Measurement 135 (2019) 782–794

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

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

Gittaly Dhingra [a,*], Vinay Kumar [b], Hem Dutt Joshi [b]

[a] Department of Electronics and Communication Engineering, Research Scholar, Thapar Institute of Engineering and Technology, Patiala, India
[b] Department of Electronics and Communication Engineering, Faculty, Thapar Institute of Engineering and Technology, Patiala, India

ARTICLE INFO

Article history:
Received 16 August 2016
Received in revised form 1 November 2018
Accepted 5 December 2018
Available online 6 December 2018

Keywords:
Leaf images
Neutrosophic logic
Texture features
Intensity features
Classifiers

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. The quality of leafs defines the degree of excellence or a state of being free from defects, deficits, and substantial variations. Also, the diseases in leafs possess threats to the economic, and production status in the agricultural industry worldwide. The identification of disease in leafs using digital image processing, decreases the dependency on the farmers for the protection of agricultural products. So, the leaf disease detection and classification is the motivation of the proposed work. In this paper, a novel fuzzy set extended form neutrosophic logic based segmentation technique is used to evaluate the region of interest. The segmented neutrosophic image is distinguished by three membership elements: true, false and intermediate region. Based on segmented regions, new feature subset using texture, color, histogram and diseases sequence region are evaluated to identify leaf as diseased or healthy. Also, 9 different classifiers are used to monitor and demonstrate the discrimination power of combined feature effectiveness, where random forest dominates the other techniques. The proposed system is validated with 400 cases (200 healthy, 200 diseased). The proposed technique could be used as an effective tool for disease identification in leafs. A new feature set is promising and 98.4% classification accuracy is achieved.

1. Introduction

Leafs are the major ingredients in traditional medicinal drugs. World Health Organization (WHO) has estimated that approximately 80% of the world population still relies on traditional medicines, which are mostly plant-based drugs [1]. Although researchers have worked intensively to identify the diseases of plant leafs using various techniques like DNA/RNA, polymerase chain reaction, sensors techniques etc. [2] but the domain of computer vision to recognize the symptoms of diseases in medicinal plant leafs, still remains less explored. The objective of this paper is to present a computer vision-based approach for detecting basil leaf healthy or disease. Basil, an ancient and popular herbal plant is characterized with significant health benefiting phytonutrients. Basil has a profound significance in medicine and religious prospective. 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 [3]. Basil is indigenous to the countries of to Iran, India as well as other tropical regions of Asia [4], contains the essential oil and oleoresin required for manufacturing perfumes, food flavours, and aromatherapy products [5]. It has been used for around 300 different herbal treatments to support healthy response to stress, energy booster, increase stamina, healing properties, promotes cardiovascular health, cancer, heart diseases etc. The proposed paper is divided into two parts; the preliminary phase develops new segmentation technique based on Neutrosophic logic while another phase describes new features extraction method. Based on these two phases, the classifier will categorize leaf as healthy or diseased. The data base for the proposed system contains healthy and infected basil leaf images. The work represented in this paper is divided into 5 sections. Section 2 gives a brief review of the literature. Section 3 represents the technique and measures including novel segmentation and feature extraction technique. The features accuracy is evaluated using nine different classifiers; it is defined in Section 4. Experimental results are illustrated in Section 5. Conclusion is discussed in Section 6.

2. Literature review

The conventional method of identification and determination of the disease in medicinal leafs is manual. However, this manual

* Corresponding author.

https://doi.org/10.1016/j.measurement.2018.12.027
0263-2241/

Table 1
Survey of leaf disease detection and classifications system [9–40].

Culture	Features	Techniques	No. of images considered	Image acquisition device/dataset	Accuracy	Researchers
Sunflower & oat leafs	Area, size Diameter	Thresholding, Tracer algorithm	Up to 20 for each disease	TMC-76 color CCD	NA	Tucker and Chakraborty et al. [9]
Maize leafs	Color index	Iterative method	720 images	Digital camera, MS3100, captured from greenhouse at Embrapa Corn & Sorghum, SeteLagoas, Brazil	94·72%	Sena et al. [10]
Citrus leafs	Intensity and texture features	Discriminant classifier	Total 40 images 20 images [training] 20 images [testing]	3 CCD camera (JAI, MV90) captured from central Florida	96%	Pydipati et al. [11]
Orchid leafs	Texture and color features	ANN	289 images 145 images [training] 144 images [testing]	CCD (coupled-charge device) color camera (XC-711, Sony, Japan) Taiwan Sugar Research Institute, Tainan.	89.6%	Huang et al. [12]
Cotton crop	Co-occurrence matrix and fractal dimension	SVM	117 images	The Department of Entomology, at the University of Georgia, USA.	90%	Camargo et al. [13]
Rice leafs	Texture features	SVM	216 images	CCD color camera (Nikon D80, zoom lens 18–200 mm, F3.5–5.6, 10.2 Mega pixels) in the rice field of China National Rice Research Institute, located in Fuyang.	97.2%	Yao et al. [14]
Citrus leafs	Color and texture features	AdaBoost	500 images	Digital camera Sony DSCP92 and Canon EOS350D. samples collected from orange plants in winter in 2005 and 2006 from Guangdong China and in spring in 2007 from Guangxi province, China	88%	Zhang et al. [15]
Wheat and Grapes leafs	Color, Texture and Shape features	GRNN, PNN	Total 185 images 110 [training] 75 [testing]	Common digital camera	94.29%	Wang et al. [16]
Cereals	CCM	ANN SVM	750 images	Images collected from University of Agricultural Sciences, Dharwad, India	68.5% and 87% [ANN] 77.5% and 91.16% [SVM]	Pujari et al. [17]
Rose, beans, lemon and banana leafs	Texture and color features	SVM	500 images	Digital color camera. Images collected from Tamil Nadu	94%	Arivazhagan et al. [18]
Sugarcane leafs	Texture features	SVM	A set of images [training] 30 images [testing]	Digital color camera. Images are collected from sugarcane fields Indonesia	80%	Ratnasari et al. [19]
Rice leaf	Color features	ANN	134 images	Common digital camera. Images captured at Greenhouse of the International Rice Research Institute located at Los Banos, Laguna, Philippine	100%	Orillo et al. [20]
Tomato leafs	Texture features	PSO	NA	Digital color camera	NA	Muthukannan et al. [21]
Tomato leafs	Wavelets based features	SVM	200 im ages 10-fold cross validation	Digital color camera, captured at different farms in baniseef	78% [Invumult] 98% [Laplacian] 88% [Cauchy]	Mokhtar et al. [22]
Pomegranate leafs	Color and CCV features	SVM	610 images	Digital color camera	82%	Bhange et al. [23]
Cucumber leafs	Global and local singular values	SVM	100 images	Agricultural demonstration zone of Northwest A&F University	92% approx.	Zhang et al. [24]
Alfalfa leaf	Texture, shape and color features	SVM	899 images	Digital color camera. Images taken from LangfangForage Experimental Base, Institute of Animal Science, Chinese Academy of Agricultural Sciences & alfalfa fields Hebei Province, China	80% approx.	Qin et al. [25]
Betel vine	Color features	Otsu thresholding	12 images	CanoScanLiDE 110 [Scanner] Dhara, Rajnandgaon district of Chhattisgarh	NA	Dey et al. [26]
Vegetable crops	Color, shape and texture features	Combination of two SVM classifiers	284 images	Digital color camera	87.80%	Youssef Es-saady et al. [27]
Citrus leaf	Color histogram and texture features	Bagged tree classifier	199 images	DSLR camera	99.9%	Ali et al. [28]

(*continued on next page*)

Table 1 (*continued*)

Culture	Features	Techniques	No. of images considered	Image acquisition device/dataset	Accuracy	Researchers
Brinjal, broad beans, cucumber, ridge guard, spinach and tomato leaf	Fractal and Color Correlogram features	KNN, PNN	500 images 250 images [training] 250 images [testing]	Digital camera Nikon D7000 (16MP)	75.04% [KNN] 71.24% [PNN]	Tippannavar et al. [29]
Various leaf	GLCM features	SVM, PSO	NA	NA	95.16–98.38%	Kaur et al. [30]
Okra leaf and bitter gourd leafs	Texture features	Naives Bayes classifier with entropy based discretization	79 images [okra] 75 images [bitter gourd]s	Nikon D5100, Agricultural deptt, Pallishree Ltd, West Bengal, India	95% [okra] 82.67% [bitter gourd]	Mondal et al. [31]
Cucumber leafs	Color features	Comprehensive color feature map	93 images	Nikon Coolpix S3100 CCD camera, 69 images from Tianjin Academy of Agricultural Sciences, China and 24 images captured from internet	NA	Ma et al. [32]
Basil and parsley leafs	Statistical features	Neural networks	15 images of each category	Digital camera	80% [classification] 100% [recognition]	AL-Otaibi et al. [33]
Apple leaf	GLCM features	SVM	NA	NA	98.46%	Manimegalai et al. [34]
Plant leaf [leaf category not mentioned]	Region growing algorithm	Radial Basis Function Neural Network	6 images for first dataset and 270 images for second dataset	Plant village diseases classification challenge	86.21%	Chouhan et al. [35]
Apple and cucumber leafs	Pyramid of histogram of orientation gradient	Super pixel clustering, k-means clustering	150 images [apple leafs] 150 images [cucumber leafs]	Agricultural demonstration of district Yangling, China	90.43% [apple] 92.15% [cucumber]	Zhang et al. [36]
Wheat	Super pixel based tile extraction	Deep convolution neural networks	8178 images	Captured with various mobile phones. Pilot sites of Spain and Germany under natural conditions	Greater than 98%	Picon et al. [37]
Different leafs, fish and Kimia	Shape features	RNN	Fish = 11000 images Kimia = 99 images Leaf = 600images Rotated leaf = 3600images Scaled leaf = 2400 images Noised leafs = 2400images	NA	Fish = 99.07% Kimia = 97.98% Leaf = 84.67% Rotated leaf = 87.67% Scaled leaf = 88.92%	Junior et al. [38]
Citrus canker	Texture features	CLAHE, SVM	70 images [training] 30 images [testing]	Digital camera	NA	Sunny et al. [39]
Oil palm	Probability function	Naïve Bayes	NA	NA	80%	Nababan et al. [40]
Tomato leaf	Intensity features	Deep convolution neural networks	5000images	Digital camera, collected from South Korea	96%	Fuentes et al. [41]

Used Abbreviations; SVM: Support Vector Machine, QDA: Quadratic Discriminant Analysis, SOM: Self-Organizing Map, ANN: Artificial Neural Networks, PNN: Probabilistic Neural Networks, LDA: Linear Discriminant Analysis, PCA: Principal Component Analysis, GRNN: Generalized Regression Networks, RNN: Randomized Neural Network.

process is time-consuming, tedious and moreover very subjective [6]. In recent years, numerous methods were developed using computer vision to detect and classify agricultural and horticultural crops diseases to overcome the problems of manual techniques [7,8]. The basic approach for all of these methods includes image acquisition, feature extraction, feature selection and then classification analysis with parametric or non-parametric statistics. For effective operation of computer vision system, choice of the image-processing methods and classification strategies are chief concern. In literature, many efforts have been made to explain different modules of disease detection techniques for different agricultural/horticultural applications. A survey of the research work done during last few years on such leaf images is summarized in Table 1. The abbreviations used are summarized in the last row of Table 1.

3. Materials and methods

3.1. Data set collection and imaging set up

In the present work, the leaf dataset consists of four types of healthy and diseased basil leaf images; these are Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (holy basil) and Ocimum gratissimum (Vana-holy basil). These were collected from the herbs garden at Punjab Agriculture University Ludhiana, National Institute of Pharmaceutical Education and Research (NIPER) Mohali and Punjabi University Patiala, India for reflective study. A pictographic assessment of the above mentioned study site is shown in Fig. 1.

The images are taken to the research laboratory and cleaned for non-uniform distribution of dust to attain similar surface condition

for all leaf categories. After cleansing of the leaf samples, leafs are then taken to an imaging station and images of leaf 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. Leafs were digitally captured in color using EOS 5D Mark III, 22.3 megapixel CMOS sensor with resolution of 5760 × 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 (45 cm) over the center of the imaging station. A camera positioned vertically from the samples to contain all the components, with best possible resolution. The camera was stipulated on a camera stand which reduces the movement of hand and capturing uniform images of basil leafs. The degree of the damage caused in leafs varied between the leaf samples. Images were captured under controlled field conditions to reduce the unfavourable 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 the imaging station surface. Lamps (bulbs) with natural light reflectors located at 45 degree angle to ensure proper lighting. Fig. 2 represents the experimental set up of proposed system.

The database consists of 400 images which include 200 healthy and 200 diseased leafs of different categories of leafs i.e. Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (holy basil) and Ocimum gratissimum (Vana-holy basil). A hundred samples each for the four classes of leafs are collected. The diseases of leaf samples investigated are downy mildew, aphids, gray mold, bacterial leaf spot and fusarium wilt. Fig. 3 represents the healthy and diseased basil leafs.

3.2. System model

The system model is comprised of four essential steps as follows:

1. Preprocessing: The aim of preprocessing is to bring out details that are obscured with contrast limited adaptive histogram equalization method [42] for better contrast.
2. Segmentation: After preprocessing transform the image into neutrosophic domain, which segments the images into three different regions: True, False and Intermediate segments.
3. Feature extraction: Design a new feature pool based on segmented three regions to distinguish healthy and diseased leafs.
4. Classification: Nine different classifiers are used for final classification decision.

These four phases have been discussed in detail in the following sub-sections. The flow chart shown in Fig. 4 represents the proposed methodology.

3.2.1. Pre-processing

Quality of image is improved by adjusting the intensities of the image in order to highlight the target areas i.e. diseased visual area after data collection is completed. Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm is deployed for image enhancement, it works on small sections of the image instead of whole image. As the name suggests CLAHE algorithm applies the histogram equalization after partition the image into contextual regions [42]. It makes the hidden features of the image clearly visible and distribution of used gray values. Bilinear interpolation is used to combine the adjacent tiles for elimination of artificially induced

Fig. 1. Study sites of basil plants, from where different basil leafs were collected.

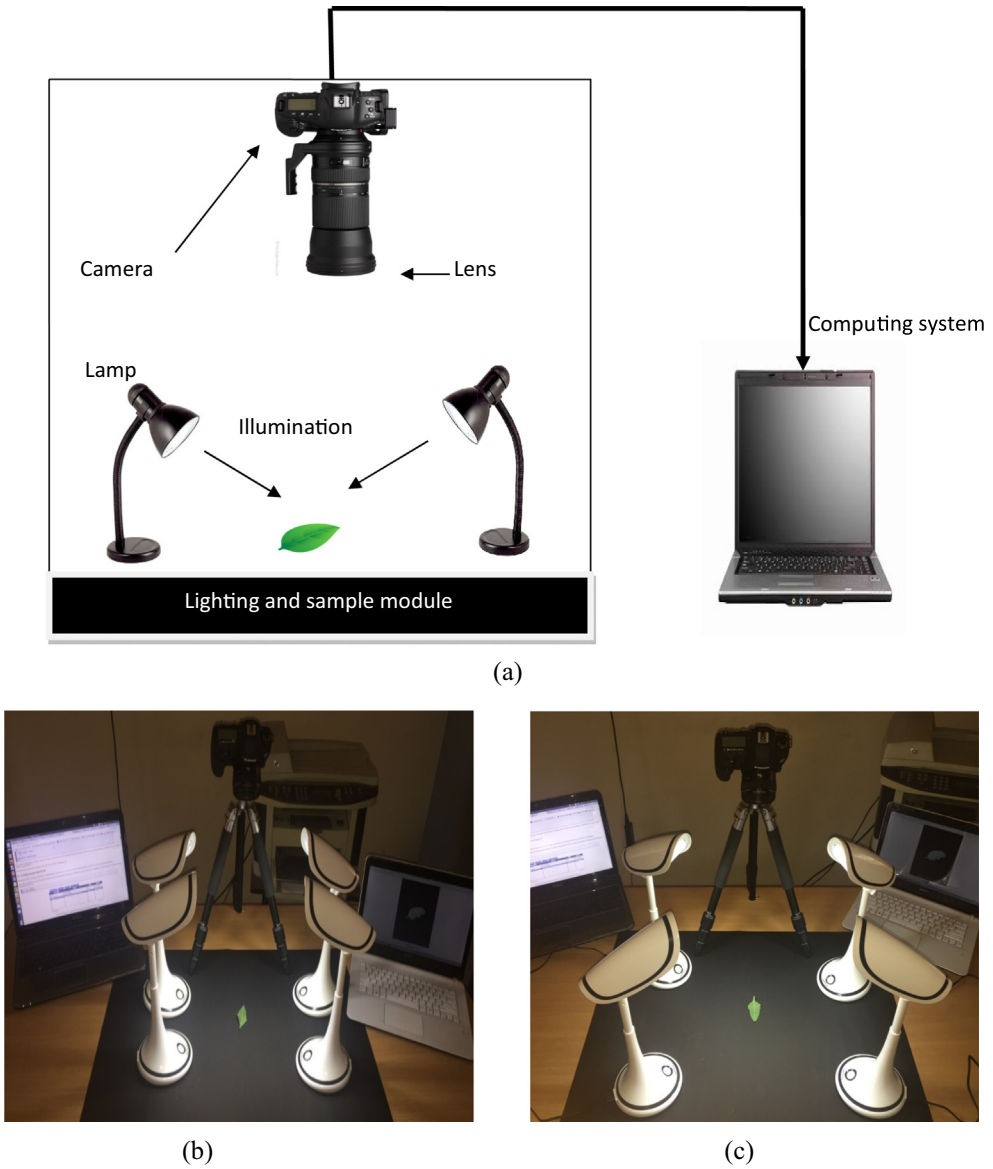

Fig. 2. a) Pictorial view of experimental set up, b) & c) Image acquisition system with different leaf structures.

boundaries. The contrast inhomogeneous' areas can be limited to avoid amplifying any noise that might be present in the image.

3.2.2. Segmentation technique

Image segmentation is a difficult task due to the complexity and diversity of images [7–43]. Factors such as illumination [44], contrast [45], and noise [46] etc. affect segmentation results. The goal of segmentation is to locate the suspicious areas to diagnose the 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 [47]. An image is represented using three different membership elements as(T), (I) and (F). Where T defines the truth scale, F as thescale of false and I represent the scale of intermediate. All elements considered are independent of each other. 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 [48].

3.2.2.1. Mapping T, F & I (Region of interest evaluation). In the proposed method, the diseased area of leaf 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. To evaluate diseased segment, the original image pixels are transformed from RGB to CIELab color space for better color perception as compared to the standard RGB space [49]. CIELab color space consists of 3 channels, as (L), (a)* and (b)*, where (L) channel represents lightness with values 0 (black) to 100 (white), positive values of (a^*) channel indicate amount of red while negative values indicate amount of green color opponent and (b*) channel, positive values indicate yellow and negative value indicates amount of blue. After enhancement and color transformation, T, I, F are mapped as follows:

1. To acquire unhealthy segment: Let the input image is$I_i(x,y)$. After contrast enhancement, it is represented as $I_c(x,y)$, then diseased segment $T_{IS}(x,y)$ is formularized as

$$T_{IS}(x,y) = I_C(x,y) \times F_a(x,y) \quad (1)$$

$F_a(x,y)$ is the binary mask obtained from (a*) chromaticity layer where, the color falls along the red-green axis. Where, (*) after L,

Fig. 3. Healthy and diseased basil leafs.

a, b pronounced star and it discriminate CIE version from Hunter's version. It performs bitwise multiplication. It separates the disease patches from different color populations. Diseases patches can be of yellow, brown, black and purple color. True section can eliminate healthy i.e. green section from the image. The flow diagram of a region of interest evaluation is shown in Fig. 5.

2. The healthy segment of leaf is evaluated as

$$F_{IS}(x, y) = 1 - T_{IS}(x, y) \tag{2}$$

where, $F_{IS}(x, y)$ represents a healthy segment of leafs. Healthy section represents the green color or section of leaf image.

3. Intermediate segment is considered as the stage which is not exactly diseased or healthy as well, we can consider it as onset disease. To evaluate intermediate portion, initially original image, $I_i(x, y)$ is transformed into CMYK color space $I_{cmyk}(x, y)$ for extracting yellow color [50] denoted as $I_y(x, y)$ in the leaf which is observed due to chemical changes, rust disease, and chlorophyll breakdown etc.

$$I_{IS}(x, y) = M_g(x, y) - M_y(x, y) \tag{3}$$

Further green color is extracted from original images$I_i(x, y)$ as $I_{green}(x, y)$. Where,$M_g(x, y)$and$M_y(x, y)$ are the masks that represent remaining portion of the leaf where yellow and green segmen are not considered. So, $T_{IS}(x, y)$. represents the degree of being a diseased segment, $F_{IS}(x, y)$ is the degree of being a healthy segment and $I_{IS}(x, y)$ is a degree of being not healthy not diseased as well. Fig. 6 represents the pictorial representation of extracted true, false and intermediate region.

3.3. Feature extraction

Feature extraction is to reduce the image data by measuring certain features or properties of each segmented regions [51]. Features are used to define the distinct characteristics of an image [52]. After image segmentation, the next important task is to extract the useful features of the image in order to diagnose the disease. We use new feature pool illustrated in the following subsections with details. Feature table exhibits histogram information content, damage structure index, disease sequence region and bin binary pattern features. The catalogue of features are illustrate in Table 2.where:

x, y = Pixel location
T_i = Pixel count of diseased area
I_i = Pixel count of on set off diseases area
F_i = Healthy region pixel count of leaf
G_K = Centre value,
G_C = Neighbourhoods pixel
K = Number of pixel in the neighbourhood.

3.3.1. Histogram information content (HIC)

Histogram is easy to compute and effective in characterizing both global and local distributions of colors in an image. Histogram information content defines the relative information content by finding the probability of occurrence of relative information about each plane in the image. Information will vary for every leaf for each plane. HIC is defined as

$$HIC = log\left(\frac{1}{histogram\ information\ contents\ of\ segmented\ regions}\right) \tag{4}$$

HIC is evaluated for all three segments T_{IS}, F_{IS}, I_{IS} for each red, green and blue plane.

3.3.2. Disease sequence region (DSR)

Disease Sequence Region (DSR) defines the correlation of individual neighbouring pixels with perceived pixel difference of the image, that is, pixel deviations between neighbouring pixels, refer Eqs. (5) and (6). We have calculated DSR for every extracted region (red, green and blue) of image vertically and horizontally. The DSR

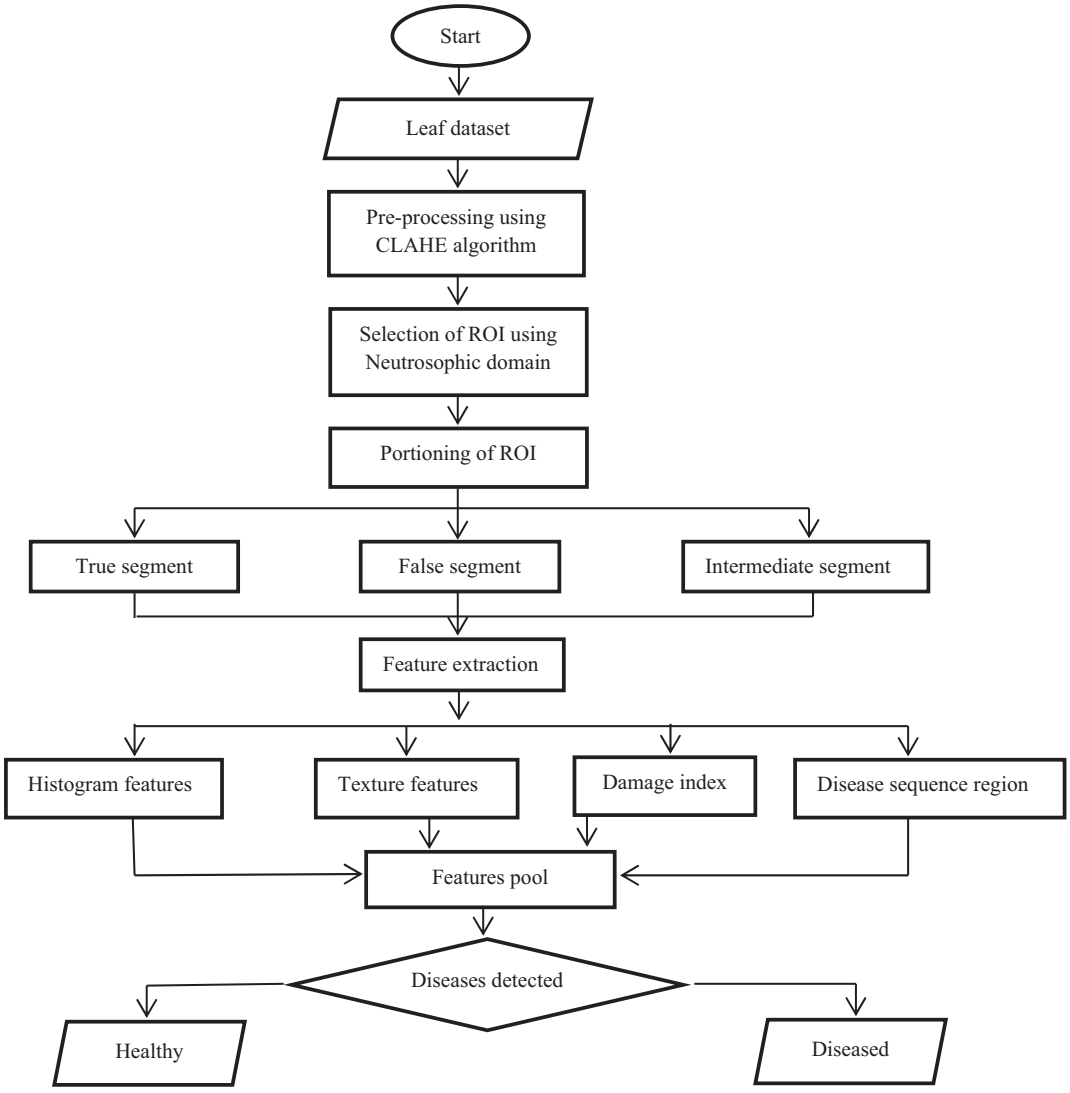

Fig. 4. Flow chart of system model for proposed method.

defined for vertical and horizontal orientation are given in Eqs. (5) and (6) as

Vertical deviation of intensity

$$DSR_{(V)} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \left(P_{(x,y+\Delta y)} - P_{(x,y)}\right) \tag{5}$$

Horizontal deviation of intensity

$$DSR_{(H)} = \sum_{x=0}^{m-1} \sum_{y=0}^{n-1} \left(P_{(x+\Delta x,y)} - P_{(x,y)}\right) \tag{6}$$

$$DSR = \left[DSR_{(V)} DSR_{(H)}\right] \tag{7}$$

where, x and y defines pixel location. Depending on horizontal and vertical deviations, we measure the deviation difference between healthy and non-healthy leaf.

3.3.3. Damage index (DI)

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

$$DI = \frac{\sum_{i=1}^{m} \sum_{i=1}^{n} (T_i + I_i)}{\sum_{i=1}^{m} \sum_{i=1}^{n} F_i} \tag{8}$$

where, T_i represents pixel count of diseased area, I_i represents pixel count of on set off diseases area and F_i represents healthy region pixel count of leaf and higher DI value indicates more diseased region. DI represents possible presence of damage (diseases) at leaf structure.

3.3.4. Bin binary pattern (BBP)

A new texture descriptor as BBP is introduced to describe the local structure information of leaf. The BBP linearly interpolates the pixel value of neighborhood to form an operator which defines the structure to distinguish all individual patterns (healthy or non-healthy leafs). 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 value and calculate the difference between the center pixel and neighboring pixels as defined in Eq. (9).

$$BBP_{(k,R)}(x_c, y_c) = \sum_{k=0}^{k} (G_k - G_c) 2^k \tag{9}$$

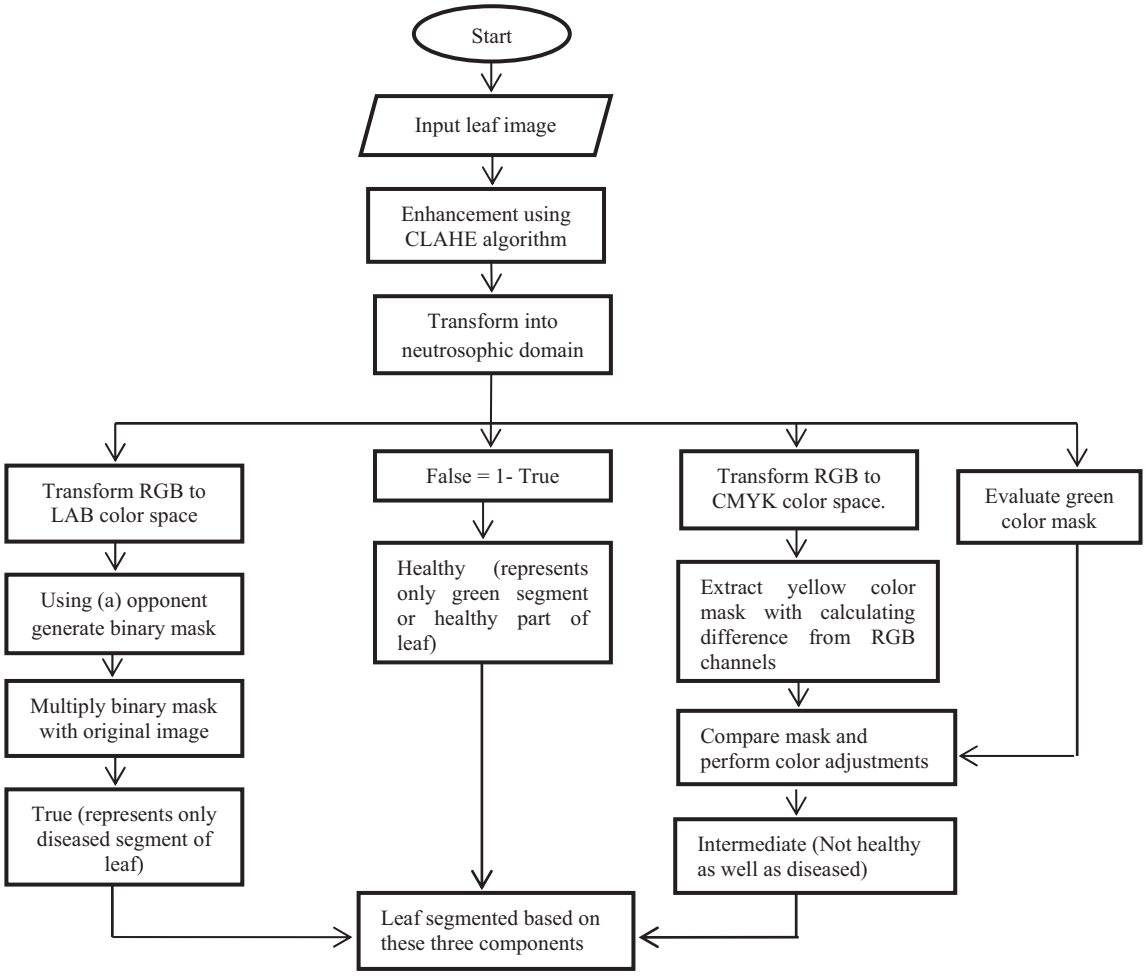

Fig. 5. Flow chart of system model for proposed segmentation method.

The weights for obtained different binary pattern are given in a clockwise direction starting from top-left and its corresponding values. Where, G_K represents centre value, G_C defines neighbourhoods pixel, x_C and y_C represents pixel value and K, defines number of pixel in the neighbourhood.

Algorithm of BBP is described as follows:

Input: *Input leaf image*
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 bins.
Step 4: Evaluate the difference of all neighbourhood bins w.r.t to centre value using Eq. (9).
Step 5: Assign weights
Step 6: Obtain unique decimal values

The whole procedure of BBP is defined briefly in Fig. 7 briefly step by step.

4. Classification

In this paper, we evaluate nine classifiers accuracies and effectiveness and select the best one. The brief summary of the classification models are listed below:

- Decision tree: Supervised learning algorithm estimates the significance of a target variable using numerous input variables [53].
- Random forest: Ensemble learning method, works using bagging method to construct a group of decision tree using random subgroup of the data [54].
- Support vector machine: Discriminative approach described by separating hyper plane that increases the boundary between the two classes [55].
- AdaBoost: Boosting approach where, multiple weak classifiers are engaged to make a single strong classifier [56].
- Linear models: Linear models analysis of covariance and single stratum analysis of variance [57].
- Naives Bayes: Supervised learning algorithm based on Bayes theorem with the naive assumption of independence between every pair of features [58]
- K-NN: Instance based learning, where data is classified based on stored and labeled instances according to some distance/similarity function [59].

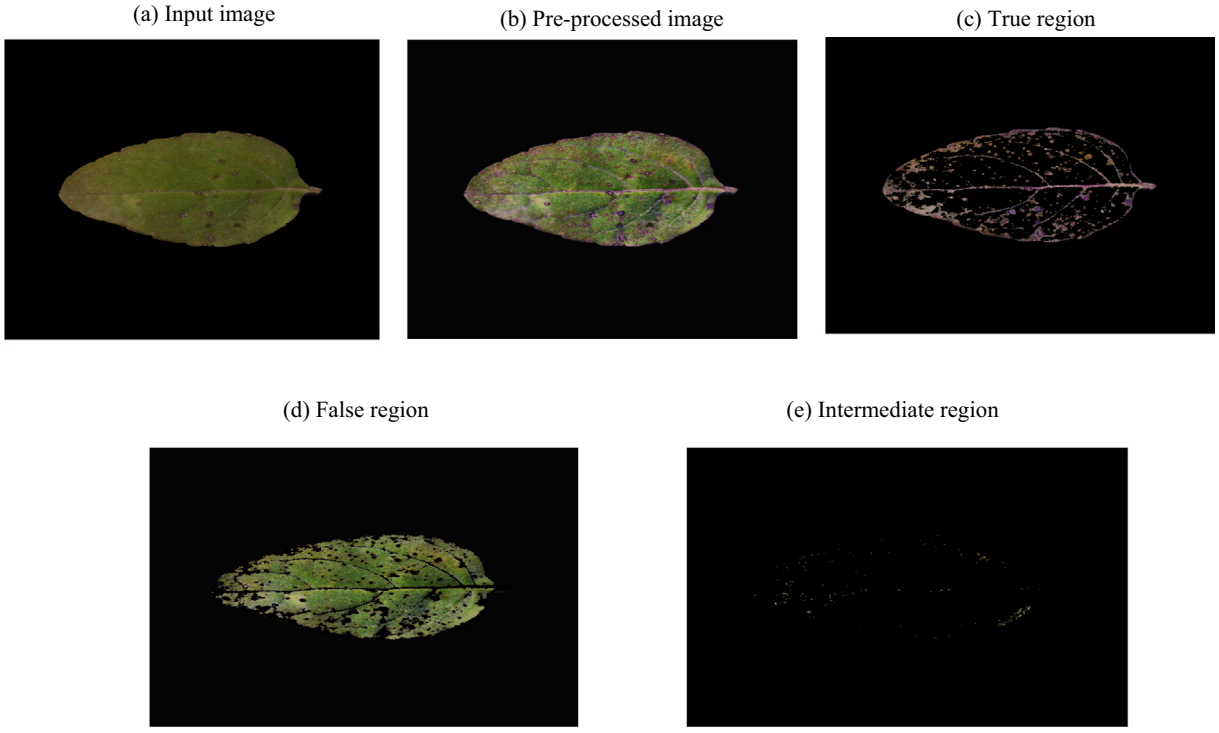

Fig. 6. Regions detection results: a) Represents original captured image, b) Pre-processing using CLAHE algorithm, c) True region which represents the diseases region, d) False region, where healthy region of leaf is presented, e) Intermediate region of the leaf.

Table 2
Catalogue of features.

Features	Expressions
Histogram information content	$HIC = log\left(\frac{1}{\text{histogram information contents of segmented regions}}\right)$
Disease sequence region	$DFR_{(H)} = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(P_{(x+\Delta x,y)} - -P_{(x,y)})$
	$DFR_{(V)} = \sum_{x=0}^{m-1}\sum_{y=0}^{n-1}(P_{(x,y+\Delta y)} - P_{(x,y)})$
Damage structure index	$DI = \frac{\sum_{i=1}^{m}\sum_{i=1}^{n}(T_i+I_i)}{\sum_{i=1}^{m}\sum_{i=1}^{n}(F_i)}$
Bin Binary pattern	$BBP_{(k,R)}(x_c, y_c) = \sum_{k=0}^{k}(G_k - G_c)2^k$

- Artificial neural networks: Mathematical model simulate data based on structure and functions of biological

  Neural networks [60,61].

- Discriminant analysis: Builds predicative model with analysis of regression and variance to define relationship between one dependent variable and one or more independent variable [62].

The tuning parameters of machine learning methods are tabulated in Table 3. It also indicates models, methods packages and platform used for calculating and finding parameters. On the basis of parameters, classifiers will categorize image as healthy or disease leafs.

5. Experimental results

5.1. Leaf images dataset

The database consists of 400 images which include 200 healthy and 200 diseased leafs of different categories ofleafs i.e. Ocimum sanctum (Kapoor basil), Ocimum tenuiflorum (Ram & Shyama basil), Ocimum basilicum (holy basil) and Ocimum gratissimum (Vana-holy basil). A hundred samples each for the four classes of leafs are collected. During our experiments, we use both Matlab (2015a) and R open (version 3.2.2) software tools on Sony Vaio Core i3, 2.2-GHz platform.

5.2. Classification model evaluation metrics

Different evaluation parameters were used to measure the performance of the classification process [63], defined as

$$\text{Precision(Positive predicted value)} = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$\text{Error rate} = \frac{FP + FN}{TP + TN + FP + FN} \tag{13}$$

$$\text{Specificity(Trueneagtiverate)} = \frac{TP}{TP + FP} \tag{14}$$

$$\text{Negative predicted valve (NPV)} = \frac{FP}{TN + FP} \tag{15}$$

where, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative

5.3. Feature discriminability test (Training-Testing Experiment)

In this section, we analyze the prediction results of nine machine learning methods on the basis of training-testing dataset described in Table 4. The distribution of data in the training-testing experiment is set to 70% and 30% respectively for all models.

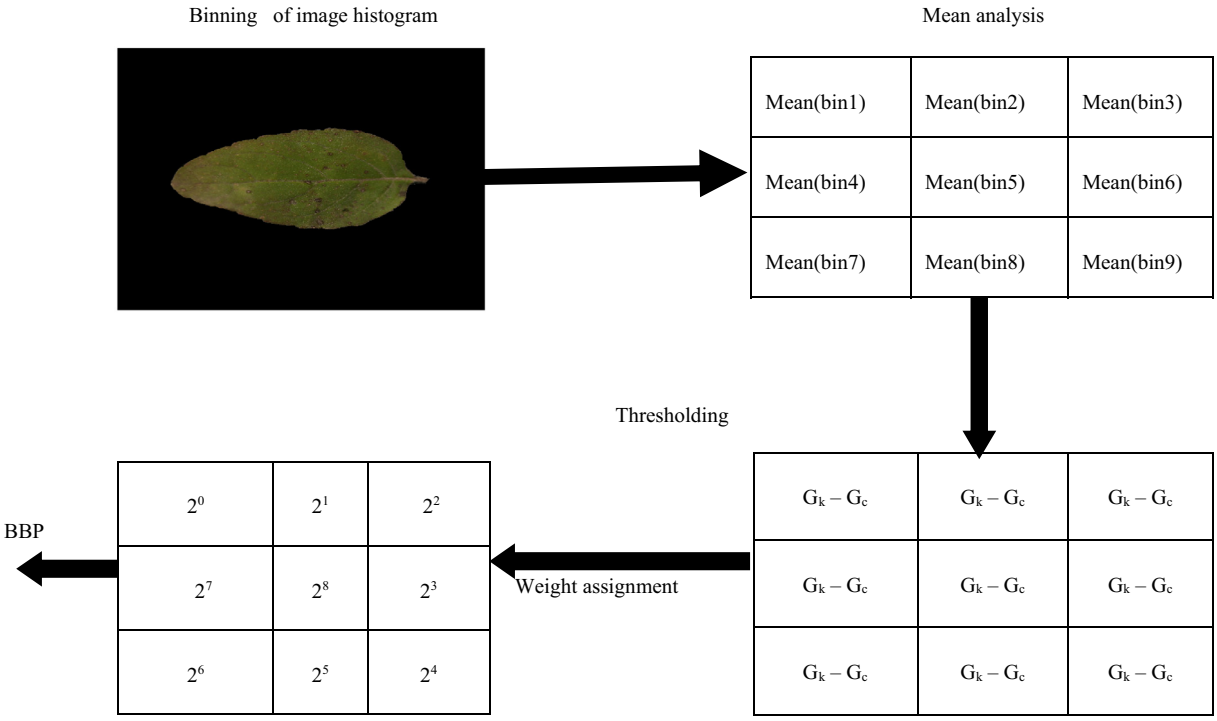

Fig. 7. Bin binary pattern.

Table 3
Tuning parameters of classifiers.

	Model	Method	Package	Tuning parameter(s)
R(version 3.2.2)	Decision Tree	Rpart	rpart	Min Split = 20, Max Depth = 30, Min Bucket
	Random forest	Rf	randomForest	Number of tree = 500
	Support Vector Machine	Svm	e1071	Kernel Radial Basis
	AdaBoost	Ada	Ada	Max Depth = 30, Min Split = 20, xval = 10,
	LM	Lm	Glm	Multinomial
	Artificial Neural networks	Neuralnet	Neuralnet	Hlayers = 10, MaxNWts = 10000, maxit = 100
Matlab(2015a)	Naives Bayes	NBModel	fitcnb	No.of observation = 150
	KNN	knn_model	fitcknn	NumObservations = 150, Distance= 'euclidean', NumNeighbours:5
	Discriminant Analysis	obj	fitcdisr	No. of observation = 150, DiscrimType: 'linear'

Table 4
Testing and training data distribution of healthy and diseased leafs.

Data set	Total samples	Training samples	Testing samples
Healthy	200	140	60
Diseased	200	140	60

Figs. 8–10 illustrates the classification accuracy of all classifiers with respect to various evaluation parameters as defined in Section 5.2. Compared to other machine learning models, random forest maintains a high accuracy.

Another performance measures are Receiver Operating Characteristics (ROC) and Area under the Curve (AUC). ROC is an efficient method for evaluating discrimination power of statistical model [64]. It plots the sensitivity versus specificity across the different

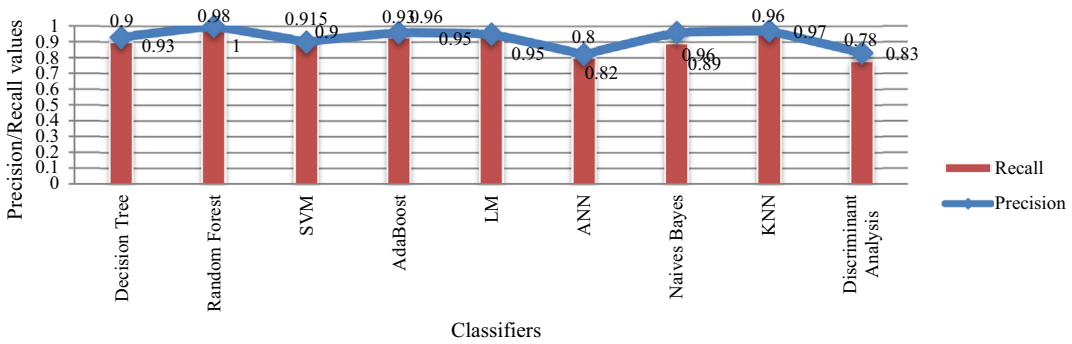

Fig. 8. Comparison of various classifiers in terms of precision and recall ratio.

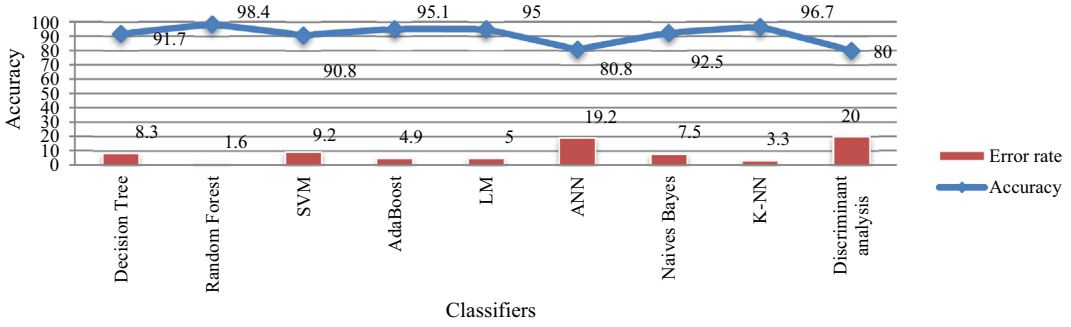

Fig. 9. Comparison of various classifiers in terms of accuracy and error rate.

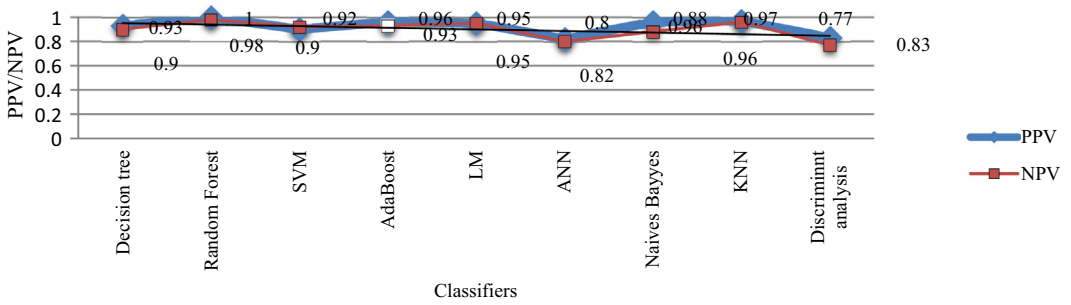

Fig. 10. Comparison of various classifiers in terms of PPV and NPV.

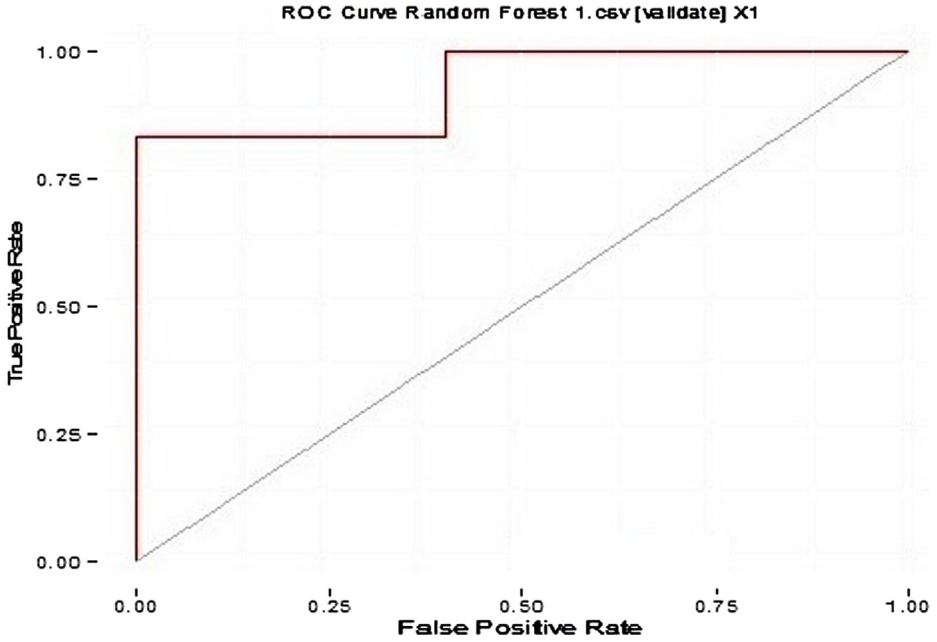

Fig. 11. ROC curve of random forest.

Table 5
Performance comparison on different testing –training partition.

Models	Training and testing partition evaluation			
Random Forest	50–50% 97.03%	60–40% 98.4%	70–30% 98.4%	80–20% 98.4%

Table 6
Performance comparison of various classifiers.

Classifiers	Proposed method (Accuracy)	Pydipati et al. [11] (Accuracy)	SFM Features [66] (Accuracy)	Law features [67] (Accuracy)
Decision tree	91.8%	88.9%	91.3%	77.5%
Naives 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%
Random Forest	98.4%	94.6%	91.9%	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%

possible threshold values. It provides the capability to access the performance of classifier. Where, AUC process the whole two dimensional area under the entire ROC curve. The AUC portray the probability that an indiscriminately selected positive example is accurately rated with greater suspicion than a randomly chosen negative example [65]. 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; an area of 0.5 represents a worthless test. The model performs better if an ROC curve is lifted up and away from the diagonal. Fig. 11 shows ROC curve of Random forest.

5.4. Performance of features

Furthermore, the accuracy is evaluated on 50–50, 60–40, 70–30 and 80–20 testing-training partition respectively to ensure its uniformity as illustrated in Table 5. Results show that Random forest performs well in all testing-training partition.

In the next experiment, we compare the classification accuracy of our proposed features with respect to the traditional feature extraction methods [11,66,67]. As shown in Table 6 proposed method gives better performance than other classifiers.

6. Conclusion

The main contribution of this paper is to successfully design new segmentation technique together with a new set of features. The whole procedure was described, respectively, from gathering images to segmentation and finally classification. Based on the segmentation new features have been extracted. These features combine the discrimination power of intensity and texture of leafs. Nine classifiers are used to measure the accuracy of proposed features. These proposed features give promising results and has been compared with existing feature extraction methods. The developed model was able to distinguish healthy and diseases leaf. Based on the graphical analysis, RF performs better than other machine learning models with 98.4% accuracy.

Future studies could focus on to extend proposed work to classify each diseases category individually and estimate the severity of the detected diseases. An undiscovered amalgamation of feature extraction, feature selection and learning methods can also be explored to enhance the efficacy of diseases detection and classification models.

Acknowledgments

The authors would like to thank Dr. J.I.S. Khattar, Professor and Head (Department of Botany), Punjabi University, Patiala, India, the faculty members of (Department of Botany) of Punjab Agriculture University, Ludhiana (India) and National Institute of Pharmaceutical Education and Research (NIPER), India for fruitful discussions.

References

[1] V. Pallavi, S. Suresh, Integration of organic farming practices in cultivation of ayurveda herbs: an innovative approach, Int. J. Herbal Med. 4 (1) (2016) 34–38.
[2] Yi Fang, R.P. Ramasamy, Current and prospective methods for plant disease detection, Biosensors 5 (3) (2015) 537–561.
[3] A.A. Izzo, F. Borrelli, R. Capasso, V.D. Marzo, R. Mechoulam, Non-psychotropic plant cannabinoids: new therapeutic opportunities from an ancient herb, Trends Pharmacol. Sci. 30 (10) (2009) 515–527.
[4] M.R. Azarakhsh, Z. Asrar, H. Mansouri, Effects of seed and vegetative stage cysteine treatments on oxidative stress response molecules and enzymes in Ocimum basilicum L. under cobalt stress, J. Soil Sci. Plant Nut. 15 (3) (2015) 651–662.
[5] A.K. Pandey, P. Singh, N.N. Tripathi, Chemistry and bioactivities of essential oils of some Ocimum species: an overview, Asian Pac J. Trop. Biomed. 4 (9) (2014) 682–694.
[6] R. Tadeusiewicz, P. Tylek, F. Adamczyk, P. Kiełbasa, M. Jabłoński, P. Pawlik, A. Pilat, J. Walczyk, J. Szczepaniak, T. Juliszewski, M. Szaroleta, Automation of the acorn scarification process as a contribution to sustainable Forest Management, Case Study Common Oak, Sustain. 9 (12) (2017) 2276.
[7] J.G.A. Barbedo, Digital image processing techniques for detecting, quantifying and classifying plant diseases, SpringerPlus 2 (1) (2013) 1–12.
[8] G. Dhingra, V. Kumar, H.D. Joshi, Study of digital image processing techniques for leaf disease detection and classification, Multimedia Tools Appl. 77 (15) (2018) 19951–20000.
[9] C.C. Tucker, S. Chakraborty, Quantitative assessment of lesion characteristics and disease severity using digital image processing, J. Phytopathol. 145 (7) (1997) 273–278.
[10] D.G. Sena Jr, F.A.C. Pinto, D.M. Queiroz, P.A. Viana, Fall armyworm damaged maize plant identification using digital images, Biosyst Eng. 85 (4) (2003) 449–454.
[11] R. Pydipati, T.F. Burks, W.S. Lee, Identification of citrus disease using color texture features and discriminant analysis, Comput. Electron. Agric. 52 (1–2) (2006) 49–59.
[12] K.Y. Huang, Application of artificial neural network for detecting phalaenopsis seedling diseases using color and texture features, Comput. Electron. Agric. 57 (1) (2007) 3–11.
[13] A. Camargo, J.S. Smith, An image-processing based algorithm to automatically identify plant disease visual symptoms, Biosyst. Eng. 102 (1) (2009) 9–21.
[14] Q. Yao, Z. Guan, Y. Zhou, J. Tang, Y. Hu, B. Yang, Application of support vector machine for detecting rice diseases using shape and color texture features, in: Proceedings of the IEEE International Conference on Engineering Computation (ICEC), 2009, pp. 79–83.
[15] M. Zhang, Q. Meng, Automatic citrus canker detection from leaf images captured in field, Pattern Recog. Lett. 32 (15) (2011) 2036–2046.
[16] H. Wang, G. Li, Z. Ma, 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 (ICNC), 2012, pp. 246–251.
[17] J.D. Pujari, R. Yakkundimath, A.S. Byadgi, Classification of fungal disease symptoms affected on cereals using color texture features, Int. J. Signal Process. 6 (6) (2013) 321–330.
[18] S. Arivazhagan, R.N. Shebiah, S. Ananthi, S.V. Varthini, Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, Agric Eng Int. CIGR J. 15 (1) (2013) 211–217.
[19] E. K. Ratnasari, M. Mentari, R. K. Dewi, 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 (ICTS), 2014, pp. 93–98.
[20] J.W. Orillo, J. D. Cruz, L. Agapito, P. L. Satimbre, I. Valenzuela, Identification of diseases in rice plant (Oryza Sativa) using back propagation artificial neural network, in: Proceedings of the IEEE International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management, 2014, pp. 1–6.
[21] K. Muthukannan, P. Latha, A PSO model for disease pattern detection on leaf surfaces, Image Anal. Stereol. 34 (2015) 209–216.
[22] U. Mokhtar, M. A. Ali, A. E. Hassenian, H. Hefny, Tomato leaves diseases detection approach based on support vector machines, in: Proceedings of the

IEEE International Computer Engineering Conference (ICENCO), 2015, pp. 246–250
[23] M. Bhange, H.A. Hingoliwala, Smart Farming: Pomegranate disease detection using image processing, in: Proceedings of the Second International Symposium on Computer Vision and the Internet (VisionNet), 2015, pp. 280–288.
[24] S. Zhang, Z. Wang, Cucumber disease recognition based on Global-Local Singular value decomposition, Neurocomputing 205 (2016) 341–348.
[25] F. Qin, D. Liu, B. Sun, L. Ruan, Z. Ma, H. Wang, Identification of alfalfa leaf diseases using image recognition Technology, PLoS ONE 11 (12) (2016) 1–26.
[26] A. K. Dey, M. Sharma, M. R. Meshram, Image processing based leaf rot disease, detection of betel vine (Piper BetleL.), in: Proceedings of the International Conference on Computational Modeling and Security (CMS), 2016, pp. 748–754.
[27] Y. Es-saady, I. El Massi, M. El Yassa, D. Mammass, A. Benazoun, Automatic recognition of plant leaf diseases based on serial combination of two SVM classifiers, in: Proceedings of the Second International Conference on Electrical and Information Technologies (ICEIT), 2016, pp. 561–566.
[28] H. Ali, M.I. Lali, M.Z. Nawaz, M. Sharif, B.A. Saleem, Symptom based automated detection of citrus diseases using color histogram and textural descriptors, Comput. Electron. Agric. 138 (2017) 92–104.
[29] S. Tippannavar, S. Soma, A machine learning system for recognition of vegetable plant and classification of abnormality using leaf texture analysis, Int. J. Sci. Eng. Res. 8 (6) (2017) 1558–1563.
[30] P. Kaur, S. Singla, S. Singh, Detection and classification of leaf diseases using integrated approach of support vector machine and particle swarm optimization, Int. J. Adv. Appl. Sci. 4 (8) (2017) 79–83.
[31] D. Mondal, D.K. Kole, K. Roy, Gradation of yellow mosaic virus disease of okra and bitter gourd based on entropy based binning and Naive Bayes classifier after identification of leaves, Comput. Electron. Agric. 142 (2017) 485–493.
[32] J. Ma, K. Du, L. Zhang, F. Zheng, J. Chu, Z. Sun, A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing, Comput. Electron. Agric. 142 (2017) 110–117.
[33] M.B. AL-Otaibi, A.S. Ashour, N. Dey, R. Abdullah, A. A. AL-Nufaei, S. Fuqian, Statistical image analysis based automated leaves classification, in: Proceedings of the 2nd International Conference on Information Technology and Intelligent Transportation Systems (ITITS), 2017, vol. 296, pp. 469.
[34] S. Manimegalai, G. Sivakamasundari, Apple leaf diseases identification using support vector machine, in: Proceedings of the International Conference on Emerging Trends in Applications of Computing (ICETAC), 2017, pp. 1–4.
[35] S. S. Chouhan, A. Kaul, U. P. Singh, S. Jain, Bacterial foraging optimization based Radial Basis Function Neural Network (BRBFNN) for identification and classification of plant leaf diseases: An automatic approach towards Plant Pathology, IEEE Access 6 (2018) pp. 8852–8863.
[36] S. Zhang, H. Wang, W. Huang, Z. You, Plant diseased leaf segmentation and recognition by fusion of superpixel K-means and PHOG, Optik 157 (2018) 866–872.
[37] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild, Comput. Electron Agric (2018).
[38] J.J.D.M.S. Junior, A.R. Backes, O.M. Bruno, Randomized neural network based descriptors for shape classification, Neurocomputing 312 (2018) 201–208.
[39] S. Sunny, M.P.I. Gandhi, An efficient citrus canker detection method based on contrast limited adaptive histogram equalization enhancement, Int. J. Applied Engg. Res. 13 (1) (2018) 809–815.
[40] M. Nababa, Y. Laia, D. Sitanggang, O. Sihombing, E. Indra, S. Siregar, W. Purba, R. Mancur, The diagnose of oil palm disease using naive bayes method based on expert system technology, J. Phys. Conf. Ser. 1007 (1) (2018) 1–5.
[41] A.F. Fuentes, S. Yoon, J. Lee, D.S. Park, High-performance deep neural network based tomato plant diseases and pests diagnosis system with refinement filter bank, Front. Plant Sci. 9 (2018).
[42] B.S. Min, D.K. Lim, S.J. Kim, J.H. Lee, A novel method of determining parameters of CLAHE Based on Image Entropy, Int. J. Softw. Eng. Its Appl. 7 (5) (2013) 113–120.
[43] Z. Wang, K. Wang, F. Yang, S. Pan, Y. Han, Image segmentation of overlapping leaves based on chan-vese model and sobel operator, Informat. Process. Agric. 5 (1) (2018) 1–10.
[44] D. Pratiwi, K.H. Kartowisastro, Object Segmentation under Varying Illumination Effects, in: D. Barbucha, N. Nguyen, J. Batubara (Eds.), New Trends in Intelligent Information and Database Systems Studies in Computational Intelligence, Springer, Cham, 2015, pp. 13–21.
[45] Y. Zhang, L. Ge, New approach to low contrast image segmentation, in: Proceedings of the Second IEEE International Conference on Bioinformatics and Biomedical Engineering (ICBBE), 2008, pp. 2369–2372.
[46] S. Rital, H. Cherifi, 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
[47] Y. Guo, H.D. Cheng, New neutrosophic approach to image segmentation, Pattern Recog. 42 (5) (2009) 587–595.
[48] U. Rivieccio, Neutrosophic logics: Prospects and problems, Fuzzy Sets Syst. 159 (14) (2008) 1860–1868.
[49] C. Connolly, T. Fleiss, A study of efficiency and accuracy in the transformation from RGB to CIELAB color space, IEEE Trans. Image Process. 6 (7) (1997) 1046–1048.
[50] R. Gonzalez, R. Woods, Digital image processing, third ed., Prentice Hall, New York, 2008.
[51] I. Guyon, S. Gunn, M. Nikravesh, L.A. Zadeh, Feature Extraction: Foundations and Applications, Springer, 2008.
[52] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: Proceedings of the IEEE Science and Information Conference (SAI), 2014, pp. 372–378.
[53] J.R. Quinlan, Induction of decision trees, Mach Learn 1 (1) (1986) 81–106.
[54] L. Breiman, Random forests, Mach Learn 45 (2001) 5–32.
[55] K. P. Soman, R. Loganathan, V. Ajay, Machine learning with SVM and other kernel methods, PHI Learning, India, 2009
[56] P.L. Bartlett, M. Traskin, AdaBoost is consistent, J. Mach. Learn. Res. 8 (2007) 2347–2368.
[57] J. Kim, A. Mowat, P. Phoole, N. Kasabov, Linear and non-linear pattern recognition models for classification of fruit from visible-near infrared spectra, Chemometr. Intell. Labor. Syst. 51 (2) (2000) 201–216.
[58] N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers, Mach. Learn. 29 (2–3) (1997) 131–163.
[59] M.F. Unlersen, K. Sabanci, The classification of diseased Trees by Using KNN and MLP Classification models according to the satellite Imagery, Int. J. Intell. Syst. Appl. Eng. 4 (2) (2016) 25–28.
[60] G.P. Zhang, Neural networks for classification: a survey, IEEE Trans. Syst. Man, Cybern. Appl. Rev. 30 (4) (2000), 451–462.I.
[61] A. Głowacz, W. Głowacz, Shape recognition of film sequence with application of sobel filter and backpropagation neural network, in: Human-Computer Systems Interaction, Springer, Berlin, Heidelberg, 2009, pp. 505–516.
[62] A.J. Izenman, A. Julian, Linear discriminant analysis, in: A.J. Izenman (Ed.), Modern Multivariate Statistical Techniques, Springer, New York, 2013, pp. 237–280.
[63] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manage. 45 (4) (2009) 427–437.
[64] T. Fawcett, An introduction to ROC analysis, Pattern Recog. Lett. 27 (8) (2006) 861–874.
[65] B. Song, G. Zhang, W. Zhu, Z. Liang, ROC operating point selection for classification of imbalanced data with application to computer-aided polyp detection in CT colonography, Int. J. Comput. Assist. Radiol. Surg. 9 (1) (2014) 79–89.
[66] C.M. Wu, Y.C. Chen, Statistical feature matrix for texture analysis, Graph. Models Image Process. 54 (5) (1992) 407–419.
[67] A.S. Setiawan J. Elysia Y. Purnama Wesley Mammogram classification using Law's texture energy measure and neural networks in: Proceedings of the International Conference on Computer Science and Computational Intelligence (ICCSCI) 2015 92 97