

SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique

Jagadeesh D.Pujari, Rajesh Yakkundimath and Abdulmunaf. Syedhusain Byadgi

¹ *SDMCET, Dharwad, India*

² *KLE. Institute of Technology, Hubli, India*

³ *UAS, Dharwar, India*

Abstract — Computers have been used for mechanization and automation in different applications of agriculture/horticulture. The critical decision on the agricultural yield and plant protection is done with the development of expert system (decision support system) using computer vision techniques. One of the areas considered in the present work is the processing of images of plant diseases affecting agriculture/horticulture crops. The first symptoms of plant disease have to be correctly detected, identified, and quantified in the initial stages. The color and texture features have been used in order to work with the sample images of plant diseases. Algorithms for extraction of color and texture features have been developed, which are in turn used to train support vector machine (SVM) and artificial neural network (ANN) classifiers. The study has presented a reduced feature set based approach for recognition and classification of images of plant diseases. The results reveal that SVM classifier is more suitable for identification and classification of plant diseases affecting agriculture/horticulture crops.

Keywords — Plant Disease, Image Processing, Feature Selection, Classifiers, Experimentation.

I. INTRODUCTION

PLANT disease diagnosis is an art as well as science. The diagnostic process (i.e. recognition of symptoms and signs), is inherently visual and requires intuitive judgement as well as the use of scientific methods. Photographic images of plant disease symptoms and signs used extensively to enhance description of plant diseases are invaluable in research, teaching and diagnostics etc. Plant pathologists can incorporate these digital images using digital image transfer tools in diagnosis of plant diseases. Most plant diseases are caused by fungus, bacteria, viral, nematodes etc [9]. These infections agents or pathogens are either saprophytes (also called necrotrophies or heterotrophies) or autotrophies (also called parasites) or facultative. Saprophytes survive on dead tissues and parasites survive on living tissues, where as facultative (in both) which can change the behavior based on circumstances. Many disease produce symptoms which are the main tools for field diagnosis of diseases showing external systems out of a series of reactions that take place between host and pathogen. As such, several important decisions regarding safe practices, the production and processing of plant have been made in the recent past. One of the main concerns of scientists is the automatic disease diagnosis and control [10]. Keeping this in back drop, work on recognition and classification of plant diseases affecting agriculture/horticulture crops has been carried out. The following papers have been cited during the literature survey to understand the different applications of computers in allied areas of the present work carried out.

II. A BRIEF OVERVIEW OF RELATED WORK

Khoje et al., [11] have proposed fruit grading system based on discrete cosine transform (DCT). The system is developed to gauge quality of fruits, namely, guava and lemon. Texture features based on DCT are extracted from the surface of normal and affected fruit sample images and fed as input to SVM and PNN classifiers. The classification results have shown that the SVM performed as better discriminator compared with PNN classifier. Razmjooy et al., [17] have proposed a hierarchical grading method for quality evaluation of defected potatoes. The system is developed in conjunction with size sorting of potatoes. Color features are extracted from defective potato images. Experimental results have shown that the SVM has very high accuracy and speed across classifiers for defect detection. Anami and Savakar [18] have presented a neural network approach to identify effect of foreign bodies (FB) on bulk food grain image samples. Different food grains, namely, green gram, groundnut, jowar, rice, and wheat are considered for recognition and classification. The color and texture features extracted from sample images are provided as input to a neural network classifier. The average classification accuracy of 90% is obtained for the tested sample images using ANN classifier. Vibhute and Bhode [21] have done a survey on several image processing techniques applied in agricultural applications. Focus has been on techniques like remote sensing, hyper-spectral imaging, fuzzy logic, neural network, genetic algorithm, wavelet, PCA, etc. The importance of image processing techniques in automatic sorting of fruits, weed detection has also been discussed. (Prasad et al., 2012) have discussed image processing methods to detect crop diseases. The proposed imaging system consists of disease spot detection using histogram based segmentation, feature extraction using Gabor wavelet transform (GWT), and classification using SVM. Classification accuracy rate of over 89% is obtained in all the cases. Landge et al., [12] have designed an image processing system to detect maize crop diseases. Color features are extracted from disease symptoms, namely, stem borer and brown stripe downy mildew found on maize crops. An ANN classifier is developed for classification of disease type. Barbedo [3] has done a survey on several image processing methods for detecting plant diseases. The techniques discussed are useful in automatic recognition, classification, and quantifying disease severity in plants. These methods are awaited to be useful for researchers providing comprehensive overview of vegetable pathology and automatic detection of plant diseases using pattern recognition techniques. Dubey et al., [5] have presented a novel approach for the detection of defected apples based on color features with k-means clustering technique. Apple scab, apple rot, and apple blotch diseases are considered for experimentation. The potential of proposed image processing approach in evaluating the quality of defected apples is observed. The simulation results revealed that the proposed approach is promising in terms of precision and computational efficiency. Rathod et al., [16] have provided various image processing and machine vision techniques for automatic

recognition of plant leaf disease. The techniques used for image preprocessing, segmentation, feature extraction, which provide fast and accurate detection of plant leaf disease have been focused. Bauer et al. [2], and Pachecho et al. [22], have proposed image processing methods for classification of leaf diseases. The high resolution multi-spectral stereo images of healthy leaves and infected with the leaf spot pathogens, namely, cercospora beticola and uromyces betae found on sugar beet are considered for the study. The classification of disease type is performed using k-NN and an adaptive Bayes classifier using Gaussian mixture models. The classification results observed indicates that an adaptive Bayes classifier using Gaussian mixture models performed better in comparison with k-NN classifier. Sungkur et al., [20] have described the design and implementation of an automated system to recognize disease spots present on leaves of sugarcane. To recognize disease spots present on leaves of sugarcane, several descriptors, namely, aspect ratio, eccentricity, circularity, and moments analysis have been tested. The system has given a classification accuracy of 95.25% using minimum distance classifier. Bandi et al., [1] have evaluated the performance of various classifiers for classifying various citrus diseases. The disease present on citrus leaves, namely, melanose, greasy spot, and scab are considered for the study. K-nearest neighbor, naive bayes (NB), LDA, and random forest tree (RFT) classifiers are used for the classification purpose. Texture features based on CCM are extracted from the disease affected citrus leaves. The experimental results reveal that the proposed work achieved 98.75% using LDA. (Fina et al., 2013) have demonstrated various image processing methods for detection and recognition of the plant pests. The imaging system used combination of k-means clustering technique and the correspondence filter. The objects (pests) from the background (pest habitat) are separated by partitioning the dataset into voronoi cells. Different attributes between the pest and its habitat (leaf, stem) are extracted and plant pests are identified using corresponding filter. The effectiveness of corresponding filter is promising in recognition and detection of pests, which can achieve rotational invariance of pests up to 360 degrees. Cui et al., [4] have developed image processing methods for quantifying severity of rust disease found on soybean leaves. Segmentation of infected areas from multi-spectral images of soybean plant leaves is done using fast manual threshold-setting method based on HSI color model. The extracted color features with ratio of infected area (RIA) and rust color index (RCI) diagnostic parameters are used as symptom indicators for quantifying rust severity. The performance of threshold-setting method and centroid-locating method are compared. It is observed that threshold-setting method works well under laboratory conditions for detecting soybean rust severity and centroid-locating method has the better performance in field application. Fenyvesi et al., [6] have demonstrated level of damage in the fruits during harvesting, transport, and manipulating. The stress developed in fruits leading to its damage is analyzed based on general material properties and unit load using finite element method (FEM) simulations. According to the model developed, it is reported that the maximum stress is developed in the middle of the fruits which results in breakage of fruits. The literature survey has revealed that there is fair amount of scope for plant disease detection in the area of agriculture and horticulture. Fungi, bacteria, virus, nematodes, etc., are the main source of diseases in plants [9]. From the literature survey presented above, it is observed that the work on classification of plant diseases like fungi, bacteria, virus, nematodes and deficiency diseases affecting agriculture/horticulture crops reported in the literature is scarce. This being the motivation, the work on classification of plant diseases in agriculture/horticulture crops is proposed to assist the farmers technologically. In the proposed work, focus has been on early detection and classification of plant diseases like fungi, bacteria, virus, nematodes and deficiency diseases affecting different agriculture/horticulture crops. The paper is organized into

five sections Section.3 gives the proposed methodology. Section.4 describes results and discussion. Section.5 gives conclusion of the work.

III. PROPOSED METHODOLOGY

In the proposed work, we have focused on early detection of plant disease affecting agriculture/horticulture crops. This useful work in the real world comprises of the tasks like image acquisition, preprocessing of images, feature selection, development of methodologies for identification of different plant diseases, finally development and classification of architecture for the computer vision system The Fig. 1 shows the schematic flow diagram of the proposed work.

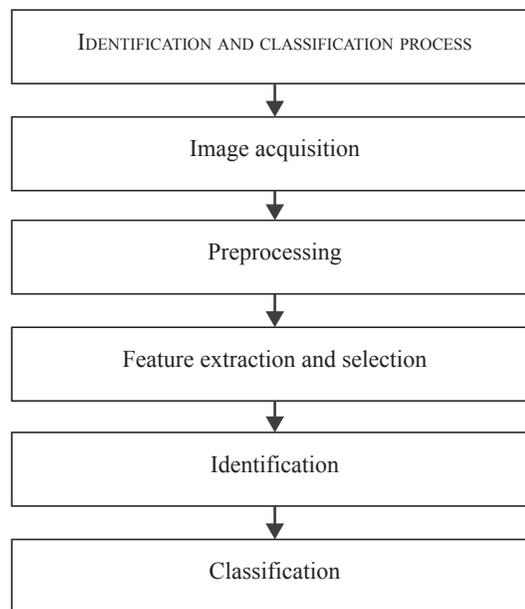


Fig. 1. Block diagram of the approach followed in the work

A. Image data set

In the present work, plant diseases found on agriculture/horticulture crops, namely, fungal, bacterial, viral, nematodes, deficiency and normal (not affected) are considered for recognition and classification. This constitutes 6 classes. The study considers 900 sample images (150 samples of each class) for recognition and classification purpose. The set of 900 sample images used in this work was obtained from department of plant pathology, at the University of Agricultural Sciences, Dharwad, INDIA. As first signs and symptoms of disease are observed on plant leaves, methodologies for identification and classification of plant diseases affecting leaves of agriculture/horticulture crops have been developed. The Fig. 2 to 6 shows the sample images affected by plant diseases in agriculture/horticulture crops used in the present work. The classification tree is shown in the Fig. 7.

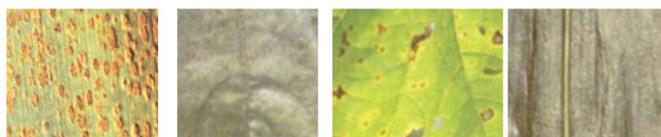


Fig. 2. Images of fungal disease affected samples a) leaf rust of wheat b) powdery mildew of sunflower c) anthracnose of grape d) downey mildew of maize

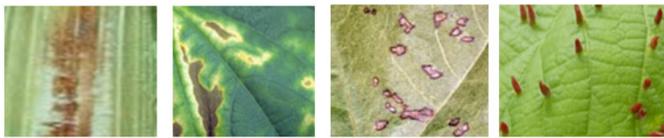


Fig. 3. Images of bacterial disease affected samples a) Stalk rot of maize b) bacterial wilt of cucumber c) angular leaf spots of cotton d) galls on lime

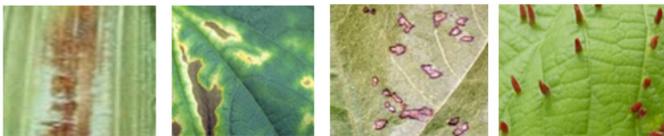


Fig. 4. Images of bacterial disease affected samples a) Stalk rot of maize b) bacterial wilt of cucumber c) angular leaf spots of cotton d) galls on lime



Fig. 5. Images of viral disease affected samples a) mosaic on cucumber b) yellow virus on grape c) vein chlorosis on tomato d) enations on sugarbeet



Fig. 6. Images of nematodes disease affected samples a) foliar nematode of tomato b) cyst disease on soybean c) ear cockle on wheat d) ear rot on maize



Fig. 7. Images of deficiency disease affected samples: a) magnesium deficiency on grapes b) zinc deficiency on citrus c) boron deficiency on sunflower d) potassium deficiency on tomato

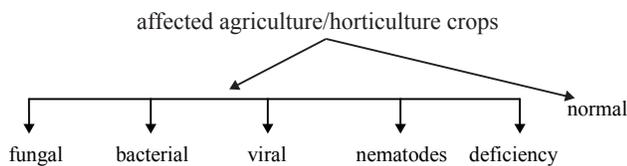


Fig. 8. Classification tree.

The symptoms of plant disease exhibit different properties like color, shape, and texture. When samples of different normal and disease affected agriculture/horticulture crops are considered, patterns vary from disease to disease. Color is an important dimension of human visual perception that allows discrimination and recognition of visual information [14]. Many natural surfaces and naturally occurring patterns reveal texture characteristic, meant to capture the granularity and repetitive forms of surfaces within an image [14]. The considered work has used some state-of-the-art color and texture features for recognition and classification of disease affected agriculture/horticulture crops to validate the accuracy and efficiency.

It is apparent from the Figs 2 to 6 that the sample images belonging shows significant difference in terms of color and texture.

B. Image acquisition and preprocessing

The image data set, used for experimentation in this thesis, comprises 9912 digital images. For image acquisition, a color camera (DXC-3000A, Sony, Tokyo, Japan) is used. The camera has a zoom lens of 10-120 mm focal length and a 72 mm close-up lens set. The camera is vertically oriented and approximately a distance of 0.5 meter is maintained while capturing the images. Image quality is definitive for the after effects of investigation, influencing both the ability to detect features under examination and accuracy of consequent estimations. Image enhancement techniques are used to emphasize features of interest and highlight certain details hidden in the image. To improve the quality of the image, preprocessing steps are applied over the image. MATLAB version 7.10 is used for implementation of the digital image processing algorithms.

Preprocessing of the image includes shade correction, removing artifacts, and formatting. Some images, originally from camera, manifest uneven lighting called shade. Due to variation in outdoor lightning conditions, some regions are brighter and some others are darker than the mean value for the whole image. This phenomenon is a consequence of inaccuracy in the system. Precise tuning of camera is done to minimize this effect. The images contain some artifacts induced like scratches, coat or mark, lumps of dust or abrasive particles. Hence, median filter and imfilter have been used to remove such artifacts. Formatting deals with storage representation and setting the attributes of the image. The images acquired from the camera are of 1920x1080 pixels and are reduced to suitable size for the reasons of reducing computational time required for feature extraction and their storage on the medium.

C. Feature selection

A pattern can denote a quantitative or morphological description of an object or some other point of interest in an image, in which some organization of underlying structure can be supposed to live. In other words, a pattern is an arrangement of descriptors. Descriptors are also called features in pattern recognition literature [15]. Only significant features are extracted from the processed image. This is where the features reduction method is adopted. In the present work, feature extraction employs color features based on RGB, HSI color models, texture features based on GLCM [15] [19].

Color feature extraction

One of the primary facets of color feature extraction is the selection of a color space. A color space is a multidimensional space, in which different dimensions represent different constituents of the color. Most color spaces are three dimensional. An instance of a color space is RGB, which attributes to each pixel a three element vector, giving the color intensities of the three primary colors, namely, red (R), green (G), and blue (B). The space covered by the R, G, and B values completely describes visible colors, which are entitled as vectors in the 3D RGB color space. As a consequence, the RGB color space offers a useful starting point for representing color features of the images [14][15].

The following method is adopted in the extraction of RGB features. The foremost step is the separation of RGB components from the original color images. The next step is the computation of mean, standard deviation, variance, and range from the separated RGB components using the following Equations (1) to (6).

$$\mu \frac{1}{N} \sum_{i=1}^N x_i = \frac{x_1 + x_2 + \dots + x_N}{N} \quad (1)$$

Where,

N is the total number of pixels,

x_i is the i^{th} pixel value

$$\text{Standard deviation } \sigma = \frac{1}{N} \sum_{i=1}^N \sqrt{(x_i - \mu)^2} \quad (2)$$

$$\text{Variance} = \sigma \times \sigma \quad (3)$$

Maximum element and minimum elements from given input color (RGB) image is calculated using Equation (2.4).

$$\text{max1} = \max(\text{image}), \text{max2} = \max(\text{max1}) \quad (4)$$

The above function returns the row vector containing maximum element from each column, similarly find minimum element from whole matrix using Equation (2.5).

$$\text{min1} = \min(\text{image}), \text{min2} = \min(\text{min1}) \quad (5)$$

Range is the difference between the maximum and minimum elements and is given in the Equation (2.6).

$$\text{Range} = \text{max2} - \text{min2} \quad (6)$$

When humans see a color object, the object is depicted by its hue (H), saturation (S), and brightness or intensity (I). Hue is a good descriptor of a pure color (pure yellow, orange or red), whereas saturation refers to the amount of pure color mixed with white light. The chromatic notion of intensity (gray level) which describes brightness is the most useful descriptor of monochromatic images. The intensity component is easily quantifiable and interpretable. The HSI color model separates the intensity component from the color carrying information (hue and saturation) in a color image. As a consequence, the HSI model is an absolute aid for developing image processing algorithms based on color descriptions that are natural and instinctive to humans, who, after all, are the developers and users of these algorithms. The hue, saturation, and intensity components are extracted from these RGB components [14] [15]. RGB color space can be transformed to HSI color space using the Equations (7) to (10).

$$H = \left\{ \begin{array}{ll} \square, & \text{if } B \leq G \\ 360^\circ - \square, & \text{if } B > G \end{array} \right\} \quad (7)$$

With

$$\theta = \cos^{-1} \left(\frac{1/2 [(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right) \quad (8)$$

The saturation component is given by

$$S = I - \frac{3}{(R + G + B)} [\min(R, G, B)] \quad (9)$$

$$I = \frac{1}{3} R (G + B) \quad (10)$$

The concept of RGB and HSI component extraction is presented in the Algorithm 1. There are 24 color features extracted from the images and they are listed in the Table 1.

TABLE 1. COLOR FEATURES

Sl.No.	Features	Sl.No.	Features
1	Red mean	13	Hue mean
2	Red variance	14	Hue variance
3	Red range	15	Hue range
4	Red standard deviation	16	Hue standard deviation
5	Green mean	17	Saturation mean
6	Green variance	18	Saturation variance
7	Green range	19	Saturation range
8	Green standard deviation	20	Saturation standard deviation
9	Blue mean	21	Intensity mean
10	Blue variance	22	Intensity variance
11	Blue range	23	Intensity range
12	Blue standard deviation	24	Intensity standard deviation

Color feature reduction

It is found through experimentation that only eight color features which are common in all the sample images are significant. Hence, these eight features contribute more to the classification of plant diseases. Therefore, eight features have been considered as first-level feature reduction. The reduction is done based on threshold and delta value. Any feature values below the threshold are discarded. The threshold is chosen based on average of minimum feature value and maximum feature value. The threshold value is empirically determined as 0.2. Delta is the minimum difference between two feature values and is empirically determined as 10-3 [8]. The Fig. 8 show the histogram plot values of eight color features obtained for images of plant diseases affecting agriculture/horticulture crops. The procedure involved in color feature reduction is given in the Algorithm 1.

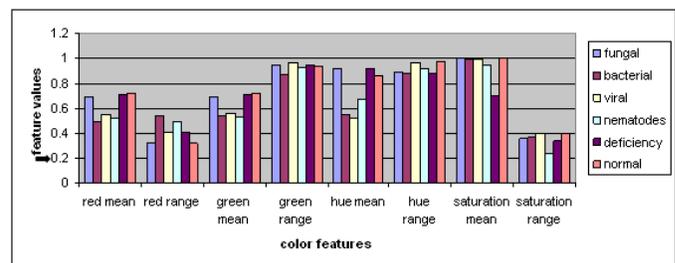


Fig. 8: color feature values of plant diseases (Threshold 0.2)

Algorithm 1: Color feature reduction

Input: Color (RGB) image.

Output: Reduced color feature vector.

Description: Delta is the minimum difference between two features and is set to 10-3. Threshold is the average of minimum and maximum feature value and is set to 0.2.

Start

Step 1: Separate the RGB components from the original 24-bit input color image

Step 2: Obtain the HSI components using the Equations (7), (8), (9), and (10)

Step 3: Compute mean, variance, and range for each RGB and HSI components using the Equations (1) through (6)

Step 4: Threshold = (minimum feature value + maximum feature value)/2

Step 5: Initialize feature vector to zeros

Step 6: For (i=1 to size of the feature vector)

If (value of feature (i) > threshold)

Select as reduced feature

Step 7: For (i=1 to size of the reduced feature vector)

Compare each feature with the other

If (features values are equal OR feature values differ by delta)

Discard the feature

Else

Select as reduced color feature

Stop.

Texture feature extraction

For texture features based on spatial domain analysis, one way to describe the descriptor is using a second order statistics of pairs of intensity values of pixels in an image using co-occurrence matrix method [89]. The co-occurrence matrix method of texture description is developed through the use of spatial gray level dependence matrices (SGDMs), which is based on repeated occurrence of some gray level configuration in the texture. This configuration varies rapidly with distance in fine textures and slowly with coarse textures. The GLCM $P_{\phi, d}(i, j)$ represents a matrix of relative frequencies describing how frequency pair of gray levels (i, j) appear in the window separated by a given distance $d = (dx, dy)$ at an angle ' ϕ ' [105]. Gray level co-occurrence matrices (GLCMs) method counts how often pairs of gray level of pixels separated by certain distance and oriented in a certain direction, while scanning the image from left-to-right and top-to-bottom. In the present work, a distance of 1 ($d=1$) when ' ϕ ' is 0° or 90° and $\sqrt{2}$ ($d=\sqrt{2}$) when ' ϕ ' is 45° or 135° is considered [19]. The Equations (11) to (16) are used to evaluate the textural features.

$$\text{Energy} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p^2 d(i, j) \tag{11}$$

$$\text{Entropy} = - \sum_{i,j} P(i, j) \log P(i, j) \tag{12}$$

$$\text{Homogeneity} = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} \frac{p_d(i, j)}{1 + |i - j|} \tag{13}$$

$$\text{Maximum Probability} = \max(P(x, y)) \tag{14}$$

$$\sigma_x = \sum_x (x - \mu_x)^2 \sum_y P(x, y) \tag{15}$$

$$\text{Contrast} = \sum_{n=0}^{Ng-1} n^2 \sum_{|i-j|=n} P_d(i, j) \tag{16}$$

$$\text{Inverse Difference Moment} = \sum_{x, y; x \neq y} \frac{P^\lambda(x, y)}{|x - y|^k} \tag{17}$$

where $\mu_x, \mu_y, \sigma_x, \sigma_y$ are means and standard deviations defined by,

$$\mu_x = \sum_x x \sum_y P(x, y), \quad \mu_y = \sum_y y \sum_x P(x, y),$$

$$\sigma_x = \sum_x (x - \mu_x)^2 \sum_y P(x, y) \quad \text{and}$$

$$\sigma_y = \sum_y (y - \mu_y)^2 \sum_x P(x, y)$$

The differentiation between sample images is carried out in the simplest way, quantifying average gray levels within the matrix, change in the gray level with respect to average level of minimum and maximum gray levels present in the matrix. Hence, basic co-occurrence features, namely, mean, variance, and range has been considered using the Equations (1) to (6). The texture features used in the work are listed in the Table 2.

TABLE 2: GLCM TEXTURE FEATURES

Sl.No.	Features	Sl. No.	Features	Sl.No.	Features
1	Red GLCM mean	11	Green GLCM mean	21	Blue GLCM mean
2	Red GLCM variance	12	Green GLCM variance	22	Blue GLCM variance
3	Red GLCM range	13	Green GLCM range	23	Blue GLCM range
4	Red GLCM energy	14	Green GLCM energy	24	Blue GLCM energy
5	Red GLCM entropy	15	Green GLCM entropy	25	Blue GLCM entropy
6	Red GLCM homogeneity	16	Green GLCM homogeneity	26	Blue GLCM homogeneity
7	Red GLCM sum mean	17	Green GLCM sum mean	27	Red GLCM sum mean
8	Red GLCM correlation	18	Green GLCM correlation	28	Blue GLCM correlation
9	Red GLCM contrast	19	Green GLCM contrast	29	Blue GLCM contrast
10	Red GLCM Maximum Probability	20	Green GLCM Maximum Probability	30	Blue GLCM Maximum Probability

Texture feature reduction

It is found through experimentation that only five texture features which are common in all the sample images are significant. Hence, these five features contribute more to the classification of plant diseases. Therefore, five features have been considered as first-level feature reduction. The reduction is done based on threshold and delta value. Any feature values below threshold are discarded. The threshold is chosen based on average of minimum feature value and maximum feature value. The threshold value is empirically determined as 100. Delta is the minimum difference between two feature values and is empirically determined as 10-3 [8]. The Fig. 9 show the histogram plot values of six texture features obtained for images of plant diseases affecting agriculture/horticulture crops. The procedure involved in texture feature reduction is given in the Algorithm 2.

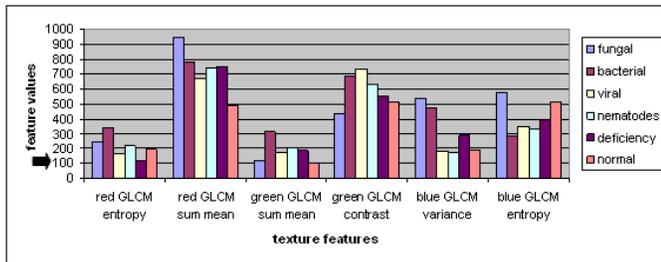


Fig. 9: texture feature values of plant diseases (Threshold = 100)

Algorithm 2: Texture feature reduction

Input: Color (RGB) image.

Output: Reduced texture feature vector

Description: P_{ϕ} , $d(x, y)$ means GLCM matrices in the direction ($\phi = 00, 450, 900, \text{ and } 1350$) and ' d ' is the distance. Delta is the minimum difference between two features and is set to 10-3. Threshold is the average of minimum and maximum feature value and is set to 100.

Start

Step 1: For all the separated RGB components, derive the co-occurrence matrices P_{ϕ} , $d(i, j)$ in four directions 00, 450, 900, and 1350 and $d=1$

Step 2: Compute mean, variance, and range for each RGB components using the Equations (1) through (6)

Step 3: Threshold = (minimum feature value + maximum feature value)/2

Step 4: Initialize feature vector to zeros

Step 5: For ($i=1$ to size of the feature vector)

If (value of feature (i) > threshold)

Select as reduced feature

Step 6: For ($i=1$ to size of the reduced feature vector)

Compare each feature with the other

If (feature values are equal OR feature values differ by delta)

Discard the feature

Else

Select as reduced texture feature

Stop.

D. Classification

The study has adopted artificial neural network based classifiers using multilayer feed-forward with back propagation support vector machine based classifiers in the recognition of images of plant disease and studied their behavior in terms of suitability of classifiers for identification of different plant diseases.

IV. RESULTS AND DISCUSSION

The classifiers are trained and tested using images of plant diseases. The sample images are divided into two halves and one half is used for training and other is used for testing. The color and texture features are used to train and test neural network model.

The percentage accuracy of recognition and classification is defined as the ratio of correctly recognized sample images to the total number of sample images and is given in the Equation (17). The process of recognition and classification is given in the Algorithm 3.

$$\text{Percentage accuracy (\%)} = \frac{\text{correctly recognized sample images}}{\text{total number of test sample images}} \times 100 \quad (17)$$

Algorithm 3: Recognition and classification of plant diseases affecting agriculture/horticulture crops

Input: Color (RGB) images of plant diseases affecting agriculture/horticulture crops.

Output: Recognized and classified images.

Start

Step 1: Apply color and texture feature extraction input color image and obtain color and texture features

Step 2: Apply color and texture feature reduction Algorithms 1 and 2 to color and texture features and obtain reduced color and texture feature vector

Step 3: Train the SVM and BPNN with reduced color and texture feature vector

Step 4: Accept test images and repeat Steps 1 and 2

Step 5: Recognize and classify the images using SVM and BPNN

Stop.

SVM based classifier

A support vector machine (SVM) is used to recognize plant disease affecting agriculture/horticulture crops. The study has chosen SVM because of its efficient implementations and performances proved to be excellent for high dimensional problems and small data sets. Viewing training input vector in an n -dimensional space, SVM constructs a hyper-plane in the space, which can be used for classification that has the highest distance to the closest training data point of any class (functional margin). To compute the margin, two parallel hyper-planes are constructed, one on every side of the isolating hyper-plane, which are pushed up in opposition to the two data sets. The aim is to determine which class a new data point belongs based on data points associated to one of the two classes. In the case of support vector machines, a data point is computed as a p -dimensional vector (a list of p numbers) and it is meant to know whether such levels can be forked by a $(p-1)$ dimensional hyper-plane. This is called a linear classifier or maximum margin classifier [123]. The core of SVM Matlab toolbox used in the present work is based on Dr. Lin's Lib SVM version 2.33. It is developed by Junshui Ma, Los Alamos National Lab and Yi Zhao, Electrical Engineering department, Ohio State University [10].

The Fig. 10 shows the graph of classification accuracy of plant diseases affecting agriculture/horticulture crops using reduced eight color features with SVM classifier. From the graph, it is observed that the maximum classification accuracy of 91% has occurred with images of normal. The minimum classification accuracy of 78% has occurred with images of nematodes disease. The average classification accuracy of 84% is achieved irrespective of the image types of plant disease affecting agriculture/horticulture crops.

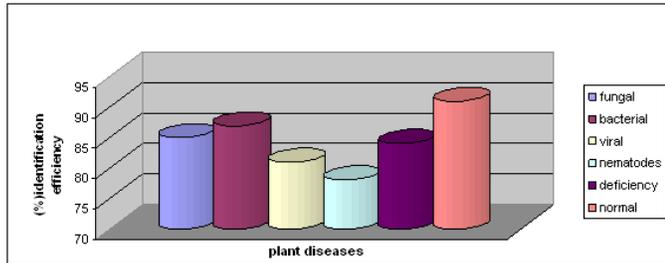


Fig.10: Classification efficiency for plant diseases using color features with SVM

The Fig. 11 shows the graph of classification accuracy of plant diseases affecting agriculture/horticulture using reduced six texture features with SVM classifier. From the graph, it is observed that the maximum classification accuracy of 96% has occurred with images of normal. The minimum classification accuracy of 83% has occurred with images of bacterial disease. The average classification accuracy of 89% is achieved irrespective of the image types of plant disease affecting agriculture/horticulture crops.

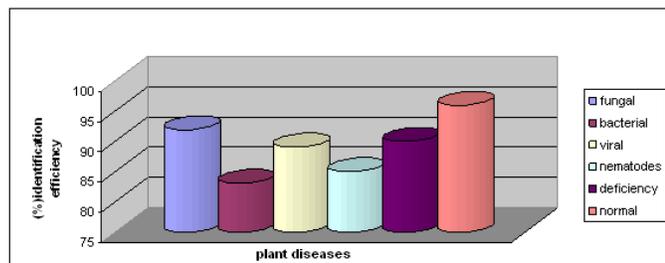


Fig. 11: Classification efficiency for plant diseases using texture features with SVM

In order to take advantages offered by both color texture features, color and texture features are combined and used as input to the SVM classifier. The training and testing are carried out with combined color and texture features. The Fig. 12 shows the graph of classification accuracy of plant diseases affecting agriculture/horticulture using combined features with SVM classifier. From the graph, it is observed that the maximum classification accuracy of 98% has occurred with images of fungal disease. The minimum classification accuracy of 86% has occurred with images of nematodes disease. The average classification accuracy of 92% is achieved irrespective of the image types of plant disease affecting agriculture/horticulture crops.

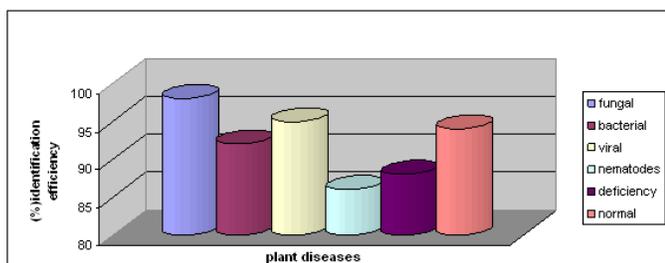


Fig. 12: Classification efficiency for plant diseases using combined features with SVM

ANN based classifier

In order to corroborate the accuracy of classification obtained from SVM classifier, the study has considered a multilayer BPNN as an alternate model to identify plant disease symptoms affecting agriculture/horticulture crops.

Multilayer back propagation neural network (BPNN) has been used as a classifier in the present work. A typical multilayer neural network comprises an input layer, output layer, and hidden (intermediate) layer of neurons. BPNN has the potential to classify different forms (patterns) of arbitrary complex input/output mappings or decision surfaces. Among the majority of multilayer feed-forward artificial neural network (ANN) algorithms, BPNN emerged as the most important algorithm for the supervised training and has been the workhorse for many classification problems in different applications of machine learning. The study has chosen BPNN because of its ease and strength in execution for large training data set. The number of neurons in the input layer corresponds to the number of input pattern vectors and the number of neurons in the output layer corresponds to the number of pattern types (classes). In the hidden layers sigmoid functions have been used. The number of nodes in the hidden layer is calculated using the Equation (2.32) [8].

$$n = \frac{(I+O)}{2} + y^{0.5} \tag{18}$$

Where, n= number of nodes in hidden layer,

I= number of input features,

O= number of outputs,

y= number of inputs pattern in the training set

The training and testing is carried out using reduced color and texture features. The number of nodes in the hidden layer is calculated using the Equation (18).

The ANN classifier has used eight input nodes and six output nodes corresponding to six chosen categories of plant diseases and the chosen eight color features respectively. The Fig. 13 shows the graph of classification accuracy of plant diseases affecting agriculture/horticulture using reduced eight color features with ANN classifier. From the graph, it is observed that the maximum classification accuracy of 94% has occurred with images of normal. The minimum classification accuracy of 72% has occurred with images of deficiency disease. The average classification accuracy of 82% is achieved irrespective of the image types of plant disease affecting agriculture/horticulture crops.

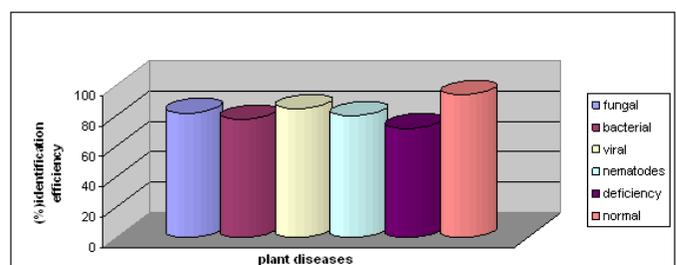


Fig. 13: Classification efficiency for plant diseases using color features with ANN

The ANN classifier has used six input nodes and six output nodes corresponding to six chosen categories of plant diseases and the chosen six texture features respectively. The Fig. 14 shows the graph

of classification accuracy of plant diseases affecting agriculture/horticulture using reduced six texture features with ANN classifier. From the graph, it is observed that the maximum classification accuracy of 92% has occurred with images of normal. The minimum classification accuracy of 78% has occurred with images of deficiency disease. The average classification accuracy of 84% is achieved irrespective of the image types of plant disease affecting agriculture/horticulture crops.

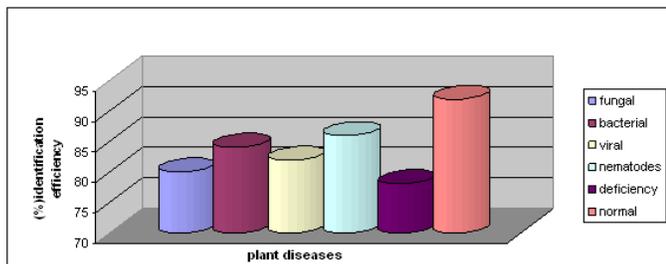


Fig. 14: Classification efficiency for plant diseases using texture features with ANN

In order to take advantages offered by both color texture features, color and texture features are combined and used as input to the BPNN classifier. The Fig. 15 shows the graph of classification accuracy of plant diseases affecting agriculture/horticulture using combined features with ANN classifier. From the graph, it is observed that the maximum classification accuracy of 92% has occurred with images of normal. The minimum classification accuracy of 82% has occurred with images of deficiency disease. The average classification accuracy of 87% is achieved irrespective of the image types of plant disease affecting agriculture/horticulture crops.

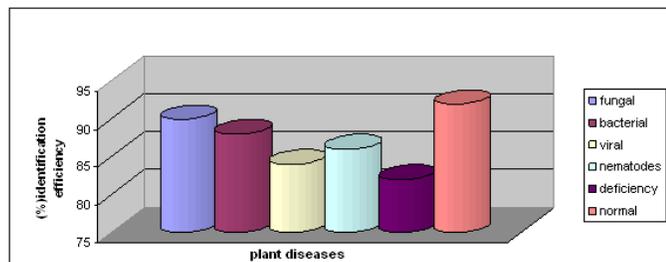


Fig. 15: Classification efficiency for plant diseases using combined features with ANN

A comparison between the performances of SVM and ANN classifiers has been done. The comparison of classification accuracies of both the classifiers using reduced color, texture and combined features is given in the Table 3. From the Table 3, it is evident that the SVM based classifier has given better classification accuracy than the ANN classifier.

TABLE 3: PERFORMANCE OF SVM AND ANN CLASSIFIERS USING COLOR, TEXTURE AND COMBINED FEATURES

classification performance rate(%)	classifiers					
	SVM			ANN		
	color	texture	comb ined	color	texture	comb ined
minimum accuracy	78	83	86	72	78	82
maximum accuracy	91	96	98	94	92	92
average identification accuracy	84	89	92	82	84	87

The performance of individual classifiers with respect to true positive rate (TPR), false positive rate (FPR), precision, recall, F-measure, and average classification accuracy (ACA) using SVM and ANN classifiers with combined features is given in the Table 4. From the Table 4, it is evident that SVM classifier has given better classification accuracy than ANN for identification and classification of plant diseases affecting agriculture/horticulture crops.

TABLE 4: PERFORMANCE EVALUATION OF SVM AND ANN CLASSIFIERS USING COMBINED FEATURES

Classifier	TP Rate	FP Rate	Precision	Recall	F-Measure	ACA (%)
SVM	0.931	0.917	0.976	0.921	0.945	92.17
ANN	0.847	0.953	0.823	0.847	0.834	87.48

V. CONCLUSION

Recognizing the plant disease is mainly the purpose of the proposed approach. Thus, the proposed algorithms are tested on five plant diseases with normal which influence the agriculture/horticulture crops. The developed classifier uses color and texture features for classification of plant diseases and affecting agriculture/horticulture crops. Even though these features have given different accuracies in isolation for different images of plant disease affecting agriculture/horticulture crops, the very combination of features has proved to be effective. The main advantage of this proposed approach is that, considerably less number of features is deployed to achieve better classification accuracy and to reduce computation time. The experimental results indicate that the proposed approach is a valuable approach, which can significantly support an accurate detection of plant disease in a little computational effort. It indicates that SVM achieves a significant improvement in the classification accuracy over ANN. SVM proved to be a powerful tool for automatic classification of plant diseases considered in the present work. But, there is scope for enhancement in the classification accuracy.

REFERENCES

- [1] Bandi S R., Vardharajan A. and Chinnasamy A, Performance Evaluation of Various Statistical Classifiers in Detecting the Diseased Citrus Leaves, International Journal of Engineering Science and Technology, Vol. 5, Issue.2 February 2013, pp.297-307.
- [2] Bauer S D., Korc F. and Wolfgang Forstner, The Potential of Automatic Methods of Classification to Identify Leaf Diseases from Multi-spectral Images, Precision Agriculture, Vol.12, 2011, pp.361-377.
- [3] Barbedo J C A, Digital Image Processing Techniques for Detecting, Quantifying and Classifying Plant Diseases, Springer Plus, Vol.2, Issue.660, 2013, pp.1-12.
- [4] Cui D., Zhang Q., Li M., Hartman G L. and Zhao Y, Image Processing Methods for Quantitatively Detecting Soybean Rust from Multi-spectral Images, Biosystems engineering, Vol.107, Issue.3, 2010, pp.186-193.
- [5] Dubey S R., Dixit P., Singh N. and Gupta J P, Infected Fruit Part Detection using K-Means Clustering Segmentation Technique, International Journal of Artificial Intelligence and Interactive Multimedia, Vol. 2, Issue.2, 2013, pp.65-72.
- [6] Fenyvesi L., Fenyvesi D. and Csatár A, (2013), Stress Analysis in Fruits, Advances in Mechanical Engineering, Hindawi Publishing Corporation, Article ID: 874673, 2013, pp.1-6.
- [7] Fina F., Birch P., Young R., Obu J., Faithpraise B. and Chris Chatwin, Automatic Plant Pest Detection and Recognition Using k-means Clustering Algorithm and Correspondence Filters, International Journal of Advanced Biotechnology and Research, Vol. 4, Issue.2, 2013, pp.189-199.

- [8] Jagadeesh D. Pujari, Rajesh Yakkundimath and A.S.Byadgi, Recognition and Classification of Normal and Affected Agriculture Produce Using Reduced Color and Texture Features, International Journal of Computer Applications, Vol.93, Issue.11, 2014, pp.17-24.
- [9] Jagadeesh D. Pujari, Rajesh Yakkundimath and A.S.Byadgi, Automatic Fungal Disease Detection based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops, International Journal of Image, Graphics and Signal processing, Vol.6, Issue.1, 2014, pp.24-31.
- [10] Jagadeesh D. Pujari, Rajesh Yakkundimath and A.S.Byadgi, Detection and Classification of Fungal Disease with Radon Transform and Support Vector Machine Affected on Cereals, International Journal of Computer Vision and Robotics, Vol.4, Issue.4, 2014, pp.261-280.
- [11] Khoje S A., Bodhe S K. and Adsul A, Automated Skin Defect Identification System for Fruit Grading Based on Discrete Curvelet Transform International Journal of Engineering and Technology, Vol.5, Issue.4, September-2013, pp.3251-3256.
- [12] Landge P S., Patil S A., Khot D S., Otari O D. and Malavkar U G, Automatic Detection and Classification of Plant Disease through Image Processing, International Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue.7, July 2013, pp.798-801.
- [13] Prasad S., Piyush Kumar., Hazra R. and Ajay Kumar, Plant Leaf Disease Detection Using Gabor Wavelet Transform, Proceedings of 3rd International Conference on Swarm, Evolutionary, and Memetic Computing, Springer-Verlag, Berlin, Heidelberg, 2012.
- [14] Pujari J D, Content Based Image Retrieval Using Color and Texture Features, Ph.D. Thesis, Gulbarga University, Gulbarga, 2008.
- [15] Rafael C G., Woods R E. and Eddins S L, Digital Image Processing, Prentice Hall, 3rd edition, 2008.(92)
- [16] Rathod A N., Tanawal B. and Shah V, Image Processing Techniques for Detection of Leaf Disease, International Journal of Advanced Research in Computer Science and Software Engineering, Vol.3, Issue.11, November 2013, pp.397-399.
- [17] Razmjooya N., Mousavib B S. and Soleymani F, A Real Time Mathematical Computer Method for Potato Inspection Using Machine Vision, Computers and Mathematics with Applications, Vol. 63, 2012, pp.268–279.
- [18] Savakar D G. and Anami B S, Improved Method for Identification and Classification of Foreign Bodies Mixed Food Grains Image Samples, International Journal of Artificial Intelligence and Machine Learning, Vol.9, Issue.1, 2009, pp.1-8.
- [19] Sonka M., Hlavac V. and Boyle R, Digital Image Processing and Computer Vision, Cengage Learning, Indian Edition, 2008.(105)
- [20] Sungkur R K., Baichoo S. and Poligadu A, 2013, An Automated System to Recognize Fungi-caused Diseases on Sugarcane Leaves, Proceedings of Global Engineering, Science and Technology Conference, Singapore, 3-4 October, 2013.
- [21] Vibhute A. and Bhode S K, Applications of Image Processing in Agriculture: A Survey, International Journal of Computer Applications, Vol. 52, Issue. 2, August 2012, pp.34-40.
- [22] Pacheco, A. et al. Reconstruction of High Resolution 3D Objects from Incomplete Images and 3D Information. International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 2, Issue 6, pp. 7-16. DOI: 10.9781/ijimai.2014.261



Abdulmunaf. Syedhusain Byadgi He has obtained Bachelor of Science (Agri.) in 1977(UAS, Bangalore, INDIA), M.Sc. (Agri.) in 1980(UAS, Bangalore, INDIA) and Ph.D. in 1994 (IARI, New Delhi, INDIA). Currently is working as Professor in the department of Plant Pathology, UAS, Dharwar, INDIA.



Jagadeesh D.Pujari He has obtained Bachelor of Engineering in Computer Science in 1990 and MS in Software Systems (BITS PILANI) and Ph.D. in Computer Science in 2008. Currently he is working as Professor & Head in the department of Information Science & Engineering, SDMCET, Dharwad, INDIA. His research areas of interests are Image Processing Pattern Recognition and Content-Based Image Retrieval.



Rajesh Yakkundimath He has obtained Bachelor of Engineering in Electronics & Instrumentation in 2004(VTU, Belgaum, INDIA) and M.Tech. in Computer Science & Engineering in 2007 (VTU, Belgaum, INDIA). He is working for his doctoral degree in Computer Science under VTU, Belgaum, INDIA. Since 2008 he is working as Assistant Professor in the department of Computer Science & Engineering, KLE.Institute of Technology, Hubli,

INDIA.