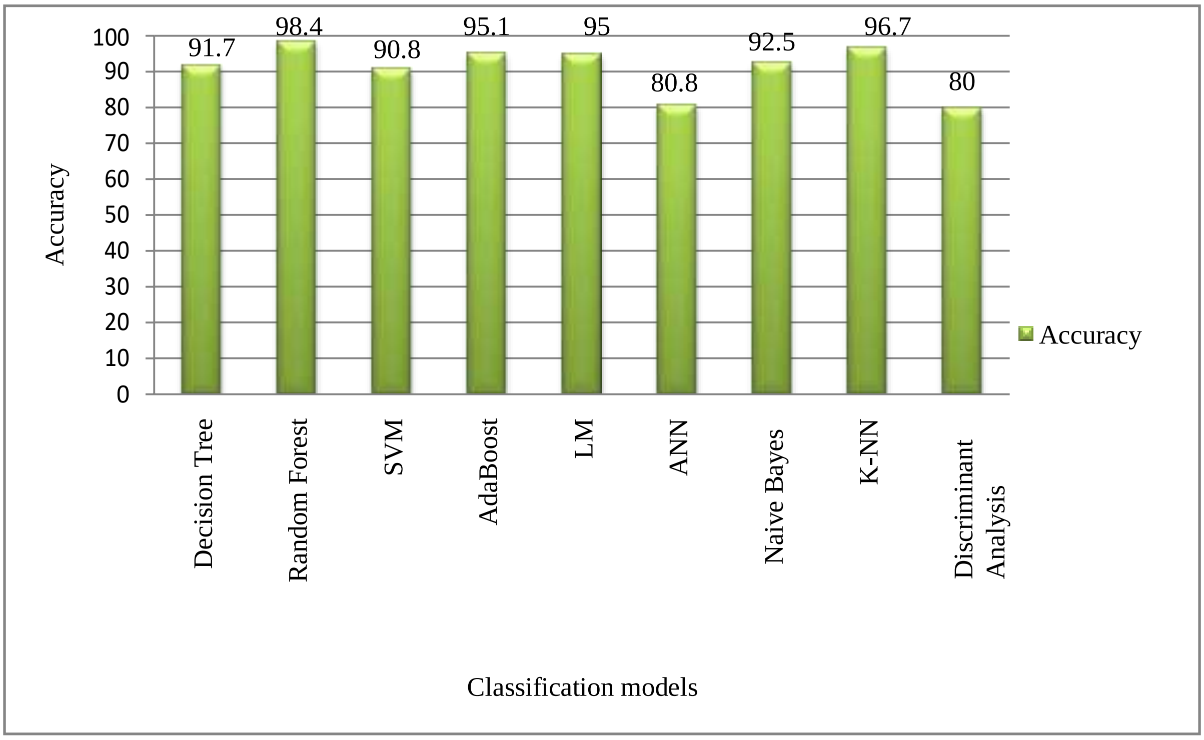


Figure 5.4: Comparative study of accuracy corresponding to classification results using new segmentation and feature extracting methodology

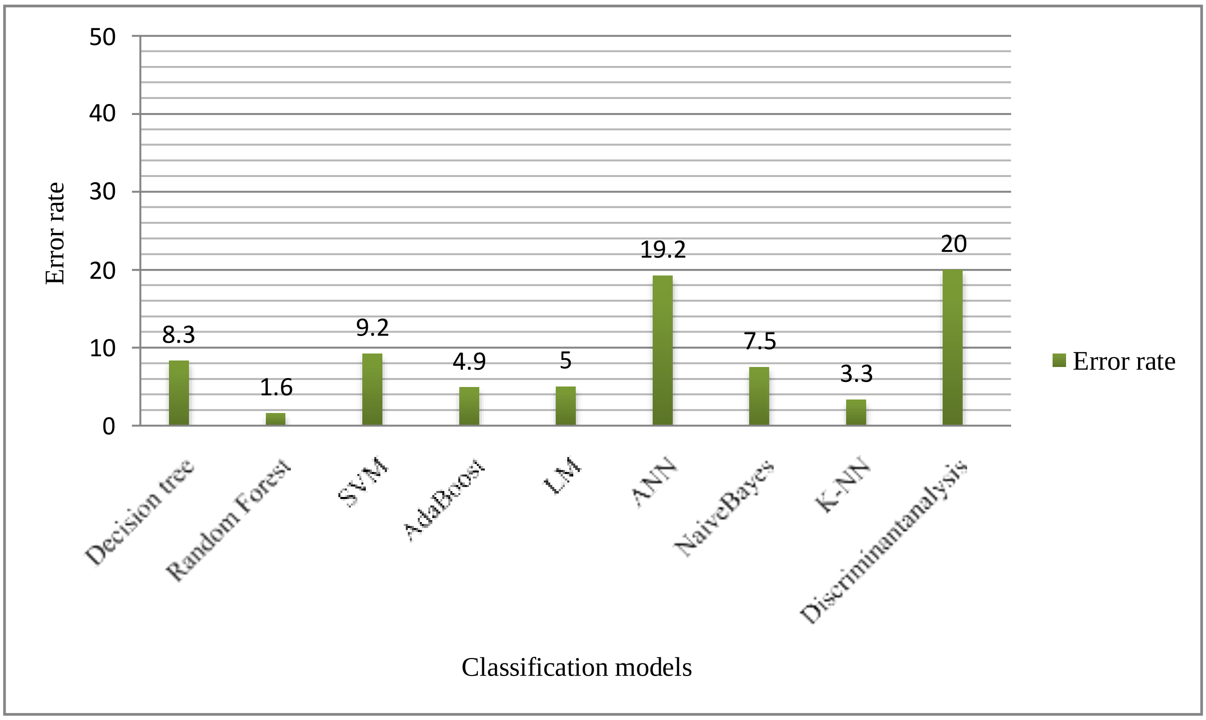


Figure 5.5: Comparative study of error rate corresponding to classification results using new segmentation and feature extracting methodology

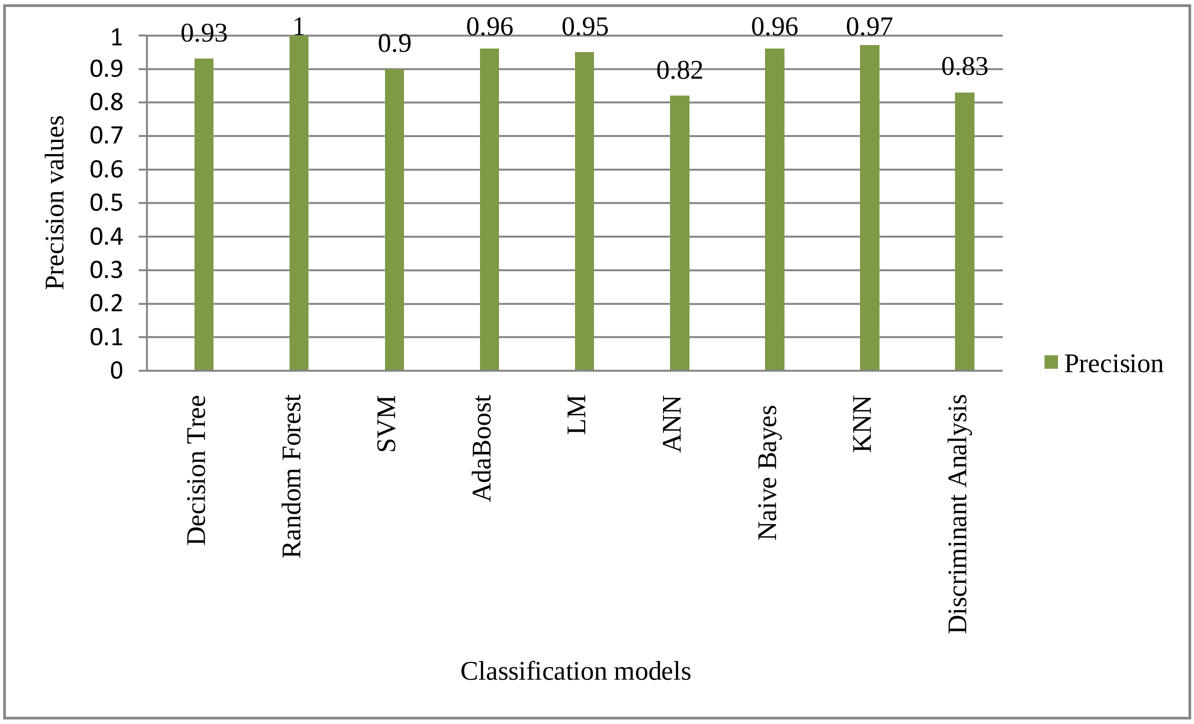


Figure 5.6: Precision graph curve corresponding to classification results using new segmentation and feature extracting methodology

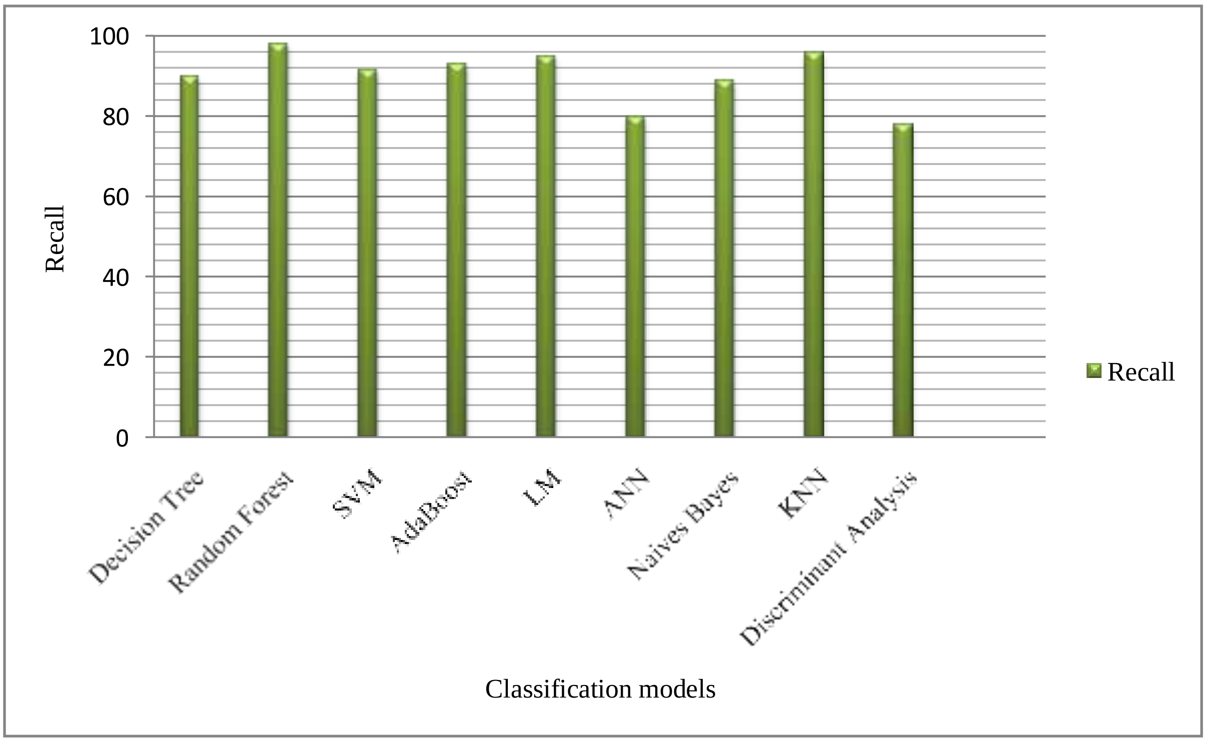


Figure 5.7: Recall graph curve corresponding to classification results using new segmentation and feature extracting methodology

As shown in above figures, compared to other machine learning models, random forest maintains high accuracy with respect to 98.4% accuracy and 1.6% error rate. Every time, new training set is drawn with substitute from original training set and after that a new tree is grown up on the new training set using random feature selection. More number of trees provides high accuracy results. RF makes an internal unbiased approximation of the error as the forest edifice processes. It may also estimate missing data. Random forest mainly works on principle of divide and conquers approach. It employs both random feature selection and bagging techniques. Figure 5.8, shows sensitivity and specificity values correspond to RF model across the entire data set. RF model exhibit highest sensitivity, that points towards minimal misclassification. Figure 5.9 illustrates relation between predicted v/s observed values. It represents the relationship among RF model prediction results and real results.

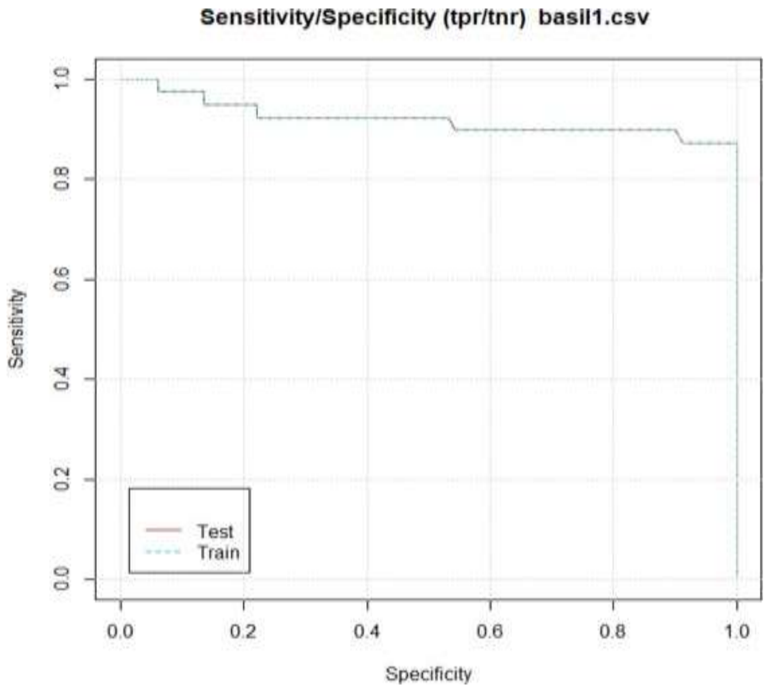


Figure 5.8: Sensitivity and specificity graph curve corresponding to RF model

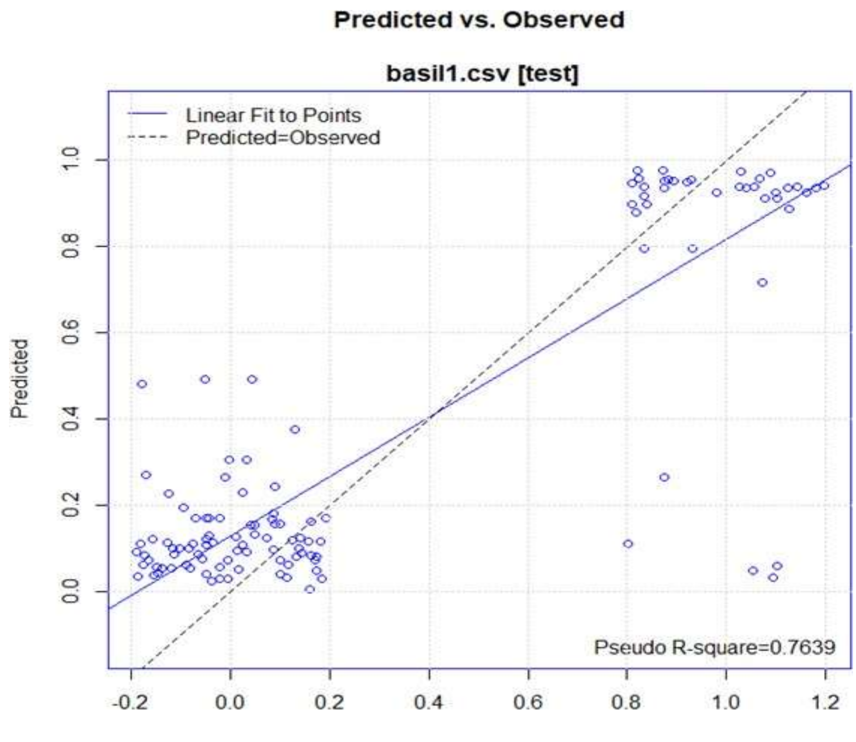


Figure 5.9: Predicted v/s observed graph curve corresponding to RF model

5.3.4 Receiver Operating Characteristics (ROC)

Receiver Operating Characteristics offers ability to access the performance of classifier. It defines distinguishing capability of classifier between diseased and healthy classes. The comparative results among sensitivity and 1 – specificity (FPR) can be assessed by receiver operating curve [157]. In these four subplots ROC, AUC, H measure and class score distribution curve as a performance indicator is shown in Figure 5.10. For ROC, AUC is measured in form of metric. Higher value of AUC corresponds to accurate classification results. H measure is another performance measure that indicates the misclassification rate of classification model.

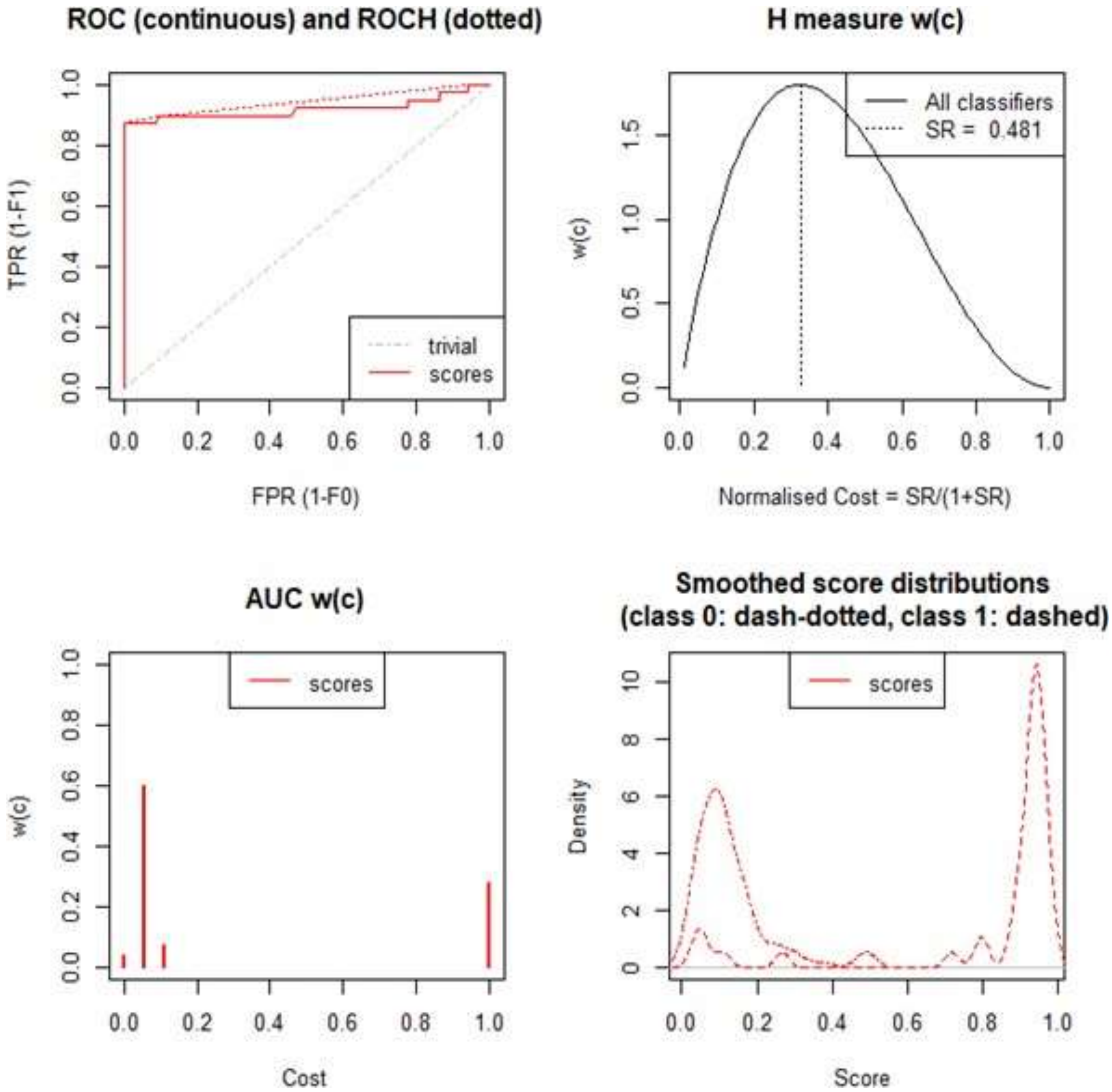


Figure 5.10: Predicted ROC and AUC curve corresponds to RF model

5.3.5 Cost curve

Figure 5.11 graphically represents cost curve of RF model. It visualizes the effectiveness of model over misclassification costs on the basis of predicted false positive values and false negative values.

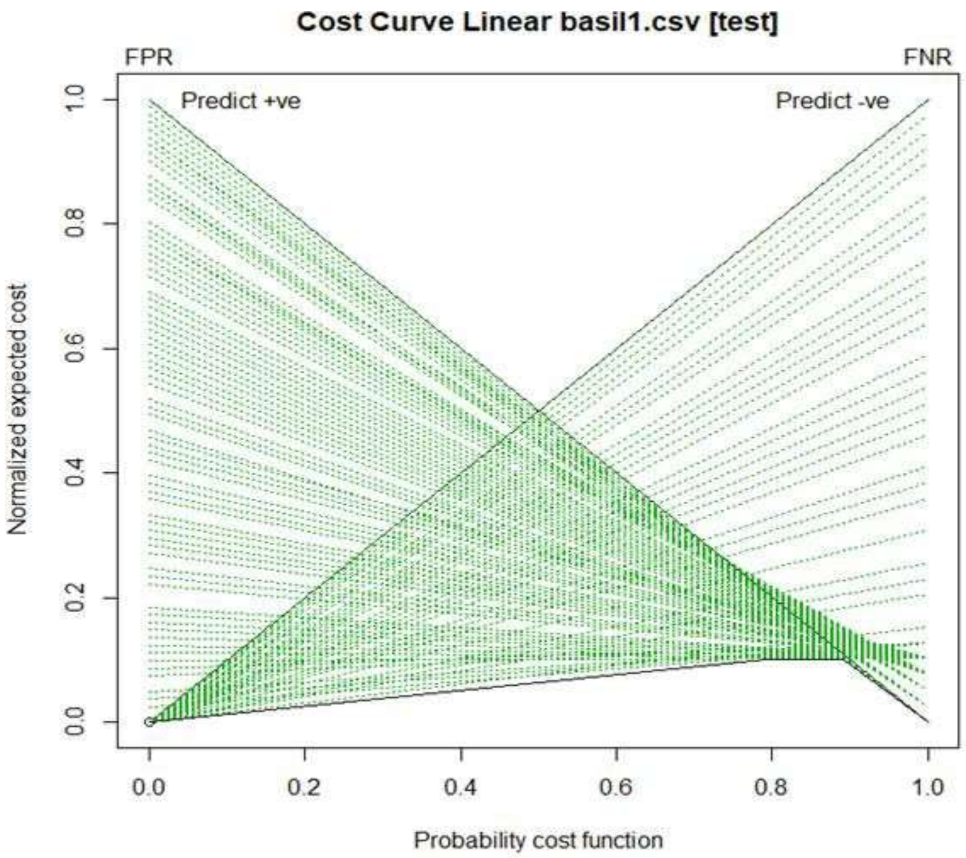


Figure 5.11: Predicted cost curve corresponds to RF model

5.3.6 Classifier performance

To check uniformity of our technique on new unobserved data, we ensure all experiments across an entire series of training and testing splits, as 50-50 (50% of data is used for training and 50% is used for testing), 60-40 (60% of data is used for training and 40% is used for testing), 70-30 (70% of data is used for training and 30% is used for testing), 80-20 (80% of data is used for training and 20% is used for testing), 40-60 (40% of the whole dataset used for training, and 60% for testing), 30-70 (30% of the whole dataset used for training, and 70% for testing) and finally 20-80 (20% of the whole dataset used for training, and 80% for testing). Results show that RF performs stable in all testing-training partition. Figure 5.12 shows performance analysis of RF model on different training testing partition.

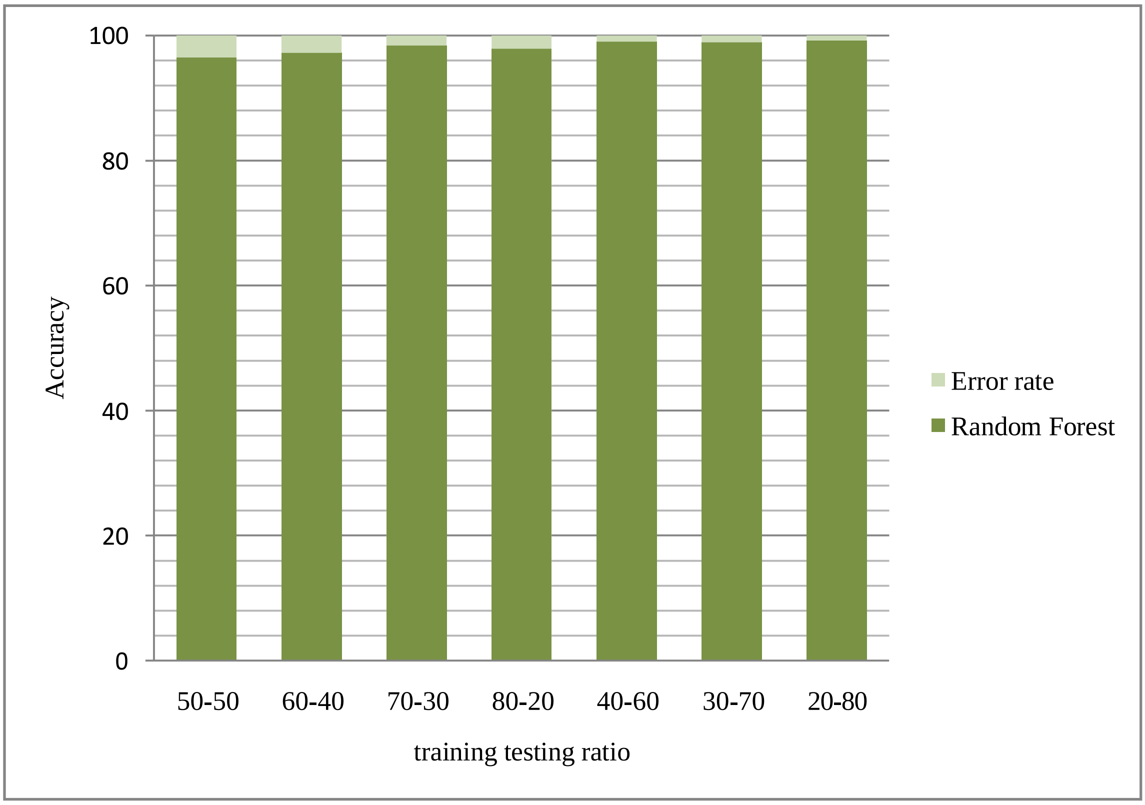


Figure 5.12: Performance comparison on different testing–training partition

5.3.7 Comparison analysis

Table 5.2 demonstrates comparison analysis of existing techniques with proposed one. Combination of color and texture features together enhanced accuracy in comparison to one kind of feature set alone. The proposed method is compared with existing three techniques. The existing method considered 40 images [27] and intensity texture features are considered. In [158], for 150 images, features are extracted using statistical feature matrix. Law texture features are evaluated on MIAS data base [159]. Proposed feature extraction technique performs well as compare to others.

Table 5.2: Comparison analysis of new techniques with existing techniques

Classifiers	Proposed Method (Accuracy)	CCM features [27] (Accuracy)	SFM Features [158] (Accuracy)	Law Features [159] (Accuracy)
Decision tree	91.8%	88.9%	91.3%	77.5%
Naive Bayes	92.31%	85.9%	76.1%	90%
KNN	96.92	70.4%	89.13%	72.5%
SVM	90.8%	77.8%	89.13%	92.5%
RF	98.4%	94.6%	91.6%	91.9%
AdaBoost	95.03%	94.6%	91.9%	91.9%
ANN	80.33%	93%	78.4%	96%
Discriminant analysis	80%	81.5%	82.6%	80%
GLM	95.03%	90%	89.9%	89.9%

5.4 Chapter summary

The whole procedure was described, respectively, evaluating feature extraction from segmenting regions and finally classification. These features combine the discrimination power of intensity and texture of leaves. Nine different classification models have been applied on extracted features. Based on the graphical analysis, RF performs better than other machine learning models with 98.4% accuracy. The developed model was able to distinguish healthy and diseases leave successfully. An effective combination of different features has been presented in this chapter. Presented chapter contributes in the area of healthcare and medicines, which plays a significant role in daily life routine.

Chapter 6

Classification

In this section, a new classification model using survival of fittest approach is outlined. This algorithm is used to classify healthy and diseased basil leave samples. A three-level hierarchical approach is developed to recognize leave. In Section 6.1, various existing classification techniques are discussed. Section 6.2 explains segmentation, feature extraction and feature selection techniques used for this study. Experiments are performed by utilizing combination of texture and color features. Random Forest is used to attain high informative features. A new classification model using survival of fittest approach with successive generation of best results is proposed in Section 6.3. Experimental results and conclusion is discussed in Section 6.4. Figure 6.1 represents flow chart highlighting steps of proposed diseases detection and classification model.

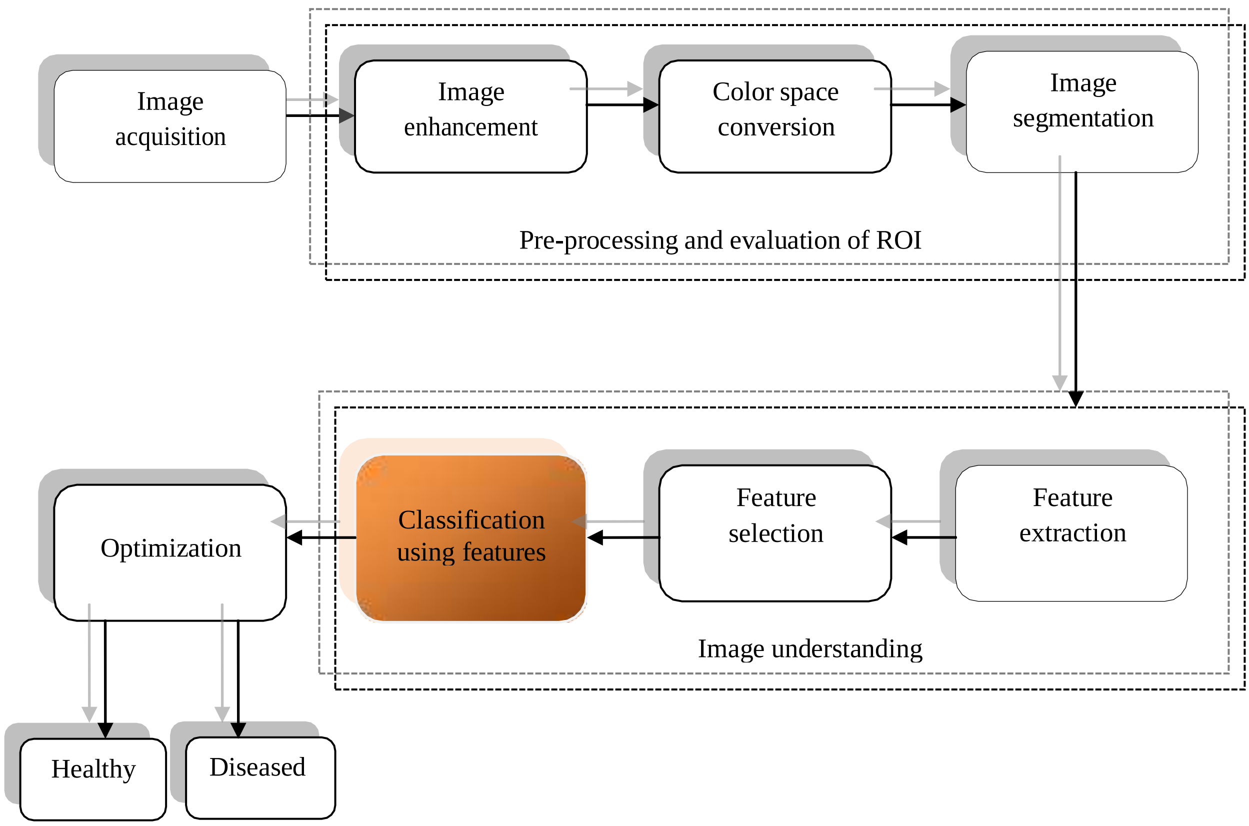


Figure 6.1: Highlighting steps of classification proposed system model

6.1 Introduction

Classification is a decision–theoretic technique used to organize data into categories. Category decision is depending on numerical characteristics of different image features expressed as

$$f\ (o) = o \rightarrow \Delta;\ o \in Z^u \tag{6.1}$$

Where, $\Delta = c_1, c_2, c_3, c_4, \ldots \ldots c_u$, $(c_1, c_2, c_{3 \ldots\ldots} c_u)$ defines numbers of class and f is a function conveying a pixel vector to a single class in the set of Δ classes. Classifier behaves as a discriminant, which prioritize one class against others [160-161]. In case of two class classification

$$b\ (c_u, o) = b\ (c_1, o),\ \mathrm{I} \neq \mathrm{u},\ \mathrm{o} \in \mathrm{c_u} \tag{6.2}$$

In case of multiclass classification

$$b\ (c_u, o) > 0,\ o \in c_{1;}\ b(cu, o) < 0,\ o \in c_{2;} \tag{6.3}$$

$b\ (c_u, o)$ defined as a discriminant function, (o) describes feature vector and (c_u) defines different classes. Classification process contains two phases; training and testing. In training phase, distinctive explanation of every arrangement, features and labels are created. To predict results for unknown dataset is known as the testing phase.

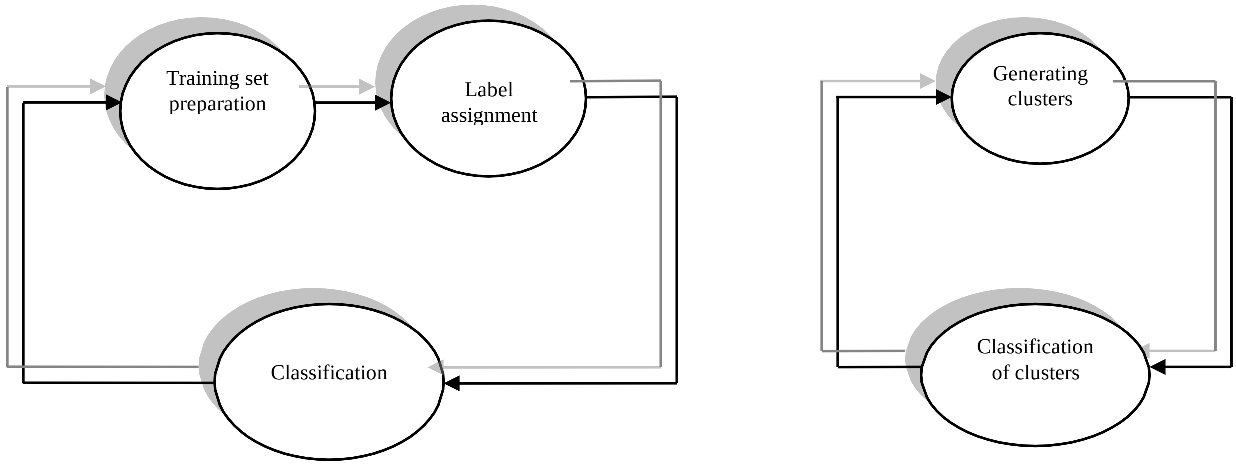
Classification broadly classified into two categories as supervised classification and unsupervised classification [162-163].

(i) Supervised classification

In supervised classification all data information is labelled and anticipates output with respect to input. It considers priori information of images. Supervised learning algorithm estimates the significance of a target variable using numerous input variables.

(ii) Un- Supervised classification

Unsupervised classification does not consider any kind of priori information. Data information is not labelled. It automatically discovers patterns and associations in the dataset by constructing clusters in it. Figure 6.2 briefly illustrated types of classification model [164].



(a) (b)

Figure 6.2: Classification methods a) Supervised classification b) Un-supervised classification

There are numerous techniques that have been originated from supervised and un-supervised classification used for diseases identification and classifications are presented in Figure 6.3.

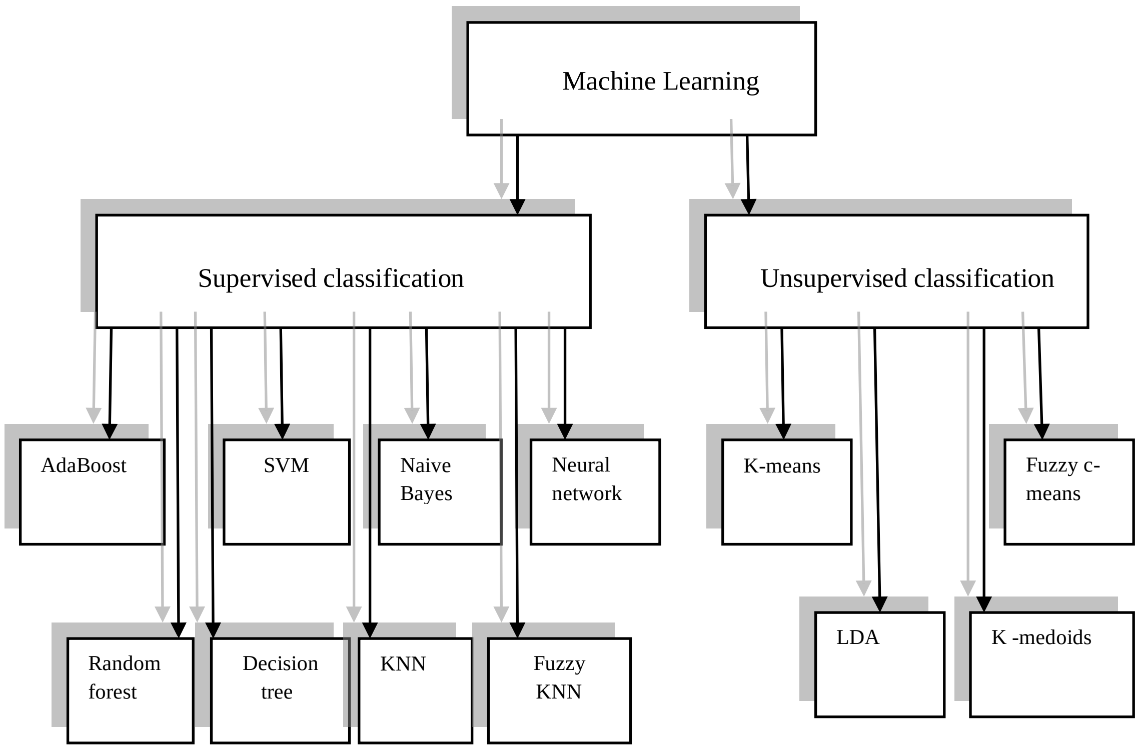


Figure 6.3: Learning prototype

(i) Decision tree

Decision tree works for mutually continuous and categorical input and output variables. It basically represents tree of decision for any computational task. It represents as a tree structure, where each internal node defined as attribute; every branch signifies possible outputs, and every terminal node embrace a class label [165]. The cost function for decision tree for classification is described as

$$g = sum\ (ck^{*}\ (1 - ck)) \tag{6.4}$$

Where, (ck) defines the ratio of class inputs of a specific group. A perfect decision structure made when a group conclude all inputs from the same class. Most important measure are Ginni index and entropy for decision tree.

(ii) Support vector machine

Supervised learning algorithm used to define hyper plane that categorize data points. It acts as a decision boundary for different classes. The vectors who characterize the hyper plane are known as support vectors [166]. For better classification, plane should have maximum margin or distance between data points. There are two types of SVM classifiers represents as 1) Linear Support Vector Machine 2) Non-linear Support Vector Machine. Linear SVM calculate straight hyper plane to divide classes with maximum margin and define maximum distance between classes [167]. But, in case of nonlinear SVM, data is separated using different tricks to maximize margin. Some benchmark kernel functions used are

- Radial Basis Function Kernel
- Polynomial(non-homogeneous) Kernel
- Polynomial (homogeneous) Kernel

Using SVM, non-Linear data can also be categorized with help of customized hyper planes built by using different kernel trick.

(iii) K-Nearest neighbor

K-NN is an instance-based learning algorithm, where data is classified based on different similarity measures [168]. A class is distinguished by majority vote of its neighbors.

K-NN is also a non-parametric algorithm which is mainly dividing into two forms, one is known as structure based NN techniques and other one is structure less NN techniques. Structure based NN technique depends on ball tree, axis tree, central line, nearest future line tree and orthogonal structure tree data structure. Entire data is divided into training and testing data. In structure less NN techniques lowest distance is calculated between sample data neighbors using similarity measures. K-NN classification is mainly performed on continuous data and its performance is determined by selection of (K) and distance metrics. Mostly Euclidean distance is used to measure the distance between neighbors. K-NN model accuracy is depending on the test instance.

(iv) Random forest

It is an ensemble learning method, works using bagging method to construct a group of decision tree using random subgroup of the data [169]. It mainly works on the principle of divide and conquers approach. Where, all weak classifiers combine together and make a strong classifier. Random Forest adds up randomness to the learning model for growing tree and explores best features between random subset of features.

(v) Naive Bayes

Classification algorithm based on Bayes theorem, used to distinguish diverse data classes using certain features [170]. It assumes that features are independent and characteristics or presence of one particular feature does not influence other features. Where, Bayes theorem is defined as

$$P(A|C) = \frac{P(C|A)P(A)}{P(C)} \qquad \textbf{(6.5)}$$

Where (P_r), defines the posterior probability, $P_r(C)$ predicts the prior probability, $P_r(A)$ measures the class prior probability and $P_r(C|A)$ is the likelihood of the probability of the predicted class. This classifier is mainly suitable, where the high dimensional inputs are used. It is also known as independence Bayes.

(vi) Linear discriminant analysis

Linear discriminant analysis classifier is a simplification of Fisher's linear discriminant method to evaluate linear grouping of features that separates entire data set into two or more classes [171]**.** LDA learns only linear boundaries and efficiently handle the data, where within class frequencies are not equal and performances has been checked on randomly produced data. To achieve maximal separability, LDA maximizes proportion of between-class variance to the with-in class variance. LDA is mainly derived from probabilistic model, which maximizes conditional probability of the model and derived from Bayes rule.

(vii) Gaussian process

Non-parametric classification method based on the principle of Bayesian methodology. It mainly focuses on the representation of the posterior probabilities with respect to smoothness properties [172]. It considers one latent variable as (f_i) for the object pattern (i). For two class separation for (c_1) and (c_2), (f_i) defines the degree of membership of particular class (c_i) and consider certain parameters as

i) (f_i) is positive→ object pattern fit into class c_1

ii) (f_i) is negative→ object pattern fit into class c_2

iii) (f_i) is zero→ class membership is less definite

For class (c_1), the posterior probability is expressed as and $(y_i = 1)$

$$P(c_1|x_i) = P\left(y_i = \frac{1}{f_i}\right) = \sigma(f_i) \tag{6.6}$$

$$= \int_{-\infty}^{f_i} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} \tag{6.7}$$

In the case of second class $y_i = -1$.

(viii) Neural networks

Neural networks are biologically-inspired programming theory which allows a computer to learn from speculative data. Neural networks are composed of highly interconnected structure known as neurons for working in alliance to resolve specific issues. They understand neurological data with help of machine perception like human brain processing [173-174]. Neural networks are trained by exemplar. They are not programmed to execute a specific task. They hold some learning rules which change the weights of connecting structure according to input patterns. Patterns commune with hidden layers with weighting methodology.

(ix) Boosting methods

Boosting also known as ensemble method used to enhance the classification model predictions by altering weak classifier to strong classifier. The motivation behind boosting is to first train weak classifier, then creating a second classifier that correct the miscalculation of first classifier [175]. The process of training weak classifiers continues until the error rate reduces minimally. AdaBoost, gradient boosting and XG boosting are boosting inspired ensemble methods. Boosting methods mainly reduce variance and enhance the robustness of classification model.

(ix) Bagging methods

Bagging method considers identical weak classifiers (decision tree), trained them autonomously in parallel and finally combines them subsequently using averaging process [176]. Compilation of dividing data set is used to train their decision trees. It is also known as bootstrap aggregation method. This technique is used to avoid over-fitting and reduce variance while maintaining bias. Average of misclassification errors provides a better estimation of predictive ability of learning model.

6.2 Image pre-processing and segmentation

Before classification, one of the important tasks is image pre-processing and segmentation. CLAHE algorithm is used to enhance the contrast of image as discussed earlier in Chapter 4.

6.2.1 Image segmentation (Mapping of healthy and diseases regions based on pixel color identification)

To segment image, original image RGB image is transformed to CIELab color space for better color perception as compared to the standard RGB space. After enhancement and color transformation, healthy and diseased regions are segmented based on pixel color identification as follows:

Case (i): Initially, to evaluate diseased segment, consider input image as $I_i(x, y)$. After contrast enhancement image is represented as $I_c(x, y)$ and diseased segment $D_{Segment}(x, y)$ is formularized as

$$D_{Segment}\,(x,y) = I_C\,(x,y) \times F_a\,(x,y)$$

(6.4)

$F_a(x, y)$ is binary mask evaluated from chromaticity layer (a^*) . Diseases pattern contains brown, dark purple and black color information. Different color pattern or clusters can be evaluated by calculating distance between colors. Figure 6.4 represent the structure of segmentation process.

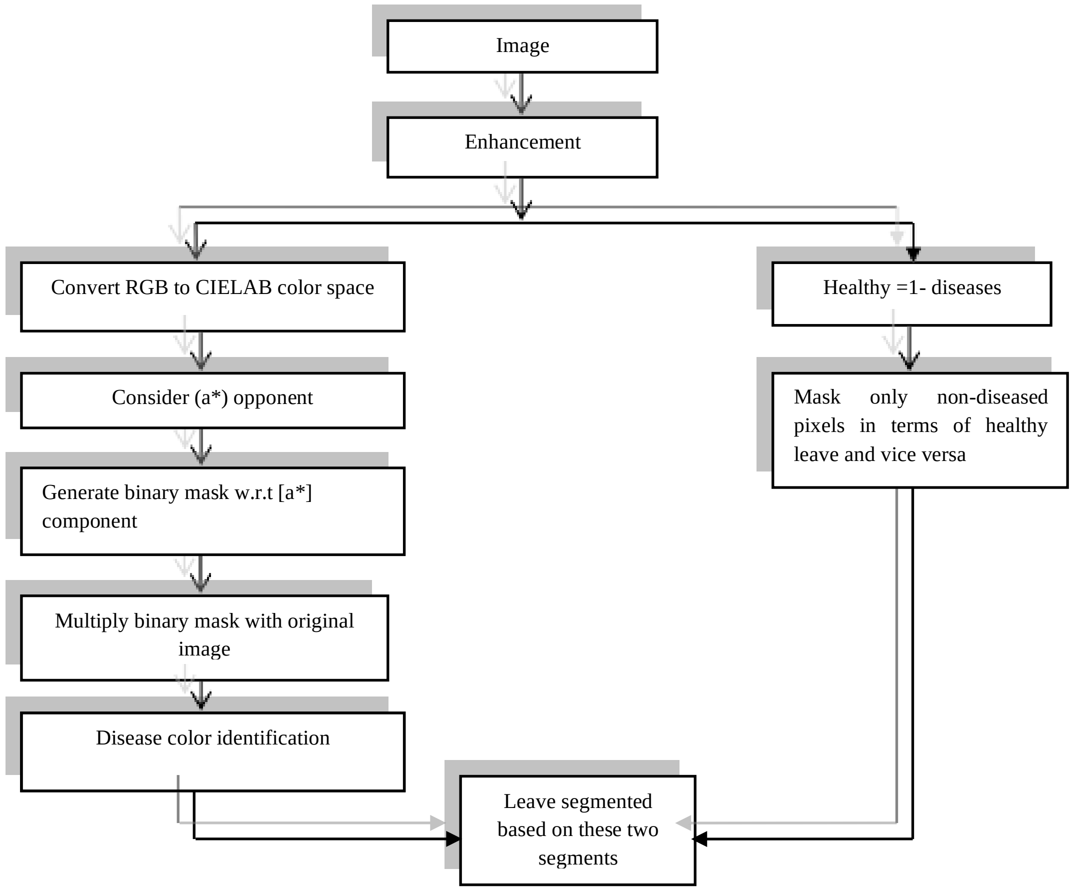


Figure 6.4: Segmentation structure

Case (ii): The Healthy segment of leave is estimated as

$$Healthy_{Segment}\,(x, y) = 1 - D_{Segment}\,(x, y)$$

(6.5)

Where, Healthy $_{Segment}$ (x, y) represents masked green pixels of image in terms of healthy green leaves. But in case of diseased leaves, it represents green and yellow color information.

Every pixel is classified as one of the two categories, which is spot pixel (diseases) and leave pixel. Every disease contains variations of colors with respect to their RGB coordinates combinations. Figure 6.5 represents the classified pixels based on different color values.

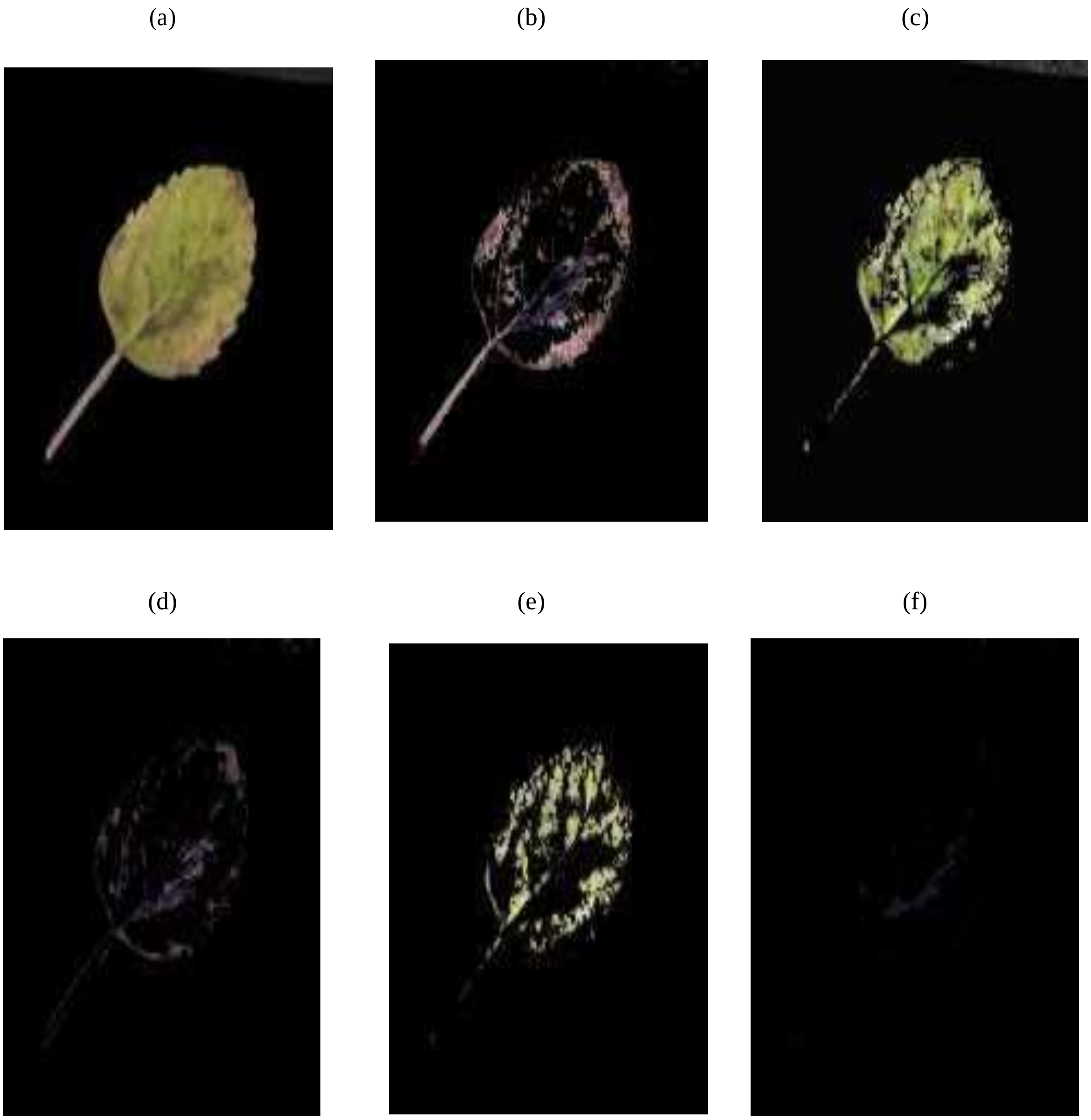


Figure 6.5: Represents the classified pixel based on the color values for visual testing (a) Original image b) Processed diseased image, (c), (d), (e) and (f) defined different diseases color clusters

6.3 Feature extraction

After segmentation, next important task is to extract useful features in order to identify and classify disease. Color and texture features are extracted to represent disease regions. Both of these features define dominating properties of the diseased region.

6.3.1 Texture features

Texture is a repetitive outline of information of the structure with usual intervals. It refers to the properties and appearance of the surface of the object. To extract texture features, Gray Level Co-occurrence Matrices (GLCM) is generated for each color pixel map of the image. The GLCM is based on the repeated occurrence of some gray-level configuration in the texture [177]. This method $P\ (x,\ y\ |d,\ \theta)$ measures occurrence of gray levels $'x',\ 'y'$ between a specific position $P\ (x,\ y)$ in the image and a neighboring pixel according to a given distance d and direction (θ). Number of columns and rows are identical to the quantization stages of the image. GLCM defines the occurrence of gray level emerging in a particular spatial linear correlation with a different gray level. Figure 6.6 represent GLCM matrix with respect to the various directions [176].

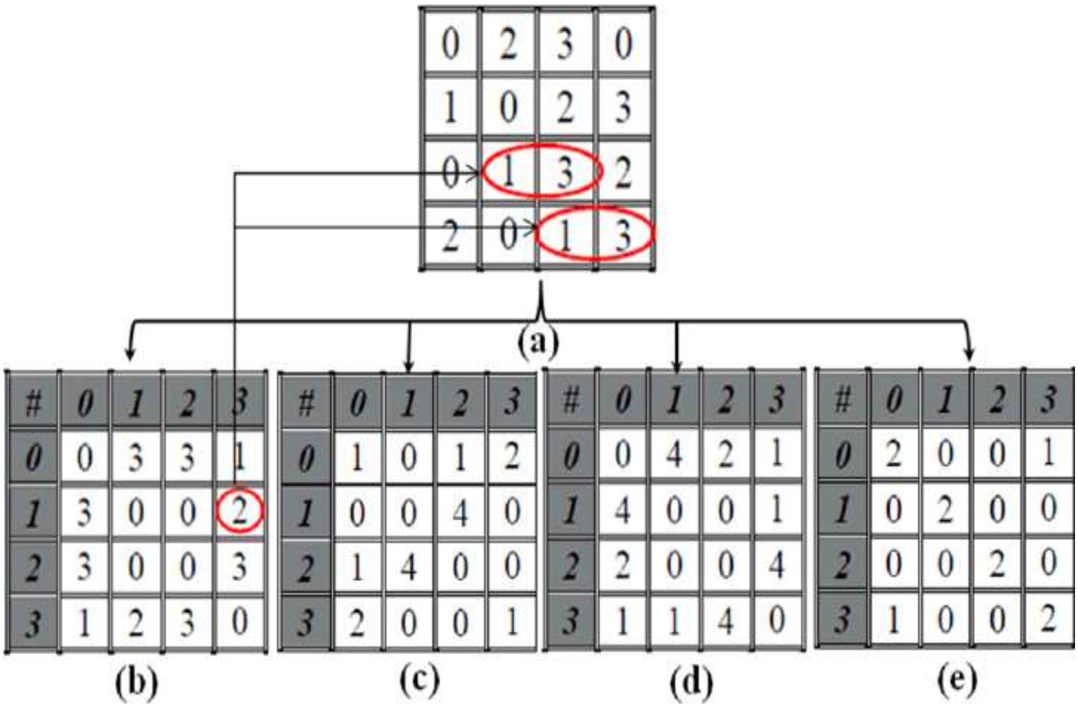


Figure 6.6: GLCM representations with respect to various directions [176]

The neighboring pixels can be defined in eight different directions generally fours directions as [0^{θ}, 45^{θ}, 90^{θ}, 135^{θ}] and in reverse directions also consider. It exhibits information regarding the location of the pixels containing same gray level information values. The features extracted from an image using GLCM are energy, entropy, homogeneity and correlation.

(i) Energy

Energy measures number of repeated pairs and also measures uniformity of the normalized matrix. It is also known as angular second moment. It detects disorder in the texture of image. Maximum value of energy is 1.

$$\text{Energy} = \sum_{x,y=0}^{n-1} {P_{xy}}^2$$

(6.6)

(ii) Entropy

Entropy evaluates the uncertainty of intensity distribution and it became large when the image textural uniformity is not there. More complex image texture has more entropy. Entropy is inversely correlated with energy.

$$\text{Entropy} = \sum_{x,y=0}^{n-1} -\ln(\mathrm{P_{xy}})\, \mathrm{P_{xy}}$$

(6.7)

(iii) Contrast

Contrast is defined as the amount of local variations in the image. Higher the values of contrast mean sharper the variations in the image.

$$\text{Contrast} = \sum_{x,y=0}^{n-1} P_{xy(x-y)^2}$$

(6.8)

(iv) Correlation

It is the difference between the highest and the lowest values of a neighboring set of pixels. It defines the linear dependence of grey levels on neighboring pixels.

$$\text{Correlation} = \sum_{x,y=0}^{n-1} P_{xy} \frac{(x-\mu)(y-\mu)}{\sigma^2}$$

(6.9)

(P_{xy}) define normalized GLCM elements of segmented image, (μ) represents mean of pixels and (σ) defines variance of the intensities of all reference pixels.

6.3.2 Color features

Color is a most important feature that humans distinguish when looking at an image. [179]. Change of the color of leave segment from the original color is a crucial feature for disease recognition. Color in diseased segment varies for every different disease. Disease index and average brightness information are considered as color features. Such color features are obtained by evaluating diseased spot color, non-diseased spot color and change of spot color with respect to the background considering each plane individually.

(i) Disease Index (DI)

The disease index can be defined as the amount of space taken by diseased segment given as by

$$\text{DI} = \frac{\sum_{i=1}^{m} \sum_{i=1}^{n} (\text{diseased segment})}{\sum_{i=1}^{m} \sum_{i=1}^{n} I_i}$$

(6.10)

Where, diseased index represents pixels count of diseased area, (I_i) represents pixel count of whole leave.

(i) Average brightness information

Average brightness information defined as the intensity value of every pixel in the image and average brightness information of RGB values of segmented diseased region.

$$\text{Average brightness information} = 0.299r + 0.587g + 0.114b$$

(6.11)

Where, (r), (g) and (b) define the mean of RGB values contained in the diseased segment. Based on texture and color analysis 56 features are extracted.

6.4 Feature importance

Before classification of leaves, removing redundant features helps to improve the classification model accuracy. In the proposed work, Random Forest (RF) model is used for feature selection task. Random forest is a method of bagging, and averaging over decision trees can substantially reduce instability [180]. Using Ginni index and mean decrease value, RF ranks the important features. In random forest, the impurity decrease from each feature can be averaged, and features are ranked according to this measure as described in Table 6.1.

Table 6.1: Features importance

Features	Mean decrease value	Ginni index value
F18	6.04	0.83
F52	5.81	0.74
F13	5.33	0.25
F6	5.38	0.32
F19	5.02	0.55
F45	4.76	0.49
F9	5.14	0.32
F10	4.25	0.33
F21	6.18	0.71
F53	5.06	0.59
F3	4.61	0.23
F56	4.15	0.40
F51	4.03	0.42
F48	4.51	0.46
F55	4.32	0.45
F14	3.45	0.26
F49	4.71	0.44
F47	4.73	0.54
F54	5.69	0.63
F11	3.14	0.22
F12	3.32	0.17
F15	2.54	0.15
F20	2.80	0.16
F17	3.08	0.16
F23	3.67	0.28
F5	2.34	0.14
F24	2.41	0.10
F2	1.42	0.05
F27	2.70	0.20
F1	1.67	0.07
F29	1.37	0.77
F28	2.56	0.21
F50	1.15	0.09
F31	2.03	0.11
F43	1.81	0.04
F35	3.12	0.09
F4	0.53	0.02
F30	1.86	0.06
F39	2.73	0.09
F22	3.05	0.11
F38	3.30	0.13

Features	Mean decrease value	Ginni index value
F8	1.80	0.07
F7	1	0.02
F34	2.14	0.06
F32	1.64	0.08
F36	2.42	0.13
F40	2.83	0.09
F42	1	0.02
F26	1.83	0.08
F33	2.51	0.20
F37	3.27	0.35
F41	1.67	0.26

Best features as selected for further classification process. Least important features are not considered for training the model. So, based on RF model most relevant features of feature vector are selected to improve the performance of proposed optimization algorithm.

6.5 Proposed classification methodology

In this part, proposed classification framework for leave disease classification has been presented.

6.5.1 Proposed classification framework using survival of fittest approach

Researchers have been investigating to devise new and efficient way of solving classification and optimization problems as the traditional methods are characterized by ineffectiveness and inadequacy. So, a new optimization algorithm or classification technique using survival of fittest approach illustrated in Figure 6.7.

The training features and the labels are given as input to the classifier and a leave diseases category achieved as output by the classifiers. Based on the important features, proposed classifier target class is mentioned below:

$$\text{Classifier}_{\text{leaves}} \sim f\left(F_{\text{features (new selected)}}\right) \quad \textbf{(6.12)}$$

In survival of the fittest data can be tested to determine the best performing members of any set, as the fittest remains and the weak becomes extinct. The fitness function rounds out the evolutionary system by testing all data and eliminates the poor performers. For example, it is a well-known fact that the animals that are better at finding food, generally live a longer life and have more children in comparison with the animals that are not as good at finding food.

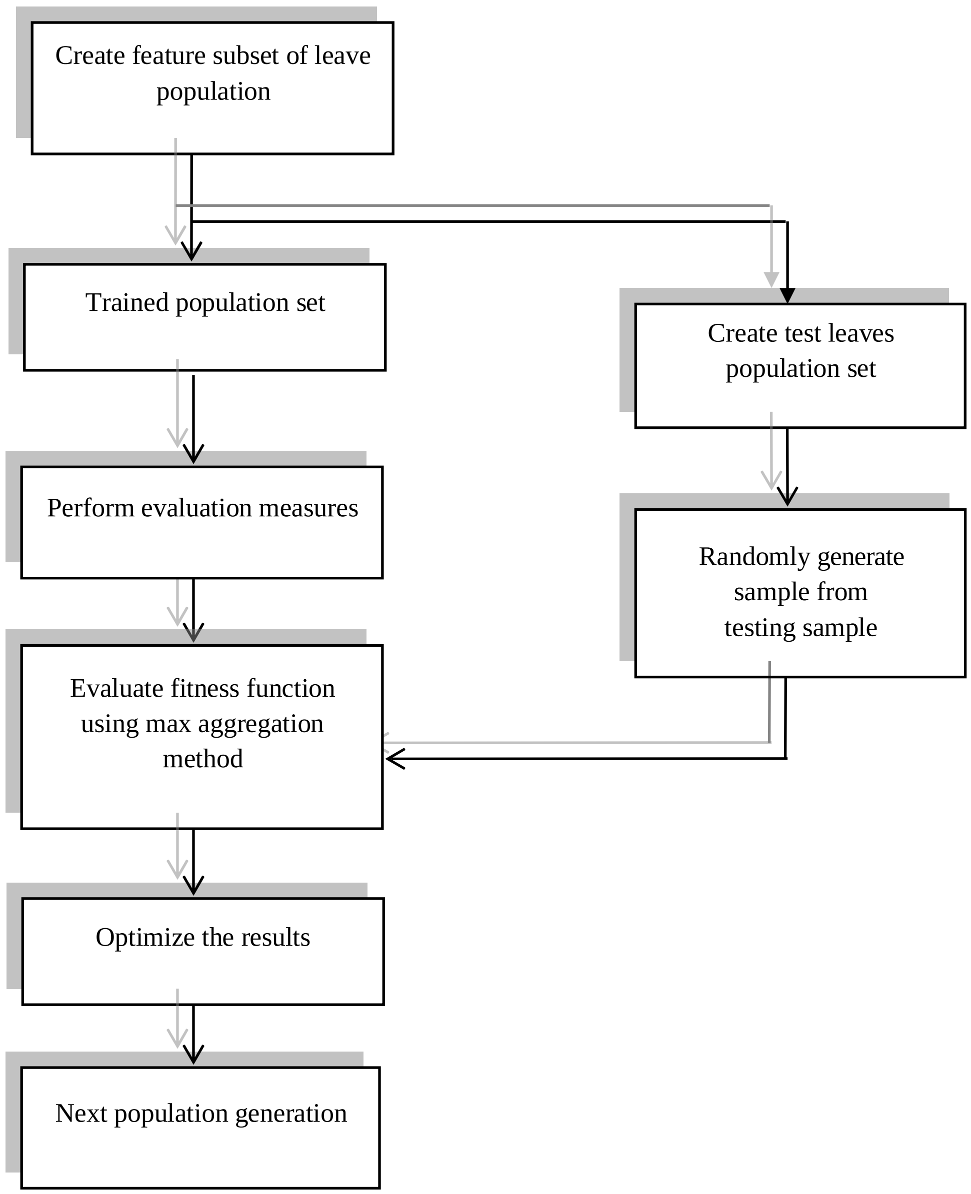


Figure 6.7: Proposed classification model

A new approach for classification of diseased and healthy leaves using survival of fittest approach w.r.t evaluating fitness function using different distance measures is proposed. The distance measures (distance and similarity) play a critical role in pattern analysis problems such as classification, clustering, etc. [181-182] as represented in Table 6.2 The adequate distance between two populations (class) is to be deciding factor. The global distance evaluated by combining number of local distance function and measured the correlation among their attributes using maximum aggregation method. Global distance function provides correlation matrix between each sample of the dataset and evaluate fitness value.

Table 6.2: Evaluation measures

S.no.	Distance/Similarity measures	Formulas
1.	Euclidean distance measure	$d_{EUC} = \sqrt{\sum_{k=1}^{n} \lvert x_k - y_k \rvert^2}$
2.	City block distance measure	$d_{CB} = \sqrt{\sum_{k=1}^{n} \lvert x_k - y_k \rvert}$
3.	Chebyshev distance measure	$d_{Cheb} = \max\lvert x_k - y_k \rvert$
4.	Cosine distance measure	$d_{cos} = \frac{\sum_{k=1}^{n} x_k\, y_k}{\sqrt{\sum_{k=1}^{n} {x_k}^2}\sqrt{\sum_{k=1}^{n} {y_k}^2}}$
5.	Jaccard measure	$d_{Jac} = \frac{\sum_{k=1}^{n} (x_{k-} y_{k)} 2}{\sum_{k=1}^{n} x_k^2 + \sum_{k=1}^{n} y_k^2 - \sum_{k=1}^{n} x_k y_k}$
6.	Minkowski measure	$d_{CB} = p\sqrt{\sum_{k=1}^{n} \lvert x_k - y_k \rvert^p}$
7.	Squared Euclidean measure	$d_{seuc} = \sum_{k=1}^{n} (x_k - y_k)^2$
8.	Dice measure	$d_{dice} = \frac{\sum_{k=1}^{n} (x_k - y_k)^2}{\sum_{k=1}^{n} x_k^2 + \sum_{k=1}^{n} y_k^2}$
9.	Pearson measure	$d_{pearson} = \sum_{k=1}^{n} \frac{(x_k - y_k)^2}{y_k}$

If the fitness value is better than the previous value, then it is set as the new best value. The sample is further chosen by the best fitness value of all as the global best. The process will be continued till maximum iteration, or minimum error criteria are not accomplished. Where, (xi) and (yi) are the two populations from the training data set. To formulate the updating formulas in the proposed algorithm, pseudo code is illustrated below:

Algorithm 1 *Pseudo code for proposed survival of fittest classification algorithm*

Input:

$\boldsymbol{F_{Train,20}}$: *Training features extracted based on the texture and color attributes.*

$\boldsymbol{F_{Test,20}}$ ***:*** *Testing features extracted based on the texture and color attributes.*

$\boldsymbol{Label}$ ***:*** *For labeling the various diseases.*

Output: *Classification of various diseases.*

For

Each sample of training set $F_{Train,20}$

Initialize the sample of testing set $F_{Test,20}$ *then*

Select a sample from testing population randomly

Calculate the fitness function according to the various distance measures w.r.t each sample

The training features are labeled according to the value of the various distance and similarity measures.

The best solution is obtained through the fitness function with minimum distance and maximum similarity values with calculating maximum aggregation.

end

6.6 Experimental results

The results are evaluated for all said criteria. The database consists of different diseased and healthy leaves of different varieties i.e. Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (Holy basil) and Ocimum gratissimum (Vana-holy basil). The disease of leave samples investigated are downy mildew, and bacterial leave spot.

6.6.1 Classification of leaves using machine learning models

Eleven classifiers are compared with proposed framework to evaluate effectiveness of classification models. On the basis of parameters, classifiers will categorize leave disease category. Matlab and R software licensed under GNU GPL are used for proposed method implementation. The tuning parameters of machine learning methods are tabulated in Table 6.3.

Table 6.3: Tuning parameters of machine learning algorithm

Platform	Model	Method	Package	Tuning Parameter(s)
R(version 3.2.5) and Matlab (2015a)	Random Forest	RF	randomForest	Number of trees=500
	Naives Bayes	NBModel	fitNaivesbayes	No. of observation=150
	KNN	knn_model	fitcknn	NumObservations=150, Distance= "euclidean", NumNeighbours=5
	SVM	Ksvm	e1071	Kernel Radial Basis
	Discriminant Analysis	Obj	Fitcdisr	No. of observation=150, DiscrimType: "linear"
	Bayesian Generalized Linear Model	Bayesglm	Arm	Family = gaussian, prior. mean = 0, prior. Scale = NULL, prior.df = 1, prior.mean.for.intercept = 0, prior.scale.for.intercept = NULL, prior.df.for.intercept =1, min.prior.scale=1e-12
	Gaussian Process	gaussprLinear	Kernlab	kpar = list(sigma = 0.1), kernel= "rbfdot"
	Extreme Gradient Boosting	xgbLinear	Xgboost	nrounds[default=100], nrounds[default=100], gamma[default= 0][range: (0,Inf)],max_depth[default=6][range:(0 ,Inf)],subsample[default=1][range: (0,1)]
	Bagging	Bagging	Ipred	No. of decision tree=25
	Conditional Inference Tree	Ctree	Party	weights = NULL, controls = ctree_control(), xtrafo = ptrafo, ytrafo = ptrafo,
	Flexible Discriminant Analysis	Fda	Mda	Default parameters
	Linear Model	LM	Glm	Multinomial
	AdaBoost	Ada	Ada	Max Depth=30, Min Split=20, xval=10,
	Artificial Neural networks	Neuralnet	Neuralnet	Hlayers=10, MaxNWts=10000, maxit=100
	Naives Bayes	NBModel	fitcnb	No. of observation=150

Figure 6.8 graphically represents the performance analysis of proposed model with respect to predicted and actual values. It represents the relationship among prediction and real results.

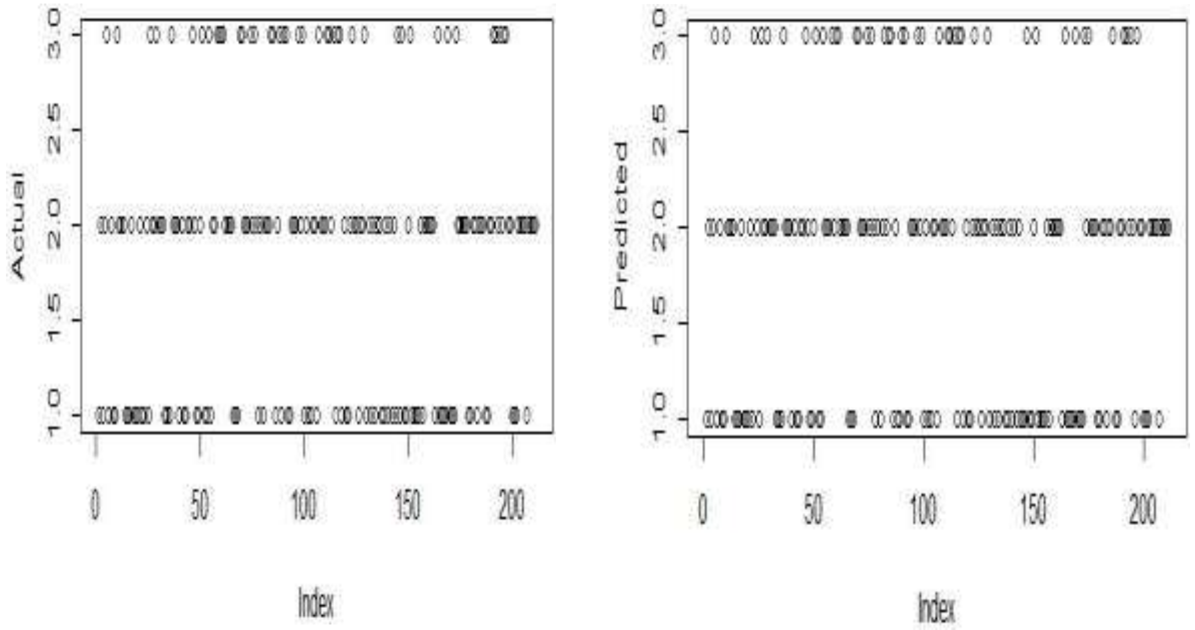


Figure 6.8: Simulated results of proposed classifier model (a) and (b) graphically represent the analysis of predicted vs. observed values

6.6.2 Analysis of confusion matrix

Table 6.4 defines the evaluated statistical values of proposed model. Confusion matrix describes the performance of proposed classifier with respect to predicted results. It also defines confidence interval, P-value, information rate and Mcneman's test P value. The overall accuracy of the proposed system is 95.73%.

Table 6.4: Confusion matrix of proposed model

Parameters of confusion matrix	
Accuracy (%)	95.73%
95%(CI)	(0.9206, 0.9803)
No information rate	0.4408
P value	<2e-16
Kappa	0.9324
Mcneman's Test P value	0.2929

6.6.3 Analysis of ROC and AUC

The comparative results among sensitivity and 1 – specificity (FPR) can be assessed by Receiver operating curve**.** Figure 6.9 illustrate ROC curve of proposed model, where value of area under the curve is 0.9814.

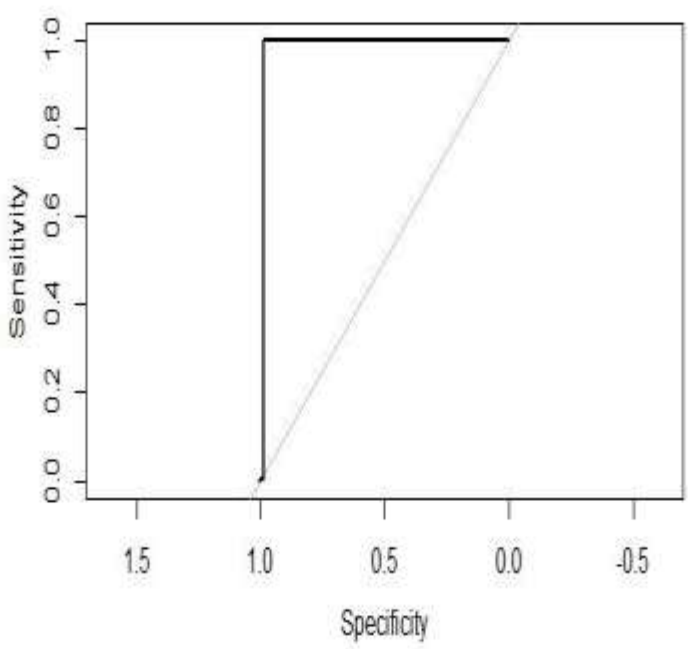


Figure 6.9: Region of convergence curve of proposed model

6.6.4 Classifier based evaluation

Proposed optimization classification algorithm is compared with 11 existing classification models for effective statistical learning. The tuning parameters of machine learning methods are tabulated in Table 6.3.

Figure 6.10 and 6.11 represents performance comparison of proposed method with existing classifier and proves the effectiveness of the proposed model in terms of accuracy and error rate results. Our method attains accuracy 95.73% with error rate 4.27%.

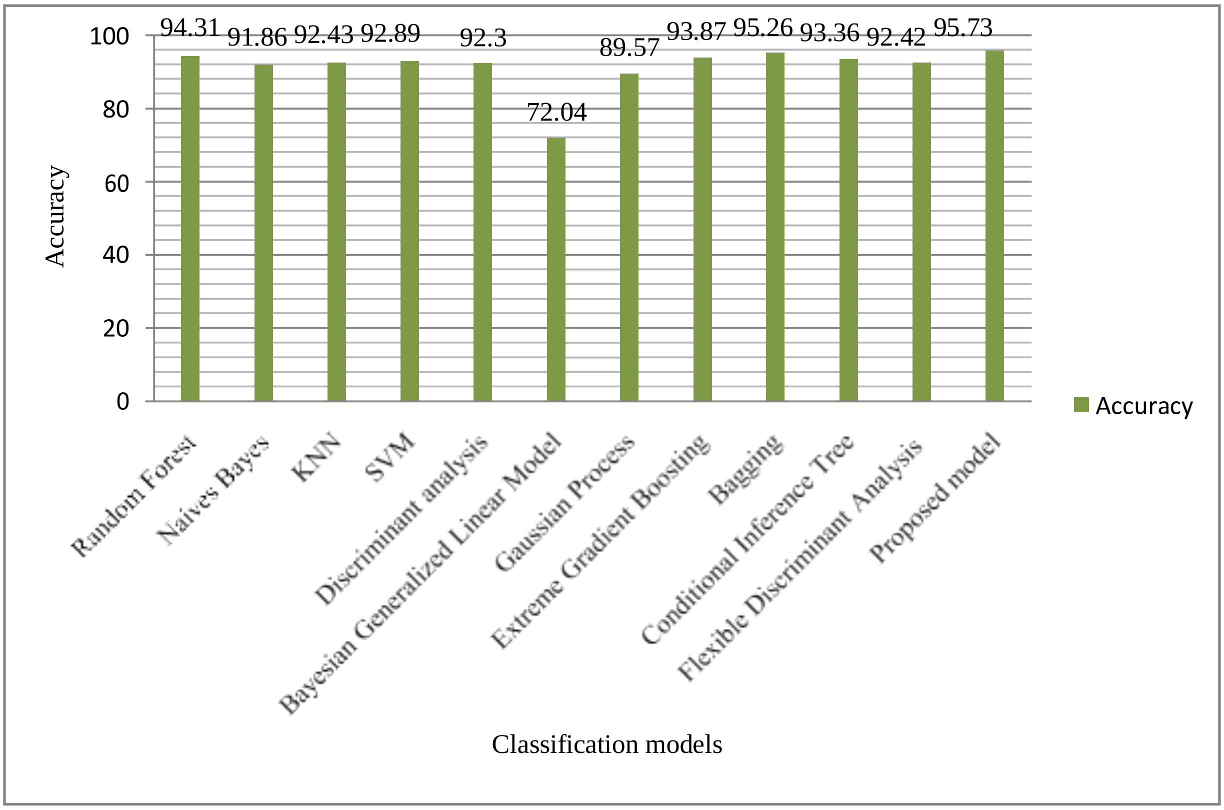


Figure 6.10: Comparison analysis of proposed model with existing classification model in terms of accuracy

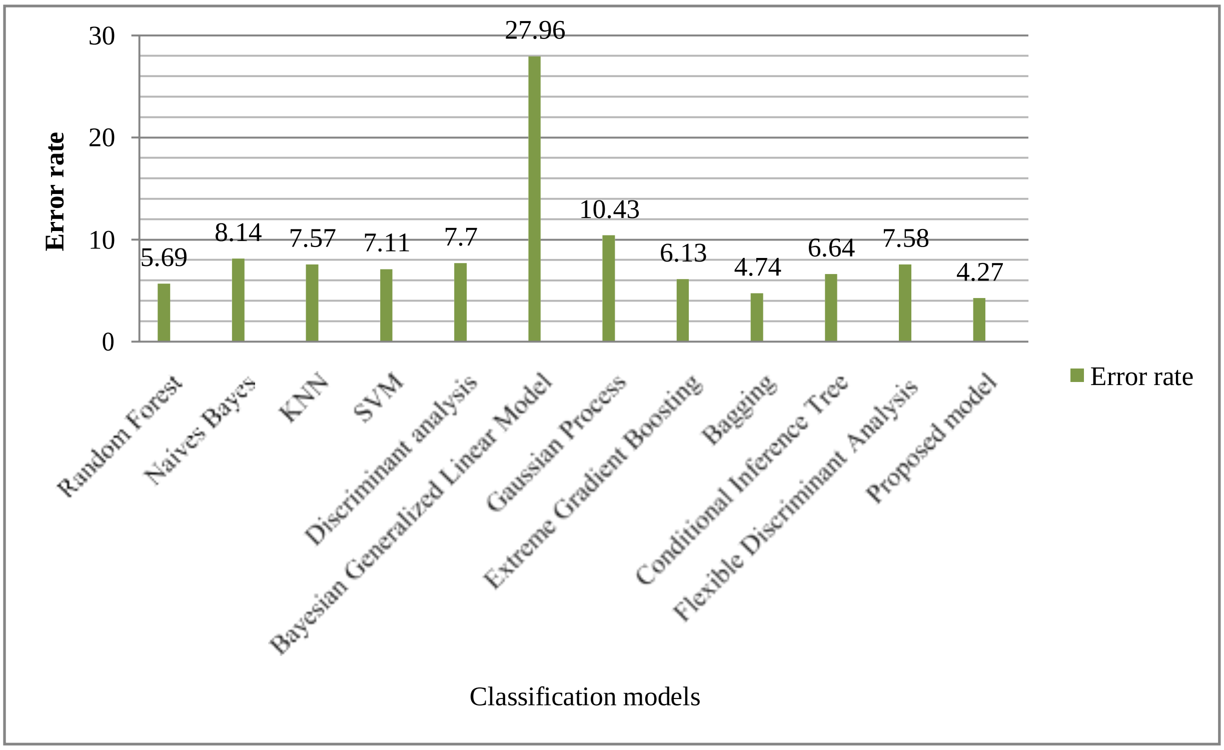


Figure 6.11: Comparison analysis of proposed model with existing classification model in terms of error rate

Table 6.5 explains the comparison analysis of evaluation parameters values i.e. sensitivity, specificity, PPV, NPV and balanced accuracy of machine learning models.

Table 6.5: Evaluation parameters comparison analysis

Classification models	Evaluation parameters (%)	Class 1 Downy mildew	Class 2 Healthy	Class 3 Bacterial leave spot
Random Forest	Sensitivity	98.80	96.47	79.07
	Specificity	94.53	98.41	97.62
	Positive predicted rate	92.13	97.62	89.47
	Negative predicted rate	99.18	97.64	94.86
	Balanced accuracy	96.66	97.44	88.34
Naive Bayes	Sensitivity	94.53	100	67.19
	Specificity	99.30	90.30	97.38
	Positive predicted rate	98.85	88.43	85.15
	Negative predicted rate	96.58	100	92.99
	Balanced accuracy	96.91	95.15	82.28

Classification models	Evaluation parameters (%)	Class 1 Downy mildew	Class 2 Healthy	Class 3 Bacterial leave spot
KNN	Sensitivity	95.62	100	67.97
	Specificity	99.53	90.80	97.90
	Positive predicted rate	99.24	88.96	88.12
	Negative predicted rate	97.25	100	93.49
	Balanced accuracy	97.58	95.40	83.72
Support Vector Machine	Sensitivity	98.68	98.92	69.02
	Specificity	98.52	88.92	100
	Positive predicted rate	97.40	87.62	100
	Negative predicted rate	99.25	99.06	92.86
	Balanced accuracy	98.60	93.85	84.53
Linear Discriminant Analysis	Sensitivity	95.62	100	69.53
	Specificity	99.30	90.55	97.99
	Positive predicted rate	98.27	88.69	87.88
	Negative predicted rate	97.24	100	93.18
	Balanced accuracy	97.46	95.27	82.94
Bayesian Generalized Linear Model	Sensitivity	77.63	100	82.71
	Specificity	99.26	51.69	100
	Positive predicted rate	98.33	62	57.89
	Negative predicted rate	88.74	100	80.09
	Balanced accuracy	88.45	75.85	50
Gaussian Process	Sensitivity	94.74	100	57.14
	Specificity	100	81.36	100
	Positive predicted rate	100	80.87	100
	Negative predicted rate	97.12	100	90.37
	Balanced accuracy	97.37	90.68	57.17
Extreme Gradient Boosting	Sensitivity	97.37	3.55	95.23
	Specificity	99.26	98.31	95.86
	Positive predicted rate	98.67	97.75	85.11
	Negative predicted rate	98.53	95.08	98.78
	Balanced accuracy	98.31	95.93	95.55
Bagging	Sensitivity	98.68	95.70	88.10
	Specificity	97.04	96.61	98.82
	Positive predicted rate	94.94	95.70	94.87
	Negative predicted rate	99.24	96.61	97.09
	Balanced accuracy	97.86	96.15	93.46

Classification models	**Evaluation parameters (%)**	**Class 1 Downy mildew**	**Class 2 Healthy**	**Class 3 Bacterial leave spot**
Conditional Inference Tree	Sensitivity	98.68	94.62	88.10
	Specificity	94.81	98.31	98.82
	Positive predicted rate	91.46	97.78	94.87
	Negative predicted rate	99.24	95.87	97.09
	Balanced accuracy	96.75	96.46	93.46
Flexible Discriminant Analysis	Sensitivity	96.05	100	69.05
	Specificity	100	88.98	98.22
	Positive predicted rate	100	87.74	90.63
	Negative predicted rate	97.83	100	92.74
	Balanced accuracy	98.03	94.49	83.65
Proposed method	Sensitivity	97.37	98.92	85.71
	Specificity	98.57	94.92	99.41
	Positive predicted rate	97.37	93.88	97.30
	Negative predicted rate	98.52	99.12	96.55
	Balanced accuracy	97.94	96.20	92.56

6.6.5 Feature extraction and selection-based evaluation

Various traditional feature extraction methods such as GLCM method [27-183] and Tamura features [184] are compared with the proposed present in Table 6.6. The performance analysis demonstrates that the proposed method is effective and obtained higher recognition rate than the other existing features extraction method. The combination of two feature extraction methods; GLCM texture features and color features are considered to be effective combination for classifying leave images.

Table 6.6: Comparison of feature extraction methods

Methods	Without feature selection				With feature selection			
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Error rate (%)	Accuracy (%)	Sensitivity (%)	Specificity (%)	Error rate (%)
GLCM features [27-183] using leave image dataset.	87.4	87.8	86.9	12.6	90.2	90.9	90.00	9.82
Existing Tamura features using leave images dataset [184]	83.8	86.11	81.6	16.2	86.3	87.1	85.5	13.7
Proposed (GLCM+ Color features) feature extraction method using leave image dataset	92.4	92.6	92.6	7.6	95.73	93.9	97.6	4.3

6.7 Chapter summary

A new method of classification using survival of fittest is explored in order to automatically classify and detect diseases from basil leave images. The developed model was able to distinguish between healthy leaves and 2 different basil diseases. The whole procedure was described, respectively, from gathering images to segmentation, where texture and color features are extracted. These features combine the discrimination power of color and texture properties. Relevant features are selected using Random Forest model to increase the effectiveness of classifier. Finally, diseases are classified using proposed classification technique. Different tests are assessed in order to evaluate the performance of classification model. Proposed classification model and combination of features gives promising results than other existing techniques. The overall accuracy of the proposed trained model was 95.73%.

Chapter 7

Conclusion and Future work

The block diagram of proposed system model has been reorganizing after including all modifications in Figure 7.1. The shaded blocks are clearly depicting the contribution reported in this study.

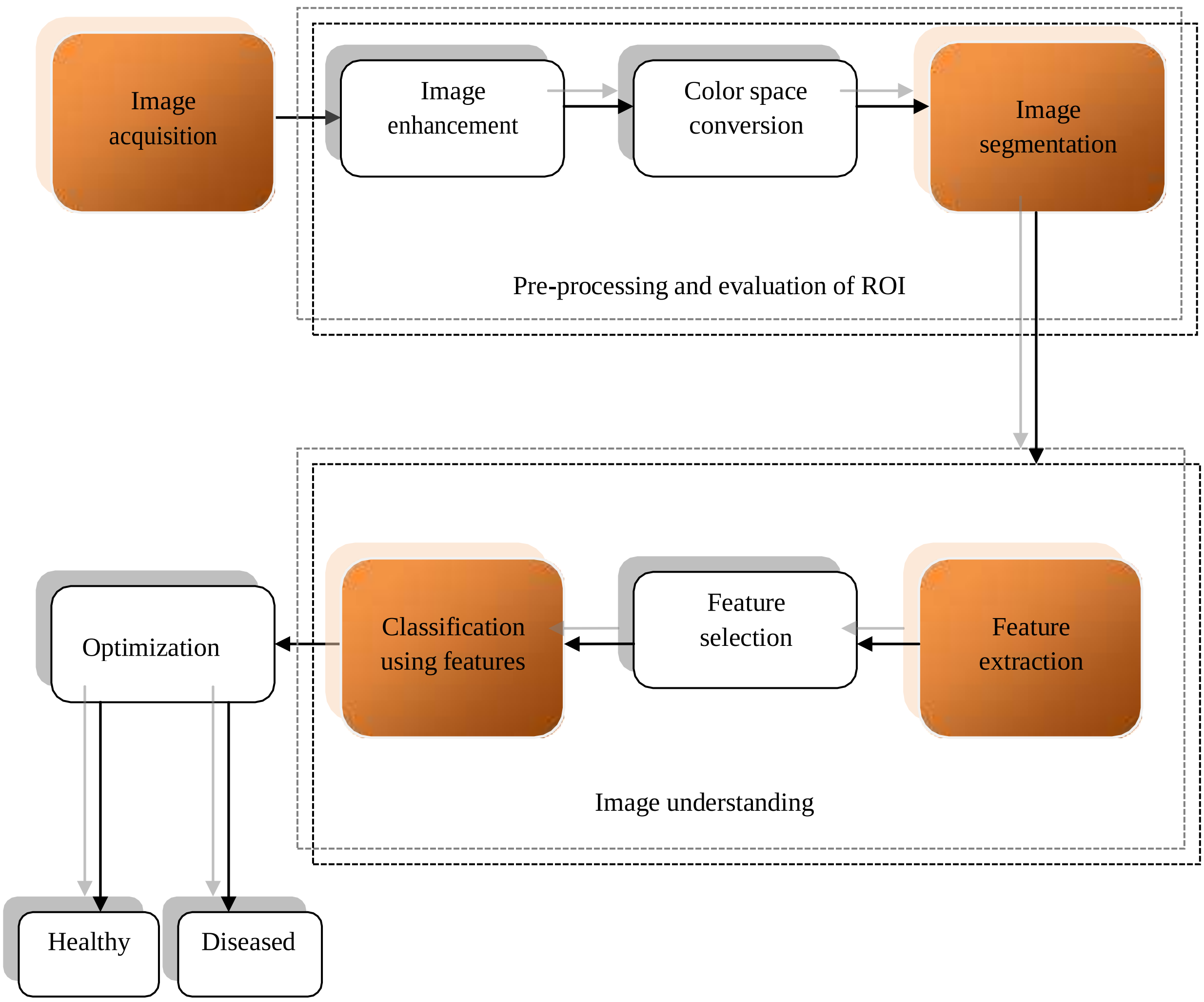


Figure 7.1: Proposed system model

7.1 Conclusion

Precise and well-timed identification and classification of basil diseases is helpful to improve the growth. Disease occurred due to various factors such as improper harvesting, climate changes, water and chemical contamination etc. Diseases can be identified by naked eye observation using continuous monitoring experience level. But it results in high cost as well as time consuming. It has been shown that computer vision and image processing prove to be effective tools for identification and classification of diseases, wherein the digital camera functions as a better substitute for human eye and human brain is superseded by a learning and optimization algorithm.

To overcome the difficulties of manual process, several techniques based on computer vision are developed in recent years to identify and recognize disease of agriculture and horticulture crops. The literature explained various methods for identification and classification for different plant leave diseases. Even though the importance of the subject of identifying leave diseases using digital image processing is tremendous and this has been studied from approximately 25 years.

This thesis addressed the issues of identifying and classifying basil leave diseases using computer vision and image processing. Experimental evidences proof the choice of computer vision tasks and machine learning algorithms considerably influence the performance. In the preceding chapters (4), (5) and (6) we have discussed existing segmentation and classification approaches and presented new theory of segmentation, feature extraction and classification.

Chapter 3 presents the process of data collection and details of camera and site selection etc. Various herb gardens of Punjab and Chandigarh are selected to collect different varieties of leaves. Healthy and diseased leave images (Downy mildew, bacterial leave spot and Fusarium wilt and aphids) are captured in research laboratory under controlled lighting conditions.

In Chapter 4, new segmentation algorithm using neutrosophic fuzzy set is introduced. The algorithm segments an image, into three regions true (diseased), false (healthy) and intermediate (onset disease). As compare to existing segmentation techniques rather than dividing into two segments, our technique introduces a new region known as intermediate. Using all these three segmented regions, features are extracted and leave is classified as healthy or diseased. The proposed segmentation framework correctly identifies diseased segment and produce highest segmentation accuracy.

In Chapter 5, combination of color and texture features is introduced. Based on segmented regions features are extracted. A new texture feature Bin Binary Pattern is introduced. These features combine the discrimination power of intensity and texture of leaves. Nine machine learning models are used to classify leaves. Comparison results show that Random Forest performs best than others.

In Chapter 6, new classification model using survival of fittest approach with successive generation of best results is proposed. The classifier is trained using various color and texture features evaluated from segmented images. The model is able to distinguish disease class of basil efficiently. Proposed method is compared to existing classification algorithm [ref] and performs better than other models. Overall, proposed model is able to categorize basil diseases efficiently.

7.2 Future work

In this study, different segmentation, feature extraction and same color diseases identification and classification techniques system have been considered. Based on our key findings from the previous studies, following future aspects can be considered for further research.

- In future, the existing algorithms can also be utilized in outdoor conditions along with the combination of leave front and leave backs into a common dataset.
- Further research can be devoted to model the detection and recognition techniques under mixed lighting conditions. It is of great importance to confirm that the accuracy of classification can be maintained under these conditions or not.
- An unexplored combination of feature extraction, selection and learning methods can also be used to increase the effectiveness of detection and classification techniques.
- The future work can also be dedicated to the automatic estimation of severity of the detected diseases.
- The instant solutions can be made available to the farmers by designing the mobile based applications.
- Online solutions related to plants diseases can be provided by using web portals.
- It is also required to increase the number of data for training and testing purposes to

achieve better accuracy.

- ❖ Existing work can also be extended to achieve high speed and accuracy by developing the advanced algorithms.

References

1. U. M. Thatte and S. A. Dahanukar, “Ayurveda and contemporary scientific thought,” *Trends in Pharmacological Sciences*, vol. 7, pp. 247–251, 1986.
2. M.M. Pandey, S. Rastogi, and A. K. S. Rawat, “Indian traditional Ayurvedic system of medicine and nutritional supplementation,” *Evidence-Based Complementary and Alternative Medicine*, vol. 2013, pp. 12, 2013.
3. S. Sen and R. Chakraborty, “Revival, modernization and integration of Indian traditional herbal medicine in clinical practice: Importance, challenges and future,” *Journal of Traditional and Complementary Medicine*, vol. 7, no. 2, pp. 234– 244, 2017.
4. O. Oyebode, N. B. Kandala, P. J. Chilton, and R. J. Lilford, “Use of traditional medicine in middle-income countries: a WHO-SAGE study,” *Health Policy and Planning*, vol. 31, no. 8, pp. 984–991, 2016.
5. S. L. Chen, H. Yu, H. M. Luo, Q. Wu, C. F. Li, and A. Steinmetz, “Conservation and sustainable use of medicinal plants: Problems, Progress, and Prospects,” *Chinese Medicine*, vol. 11, no. 1, pp. 37, 2016.
6. L. C. Tapsell, I. Hemphill, L. Cobiac, D. R. Sullivan, M. Fenech, C. S. Patch, S. Roodenrys, J. B. Keogh, P. M. Clifton, P. G. Williams, V. A. Fazio, and K. E. Inge, “Health benefits of herbs and spices: the past, the present, the future,” *Medical Journal of Australia*, vol. 185, no. 4, pp. 1–24, 2006.
7. M. M. Cohen, “Tulsi-Ocimum sanctum: A herb for all reasons,” *Journal of Ayurveda and Integrative Medicine*, vol. 5, no. 4, pp. 251, 2014.
8. K. W. Singletary, “Basil: a brief summary of potential health benefits,” *Nutrition Today*, vol. 53, no. 2, pp. 92–97, 2018.
9. J. C. Chalchat and M. M. Ozcan, “Comparative essential oil composition of flowers, leaves and stems of basil (Ocimum basilicum L.) used as herb,” *Food Chemistry*, vol. 110, no. 2, pp. 501–503, 2008.
10. A. A. Izzo, F. Borrelli, R. Capasso, V. D. Marzo, and R. Mechoulam, “Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb,” *Trends in Pharmacological Science*, vol. 30, no. 10, pp. 515–527, 2009.
11. A. K. Pandey, P. Singh, and N. N. Tripathi, “Chemistry and bioactivities of essential oils of some Ocimum species: an overview,” *Asian Pacific Journal of Tropical Biomedicine*, vol. 4, no. 9, pp. 682–694, 2014.

12. R. R. Alves and I. M. Rosa, “Biodiversity, traditional medicine and public health: where do they meet?,” *Journal of Ethnobiology and Ethnomedicine,* vol.3, no.1, pp. 14, 2007.
13. J. G. A. Barbedo, “Digital image processing techniques for detecting, quantifying and classifying plant diseases,” *SpringerPlus,* vol. 2, no. 1, pp. 1–12, 2013.
14. Y. Fang and R. P. Ramasamy, “Current and prospective methods for plant disease detection,” *Biosensors,* vol. 5, no. 3, pp. 537–561, 2015.
15. S. Kaur, S. Pandey, and S. Goel, “Plants disease identification and classification through leaf images: A survey,” *Archives of Computational Methods in Engineering,* vol. 26, no. 2, pp. 507–530, 2018.
16. S. B. Dhaygude and N. P. Kumbhar, “Agricultural plant leaf disease detection using image processing,” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 1, pp. 599–602, 2013.
17. G. Nagy and A. Horvath, “Occurrence of downy mildew caused by Peronospora belbahrii on sweet basil in Hungary,” *Plant Disease*, vol. 95, no. 8, pp.1034–1034, 2011.
18. Y. Cohen, M. Vaknin, Y. Ben-Naim, A.E. Rubin, M. Galperin, D. Silverman, S. Bitton, and U. Adler, “First report of the occurrence and resistance to mefenoxam of Peronospora belbahrii, causal agent of downy mildew of basil (Ocimum basilicum) in Israel,” *Plant Disease*, vol. 97, no. 5, pp. 692–692, 2013.
19. G. Sepulveda-Chavera, W. Huanca-Mamani, S. Cardenas, M. Arismendi, F. Salinas, E. E. Ferrada, and B. A. Latorre, “First Report of gray mold (Botrytis cinerea) on sweet basil (Ocimum basilicum var. pilosum) in northern Chile,” *Plant Disease*, vol. 100, no. 2, pp. 520, 2016.
20. Z. P. M. Moreira, P. O. D. Santos, T. A. S. D. Oliveira, and J. T. D. Souza, “Occurrence of basil leaf spot caused by Pseudomonas cichorii in Bahia State, Brazil,” Summa *Phytopathologica*, vol. 41, no.1, pp. 73–73, 2015.
21. A. Garibaldi, M. L. Gullino, and G. Minuto, “Diseases of basil and their management,” *Plant Disease*, vol. 81, no. 2, pp. 124–132, 1997.
22. M. G. Moya, M. I. Aguilar, R. Blanco, A. Kenig, J. Gomez, and J. C. Tello, “Fusarium wilt on sweet basil: cause and sources in south eastern Spain,” *Phytoparasitica*, vol. 32, no. 4, pp. 395–401, 2004.
23. R. Ben-Issa, L. Gomez, and H. Gautier, “Companion plants for aphid pest management,” *Insects*, vol. 8, no. 4, pp. 112, 2017.

24. K. Lin, J. Chen, H. Si, and J. Wu, "A review on computer vision technologies applied in greenhouse plant stress detection," in *Chinese Conference on Image and Graphics Technologies*, 2013, pp. 192–200.
25. S. P. Mohanty, D. P. Hughes, and M. Salathe, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, pp. 1419, 2016.
26. T. V. Price and C.F. Osborne, "Computer imaging and its application to some problems in agriculture and plant science," *Critical Reviews in Plant Sciences*, vol. 9, no.3, pp. 235-266
27. P. Sajda, "Machine learning for detection and diagnosis of disease," *Annual Review of Biomedical Engineering*, vol. 8, pp. 537–565, 2006.
28. M. Fatima and M. Pasha, "Survey of machine learning algorithms for disease diagnostic," *Journal of Intelligent Learning Systems and Applications*, vol. 9, no. 01, pp. 1-16, 2017.
29. R. Pydipati, T.F. Burks, and W.S. Lee, "Identification of citrus disease using color, texture features and discriminant analysis," *Computers and Electronics in Agriculture,* vol. 52, no. 1, pp. 49–59, 2006.
30. A. Camargo and J. S. Smith, "An image-processing based algorithm to automatically identify plant disease visual symptoms," *Biosystems Engineering*, vol. 102, no.1, pp. 9–21, 2009.
31. H. Wang, G. Li, Z. Ma, and X. Li, "Image recognition of plant diseases based on principal component analysis and neural networks," *in Proceedings of the IEEE International Conference on Natural Computation, Chongqing*, 2012, pp. 246–251.
32. Q. He, B. Ma, D. Qu, Q. Zhang, X. Hou, and J. Zhao, "Cotton pests and diseases detection based on image processing," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 11, no. 6, pp. 3445–3450, 2013.
33. A. Asfarian, Y. Herdiyani, A. Rauf, and K. H. Mutaqin, "Paddy diseases identification with texture analysis using fractal descriptors based on Fourier spectrum," in *International Conference on Computer, Control, Informatics and its Applications,* 2013, pp. 77–81.
34. P. R. Narvekar, M. M. Kumbhar, and S. N. Patil, "Grape leaf diseases detection & analysis using SGDM matrix method," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 2, no. 3, pp. 3365–3372, 2014.

35. V. Tajane and N. J. Janwe, "Medicinal plants disease identification using canny edge detection algorithm histogram analysis and CBIR," *International Journal of Advanced Research in Computer Science and Soft Engineering*, vol. 4, no. 6, pp. 530–536, 2014.
36. J. F. Molina, R. Gil, C. Bojaca, F. Gomez, and H. Franco, "Automatic detection of early blight infection on tomato crops using a color-based classification strategy," in *Nineteenth IEEE Symposium on Image, Signal Processing and Artificial Vision*, 2014, pp. 1–5.
37. C. C. Tucker and S. Chakraborty, "Quantitative assessment of lesion characteristics and disease severity using digital image processing," *Journal of Phytopathology*, vol. 145, no.7, pp. 273–278, 1997.
38. D. G. Sena Jr, F. A. C. Pinto, D. M. Queiroz, and P. A. Viana, "Fall armyworm damaged maize plant identification using digital images," *Biosystems Engineering*, vol. 85, no. 4, pp. 449–454, 2003.
39. S. Weizheng, W. Yachun, C. Zhanliang, and W. Hongda, "Grading method of leaf spot disease based on image processing," in *IEEE International Conference on Computer Science and Software Engineering,* 2008, vol. 6, pp. 491–494.
40. J.Sekulska-Nalewajko and J. Goclawski, "A semi-automatic method for the discrimination of diseased regions in detached leaf images using fuzzy c-means clustering," in *Perspective Technologies and Methods in MEMS Design*, 2011, pp. 172– 175.
41. Z. B. Husin, A. Y. Shakaff, A. H. Aziz, and R. B. Farook, "Feasibility study on plant chili disease detection using image processing techniques," in *Third IEEE International Conference on Intelligent Systems, Modeling and Simulation*, 2012, pp. 291–296.
42. M. S. Al-Tarawneh, "An empirical investigation of olive leaf spot disease using auto-cropping segmentation and fuzzy c-means classification," *World Applied Sciences Journal*, vol. 23, no. 9, pp. 1207–1211, 2013.
43. D. Zhihua , W. Huan, S. Yinmao, and W. Yunpeng, "Image segmentation method for cotton mite diseases based on color feature and area thresholding," *Journal of Theoretical and Applied Information Technology,* vol. 48, no.1, pp. 527–533, 2013.
44. M. Krishnan and M. G. Sumithra, "A novel algorithm for detecting bacterial leaf scorch (BLS) of shade trees using image processing," in *Eleventh IEEE Malaysia International Conference on Communications*, 2013, pp. 474–478.

45. E. Hitimana and O. Gwun, "Automatic estimation of live coffee leaf infection based on image processing techniques," in *Second International Conference on Signal, Image Processing and Pattern Recognition*, 2014.
46. J. Gui, L. Hao, Q. Zhang, and X. Bao, "A new method for soybean leaf disease detection based on modified salient regions," *International Journal of Multimedia and Ubiquitous Engineering*, vol.10, no. 6, pp. 45–52, 2015.
47. K.W.V. Sanjaya, H.M. Vijesekara, I.M. Wickramasinghe, and C.R.Amalraj, "Orchid classification, disease identification and healthiness prediction system," *International Journal of Scientific and Technology Research*, vol. 4, no.3, pp. 215–220, 2015.
48. S. Shrivastava, S. K. Singh, and D. S. Hooda, "Color sensing and image processing-based automatic soybean plant foliar disease severity detection and estimation," *Multimedia Tools and Applications*, vol. 74, no. 24, pp. 11467–11484, 2015.
49. A. K. Dey, M. Sharma, and M. R. Meshram, "Image processing-based leave rot disease, detection of betel vine (Piper BetleL.)," *Procedia Computer Science*, vol. 85, pp. 748–754, 2016.
50. G. Anthonys and N. Wickramarachchi, "An image recognition system for crop disease identification of paddy fields in Sri Lanka," in *Fourth IEEE international conference on industrial and information systems*, 2009, pp. 403–407.
51. D. Pixia and W. Xiangdong, "Recognition of greenhouse cucumber disease based on image processing technology," *Open Journal of Applied Sciences*, vol. 3, no. 01, pp. 27–31, 2013.
52. K. Y. Huang, "Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features," *Computers and Electronics in Agriculture*, vol. 57, no. 1, pp. 3–11, 2007.
53. N. E. Abdullah, A. A. Rahim, H. Hashim, and K. Kamal, "Classification of rubber tree leaf diseases using multilayer perceptron neural network." in *Fifth Student Conference on Research and Development*, 2007, pp. 1–6.
54. S. Phadikar and J. Sil, "Rice disease identification using pattern recognition techniques," in *Eleventh International Conference on Computer and Information Technology*, 2008, pp. 420–423.
55. D. A. Bashish, M. Braik, and S. Bani-Ahmad, "A framework for detection and classification of plant leaf and stem diseases," in *IEEE International Conference on Signal and Image Processing*, 2010, pp.113–118.

56. H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, "Fast and accurate detection and classification of plant diseases," *International Journal of Computer Applications*, vol. 17, no. 1, pp. 31–38, 2011.
57. W. Abdullakasim, K. Powbunthorn, J. Unartngam, and T. Takigawa, "An images analysis technique for recognition of brown leaf spot disease in cassava," *Journal of Agricultural Machinery Science*, vol. 7, no. 2, pp. 165–169, 2011.
58. S. Kai, L. Zhikun, S. Hang, and G. Chunhong, "A research of maize disease image recognition of corn based on BP networks," in *IEEE Third International Conference on Measuring Technology and Mechatronics Automation*, 2011, pp.246–249.
59. D. S. Guru, P. B. Mallikarjuna, and S. Manjunath, "Segmentation and classification of tobacco seedling diseases," in *Fourth Annual ACM Bangalore Conference*, 2011, pp. 32.
60. W. Zhang, T. Guifa, and W. Chunshan, "Identification of jujube trees diseases using neural network," *Optik International Journal for Light and Electron Optics*, vol. 124, no. 11, pp. 1034–1037, 2013.
61. P. V. Keskar, S. M. Masare, M. S. Kadam, and S. M. Deoghare, "Leaf disease detection and diagnosis," *International Journal of Emerging Trends in Electrical and Electronics,* vol. 2, no. 2, pp. 28–31, 2013.
62. S. S. Sannakki, V. S. Rajpurohit, V. B. Nargund, and P. Kulkarni, "Diagnosis and classification of grape leaf diseases using neural networks," in *Fourth International Conference on Computing, Communications and Networking Technologies*, 2013, pp. 1–5.
63. K. Majid, Y. Herdiyeni, and A. Rauf, "I-PEDIA: Mobile application for paddy disease identification using fuzzy entropy and probabilistic neural network," in *IEEE International Conference on Advanced Computer Science and Information Systems*, 2013, pp. 403–406.
64. P. Revathi and M. Hemalatha, "Cotton leaf spot diseases detection utilizing feature selection with skew divergence method," *International Journal of Scientific Engineering and Technology,* vol. 3, no. 1, pp. 22–30, 2014.
65. S. B. Jagtap and M. H. Shailesh, "Agricultural plant leaf disease detection and diagnosis using image processing based on morphological feature extraction," *IOSR Journal of VLSI and Signal Processing*, vol. 4, no. 5, pp. 24–30, 2014.
66. J. W. Orillo, J. D. Cruz, L. Agapito, P. L. Satimbre, and I. Valenzuela, "Identification of diseases in rice plant (Oryza Sativa) using back propagation artificial neural

network," in *IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management,* 2014, pp.1–6.

67. J. R. Aduwo, E. Mwebaze, and J. A.Quinn, "Automated vision-based diagnosis of cassava mosaic disease," in *Industrial Conference on Data Mining-Workshops,* 2010, pp. 114–122.
68. D. Mondal and D. K. Kole, "Detection and classification technique of yellow vein mosaic virus disease in okra leaf images using leaf vein extraction and Naive Bayesian classifier," in *IEEE International Conference on Soft Computing Techniques and Implementations,* 2015, pp. 166–171.
69. M. T. Bin Mohamad Azmi and N. M. Isa, "Orchid disease detection using image processing and fuzzy logic," in *International Conference on Electrical, Electronics and System Engineering,* 2013, pp. 37–42.
70. M. Billah, M. B. A. Miah, A. Hanifa, and R. Amin, "Adaptive neuro fuzzy inference system based tea leaf disease recognition using color wavelet," *Communications on Applied Electronics,* vol. 3, no. 5, pp.1–4, 2015.
71. K. Muthukannan and P. Latha, "Fuzzy inference system based unhealthy region classification in plant leaf image," *International Journal of Computer, Electrical, Automation, Control and Information Engineering,* vol. 8, no. 11, pp. 2103–2107, 2014.
72. Z. Zhang, Y. Li, F. Wang, and X. He, "A particle swarm optimization algorithm for neural networks in recognition of maize leaf diseases," Sensors *and Transducers,* vol. 166, no. 3, pp. 181–189, 2014.
73. K. Muthukannan and P. Latha, "A PSO model for disease pattern detection on leaf surfaces," *Image Analysis and Stereology,* vol. 34, no. 3, pp. 209–216, 2015.
74. Y. Tian, C. Zhao, S. Lu, and X. Guo, "Multiple classifier combination for recognition of wheat leaf diseases," *Intelligent Automation & Soft Computing,* vol. 17, no. 5, pp. 519-529, 2011.
75. I. E. Massi, Y. E. Saddy, M. E. Yassa, D. Mammass, and A. Benazon, "Serial combination of two classifiers for automatic recognition of the damages and symptoms on plant leaf," in *IEEE Third World Conference on Complex Systems,* 2015, pp. 1–6.
76. Y. Es-saady, I. El Massi, M. El Yassa, D. Mammass, and A. Benazoun, "Automatic recognition of plant leaf diseases based on serial combination of two SVM

classifiers," in *International Conference on Electrical and Information Technologies*, 2016, pp. 561–566.

77. S. Zhang and C. Zhang, "Orthogonal locally discriminant projection for classification of plant leaf diseases," in *Ninth IEEE International Conference on Computational Intelligence and Security*, 2013, pp. 241–245.

78. D. Wu, C. Xie, and C. Ma, "The SVM classification leaf miner-infected leaves based on fractal dimension," in *IEEE Conference on Cybernetics and Intelligent Systems*, 2008, pp. 147–151.

79. V. Surendrababu, C. P. Sumathi, and E. Umapathy, "Detection of rice leaf diseases using chaos and fractal dimension in image processing," *International Journal on Computer Science and Engineering*, vol. 6, no. 1, pp. 69–74, 2014.

80. A. Meunkaewjinda, P. Kumsawat, K. Attakitmongcol, and A. Srikaew, "Grape leaf disease detection from color imagery using hybrid intelligent system," in *Fifth IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2008, pp. 513–516.

81. T. Youwen, Li. Tianlai, and N. Yan, "The recognition of cucumber disease based on image processing and support vector machine," in *IEEE Conference on Image and Signal Processing*, 2008, pp. 262–267.

82. Z. Jian and Z. Wei, "Support vector machine for recognition of cucumber leaf diseases," in *Second IEEE international conference on advanced computer control*, 2010, vol. 5, pp. 264–266.

83. H. M. Asraf, M. T. Nooritawati, and M.S. Rizam, "A comparative study in kernel-based support vector machine of oil palm leafs nutrient disease," *Procedia Engineering*, vol. 41, pp. 1353–1359, 2012.

84. L. Cunlou, S. Gao, and Z. Zhou, "Maize disease recognition via fuzzy least square support vector machine," *Journal of Information and Computing Science*, vol. 8, no. 4, pp. 316–320, 2013.

85. S. Arivazhagan, R. N. Shebiah, S. Ananthi, and S. V. Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Agricultural Engineering International CIGR Journal*, vol. 15, no. 1, pp. 211–217, 2013.

86. R. Zhou, S. Kaneko, F. Tanaka, M. Kayamori, and M. Shimizu, "Matching-based Cercospora leaf spot detection in sugar beet," *International Journal of Nutrition and Food Engineering*, vol.7, no.7, pp. 712-718, 2013.

87. E. K. Ratnasari, M. Mentari, R. K. Dewi, and R. H. Ginardi, "Sugarcane leaf disease detection and severity estimation based on segmented spots image," in *Proceedings of the IEEE International Conference on Information, Communication Technology and System*, 2014, pp. 93–98.
88. U. Mokhtar, M. A. Ali, A. E. Hassenian, and H. Hefny, "Tomato leaves diseases detection approach based on support vector machines," in *Proceedings of the IEEE International Computer Engineering Conference*, 2015, pp. 246–250.
89. S. Zhang and Z. Wang, "Cucumber disease recognition based on Global-Local Singular value decomposition," *Neurocomputing*, vol. 205, pp. 341–348, 2016.
90. F. Qin, D. Liu, B. Sun, L. Ruan, Z. Ma, and H. Wang, "Identification of alfalfa leaf diseases using image recognition Technology," *PloS one*, vol. 11, no. 12, pp. 1–26, 2016.
91. S. R. Bandi, A. Varadharajan, and A. Chinnasamy, "Performance evaluation of various statistical classifiers in detecting the diseased citrus leaf," *International Journal of Engineering Science and Technology*, vol. 5, no. 2, pp. 298–307, 2013.
92. O. M. Kruse, J. M. Prats-Montalban, U. G. Indahl, K. Kvaal, A. Ferrer, and C. M. Futsaether, "Pixel classification methods for identifying and quantifying leaf surface injury from digital images," *Computers and Electronics in Agriculture*, vol. 108, pp. 155–165, 2014.
93. M. Zhang and Q. Meng, "Automatic citrus canker detection from leaf images captured in field," *Pattern Recognition Letters*, vol. 32, no. 15, pp. 2036–2046, 2011.
94. K. J. Mohan, M. Balasubramanian, and S. Palanivel, "Detection and recognition of diseases from Paddy Plant leaf images," *International Journal of Computer and Applications*, vol. 141, no. 12, pp. 34–41, 2016.
95. S. Phadikar, S. Jaya, and A. S. Das, "Rice diseases classification using feature selection and rule generation techniques," *Computers and Electronics in Agriculture*, vol. 90, pp. 76–85, 2013.
96. S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1–11, 2016.
97. B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," *Symmetry*, vol. 10, no. 1, pp. 1–11, 2017.

98. Y. Toda and F. Okura, “How convolutional neural networks diagnose plant disease. *Plant Phenomics*, vol. 2019, pp. 1–14, 2019.
99. M. B. AL-Otaibi, A.S. Ashour, N. Dey, R. Abdullah, A. A. AL-Nufaei, and S. Fuqian, “Statistical image analysis based automated leaves classification,” in *Proceedings of the Second International Conference on Information Technology and Intelligent Transportation Systems*, 2017, vol. 296, pp. 469.
100. M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009.
101. J. Huan and C. X. Ling, “Using AUC and accuracy in evaluating learning algorithms,” *IEEE Transactions on knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.
102. T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
103. Y. Jiang, B. Cukic, and T. Menzies, “Cost curve evaluation of fault prediction models,” in *nineteenth IEEE International Symposium on Software Reliability Engineering*, 2008, pp. 197–206.
104. J. Sim and C. C. Wright, “The kappa statistic in reliability studies: use, interpretation, and sample size requirements,” *Physical therapy*, vol. 85, no.3, pp. 257–268, 2005
105. J. Penczek, P. A. Boynton and J. D. Splett, “Color error in the digital camera image capture process,” *Journal of Digital Imaging*, vol. 27, no.2, pp.182-191, 2014.
106. O. Makri and S. Kintzios, “Ocimum sp.(basil): Botany, cultivation, pharmaceutical properties, and biotechnology,” *Journal of Herbs, Spices & Medicinal Plants*, vol. 13, no. 3, pp. 123–150, 2008.
107. A. I. Hussain, F. Anwar, S. T. H. Sherazi, and R. Przybylski, “Chemical composition, antioxidant and antimicrobial activities of basil (Ocimum basilicum) essential oils depends on seasonal variations,” *Food Chemistry*, vol. 108, no. 3, pp. 986–995, 2008.
108. R. M. Haralick and L. G. Shapiro, “Image segmentation techniques,” *Computer Vision, Graphics, and Image Processing*, vol. 29, no.1, pp. 100–132, 1985.
109. J. Kaur, S. Agrawal, and R.Vig, “A comparative analysis of thresholding and edge detection segmentation techniques,” *International Journal of Computer Applications*, vol. 39, no.15, pp. 29–34, 2012.
110. S. K. Fu and J. K. Mui, “A survey on image segmentation,” *Pattern Recognition*, vol. 13, no. 1, pp. 3–16, 1981.

111. D. Pratiwi and K. H. Kartowisastro, “Object Segmentation under Varying Illumination Effects,” *New Trends in Intelligent Information and Database Systems, Studies in Computational Intelligence*, 2015, pp. 13–21.
112. Y. Zhang and L. Ge, “New approach to low contrast image segmentation,” in *Proceedings of the Second IEEE International Conference on Bioinformatics and Biomedical Engineering*, 2008, pp. 2369–2372.
113. S. Rital, H. Cherifi, and S. Miguet, “A segmentation algorithm for noisy images,” in *Proceedings of the International Conference on Computer Analysis of Images and Patterns*, 2005, pp. 205–212.
114. J. Dogra, S. Jain, and M. Sood, “Glioma extraction from MR images employing Gradient Based Kernel Selection Graph Cut technique,” *The Visual Computer*, vol. 35, no. 285, pp. 1–17, 2019.
115. R. J. Hemalatha, T. R. Thamizhvani, A. J. A. Dhivya, J. E. Joseph, B. Babu, and R. Chandrasekaran, “Active contour based segmentation techniques for medical image analysis,” *Medical and Biological Image Analysis*, IntechOpen, 2018.
116. L. Shen and R. M. Rangayyan, “A segmentation-based lossless image coding method for high-resolution medical image compression,” *IEEE Transactions on Medical imaging*, vol. 16, no. 3, pp. 301–307, 1997.
117. R.C. Gonzalez, “RE Woods Digital Imaging Processing 3rd edition,” *Pearson Prentice Hall*, 2008.
118. G. E. Sujji, Y. V. S. Lakshmi, and G. W. Jiji, “MRI brain image segmentation based on thresholding,” *International Journal of Advanced Computer Research*, vol. 3, no. 1, pp. 97–101, 2013.
119. R. Adams and L. Bischof, “Seeded region growing,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.
120. J. Tang, “A color image segmentation algorithm based on region growing,” in *Second IEEE International Conference on Computer Engineering and Technology*, 2010, vol. 6, pp. V6–634.
121. S. Y. Chen, W. C. Lin, and C. T. Chen, “Split-and-merge image segmentation based on localized feature analysis and statistical tests,” *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 5, pp. 457–475, 1991.
122. G. B. Coleman and H. C. Andrews, “Image segmentation by clustering,” *Proceedings of the IEEE*, vol. 67, no. 5, pp. 773–785, 1979.

123. T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithm: Analysis and implementation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 88–892, 2002.

124. K. S. Chuang, H. L. Tzeng, S. Chen, J. Wu, and T. J. Chen, “Fuzzy c-means clustering with spatial information for image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 30, no. 1, pp. 9–15, 2006.

125. P. Suetens, R. Verbeeck, D. Delaere, J. Nuyts, and B. Bijnens, “Model-based image segmentation: methods and applications,” *Artificial Intelligence in Medicine*, 1991, pp. 3–24.

126. S. Caraiman and V. I. Manta, “Histogram-based segmentation of quantum images,” *Theoretical Computer Science*, vol. 529, pp. 46–60, 2014.

127. R. Muthukrishnan and M. Radha, “Edge detection techniques for image segmentation,” *International Journal of Computer Science and Information Technology*, vol. 3, no. 6, pp. 259, 2011.

128. W. Khan, “Image segmentation techniques: A survey,” *Journal of Image and Graphics*, vol. 1, no. 4, pp. 166–170, 2013.

129. B. S. Min, D. K. Lim, S. J. Kim and J. H. Lee, “A novel method of determining parameters of CLAHE Based on Image Entropy,” *International Journal of Software Engineering and its applications*, vol. 7, no. 5, pp. 113–120, 2013.

130. K. Zuiderveld, “Contrast limited adaptive histogram equalization,” in *Graphics gems IV*, Academic Press Professional, Inc., 1994.

131. C. Connolly and T. Fleiss, “A study of efficiency and accuracy in the transformation from RGB to CIELAB color space,” *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 1046–1048, 1997.

132. Y. Guo and H. D. Cheng, “New neutrosophic approach to image segmentation,” *Pattern Recognition*, vol. 42, no. 5, pp. 587–595, 2009.

133. U. Rivieccio, “Neutrosophic logics: Prospects and problems,” *Fuzzy Sets Systems*, vol. 159, no. 14, pp. 1860–1868, 2008.

134. Y. Guo and A. Şengur, “A novel image segmentation algorithm based on neutrosophic similarity clustering,” *Applied Soft Computing*, vol. 25, pp. 391–398, 2014.

135. A. T. Lehrer, and E. Komor, “Symptom expression of yellow leaf disease in sugarcane cultivars with different degrees of infection by Sugarcane yellow leaf virus,” *Plant Pathology*, vol.*57, no.*1, pp.178-189, 2008.

136. F. Wang, S. Xing, Y. Ji, L. Yan, F. Liu, Y. Dong, and J. Wang, “Leaf structural reddening in smoke tree and its significance,” *Urban Forestry & Urban Greening*, vol. 14, no. 1, pp. 80–88, 2015.

137. P. Matile, “Biochemistry of Indian summer: physiology of autumnal leaf coloration,” *Experimental gerontology*, vol. 35, no. 2, pp. 145-158, 2000.

138. Guyon, Isabelle, and Andre Elisseeff, “An introduction to feature extraction,” *Feature extraction*, pp. 1–25, 2006.

139. O. D. Trier, A.K. Jain, and T. Taxt, “Feature extraction methods for character recognition-a survey,” *Pattern recognition*, vol. 29, no. 4, pp. 641–662, 1996.

140. J. Sklansky, “Image segmentation and feature extraction,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol.8, no. 4, pp.237-247, 1978.

141. H. Liu and H. Motoda, “Feature extraction, construction and selection*,” A data mining perspective*, Springer Science and Business Media, vol. 453, no. 2, pp. 26–28, 1998.

142. M. Nixon and A. S. Aguado, “Feature extraction and image processing for computer vision,” *Academic Press*, 2012.

143. D. A. Lisin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, “Combining local and global image features for object class recognition,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 2005, pp. 47–47

144. J. Lopez-Labraca, M.A. Fernandez-Torres, I. Gonzalez-Diaz, F. Diaz-de-Maria, and A. Pizarro, “Enriched dermoscopic-structure-based cad system for melanoma diagnosis,” *Multimedia Tools and Applications*, vol. 77, no.10, pp. 12171–12202, 2018.

145. S. C. Pei and C. M. Cheng, “Extracting color features and dynamic matching for image database retrieval,” *IEEE Transactions On circuits and systems for video technology,* vol. 9, no. 3, pp. 501–512, 1999.

146. Q. Zhao, J. Yang, J. Yang, and H. Liu, “Stone images retrieval based on color histogram,” *IEEE International Conference on Image Analysis and Signal Processing*, 2009, pp. 157–161.

147. R. M. Haralick and K. Shanmugam, “Textural features for image classification,”

IEEE Transactions on systems, man, and cybernetics, vol. 6, pp. 610– 621, 1973.

148. D. C. He and L. Wang, “Texture features based on texture spectrum,” *Pattern Recognition*, vol. 24, no. 5, pp. 391–399, 1991

149. F. Liu, F and R.W. Picard, “Periodicity, directionality, and randomness: Wold features for image modeling and retrieval,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.18, no.7, pp. 722–733, 1996.

150. F. A. Andalo, P. A. V. Miranda, R. S. Torres, and A. X. Falcao, “Shape feature extraction and description based on tensor scale,” *Pattern Recognition*, vol. 43, no. 1, pp. 26–36, 2010.

151. M. Yang, K. Kpalma, and J. Ronsin, “A survey of shape feature extraction techniques,” pp. 43–90, 2008.

152. S. Sergyan, “Color histogram features based image classification in content-based image retrieval systems,” in Sixth International Symposium on Applied Machine Intelligence and Informatics, 2008, pp. 221–224.

153. S. R. Maniyath, P. V. Vinod, M. Niveditha, R. Pooja, N. Shashank, and R. Hebbar, “Plant disease detection using machine learning,” in *International Conference on Design Innovations for 3Cs Compute Communicate Control*, 2018, pp. 41–45.

154. B. Sandika, S. Avil, S. Sanat, and P. Srinivasu, “Random forest-based classification of diseases in grapes from images captured in uncontrolled environments,” in *IEEE Thirteenth International Conference on Signal Processing*, 2016, pp. 1775–1780.

155. J. Kim, A. Mowat, P. Phoole, N. Kasabov, “Linear and non-linear pattern recognition models for classification of fruit from visible–near infrared spectra,” *Chemometrics and Intelligent Laboratory Systems*, vol. 51, no. 2, pp. 201–216, 2000.

156. E. Hossain, M. F. Hossain, and M. A. Rahaman, “A color and texture-based approach for the detection and classification of plant leaf disease using KNN classifier,” in *International Conference on Electrical, Computer and Communication Engineering*, 2019, pp. 1–6.

157. K.Hajian-Tilaki, “Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation,” *Caspian Journal of Internal Medicine*, vol. 4, no. 2, pp. 627, 2013.

158. C. M. Wu, and Y. C. Chen, “Statistical feature matrix for texture analysis,” *Graphical Models and Image Processing, vol.* 54, no. 5, 407-419, 1992.

159. A. S. Setiawan, Elysia, J. Wesley and Y. Purnama, “Mammogram classification using Law's texture energy measure and neural networks,” Procedia Computer Science, vol.

59, pp. 92–97, 2015.

160. W. Chong, D. Blei, and F. F. Li, “Simultaneous image classification and annotation,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1903–1910.

161. A. Vailaya, M.A. Figueiredo, A. K. Jain, and H.J. Zhang, “Image classification for content-based indexing,” *IEEE transactions on image processing*, vol.10, no.1, pp. 117–130.

162. D. Lu and Q. Weng, “A survey of image classification methods and techniques for improving classification performance,” *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, 2007.

163. M. Mohammady, H. R. Moradi, H. Zeinivand, and A. J. A. M. Temme, “A comparison of supervised, unsupervised and synthetic land use classification methods in the north of Iran,” *International Journal of Environmental Science and Technology*, vol. 12, no. 5, pp. 1515–1526, 2015.

164. H. K. Turesson, S. Ribeiro, D. R. Pereira, J. P. Papa, and V. H. C. de Albuquerque, “Machine learning algorithms for automatic classification of marmoset vocalizations,” *PloS one*, vol. 11, no. 9, pp. 1–14, 2016.

165. J. R. Quinlan, “Induction of decision trees,” *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.

166. C. J. Burges, “A tutorial on support vector machines for pattern recognition,” *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

167. M. Tanveer, A. Tiwari, R. Choudhary, and S. Jalan, “Sparse pinball twin support vector machines,” *Applied Soft Computing*, vol.78, pp. 164–175, 2019.

168. M. F. Unlersen and K. Sabanci, “The classification of diseased Trees by Using KNN and MLP Classification models according to the satellite Imagery,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 4, no. 2, pp. 25–28, 2016.

169. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

170. N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine Learning*, vol. 29, no. 2–3, pp. 131–163, 1997.

171. A. J. Izenman, “Linear discriminant analysis,” in *Modern Multivariate Statistical Techniques*, Springer, New York, 2013, pp. 237–280.

172. M. Seeger, “Gaussian processes for machine learning,” *International Journal of Neural Systems*, vol. 14, no. 02, pp. 69–106, 2004.

173. G. P. Zhang, “Neural networks for classification: A survey,” IEEE Transactions on System, Man, Cybernetics,” *Applications and Reviews*, vol. 30, no. 4, pp. 451–462, 2000.
174. A. Bhardwaj, A. Tiwari, H. Bhardwaj and A. Bhardwaj, “A genetically optimized neural network model for multi-class classification,” *Expert Systems with Applications*, vol. 60, pp. 211–221, 2016.
175. P. Buhlmann, “Bagging, boosting and ensemble methods,” *Handbook of Computational Statistics*, 2012.
176. T. G. Dietterich, “Ensemble learning,” *The handbook of brain theory and neural networks*, 2002, vol. 2, pp. 110–125.
177. S. Ozturk and B. Akdemir, “Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA,” *Procedia Computer Science*, vol. 132, pp. 40–46, 2018.
178. B. Pathak and D. Barooah, “Texture analysis based on the gray-level co-occurrence matrix considering possible orientations,” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 2, no. 9, pp. 4206–4212, 2013.
179. A. F. Aji, Q. Munajat, A. P. Pratama, H. Kalamullah, J. Setiyawan and A. M. Arymurthy, “Detection of palm oil leave disease with image processing and neural network classification on mobile device,” *International Journal of Computer Theory and Engineering*, vol. 5, no. 3, pp. 528–532, 2013.
180. A. Hapfelmeier and K. Ulm, “A new variable selection approach using random forests,” *Computational Statistics and Data Analysis*, vol. 60, pp. 50–69, 2013.
181. S. H. Cha, “Comprehensive survey on distance/similarity measures between probability density functions,” *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 1, no. 2, pp. 300–307, 2007.
182. S. S. Choi, S. H. Cha, and C. C. Tappert, “A survey of binary similarity and distance measures, Journal of Systemics,” *Cybernetics and Informatics*, vol. 8, no. 1, pp. 43–48, 2010.
183. C. Xie and Y. He, “Spectrum and image texture features analysis for early blight disease detection on eggplant leaves,” *Sensors*, vol. 16, no. 5, pp. 676, 2016.
184. N. Bagri and P. K. Johari, “A comparative study on feature extraction using texture and shape for content-based image retrieval,” *International Journal of Advanced Science and Technology*, vol. 80, no. 4, pp. 41–52, 2015.