CrossMark

Image processing based rice plant leaves diseases in Thanjavur, Tamilnadu

T. Gayathri Devi[1] · P. Neelamegam[2]

Abstract

India is a highly populated agricultural country which has land area of 60.3% for agriculture purpose. The production of rice plant is decreased up to 20–30% because of various diseases. The most frequent diseases occurred in paddy leaves are leaf blast, leaf blight, false smut, brown spot and leaf streak. This paper mainly considers a method to detect the leaf diseases automatically using image processing techniques. To determine these diseases, the proposed methodology involves image Acquisition, image pre-processing, segmentation and classification of paddy leaf disease. In this proposed system, the features are extracted using hybrid method of discrete wavelet transform, scale invariant feature transform and gray scale co-occurrence matrix approach. Finally, the extracted features are given to various classifiers such as K nearest neighborhood neural network, back propagation neural network, Naïve Bayesian and multiclass SVM to categorize disease and non-disease plants. Many classification techniques are examined to classify the leaf disease. In experimental result, the proposed work is implemented in MATLAB software and performance of this work is measured in terms of accuracy. It is observed that multi class SVM provides the better accuracy of 98.63% when compared to other classifiers.

Keywords Rice plant leaf diseases · K-means clustering · DWT · SIFT · GLCM · Multi class SVM

1 Introduction

Agriculture is the backbone of our Indian economy. Food is essential for all humans. Population of our country is growing rapidly and so the need of food is also increased. Improvement of our agriculture production is needed to satisfy the food scarcity [1]. The quantity and quality of the plant gets affected due to diseases. Some of the diseases are leaf blast, leaf blight, false smut, brown spot and leaf streak. These diseases are occurred due to bacteria, virus and fungus. Sometimes it may occur due to lack of essential minerals [2].

It is essential to detect and classify the rice leaf disease in early stages before it become severe. Rice leaf disease detection requires huge amount of work, knowledge in the plant diseases and also requires more pre-processing time. Incorrect identifying plant disease leads to huge loss of yield, time, money and quality of product. Farmers have to do constant monitoring of the plants. For successful cultivation the identification is an important task and to change this monitoring ease, automatic analyzation of field is essential.

Identification and categorization is done through physical techniques; requires the experts help, sometimes it will be error prone and who are less available. The classification by labours based on color, size etc. So they should be tested via non-destructive techniques. Hence, the classification of leaf disease is necessary in evaluating agricultural produce, increasing market value and meeting quality standards. Identifying and taking further dealings for further diffusion of the diseases is also helpful. Image processing techniques are used for detection of diseases at early stage [3].

✉ T. Gayathri Devi
    devigayathri77@src.sastra.edu

    P. Neelamegam
    neelekeer@eie.sastra.edu

1   Department of ECE, School of EEE, SRC, SASTRA Deemed University, Kumbakonam, India

2   Department of E&I, School of EEE, SASTRA Deemed University, Thanjavur, India

This paper describes the proposed system. In Section 3, which involves pre-processing, Segmentation details, the process of feature extraction. The various classifiers are detailed in Section 4. The results and discussion are discussed in Section 5. Conclusion of the work is described in Section 6.

1.1 Paddy diseases

Disease in rice plant can highly minimizes the yield. The diseases are mainly occurred by bacteria, viruses, or fungi. The common diseases of rice plants are bacterial leaf blight, leaf blast, leaf brown spot, narrow brown spot, leaf scald, bacterial leaf streak, false smut, red stripe, stem rot, sheath blight, etc. In this paper concentrates the five diseases namely, leaf blast, bacterial leaf blight, brown spot, leaf streak and false smut. Figure 1 depict the visual symptoms of the leaf diseases.

1.1.1 Bacterial leaf blight

Bacterial Blight disease in paddy is produced by *Xanthomonas oryzae* pv. oryzae. which induces wilting of shrubs and leaves to dry and yellowing. This disease is mainly developed in the weeds and stubbles of affected region of the plant. It causes in the areas of lowland and irrigated regions.

In young stages, infected leaves become grayish green and roll up. Then at the progresses of disease, the leaves changes from yellow to straw-colored and wilt, whole seedlings tends to make whole seedling dry up and lead to death.

1.1.2 Leaf brown spot

Brown Spot is created by a fungus called *Cochliobolus miyabeanus*. Its observable symptom is the number of big spots on the leaves which can damage the entire leaf. If the infection occurs in young seed the unfilled or spotted or discolored seeds are generated. As the lesions become large, they retains circle, with central area necrotic gray and the lesion borders change from reddish-brown to dark brown.

1.1.3 Leaf blast

Leaf Blast is created by the *Magnaporthe oryzae* fungus which infect all the fields of a rice plant. Blast disease is created by places where the spores of blast are present. It causes low soil moisture areas, prolonged rain shower, and cool temperature parts. At initial stages it seems white to gray-green lesions or spots, with dark green margins. After grown the lesions become elliptic or rod shaped. These lesions have the border of white gray centers with red brown which looks in diamond shape.

1.1.4 Leaf streak

Leaf streak is generated by *Xanthomonas oryzae* pv. *oryzicola.* The affected rice plants leaves seems to be dry and brown. Bacterial leaf streak causes in high temperature and high humidity areas. Initial stages it is formed as small, water-soaked, linear lesions between leaf veins which are dark green in beginning after that changes to light brown to yellowish gray.

1.1.5 False smut

False smut created by grains silkiness which tends to decrease the grain weight. It also minimizes the germination of seeds. The false smut is developed in the areas of rain, high humidity, and high nitrogen soils. Wind also spread the disease spores from one plant to other plant. This is visible only after panicle exertion. It affects the plant at the flowering stage.

2 Related works

Phelps et al. proposed that the Studies have been undertaken in a 1000 ha area of irrigated rice (*Oryzica sativa*) at Caroni (1975) Limited to determine the effect of the

Fig. 1 Visual symptoms of paddy diseases

fertilizer programme on the incidence of important diseases. Over a period of 3 years higher levels of brown leaf spot (*Cochliobolus miyabeanus*) on rice varieties Oryzica 1 and Oryzica 5 on three different soil types were associated with increasing levels of leaf P, from a low of 0.149% of dry matter (DM) to a high of 0.396% DM. On the Washington silty clay loam series (Inceptisol) brown leaf spot incidence was lowest when leaf P was between 0.135 and 0.149% of DM. However, disease incidence was higher when leaf P levels fell to 0.133% of DM or rose above 0.149%, under conditions where N was more than adequate. The moderate levels of the disease experienced over the period had no effect on yield, as grain infection was minimal. The results support the conclusion that the incidence of brown leaf spot on irrigated rice at Caroni is influenced by sub-optimal levels particularly of P [4].

Arnal Barbedo presents a survey on methods that use digital image processing techniques to detect, quantify and classify plant diseases from digital images in the visible spectrum. Although disease symptoms can manifest in any part of the plant, only methods that explore visible symptoms in leaves and stems were considered. Methods dealing with roots, seeds and fruits have some peculiarities that would warrant a specific survey. The selected proposals are divided into three classes according to their objective: detection, severity quantification, and classification. Each of those classes, in turn, are subdivided according to the main technical solution used in the algorithm. This paper is expected to be useful to researchers working both on vegetable pathology and pattern recognition, providing a comprehensive and accessible overview of this important field of research [5].

Liu et al. proposed that BP neural network classifiers were designed for classifying the healthy and diseased parts of rice leaves. This paper select rice brown spot as study object, the training and testing samples of the images are gathered from the northern part of Ningxia Hui Autonomous Region. The result shows that the scheme is feasible to identify rice brown spot using image analysis and BP neural network classifier [6].

Orillo et al. proposed that digital image processing was incorporated to eliminate the Subjectiveness of manual inspection of diseases in rice plant and accurately identify the three common diseases to Philippine's farmlands: (1) bacterial leaf blight, (2) brown spot, and (3) rice blast. The image processing section was built using MATLAB functions and it comprises techniques such as image enhancement, image segmentation, and feature extraction, where four features are extracted to analyze the disease: (1) fraction covered by the disease on the leaf; (2) mean values for the R, G, and B of the disease; (3) standard deviation of the R, G, and B of the disease and; (4) mean values of the H, S and V of the disease. The back propagation Neural

Network was used in this project to enhance the accuracy and performance of the image processing. The database of the network involved 134 images of diseases and 70% of these were used for training the network, 15% for validation and 15% for testing. After the processing, the program will give the corresponding strategic options to consider with the disease detected. Overall, the program was proven 100% accurate [7].

Joshi et al.: a new technique to diagnose and classify rice diseases has been proposed in this paper. Four diseases namely rice bacterial blight, rice blast, rice brown spot and rice sheath rot have been identified and classified. Different features like shape, the color of a diseased portion of the leaf have been extracted by developing an algorithm. All the extracted features have been combined as per the diseases and diseases have been classified using minimum distance classifier (MDC) and k-nearest neighbor classifier (k-NN). The performance of the proposed technique has been evaluated with the help of 115 rice leaf images of four diseases and 70% image data has been used for training the classifier and 30 percent has been used for testing. Classification accuracy has been calculated for each disease using both classifiers. The overall accuracy achieved by using k-NN and MDC is 87.02 and 89.23% respectively [8].

Phadikar describes a software prototype system for rice disease detection based on the infected images of various rice plants. Images of the infected rice plants are captured by digital camera and processed using image growing, image segmentation techniques to detect infected parts of the plants. Then the infected part of the leaf has been used for the classification purpose using neural network. The methods evolved in this system are both image processing and soft computing technique applied on number of diseased rice plants [9].

Phadikar et al. proposed that the aims are at classifying different types of rice diseases by extracting features from the infected regions of the rice plant images. Fermi energy based segmentation method has been proposed in the paper to isolate the infected region of the image from its background. Based on the field experts' opinions, symptoms of the diseases are characterized using features like colour, shape and position of the infected portion and extracted by developing novel algorithms. To reduce complexity of the classifier, important features are selected using rough set theory (RST) to minimize the loss of information. Finally using selected features, a rule base classifier has been built that cover all the diseased rice plant images [10].

Shang et al. proposed a plant leaf recognition to overcome the drawbacks of large database, a utilizing two-stage local similarity based classification learning (LSCL). In the author summarized the different authentication techniques using local patterns in hand held devices. A leaf diseases

detection which using k means clustering to identify the affected areas and classify the disease type using neural network. Image segmentation for leaf region is described using Otsu method in and finally it is graded [11].

3 Proposed methodology

This proposed methodology is shown in the Fig. 2. The various processes are explained below. The core idea of this paper is to detect the various rice plant leaf diseases which lead to death of crop. The images are taken from the field is processed to extract the features and then it classified under various classification methods like KNN, ANN, Bayesian and multi class SVM classifiers [12]. Before that pre-process the image using median filter and segmented using k means clustering method for the process of feature extraction. The features are retrieved by using DWT and SIFT method. The GLCM method is also accompanied for the texture analysis of the leaf.

3.1 Input image

It is the initial process to obtain the rice plant leaf images. The rice plant images are captured with high resolution digital camera. Then those images are cropped into smaller
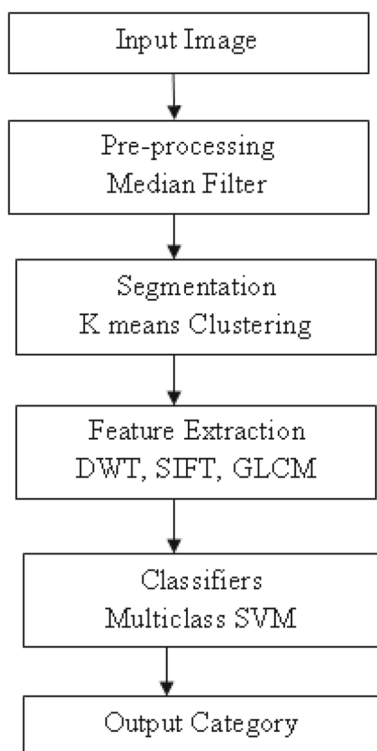
size. The images are stored in JPEG format. The Matlab image library is used to keep the image.

3.2 Image pre-processing

The image pre-processing is the process used to filter out the various noises and other object removal from the image. It enhances the image quality. It also used to resize the image for the purpose of display, storage and transformation of image. To enhance the quality of acquired image the median filter is used. Figure 3 shows input image and the image after the removal of noises from the captured image.

3.2.1 Median filter

Median filter is a kind of nonlinear digital filter which is often used to remove noises from the image. This type of noise filter effectively eliminates the type of noise known as salt and pepper noise. The principle of working is, the filter replaces each pixel in the input image by the median of gray level value of neighbor pixels. This filter is called smoothing spatial filter. In this filtering process, the median value is obtained by.

- Pixel values in the window are arranged into the order of numbers.
- After that he pixel value is replaced by median.

After filtration process of image segmentation is done which is explained below.

3.3 Segmentation

This is the process which can classify the image into different groups or clusters for the analyzation. K-means clustering technique is a frequently used method for segmentation [13] which is utilized in this research.

It is observed from Figure 4 the cluster 1 gives the exact region of interest for the infected area.
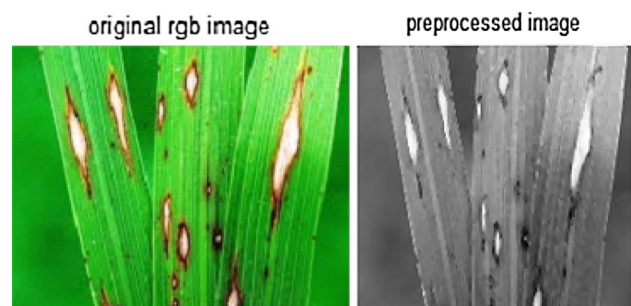

Fig. 2 Proposed methodology
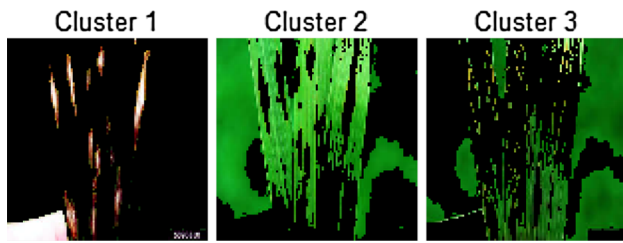

Fig. 3 Input and pre-processed image

Fig. 4 Segmentation using K-means clustering

3.3.1 K-means clustering

In this method, a collection of data is grouped into a distinct number by dividing it. Then they are classified into a k number of disjoint cluster. There are two phases in this K-means [14] clustering. The first phase is to calculate the k centers. For each cluster these centers should be placed far away from each other and in the next step consider each point belonging to the cluster and associate it to the nearest centre when no point is at demand. The commonly used method to define a nearest centroid is Euclidian distance. After the grouping process, the recalculation is done for new centroid of all clusters. Then Euclidian distance is calculated for each and every cluster center and data point which is mathematically described in Eqs. (1) and (2)

$$(E.d)^2 = \min(\mu_1 - \mu_k) \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \left| X(i,j) - \mu^2 \right|. \tag{1}$$

$$\mu_k = \frac{\sum_{i=1}^{i=n} P(i) + i}{\sum_{i=1}^{i=n} P(i)}. \tag{2}$$

The minimum distance is selected as point and assigned in cluster. All clusters are defined through its member objects and its centroid points. So this k- means clustering minimizes the distance between cluster object and centroid and this is an iterative method for segmenting.

3.4 Feature extraction

There are many features in the input images, and they are extracted to classify the objects. These image features are used to describe the location and character feature of objects. The various features extraction are depicted in Fig. 5. In this paper, they are extracted using hybrid methods of SIFT, DWT and GLCM. The discrete wavelet transform used in this method educes the size of test images.

The DWT is not shift-invariant. Because the DWT downsamples, a shift in the input signal does not manifest itself as a simple equivalent shift in the DWT coefficients at all levels. A simple shift in a signal can cause a significant realignment of signal energy in the DWT coefficients by scale. So after the scale invariant transform it is

reinforced to DWT. The grey level co-occurrence matrix (GLCM) is applied for all test images of low level components. These images are decomposed images used to extract the texture feature of the images.

Discrete wavelet transform (DWT) [15] is usually nonredundant. So this is applied for the application of nonstationary signal processing. Grey level co-occurrence matrix (GLCM) in each sub-bands of DT-DWT utilizing the texture features. Similarity measure is achieved by the Euclidian and Canberra distances.

3.4.1 Scale invariant feature transform algorithm

Scale invariant feature transform (SIFT) is a strategy for detecting salient, stable feature points in the image which are not have rotations and scaling. This SIFT features are very resilient that is back to its original, from the effects of noise in image. SIFT features approach provides features characterizing a stable point that remains invariant to changes in scaling and rotation. Then extraction process using SIFT calculation involves four stages of filtering operation. They are scale space extrema detection, finding keypoint, orientation assignment to keypoints, create keypoint descriptor. These stages are described as follows:

3.4.1.1 Scale-space extrema detection This step is to identify the acquired image Locations and scales which are obtained by filtering of image. They identify the images in different views. For this process scale space function of a image is achieved when it be a Gaussian function. This is given by the equation,
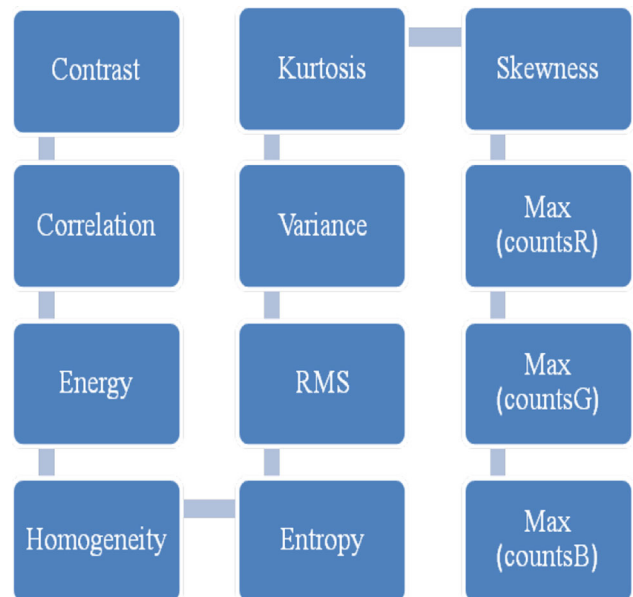

Fig. 5 Feature extraction

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (3)$$

where I(x, y) is the Acquired image and G(x, y, σ) is a variable-scale Gaussian. * is the convolution operator,

In the scale-space, stable keypoint locations are detected by different techniques. The other technique is Difference of Gaussian (DoG) method, where the scale space extrema location D(x, y, σ) is calculated using subtraction of two images.

$$D(x, \sigma) = L(x, y, \sigma) - L(x, y, \sigma). \qquad (4)$$

The local Maxima and minima of Difference of Gaussian is detected by comparing a pixel to its neighbors at the current and adjacent scales. If this data value is greater or lesser of all these points. Then that point is an extrema.

3.4.1.2 Finding keypoint This step requires to allocate with a few Key points. The keypoints extracted by SIFT is limited to an edge, i.e. discontinuity points of the gradient function by a DoG (Difference of Gaussians). This is achieved by subtraction of subsequent images. In each octave, the computation of the DoG is very fast and efficient.

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x. \qquad (5)$$

where its derivatives and D are obtained at the sample point and x is offset from this point. The location of the extrema, $\widehat{x}$ s calculating from its derivative. Then,

$$\widehat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}. \qquad (6)$$

3.4.1.3 Orientation assignment to keypoint This method supposes a constant orientation to the key points in local image intensities with all image property. The Gaussian smoothed image is selected using the scale of the keypoint, A region around each keypoint is selected to remove effects of scale and rotation. Gradient position and magnitude is enumerated using finite distance. L(x, y) is a scale that produce the gradient magnitude and orientation. Pixrl difference is used pre describe this.

$$m(x, y) = \sqrt{(L(x + 1, y) - L(x, y + 1) - L(x, y - 1))^2}. \qquad (7)$$

$$\theta(x, y) = \tan^{-1} \frac{(L(x, y + 1) - L(x, y - 1))}{(L(x + 1, y) - L(x - 1, y))}. \qquad (8)$$

3.4.1.4 Create keypoint descriptor At this stage, each keypoint has location, scale and orientation. Each

keypoints rotate the window to standard orientation. For descriptor selection, the gradient magnitudes, orientations and level of Gaussian blurs of image are being sampled around keypoint location. These gradients are also determined for all levels, for efficiency.

3.4.2 GLCM

Gray-level co-occurrence matrix (GLCM), also known as the gray-level co-occurrence matrix (GLCM) is a widely used method of examining texture which holds the spatial the relationship of pixels The GLCM functions describes the statistical texture feature of an image.

The textural features are feature which abnormality was widely spread in the image, the textural orientation of each class is different, which aid in better classification accuracy. The texture features extracted from the first order features such as mean, standard Deviation, Kurtosis, and Skewness, and second order features namely contrast, correlation entropy, energy, homogeneity, variance and RMS are obtained from GLCM. The various extracted texture features are depicted in Fig. 6.

3.4.2.1 Mean The average intensity of image can be denoted by mean value. It can be calculated as,

$$S_M = \bar{b} = \sum_{b=0}^{L-1} b p(b). \qquad (9)$$

3.4.2.2 Standard deviation The variance or deviation between the pixels in input image is represented by standard deviation.

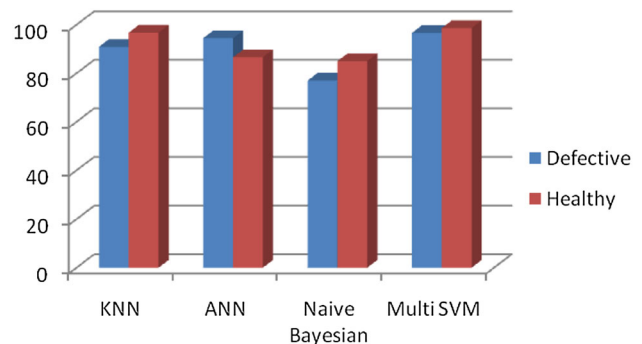$$S_D = \sigma_b = \left[\sum_{b=0}^{L-1} (b - \bar{b})^2 p(b) \right]^{1/2}. \qquad (10)$$

Fig. 6 Disease classification using two categories

3.4.2.3 Skewness The variable can be either positive or negative, this skewness is the measure of degree of asymmetry mean probability distribution.

$$S_S = \frac{1}{\sigma_b^4} \sum_{b=0}^{L-1} (b - \bar{b})^3 p(b). \tag{11}$$

3.4.2.4 Kurtosis The peak value of real valued random variable is measured by Kurtosis. It shape descriptor of a probability distribution.

$$S_K = \frac{1}{\sigma_b^4} \sum_{b=0}^{L-1} (b - \bar{b})^4 p(b) - 3. \tag{12}$$

3.4.2.5 Contrast Contrast is the difference between objects and other objects of the image and from the background. In visual properties, contrast is established to be low when the gray levels of pixels are same. It can be computed by the difference between the colour and brightness of the object.

$$S_C = \sum_i \sum_j (i - j)^2 p(i, j). \tag{13}$$

3.4.2.6 Inverse difference moment This feature explains the smoothness of the image. This measure relates inversely to the contrast measure. For the similar pixel values this value is high.

$$S_I = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i, j). \tag{14}$$

3.4.2.7 Correlation Pixels gray level linear dependence at specified locations is given by correlation.

$$S_O = \frac{\sum_i \sum_j (ij) p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y}. \tag{15}$$

3.4.2.8 Variance Variance is given by the dispersion value over the mean It is used to measure gray level contrast to establish the relative component descriptors.

$$S_V = \sum_{i,j=1}^{N} (i - j)^2 p(i, j). \tag{16}$$

3.4.2.9 Cluster shade Cluster shade feature is used to define the lack of symmetry. The image is asymmetric for the high cluster shade.

$$S_{CS} = \sum_{i,j=1}^{N} \left(i - M_x + j - M_y\right)^2 p(i, j). \tag{17}$$

3.4.2.10 Cluster prominence When low cluster prominence the co-occurrence matrix of the mean values are high,

$$S_{CP} = \sum_{i,j=1}^{N} \left(i - M_x + j - M_y\right)^4 p(i, j). \tag{18}$$

where

$$M_x = \sum_{i,j=1}^{N} ip(i, j). \tag{19}$$

$$M_y = \sum_{i,j=1}^{N} jp(i, j). \tag{20}$$

3.4.2.11 Homogeneity Homogeneity measures the pixel's similarity. GLCM diagonal and range is measured using homogeneity. For diagonal, homogeneity is 1 which gives all the information of smoothness.

$$S_H = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \left(\frac{P_{ij}}{(1 + |i - j|)}\right) \tag{21}$$

3.4.2.12 Energy The uniformity of the texture is described by this feature. So the energy of image is high when mage is more homogeneous. There is very little dominant gray-tone transition in homogeneous image. The image is considered as constant when energy equal to 1.

$$S_N = \sum_{b=0}^{L-1} [p(b)]^2 \tag{22}$$

3.4.2.13 Entropy The entropy is used to represent the gray level distribution randomness. The gray levels are distributed randomly throughout the image then the entropy is 1

$$S_N = - \sum_{b=0}^{L-1} p(b) \log_2 \{p(b)\} \tag{23}$$

Then the features are given as input to the classifiers that process methods are being analyzed in next section.

4 Classifier

Classifier is used to classify the images i based on its spectral features like density, texture etc. In this paper, some of the popular classifier such as KNN, ANN, Naïve

Bayesian and SVM are analyzed for rice plant leaf classification.

4.1 KNN classifier

K-nearest neighbors' classifier performs classification of unknown instances based on a distance or similarity function. [14, 16–18]. Training and testing is performed at the same using this classifier.re retrieved. To compare the similarity examples a distance function is required.

$$\text{Euclidean distance } \big(d(x_j, x_k)\big) = \sqrt{\sum_i \big(x_{j,i} - x_{k,i}\big)^2} \tag{24}$$

$$\text{Mahnattan distance } \big(d(x_j, x_k)\big) = \sum_i \big|x_{j,i} - x_{k,i}\big| \tag{25}$$

The direct algorithm has a cost O(n log(k)), which is not good when the dataset is large.

4.2 ANN classifier

Artificial neural networks (ANN) [19] optimized many different coefficients and this can produce predictive modeling which is rarely used. Because, it is usually tries to over-fit the relationship. This ANN is use at application weather the past is repeated almost exactly in same procedure. Therefore it requires more variability as compared to traditional models and handles it. Also it can help to perform real time operations, adaptive learning, self-organization, and fault tolerance.

4.3 Bayesian classifier

Bayesian classifier is a logic which applied to decision making and inferential statistics using details of prior events to predict feature events with probability inference. The Bayesian classifier idea is to predict the values of the other features, if the class is known to its agent. In case if the class have no knowledge about the class then Bayes rule is applied for class prediction which calculate the feature value. Naive Bayes classifiers are classifiers with the assumption that features are statistically independent of one another. It explicitly models the features as conditionally independent given class [20].

4.4 Multiclass SVM

Classification is done using multiclass SVM [21] which is the improved technique for high accuracy of output. The strategy "one against all" is used in multiclass SVM,

which consist of having one SVM for a class which are trained for differentiate the samples of different classes [22, 23].

4.5 Probability estimation

Posterior probability is estimated to separate each output using one against all method. Additional sigmoid is given by SVM classifier,

$$\hat{P}\big(\omega_j | f_j(x)\big) = \frac{1}{1 + \exp\big(A_j f_j(x) + B_j\big)} \tag{26}$$

In the above equation, $f_j(x)$ represents the output of SVM for training, to separate the class ω_j all other. A_j and B_j are computed by minimizing the below local likelihood,

$$-\sum_{k=1}^{n} t_k \log(p_k) + (1 + t_k) \log(1 - p_k), \tag{27}$$

where p_k is the sigmoid output and t_k is target of probability.

Posterior probability is the "one against all" strategy is estimated using another approach is used, which would be output of each SVM for estimation of overall probabilities. For multiclassw SVM Softmax function is the generalization of the sigmoid function That is,

$$\hat{P}\big(\omega_j | x\big) = \frac{\exp\big(A_j f_j(x) + B_j\big)}{\sum_{j=1}^{c} \exp\big(A_j f_j(x) + B_j\big)}. \tag{28}$$

Therefore it is important to create a dataset of SVM outputs, using fixed parameters of sigmoid and softmax functions. Thus it is preferable to derive an unbiased training set of the SVM outputs. Validation dataset is the first solution to use; but, in our method, the number of samples taken for each class is not proportional to the prior probability. So, cross-validation is preferable. Then training data set is divided into four parts. The SVM output of these four sets used to form an unbiased dataset, to fix the parameters of functions. After fixing the parameters the SVM is trained for the entire training set.

5 Results and discussions

The rice plant leaf image is processed and its statistical and textural features are extracted by using hybrid SIFT, DWT and GLCM techniques. Then the features are given to various classifiers to predict the disease in leaf. The rice plant leaf images are taken from Indian Rice Research Institute (IRRI) website. Images are taken from Aduthurai research center at Thanjavur district. 500 images are used for this experiment. In that, 350 images are taken to train the classifier and 150 images are used for testing purpose and

the same images are classified with classifiers such as KNN, ANN, Naïve Bayesian classifiers and multiclass SVM.

This Multiclass SVM classification gives output based on their category. The categories are: two category (healthy and defect), three category (healthy and 2 diseases) and multi category (healthy and 5 diseases). The original input images, their processed images and outputs of multiclass SVM with various leaf diseases are given below.

5.1 Two category classification (healthy and defective)

Two-category classification defines one healthy and one defective quality categories for paddy disease classification. This is because of the problems of collection of database and classification processes. So, two category grading is first introduced. Several tests are performed for two-category classification. Disease grading is done with each classifier. Figure 6 displays the best accuracy rates achieved for this test. As observed that the SVM produces the highest accuracy rate of 98.63 (Table 1).

From the Table 2 it is clearly understood that the accuracy on healthy leaf is 98.6%, while on defective leaf is (96.7%).

5.2 Multi-category classification

A practical disease classification system should not only classify the leaves are defective or healthy, but also present a more detailed classification. In order to attain such a multi-category grading, the leafs are classified into three category (healthy, blast, brown spot) and multi categories (healthy, blight, blast, brown spot, leaf strake, and false smut) manually. These are described the severity of defects (Fig. 7).

5.3 Three category disease: (healthy, brown spot and blast)

Multi-category (5 category) grading using a different classifier is done. We examine the results of each classifier.

Table 1 Performance analysis of two category classification

Categorized in	KNN	ANN	Naive Bayesian	Multi SVM
Defective	90.9091	94.5	77	96.7
Healthy	96.778	86.63	85	98.63

Table 2 Two category grading confusion matrix (SVM)

Categorized in	Defective	Healthy
Defective	483	7
Healthy	17	463
	500	470
Accuracy (%)	96.7	98.6

Like in the two-category results the highest accuracy rate is obtained by SVM with 98.6% and it exposes that this Multi SVM output is convincing result than from others (Fig. 8).

5.4 Multi category: (healthy, brown spot, leaf blast and false smut)

Figures 9, 10, 11, 12, 13, and 14, are the output of multi-class SVM classifier for two, three and multi category. The classifiers such as KNN, ANN, Naïve Bayesian classifiers and multiclass SVM are also used for the analyzation of classification in terms of accuracy. Accuracy is the quality to which a result of calculation specifies the correct value.

In all the figures, the inlet figures, Fig. 1 depict the e original image. Figure 2 gives grey scale image. Figures 3, 4, 5, and 6 emphasis the combination of discrete wavelet with scale invariant and GLCM feature extraction information with horizontal, vertical and diagnosis. SVM performs relatively best for healthy and defective categories.

By comparative study SVM-based approach proved the best solution for our classification problem (Smoother accuracies for defect categories).

5.5 Performance analysis of multi category

5.5.1 Multi category: (healthy, brown spot, leaf blast and false smut)

5.5.1.1 KNN classifier The performance accuracy of KNN for all paddy diseases is shown in the figure. The KNN achieves 97.5% accuracy for healthy leaf which is higher than all disease in paddy leaf (Fig. 15).

5.5.1.2 ANN classifier This ANN classifier performance in terms of accuracy is drawn in the above graph. It shows that it finds all diseases than the healthy leaf with high accuracy (Fig. 16).

5.5.1.3 Naïve Bayesian classifier This graph is the accuracy performance of the Naïve Bayesian classifier. In this healthy leaf is detected with highest accuracy of 85% than other diseases (Fig. 17).

5.5.1.4 Multiclass SVM classifier The multiclass SVM produces all the diseases and healthy leaf effectively. It is shown in the above graph. Healthy leaf achieves highest accuracy of 98.63% (Fig. 18).

5.6 Comparison of proposed methods

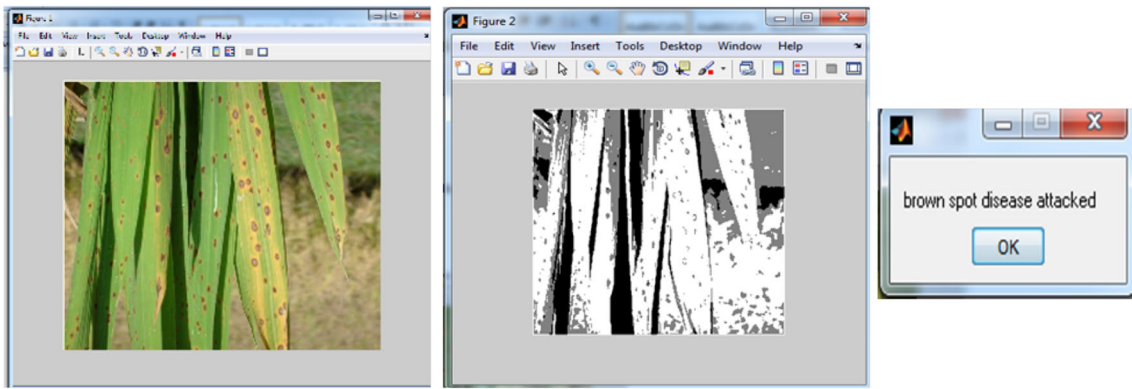The accuracy (in percentage) of all classifiers is tabulated below (Fig. 19).

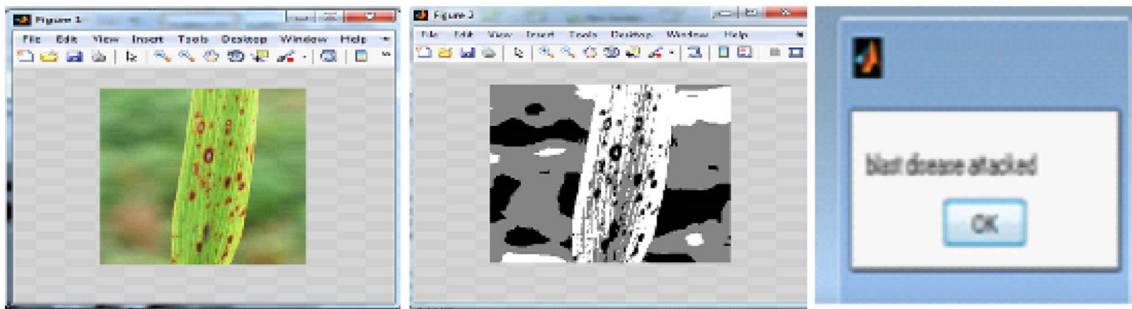**Fig. 7** Three category classification
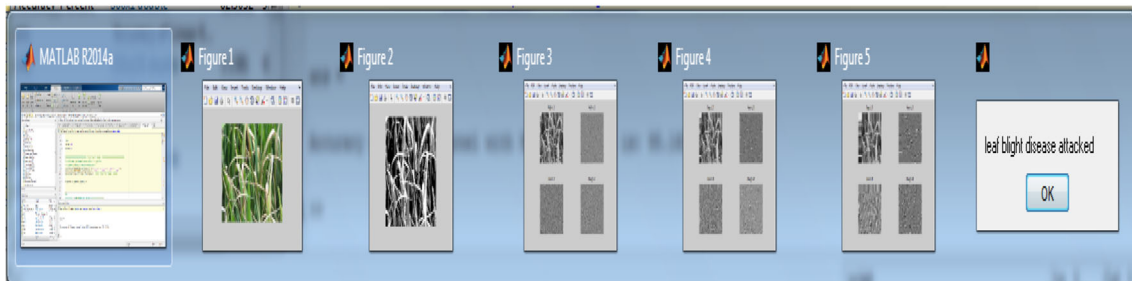

Fig. 8 Three category classification

Fig. 9 Leaf blight disease output

Fig. 10 Leaf blast disease output

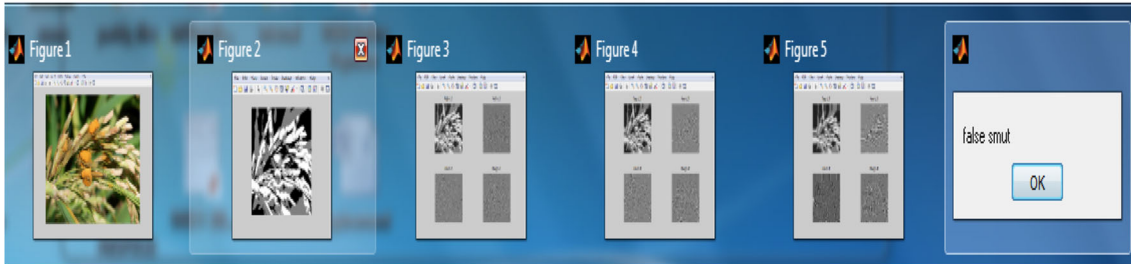**Fig. 11** Brown spot disease output

Fig. 12 False smut disease output
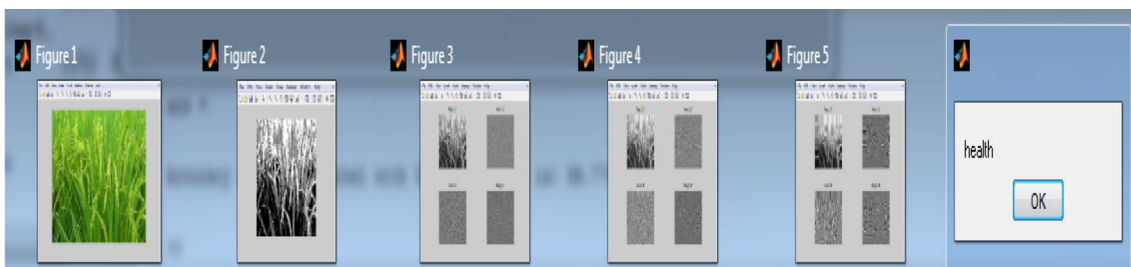

Fig. 13 Leaf streak disease output

Fig. 14 Healthy leaf output

Table 3 shows that the accuracy of various diseases and healthy leaf. The KNN produce 90.90% for leaf blast, 91% for brown spot, false smut and leaf blight, 91.2% for leaf streak and 96.77% for healthy leaf. The ANN classifier produces 94.5% for leaf blast, 94.2% for brown spot, 94% for false smut and leaf blight, 94.4% for leaf streak and 86.63% for healthy leaf. In the Naïve Bayesian classifier, the accuracies are 77% for leaf blast, 76% for brown spot
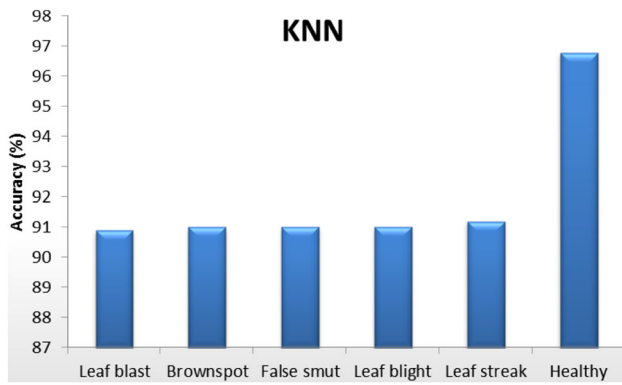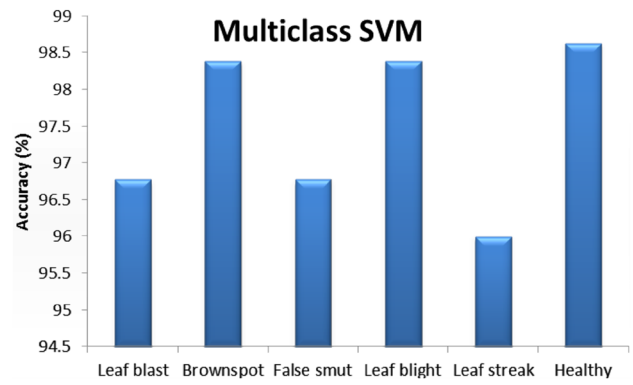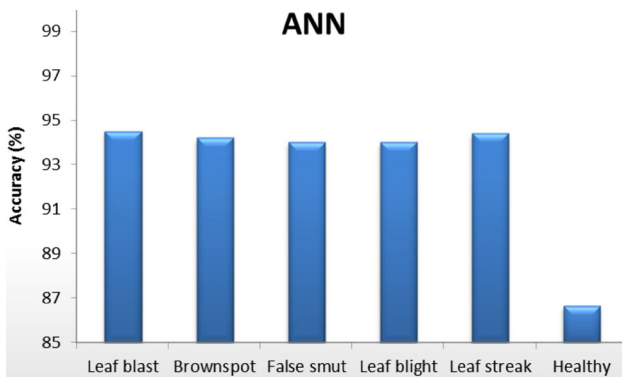
Fig. 15 Accuracy of KNN

Fig. 16 Accuracy of ANN
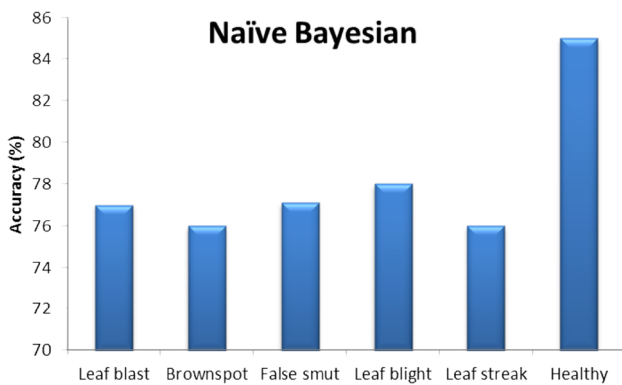

Fig. 17 Accuracy of Naïve Bayesian

Fig. 18 Accuracy of multiclass SVM

and leaf streak, 77.1% for false smut, 78% for leaf blight and 85% for healthy leaf. The multiclass SVM produces 96.77% for leaf blast, 98.38% for brown spot and leaf blight, 96.77% for false smut, 96% for leaf streak and 98.63% for healthy leaf. It is observed that the multiclass SVM gives the high accuracy among all other classifiers because of its advantages. Its working is robust when training examples contain errors and its simple geometric elucidation and a scarce solution.

The multiclass SVM classifies all the diseases and healthy leaf effectively. It is shown in the above graph. Healthy leaf achieves highest accuracy of 98.63%.

6 Conclusion

The paddy leaf images are collected and done various processes such as preprocessing, segmentation, feature extraction and classification. Various diseases and healthy leaf are detected through classifiers as in the result. By using hybrid DWT, SIFT and GLCM, more features are extracted reliably. In this paper various classifiers are observed through its accuracy range. The implementation result shows that Multiclass SVM produce high accuracy of 98.63% than other classifiers.

Fig. 19 Performance analysis of classifiers
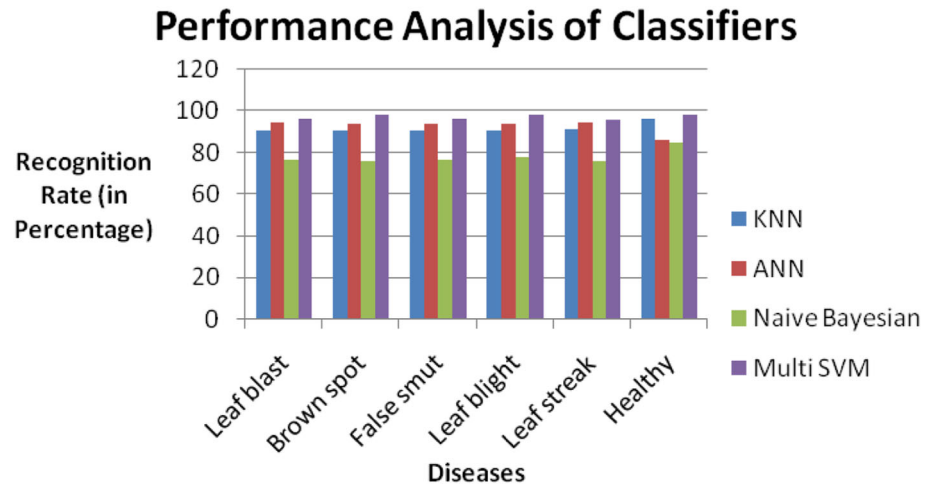
Table 3 Performance analysis of classifiers (multi category)

Diseases	KNN	ANN	Naive Bayesian	Multi SVM
Leaf blast	90.9091	94.5	77	96.7742
Brown spot	91	94.2	76	98.3871
False smut	91	94	77.1	96.7742
Leaf blight	91	94	78	98.3871
Leaf streak	91.2	94.4	76	96
Healthy	96.778	86.63	85	98.63

References

1. Saha, J.K., Selladurai, R., Coumar, M.V., Dotaniya, M.L., Kundu, S., Patra, A.K.: Soil pollution—an emerging threat to agriculture. Agric. Soil Environ. **10**, 1–9 (2017)
2. Singh, D.P.: Plant nutrition in the management of plant diseases with particular reference to wheat, pp. 273–284. Recent Adv. Diagn. Manag. Plant Dis., Springer (2015)
3. Singh, V., Misra, A.K.: Detection of plant leaf diseases using image segmentation and soft computing techniques. Inf. Process. Agric. **4**, 41–49 (2017)
4. Phelps, R.H., Shand, C.R.: Brown leaf spot disease and fertilizer interaction in irrigated rice growing on different soil types. In: Nitrogen Economy in Tropical Soils, pp. 117–121. Springer, New York (1995)
5. Barbedo, J.G.A.: Digital image processing techniques for detecting, quantifying and classifying plant diseases. Springer Plus **2**, 1–12 (2013)
6. Liu, L., Zhou, G.: Extraction of the rice leaf disease image based on BP neural network. In: International Conference on Computational Intelligence and Software Engineering, IEEE, pp. 1–3 (2009)
7. Orillo, J.W., Cruz, J.D., Agapito, L., Satimbre, P.J., Valenzuela, I.: Identification of diseases in rice plant (*Oryza sativa*) using back propagation artificial neural network. In: International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), IEEE, pp. 1–6 (2014)
8. Joshi, A.A., Jadhav, B.D.: Monitoring and controlling rice diseases using image processing techniques. In: International Conference on Computing, Analytics and Security Trends (CAST), pp. 471–476 (2016)
9. Phadikar, S., Sil, J.: Rice disease identification using pattern recognition techniques. In: 11th International Conference on Computer and Information Technology, IEEE, pp. 420–423 (2008)
10. Phadikar, S., Sil, J., Das, A.K.: Rice diseases classification using feature selection and rule generation techniques. Comput. Electron. Agric. **90**, 76–85 (2013)
11. Bakshi, S., Sa, P.K., Wang, H., Barpanda, S.S., Majhi, B.: Fast periocular authentication in handheld devices with reduce phase intensive local pattern. Multimed. Tools Appl. **16**, 1–29 (2017)
12. Cho, G.-S., Gantulga, N., Choi, Y.-W.: A comparative study on multi-class SVM & kernel function for land cover classification in a KOMPSAT-2 image. KSCE J. Civil Eng. **21**(5), 1894–1904 (2017)
13. Kanungo, T., Netanyahu, N.S.: An efficient K-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 881–892 (2002)
14. Dubey, A.K., Gupta, U., Jain, S.: Analysis of k-means clustering approach on the breast cancer Wisconsin dataset. Int. J. Comput. Assist. Radiol. Surg. **11**(11), 2033–2047 (2016)
15. Weeks, M., Bayoumi, M.: Discrete wavelet transform: architectures, design and performance issues. J. VLSI Signal Process. Syst. Signal Image Video Technol. **35**(2), 155–178 (2003)
16. Gong, A., Liu, Y.: Improved KNN classification algorithm by dynamic obtaining K. In: Advanced Research on Electronic Commerce, Web Application, and Communication, pp. 320–324. Springer, New York (2011)
17. Wang, H., Wang, J.: An effective image representation method using kernel classification. In: Tools with Artificial Intelligence (ICTAI), 2014 IEEE 26th International Conference, pp. 853–858 (2014)
18. Keerthana, P., Geetha, B.G., Kanmani, P.: Crustose using shape features and color histogram with K nearest neighbor classifiers. Int. J. Innov. Sci. Eng. Res. **4**(9), 199–203 (2017)
19. Guo, Y., De Jong, K., Liu, F., Wang, X., Li, C.: A comparison of artificial neural networks and support vector machines on land cover classification. In: Computational Intelligence and Intelligent Systems, pp. 531–539. Springer, New York (2012)

20. Zhang, H., Su, J.: Naive Bayesian classifiers for ranking. In: European Conference on Machine Learning, Springer, pp. 501–512 (2004)
21. Übeyli, E.D.: ECG beats classification using multiclass support vector machines with error correcting output codes. Digit. Signal Process. **17**(13), 675–684 (2007)
22. Aiolli, F., Sperduti, A.: Multiclass classification with multi-prototype support vector machines. J. Mach. Learn. Res. ACM **6**, 817–850 (2005)
23. Zhang, S., Wang, H., Huang, W.: Two stages plant species recognition by local mean clustering and weighted sparse representation classification. Cluster Comput. **20**, 1517–1525 (2017)

P. Neelamegam joined in 2005 as Professor in Department of Electronic Instrumentation Engineering, School of Electrical and Electronics Engineering, SASTRA Deemed university. He received his M.Sc. Degree from Madras University and PhD from Bharathidasan University in 1991. Dr P. Neelamegam has over 80 publications in International journals. He has guided 8 Ph.D. students. His field of Interest is Instrumentation and Image Processing and now he is working in Biomedical Instrumentation.

T. Gayathri Devi joined in 2007 as a Lecturer in Department of Electronics and communication Engineering, School of Electrical and Electronics Engineering, Srinivasa Ramanujan Centre, Sastra Deemed university, Kumbakonam. She received her M.Sc., Degree from Bharathidasan University and M.Tech, from SASTRA University in 2009. Currently she is doing PHD in SASTRA Deemed University. Her field of Interest is Image Processing.