❐ 4077

Soybean leaf disease detection and severity measurement using multiclass SVM and KNN classifier

Sachin B. Jadhav[1], Vishwanath R. Udupi[2], Sanjay B. Patil[3]
[1]Bharati Vidyapeeths College of Engineering Kolhapur, India
[2]Maratha Mandala Engineering College Belagavi, India
[3]Shri. Chattrapati Shivajiraje College of Engineering Bhor, India

Article Info

ABSTRACT

Soybean fungal diseases such as Blight, Frogeye leaf spot and Brown Spot are a significant threat to soybean plant due to the severe symptoms and lack of treatments. Traditional diagnosis of the thease diseases relies on disease symptom identification based on neaked eye observation by pathalogiest, which can lead to a high rate of false-recognition. This work presents a novel system, utilizing multiclass support vector machine and KNN classifiers, for detection and classification of soybean diseases using color images of diseased leaf samples. Images of healthy and diseased leaves affected by Blight, Frogeye leaf spot and Brown Spot were acquired by a digital camera. The acquired images are preprocessed using image enhancement techniques. The background of each image was removed by a thresholding method and the Region of Interest (ROI) is obtained. Color-based segmentation technique based on Incremental K-means clustering is applied to the region of interest for partitioning the diseased region. The severity of disease is estimated by quantifying a number of pixels in the diseased region and in total leaf region. Different color features of segmented diseased leaf region were extracted using RGB color space and texture features were extracted using Gray Level Co-occurrence Matrix (GLCM) to compose a feature database. Finally, the support vector machine (SVM) and K-Nearest Negbiour (KNN) classifiers are used for classifying the disease. This proposed classifers system is capable to classify the types of blight, brown spot, frogeye leaf spot diseases and Healthy samples with an accuracy of 87.3% and 83.6% are achieved.

Corresponding Author:

Sachin Balkrushna Jadhav,
Department of Electronics and Telecommunication Engineering,
Bharati Vidyapeeths College of Engineering,
Near Chitranagari, Kolhapur, M.S, 416013.
Email: sachinbjadhav84@gmail.com

1. INTRODUCTION

Soybean is one of the oldest crops of the world. Today, soybean is the largest source of oil for human consumption as well as protein for livestock feed. It is reported that soybean has helped in improving the social and economic conditions of a large number of small and marginal farmers in India. The major soybean-growing states are Madhya Pradesh, Maharashtra, Rajasthan, Karnataka, Andhra Pradesh, and Chattisgarh. These states together contribute to about 98% of the total soybean production in the country [1]. In recent decades due to some factors such as natural disasters, soil erosion etc., lead to the incidence of crop diseases. These diseases can result in significant reduction in both quality and quantity of agricultural yields. The consequence of this is it causes production and economic losses. Most of the time farmers come across great difficulties in identifying and controlling plant diseases. Thus, it is very important to diagnose the plant diseases at early or superior stages so that proper and suitable action can be taken by the farmers to avoid further loss.

Soybean Leaf diseases like Bacterial Leaf Blight, Septoria Brown Spot, and Soybean rust are cause significant reduction in yield loss and lead to affect quality of soybean Products [2], thus influence farmers life. An effective way to control soybean foliar diseases is by applying fungicides.The identification of disease with traditional expert system are always subjective and this will lead to an inaccurate diagnosis. In recent years there is rapid development of computer vision techniques and pattern recognition techniques, which can intelligently diagnose the crop diseases and accurately identify the types of diseases [3-8].

Image processing technique and computer vision system have been used for automatic detection and classification of plant disease from extracted color, texture and shape features [9-11]. Studied show that image processing and computer vision tools can be successfully used as automatic and accurate disease detection mechanism [12]. For this work, we had selected Bacterial Blight, Septoria Brown Spot and Frogeye Leaf spot disease samples, whose symptoms are specified in Table 1 and these disease samples, which are selected from our collected dataset of soybean diseases, are depicted in Figure 1.

Table 1. Symptoms of Soybean Leaf Diseases

Disease Name	Casual Pathogen	Symptoms	Conditions and Timing of Disease occurrence	Yield Loss in %
Bacterial Blight	Pseudomonas syringae pv. glycinea,	a) Causes small, angular, yellow-to-brown Spots. b) Surrounded by yellow halos. c) The angular spots enlarge and join together producing large, irregular dead areas. d) The centers of old dead areas tear away so infected leaves have a ragged appearance.	a) Plants can be infected at any time during the growing season. b) Cool, wet weather and rain storms favor disease development. c) It is spread by wind and rain and by cultivation when foliage is wet.	severe defoliation
Cercospora Leaf Spot	Cercospora sojina.	a) Lesions on leaves begin as small, dark, water-soaked spots. b) They develop into brown spots surrounded by a darker reddish-brown or purple ring. c) The center of spots may turn white with black specks visible	a) At any time during the growing season, but seems to typically occur after flowering. b) Young leaves on the tops of plants tend to be most susceptible. c) Infection and disease development is favored by warm, humid weather.	severe defoliation
Brown Spot	Septoria glycines i	a) The first symptoms usually appear on the lower leaves and then progresses to the mid-to-upper canopy throughout the summer. b) Initial symptoms are small dark brown spots. c) The brown spots often enlarge and grow together into irregular brown, yellow areas.	a) It is most common when conditions and leaves are wet and warm, ideally for extended periods of time. disease.	5-8%

Figure 1. Soybean diseased leaf samples (a) Blight, (b) Frogeye leaf spot, (c) Brown spot

The various researchers investigated their methods for disease detection, classification and assessment key of disease severities for different plant diseases which are outlined as follows: Dheeb Al Bashish, Malik Braik, Sulieman [13], proposed a framework for detection of plant diseases present on leaves and stem. The proposed framework is composed of K-Means segmentation technique and the segmented images are classified using neural network classifier. Dhiman Mondal, Dipak Kumar Kole [14] proposed a method to detect and classify the presence of yellow vein mosaic virus disease of okra leaf with the aid of K-means and Naive Bayesian classifier. He has experimented on 79 standard diseased and

non-diseased okra leaf images. Classification showed an average accuracy of 87%.

Evy Kamilah Ratnasari [15] proposes a model to identify the severity of spot disease which appears on leaves. He used thresholding a* component of L*a*b* color space for disease spot detection. Then features of diseases spots are extracted using Support Vector Machine (SVM) classification techniques, first he uses L*a*b* color space for its color features and then applied Gray Level Co-Occurrence Matrix (GLCM) to obtain texture features. Study revealed that proposed model is capable to determine the types of spot diseases with an accuracy of 80% and 5.73 error severity estimation average. It is reported that for severity measurement; percentage of the infected area estimated by quantifying number pixels in diseased symptoms area and number of pixels in total leaf area of the leaf has been often used in disease assessment techniques.

Also while using the manual assessment method it is found that such methods contain some degree of subjectivity; hence we may not be considered it as ground truth. Above discussion concludes that most of the researchers investigated disease segmentation method using K-means clustering techniques which has the limitation on cluster selection (local cluster selection). Therefore these issues along with enhancement of performance accuracies demand automatic and accurate method to detect plant diseases is of great realistic significance [2, 7, 9]. In this work, we present a novel Incremental K-means segmentation algorithm [16] (global cluster selection) for soybean leaf disease detection and classification using multiclass SVM and KNN which may be used to overcome these limits.

This paper is organized as follows: Section 2 methodology used for soybean leaf disease segmentation using novel Incremental means clustering and SVM and KNN Classifiers followed by disease identification, data acquisition, image processing technique, classification method. Section 3 reveals the experimental results and discussion using SVM and KNN Classifiers and finally, Section 4 presents the conclusion.

2.   RESEARCH METHOD

This study reports on a framework for the detection, classification and severity measurement of soybean plant disease visual symptoms by the analysis of colored images. The framework was divided into two phases: (I) Training Phase and (II) Testing Phase. For both phases the first 4 stages are common that includes:
− Training Database image set and input test Image.
− Image pre-processing: to resizing the database images in to suitable form; to eliminate leaf object background regions; to specify a suitable color transformation that best highlighted the diseased regions shown in the image; image enhancement: to filtering unwanted noise in the input image and enhanced highlighted diseased regions considered targets (possible diseased area);
− Image segmentation: to partitioning the identified leaf regions in the image into 3 clusters that out of which one cluster were likely to qualify as diseased region;
− Feature Extraction: to extract the significant information from the given input sample; color and texture features were extracted and the same is used for further image analysis.
− Classification: Classifier is trained using extracted feature values and its respective target values using the SVM and KNN classifiers and then this trained classifiers is used to classify test images. A flowchart of the complete process is shown in Figure 2.

2.1. Image database

The set of images used in this study were obtained from the Department of Plant Pathology College of agriculture; Agriculture research Centre, Sub-mountain Zone, Kolhapur, affiliated to Mahatma Fule Krishi Vidyapeeth Rahuri, Maharashtra, India. The Department of plant pathology, at the college of agriculture, India supplied the sample of diseased soybean leaves. Then diseased soybean leaves are placed on the white base to remove complexity and then images are taken by using a digital camera which is used for both training and testing the system. In all cases, the standard JPG image format was used to store these images. Stored images include the soybean leaves infected by Septoria Brown Spot, Cercospora leaf spot, and Bacterial Blight.

2.2. Image preprocessing

The set of images from a database is given and then image preprocessing is done on gathered images for improving the image quality. The preprocessing of image includes: image resizing-in this stage due to the large size of the captured image the input image is first resized to size 256*256; image smoothing and contrast enhancement–due to inadequate or non-uniform illumination input image may get blurred, so by adjusting the color map of image; contrast enhancement in R, G, B color plane in which the adjustment on

each plane (red, green, and blue) of the RGB intensity image RGB take place as shown in Figure 3. The result causes balance uneven contrast of the image and then any undesired distortion if there will be suppressed it out. After image enchantment, next stage is- image filtering: in which median filter is used to enhance highlighted diseased regions considered targets (possible diseased area) by better contrast and brightness balancing also by which noise elimination (background noise) is also removed. Next stage is Background Removal: in this stage mostly green colored pixels, are masked and based on it we computed threshold value of these pixels. Then green pixels are masked based on: if pixel intensity of the green component is less than the pre-computed threshold value, then zero value is assigned to the red components of this pixel and the region of interest (ROI) is obtained.
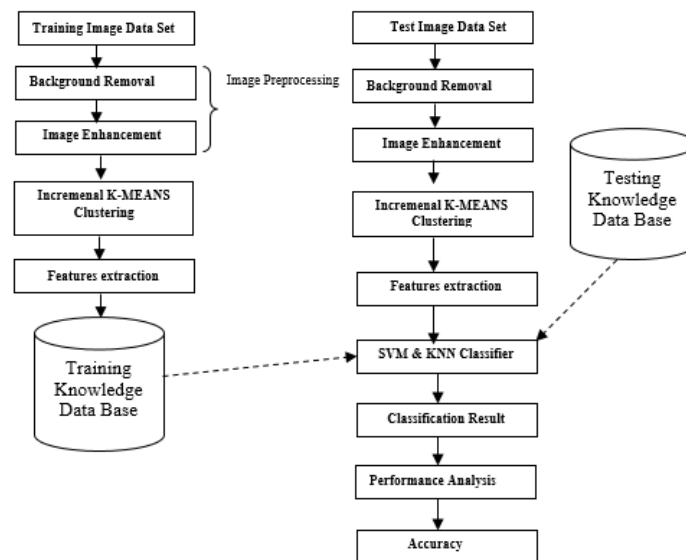
Figure 2. Framework for plant leaf disease detection and classification

Figure 3. ROI of blight diseased leaf sample: (a) Original image (b) ROI of original image

2.3. Leaf Disease Segmentation

We proposed a new system to detect and estimate the soybean leaf disease blight and leaf spots using threshold-based segmented spots image and classify the colors in (R, G, B) color space using incremental k-means clustering technique. The Otsu method is used for spot disease lesion segmentation from (R, G, B) chnnel of color space [17]. The threshold value of the pixels was computed according to the masked pixels. The green pixels were masked according to the following condition: If the pixel intensity of the green component was less than the pre-computed threshold value, a value of zero was assigned to the red components of this pixel and the region of interest (ROI) was obtained.

2.3.1. Incremental k-means clustering algorithm

The conventional k-means clustering method is an extensively used clustering technique that seeks to minimize the average squared distance between points in the same cluster. Although this method offers no

accuracy guarantees, its simplicity and speed are very appealing in practice. By Incremental k-means with a simple, randomized seeding technique, we devloped an algorithm that is competitive with the optimal K-means clustering [16]. Experimental results show our incremental k-means clustering improves both the speed and the accuracy of k-means, often quite intensely. Algorithm:

Let D(x) denote the shortest distance from a data point to the nearest center we have already chosen. Then, we define the following algorithm, which names as k-means++.

1a.   Take one center c1, chosen uniformly at random from X.
1b.   Take a new center ci, choosing x ∈ X with probability $D(x)^2/\sum_{x \in X} D(x)^2$
1c.   Repeat Step 1b until we have taken k centers altogether.
2.   For each i ∈ {1, . . . , k}, set the cluster Ci to be the set of points in X that are closer to ci than they are to cj for all j ≠ i.
3.   For each i ∈ {1, . . . , k}, set ci to be the center of mass of all points in Ci:
      ci = 1/|Ci|$\sum$ x ∈ Ci x.
4.   Repeat Steps 2 and 3 until C no longer changes euclidean distance.
where d (i, j) is the distance between ith and jth pixels of corresponding cluster.

$$d(i, j)=\sqrt{\left|x_{i1} - x_{j1}\right|^2 + \left|x_{i2} - x_{j2}\right|^2 .... + \left|x_{ip} - x_{jp}\right|^2} \tag{1}$$

An example of the output of incremental K-means clustering for a soybean leaf infected with Bacterial Blight depicted in Figure 4.

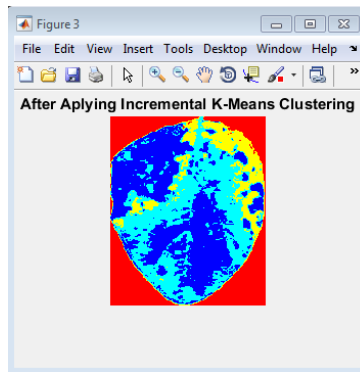
Figure 4. Output of incremental K-Means clustering for a bacterial blight leaf disease
which is shown in 4 clusters

2.3.2. Soybean leaf disease detection and severity measurement system

System is composed of two subsystems a) Disease detection b) severity measurement.
a) Disease detection

The system for detection of soybean leaf disease is shown in Figure 5. It is used for disease detection, and concurrently provides the measurement of disease severity. The proposed system is divided into two stages: the calculation of the leaf disease severity and classification of leaf disease. The first stage is the measurement of disease severity which is performed only on testing data in a certain standard size of the leaves.

Whereas the second stage uses the data from the testing and training spot and blight disease (spot and blight disease use Otsu segmentation method from RGB color space which gives lesion region area (Diseased area-Ad) by keeping only disease pixels on also called lesion region extraction, this region (area) is computed by using regionprops Matlab command. Simultaneously total leaf area (Al) is calculated by keeping all pixels belongs leaf object high (1) and pixels belongs to background keeping it at low (0) using Otsu thresholding segmentation im2bw with greythresh Matlab command and this total leaf region (area) is computed by using regionprops Matlab command shown in Figure 6. Then after disease segmentation using incremental k-means clustering, whose extracted features from knowledge database are further used for disease classification; this is depicted in Figure 5 testing Phase system diagram. Classification result shows the type of disease class and its computed features.
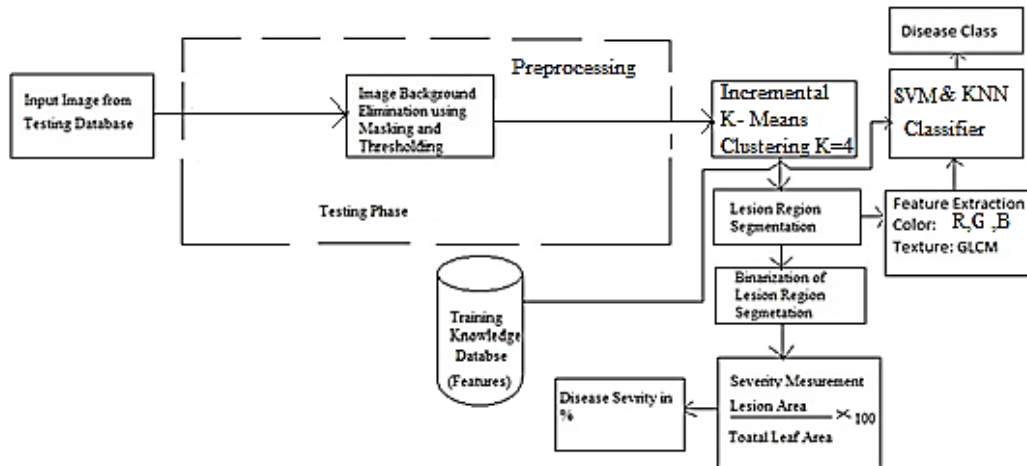
Figure 5. Soybean Leaf Disease Detection and Severity measurement system diagram at Training Phase
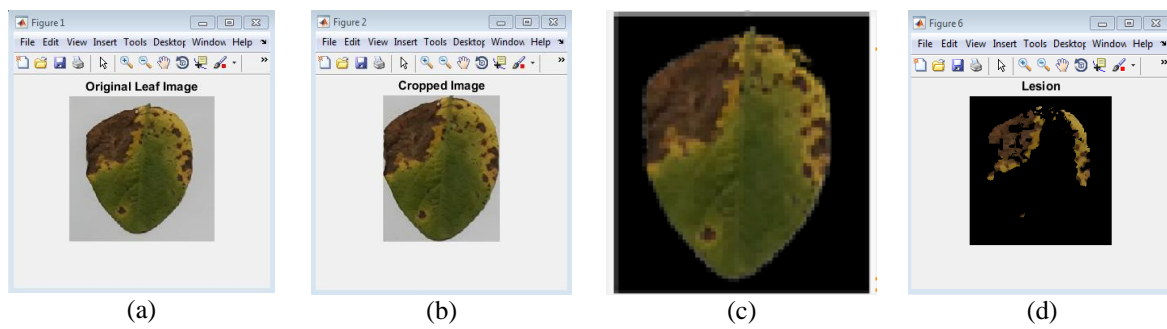

| (a) | (b) | (c) | (d) |

Figure 6. Otsu thresholding lesion segmentation: (a) Original leaf image, (b) cropped leaf image,
(c) ROI of leaf image, and (d) Blight leaf disease lesion segmentation

b) Disease Severity Measurement
Algorithm:
−   Image Acquisition.
−   Images resize.
−   Image preprocessing: Image correction, shadow and background removal
−   Color image segmentation using Incremental k-means clustering
−   Disease lesion rgion extracted and calculating disease area (A_d).
−   Convert the original image to B/W image.
−   Calculate total leaf area (A_l)
−   Calculating from extracted leaf region area (A_l). (Binarization of lesion area)
−   Applying formula to measure disease severity:

$$Ds = (A_d/A_l)*100.$$

One of the objectives of this research work aims at a measurement of the severity of a soybean leaf disease. In this stage Otsu thresholding algorithm from RGB channel color space used to isolate the disease symptoms called lesion region (Ad); from which features can be extracted and properly processed in order to provide an estimate for the severity of the disease. Disease severity is the lesion area of the leaves showing symptoms of spot disease and it is most often expressed as a percentage [15]. The disease severity of the soybean leaf is measured by quantifying the number of white (on) pixel belongs to lesion region called diseased area (Ad) and comparing it with a quantified total number of leaf object pixels. In our work, the binarization method is used to calculate the total leaf area of soybean leaf object (Al) [18]. The severity of disease is given in percentage which is computed using (2). Adopting image processing method to measure disease severity can be expressed as the following formula.

$$DS = \frac{A_d}{A_l} = \frac{P\sum\limits_{(x,y)\in R_d} 1}{P\sum\limits_{(x,y)\in R_l} 1} = \frac{\sum\limits_{(x,y)\in R_d} 1}{\sum\limits_{(x,y)\in R_l} 1} \tag{2}$$

A_d is Lesion Region Area; A_l is Leaf Region Area; P is Unit Pixel Expressed Area; R_d is Lesion Region; R_l is Leaf Region. Unit pixel in the same digital image represent the same size, so ratio DS can be obtained by segmenting lesion region from leaf Region and quantifying pixel number $\sum\limits_{(x,y)\in R_d} 1$ of Disease region and $\sum\limits_{(x,y)\in R_l} 1$ of leaf region in the cluster output of k-means segmented image.

2.3.3. Feature Extraction

In this research total of 8 featutes are extracted out of which 6 are color feaurtures such as Rmean, Gmean, Bmean, Rstd, Gstd, Bstd of R, G, B color space and 2 are texture feature such as skewness and kurtoisis are extracted. For texture features extraction GLCM method is used [19].
- Color feature:
Color feaurtures such as Rmean, Gmean, Bmean, Rstd, Gstd, and Bstd of R, G, B channel color space are extracted thease six features computed using following statisical expression:
i) Mean

$$\text{Mean: } X' = \frac{1}{N} X_i \tag{3}$$

where X_i is the pixel intensity and N is the total number of pixels. Here mean is considered as one of the feature.
ii) Standard Deviation
Standard Deviation is the square root of the variance of the distribution. It is calculated using following formula:

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{N} \sum_{i=1}^{N}(X_i - X')^2} \tag{4}$$

- GLCM texture feature
Texture feature in this research uses GLCM with two feature vectors:
i) Skewness
The skewness is used to judge the image surface. It is used to detect edges in dark objects on white background, having a sign change at luminance changes in images based on degree 3 and 4 moment, so these are termed higher order statistics [20].

$$\text{Skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}(X_i - X')^3}{\left(\frac{1}{N}\sum_{i=1}^{N}(X_i - X')^2\right)^{3/2}} \tag{5}$$

ii) Kurtoisis
Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. Then for univariate data $X_1, X_2, ..., X_N$, the kurtosis is:

$$\text{kurtosis} = \sum X_i = 1(X_i - X^-)4/Ns4 \tag{6}$$

where X^- is the mean, s is the standard deviation, and N is the number of data points. Note that in computing the kurtosis, the standard deviation is computed using N in the denominator rather than N - 1.

2.3.4. Multiclass Support Vector Machine (SVM)

In our research work a Multiclass Support Vector Machine is used for classification of leaf disease. This classifier method is used to identify the 3 different class of soybean leaf disease including 1 healthy class. Classification is based on future extraction result of target class, which deals with associating a given input pattern with one of the defined class. The soybean diseased leves samples were considered for testing and training data in this study is on segmented ROI, leaf spot image.

2.3.5. K-Nearest Neighbour Classifier (KNN)

The leaf sample from testing database is to be identified undergoes feuture extraction process as the ones in training database.Then it will be compared to each leaf fetures extracted in the training dataset one by one The sum of Euclidiean distance between the leaf sample in the testing dataset and those in the training

database are calculated. Each ratio is then used as fetures in KNN classifier.The closest matches are estimated using KNN algorithm [14].

3. RESULTS AND ANALYSIS
3.1. Image database

Classification task consists two phases, training phase and testing phase using the multiclass SVM. In the training phase, a SVM and KNN are trained using extracted leaf object feature values and its relative target class values. This trained SVM is then used to classify test images samples. Raw data of soybean diseased leaves were gathered from soybean fields in Agriculture Research Centre Kolhapur, and from Sangli, Kolhapur district Maharashtra, India and sankeshwar, karnataka, India. There are total 244 training samples of soybean leaf images as data set for defined diseae class testing data set and 80 samples were considered for testing data set (including both early stage as well as Lateral stage images) are used. They cover 55 images with Blight disease, 49 images with frogeye leaf spot disease, and 40 images with Brown spot disease and 100 images with healthy leaf. Figure 1 shows the leaves sample of testing data. The details of the training database are shown in Table 2.

Table 2. Image Database for Training and Testing SVM and KNN classifier

Leaf Species	Database Sample	Total Training Samples	Total Testing Samples
Soybean	Bacterial Blight	60	20
	Ftogeye Leaf Spot	60	20
	Brown Spot	50	20
	Healthy	100	20
Total		270	80

3.2. Incremental K-means clustering

First step is ROI image shown in Figure 7(a) is partitioned into four clusters using Incremental K-means technique. Figure 7(b) shows 4 clusters formed using Incremental K-means techniue. Then second step is selection of diseased lesion rgion cluster containing ROI which is a final segmented image whose feature is extracted for disease analysis. Figure 8 shows diseased lesion rgion cluster containing ROI selection.
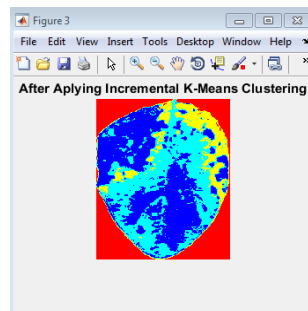
(a)   (b)

Figure 7. Output of Incremental K-Means clustering for a Bacterial Blight disease class which is shown in 4clusters, (a) ROI of leaf image, (b) Color image segmentation using Incremental k-means Clustering
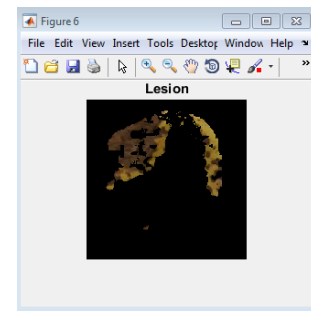
Figure 8. Diseased lesion rgion extraction

3.3. Otsu thresholding and binarisation

The next step is a measurement of disease severity from extracted lesion region using binarization method applying over lesion region and total leaf area as shown in Figure 9. In this stage Otsu thresholding algorithm from R, G, B color space used to isolate the disease symptoms called lesion region (Ad); from which features can be extracted and properly processed in order to provide an estimate for the severity of the disease. Disease severity is the lesion area of the leaves showing symptoms of blight disease and it is most often expressed as a percentage.
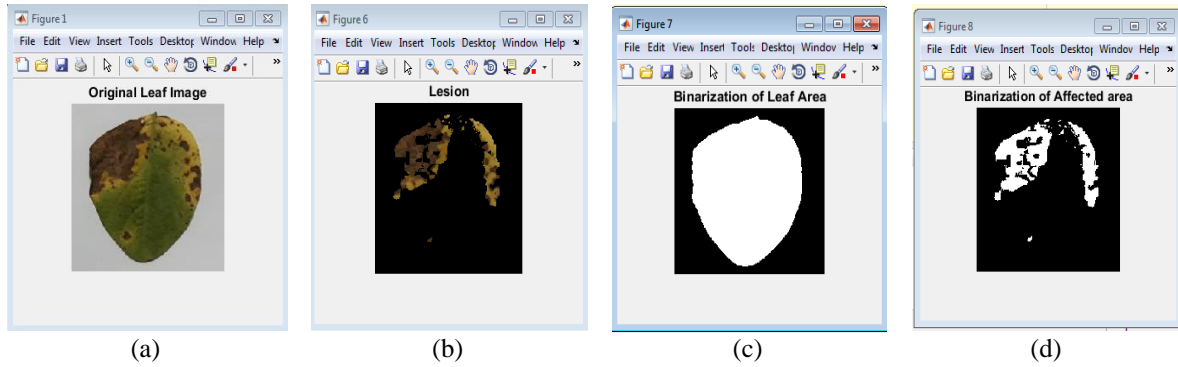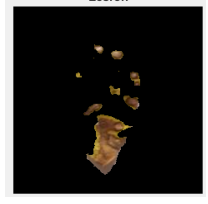
Figure 9. Example of binarisation method applied over blight disease. This shows (a) original image
(b) Extracted lesion region (c) Binarization of Leaf Area (d) Binarisation of lesion region

3.4. Disease Severity Measurement

The disease severity of the soybean leaf is measured by quantifying the number of white (on) pixel belongs to lesion region called diseased area (Ad) and comparing it with a quantified total number of leaf object pixels. In our work, the binarization method is used to calculate the total leaf area of soybean leaf object (Al). The lesion percentage of the leaf is computed using equation (2). After final clustering of ROI, the number of pixels $P\sum 1_{(x,y)\epsilon R_d}$ in the disease region is **6220** and the number of pixels $P\sum 1_{(x,y)\epsilon R_l}$ in the leaf region is **25798** for (**Bacterial Blight**). Thus it can be calculated that the ratio DS of the diseased and leaf area is **0.1460** and its severity is **24.11%.** Table 3 shows estimated soybean disease severities for Bacterial blight, Frogeye leaf spot, and Septoria brown spot disease.

Table 3. Disease severity measurement

Database Sample	Segmented Disease (Lesion) Region	(A_d) Total pixel in diseased area	(A_l) Total pixel in leaf area	(**DS**) Disease severity in%
Bacterial Blight		6220	25798	24.11
Frogeye leaf Spot		2947	12614	23.36
Brown Spot		1328	33078	4.01

3.5. Disease classification

The final step is detecting the bacterial blight, frogeye leaf spot and a brown spot disease type of the segmented images using the multiclass SVM classifier. Furthermore segmented images were used to extract texture and color features. A total of 8 features are estimated for all three partitioned parts of single leaf image. This feature values, collectively called a feature vector, is given to trained multiclass SVM and KNN classifier which classifies the input leaf image into 4 classes Bacterial Blight, frogeye Leaf Spot, Septoria Brown spot and healthy, depending upon its feature values. We labeled black blight disease as class 1, Brown spot disease as class 2, Frogeye leaf spot disease as class 3, and healthy as class 4. Total 241data samples for 4 data class are considered to train the system. From 241 data samples, there are 28 data misclassified, 10 data in class1 misclassified, 15 data with class2 misclassified and 3 data with class3 misclassified shown in Figure 8 of confusion matrix of SVM. So, classification accuracy for class 1 is 78.7%, class 2 is 55.9% class 3 is 95% and class 4 is 100% respectively for leaf with Bacterial Blight, leaf with Brown, spot and leaf with healthy. Average accuracy of SVM classifier is 88.38%.

Similarly for KNN- classifier out of 241 data samples, there are 37 data misclassified, 10 data in class1 misclassified, 16 data with class2 misclassified and 8 data with class3 misclassified shown in Figure 8 of confusion matrix of KNN. So, classification accuracy for class 1 is 78.7%, class 2 is 52.9% class 3 is 81.7% and class 4 is 100% respectively for leaf with Bacterial Blight, leaf with Brown, spot and leaf with healthy.Average accuracy of KNN classifier is 84.64%. Table 4 shows the result of multiclass SVM and KNN classifier out of which SVM perform extremely well with Bacterial black blight class and leaf with a brown spot of soybean leaves as compared with Septoria brown class.

Table 4. Performance analysis of SVM and KNN classifier

Image Database	Total Samples	Correctly Classified (SVM Classifier)	% Accuracy	Correctly Classified (KNN Classifier)	% Accuracy
Bacterial Blight	47	37	78.7	37	78.7
Septoria Brown Spot	34	19	55.9	18	52.9
Frog Leaf Spot	60	57	95	49	81.7
Healthy	100	100	100	100	100
Combined 4 Data Class	241	213	88.38	204	84.64

4. PERFORMANCE EVALUATION

4.1. Confusion Matrix and ROC for SVM Classifier

Figure 10 and Figure 11 shows confusion matrix of proposed diseasae classification system, it shows overall correctly classified and missclassified result of defined disease class, it shows percentage per true class including true positive rates (TPR) and False negative rates (FNR) [20]. The overall sucess rate of multilass SVM classifier is 88.38% shown in Figure 12.
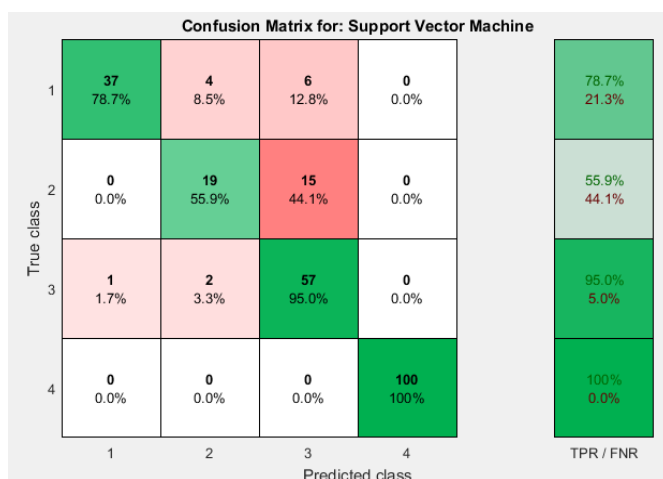

Figure 10. Confusion matrixes for SVM show percentage per true class
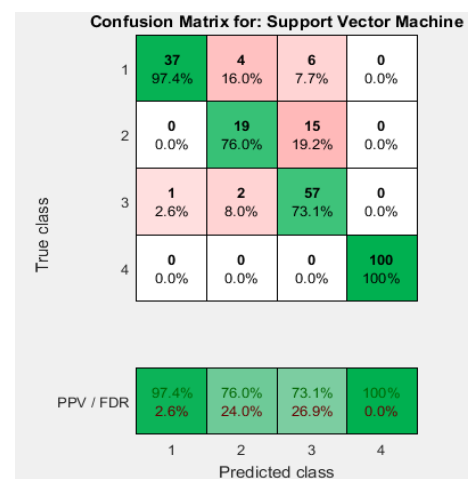

Figure 11. Confusion matrixes for SVM show percentage per predicted class

Figure 10 shows confusion matrix of proposed diseasae classification system, it shows overall percentage per predicted class including positive predictive values (PPV) and False Discovery Rates (FDR). Figure 12 shows confusion matrix of proposed diseasae classification system, it shows percentages over the entire confusion matrix.
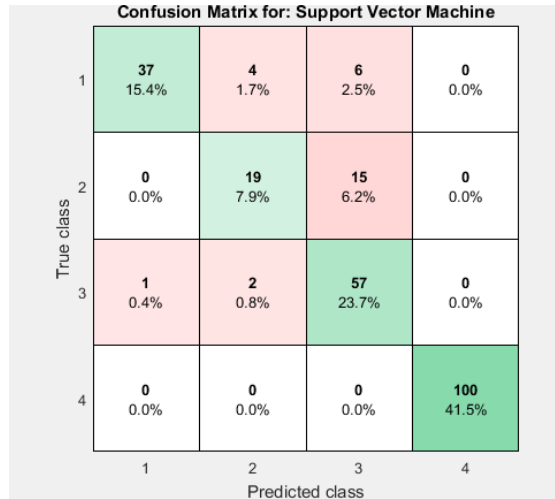

Figure 12. Percentage view of overall entire Confusion Matrix for SVM

4.2. Confusion Matrix and ROC for KNN Classifier
4.2.1. Confusion Matrix

Figure 13 and Figure 14 shows confusion matrix of proposed diseasae classification system, it shows overall correctly classified and missclassified result of defined disease class, it shows percentage per true class including true positive rates (TPR) and False negative rates (FNR) [20]. The overall sucess rate of multilass SVM classifier is 84.64 % shown in Figure 15.
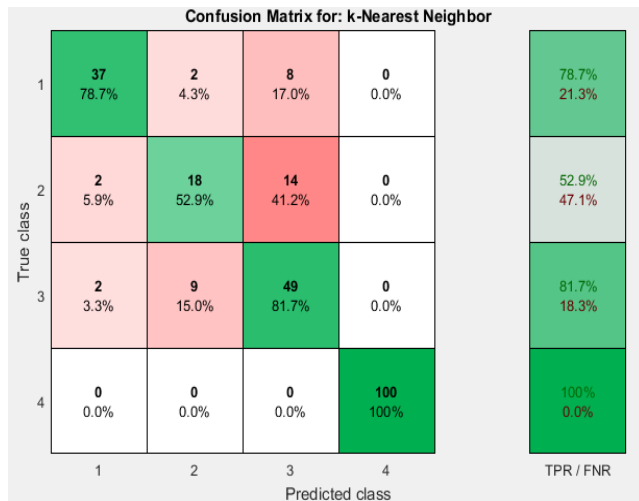
Figure 13. Confusion matrixes for KNN show percentage per true class
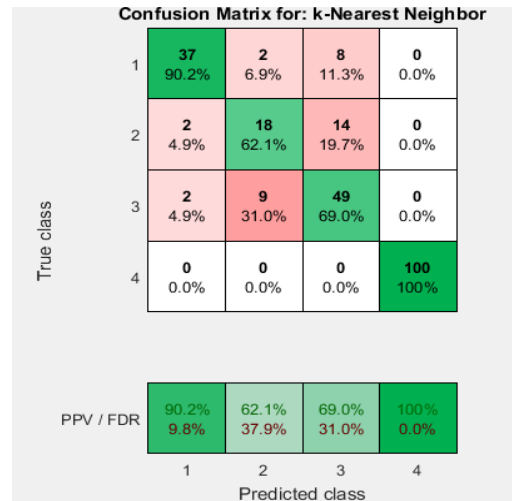
Figure 14. Confusion matrixes for KNN show percentage per predicted class

Confusion Matrix for: k-Nearest Neighbor

Figure 15. Percentage view of overall entire Confusion Matrix for KNN

4.2.2. ROC

It shows Area Under curve for each defined disease class and heathy class. Acording to classification of positive class accuracy of each disease class area under curves varies which is shown in Figure 16 AUC positive class1, Figure 17 AUC positive class2, Figure 18 AUC positive class3 and Figure 19 AUC positive class4 respectivelyfor SVM classifier.
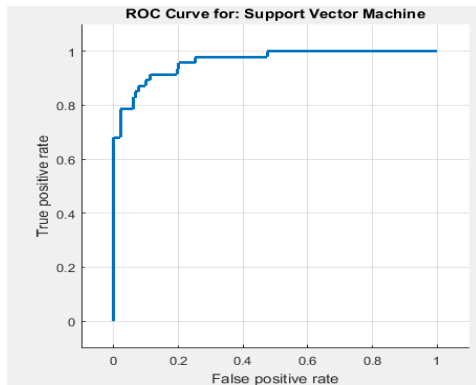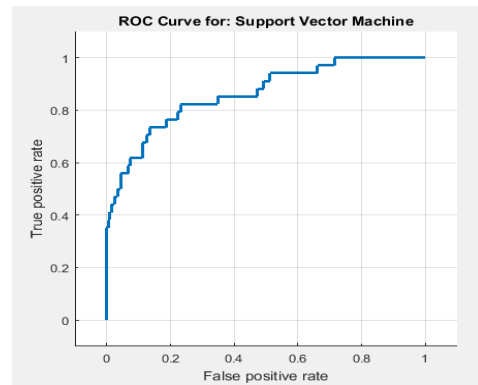

Figure 16. AUC posıtıve class 1(0.963)
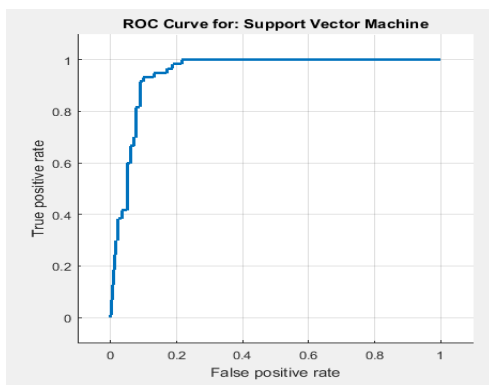

Figure 17. AUC posıtıve class 2(0.863)
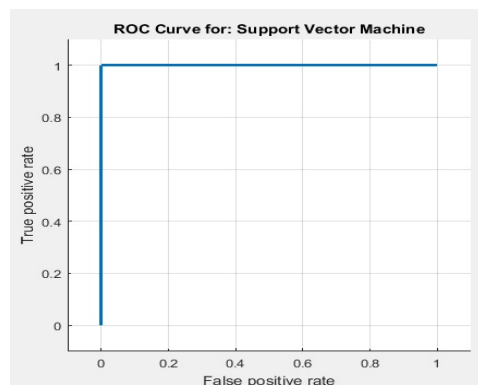

Figure 18. AUC posıtıve class 3(0.947)

Figure 19. AUC posıtıve class 4(1)

4.3. ROC Curve for KNN-Classifier

Similarly Area Under curve for KNN classifier which shows ROC characteristics for defined disease class and heathy class. Acording to classification of positive class accuracy of each disease class area under curves varies which is shown in Figure 20 AUC positive class1, Figure 21 AUC positive class2, Figure 22 AUC positive class3 and Figure 23 AUC positive class4 respectivelyfor KNN classifier.
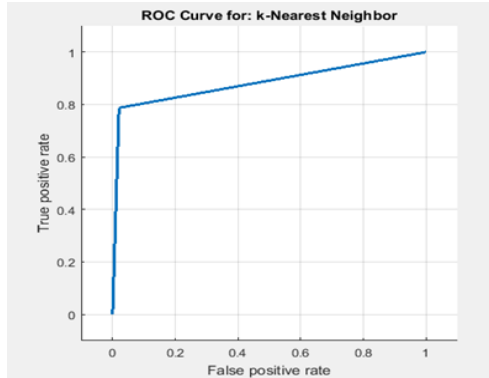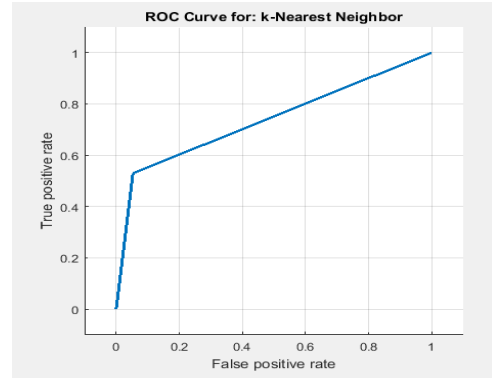

Figure 20. AUC posıtıve class 1 (0.883)
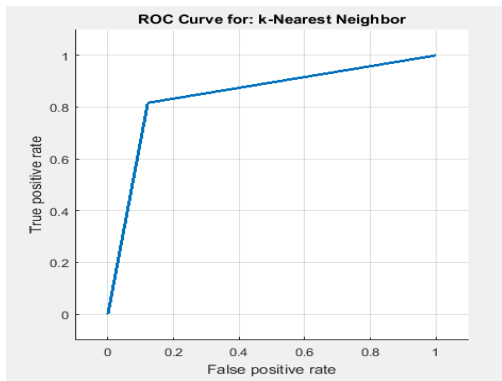

Figure 21. AUC posıtıve class 2 (0.7381)

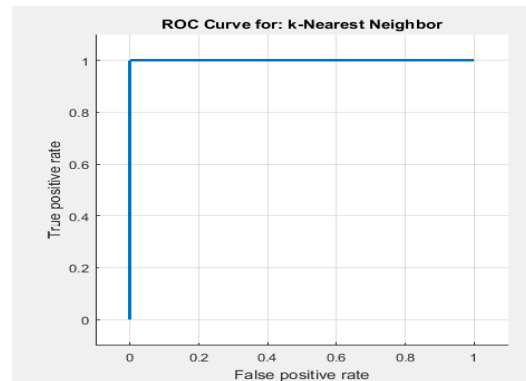Figure 22. AUC posıtıve class 3 (0.847)

Figure 23. AUC posıtıve class 4 (1)

4.4. Scatter Plot for SVM and KNN classiifer

Figure 24 and Figure 25 shows scatter plot of disease classification system, for SVM and KNN classifiers [20] it shows classwise data sanples correctly classified and misclassified for target groupover train data.
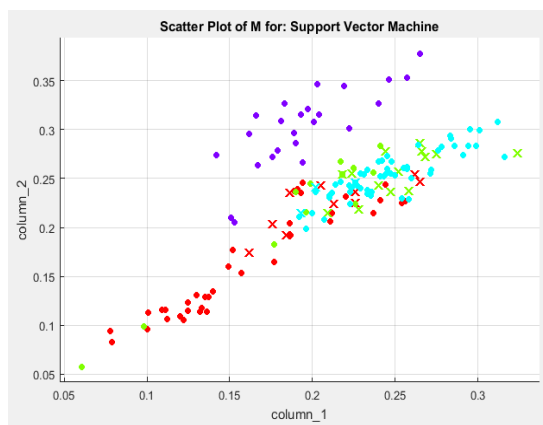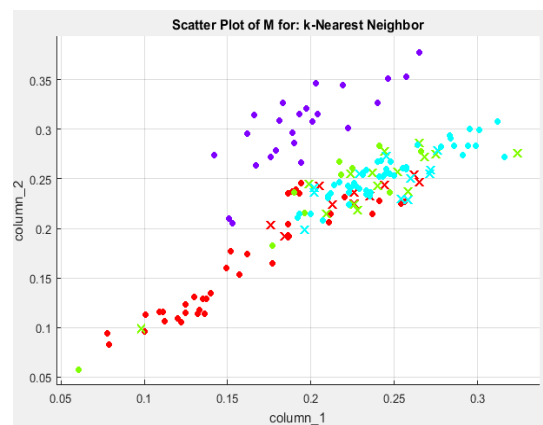

Figure 24. Scatter Plot for SVM classifier

Figure 25. Scatter Plot for KNN classifier

Soybean leaf disease detection and severity measurement using multiclass SVM and ... (Sachin B. Jadhav)

The overall performance of SVM and KNN classifier is summarized into Figure 26 and Figure 27 repsectively. Table 5 shows comparative study of our SVM classifier system with previous system.
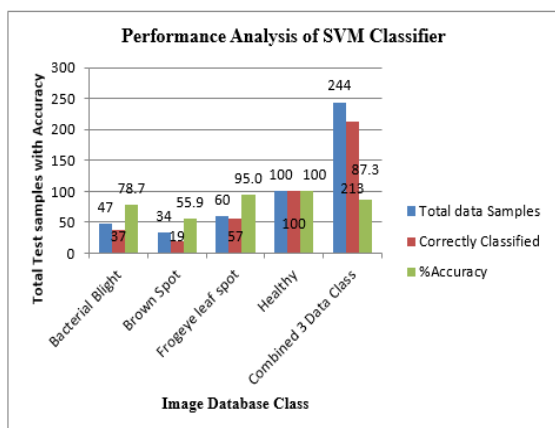

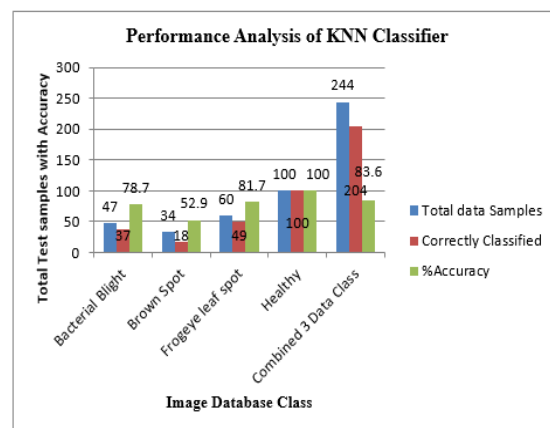Figure 26. Performance analysis of SVM classifier

Figure 27. Performance analysis of KNN classifier

Table 5. Comparative study of Proposed SVM and KNN classifier system

Particulars	Previous Implemented System by Ratnasari et al. [7]	Proposed SVM and KNN Classifier System
Crop Species	Sugarcane	Soybean
Disease Detected Class	Leaf brown spot	Bacterial Blight, Brown spot & Cercospora Leaf spot
Feature extracted	Color & Texture	Texture & Color
Classifier	SVM	Multiclass Linear SVM
Accuracy	80%	87. 3 % & 83.4%

5.   CONCLUSION

We propose a system to identify soybean leaf disease including its severity estimation and classification. We test our algorithms on three disease class of soybean leaves were considered namely, Bacterial Blight, Frogeye Leaf Spot, and Septoria Brown Spot.The given system uses image resizing, and thresholding for image preprocessing. To segment the lesion leaf area, the Incremental K-means clustering technique is used then both color and texture features extraction are done using R,G,B color space and Gray Level Co-occurrence Matrix (GLCM) respectively. Then finally the SVM and KNN classification technique is used to detect the defined type of leaf disease. Our experimental results indicate that the SVM classifier algorithm outperforms over KNN algorithm with an accuracy of 87.3% and 83.4% respectively which can significantly support accurate and automatic detection including classification of leaf diseases. The performance of our proposed system is dependent on the size of a database. In future, we plan to increase random database and clean database to cover more disease types and still better result.

REFERENCES

[1] D. K. Agarwal, et al., "Soybean: Introduction, Improvement, and Utilization in India-Problems and Prospects," *Agric Res,* vol/issue: 2(4), pp. 293-300, Dec 2013.

[2] C. V. Godoy, et al., "Diagrammatic Scale for Assessment of Soybean Rust Severity," *Fitopatol. Bras.,* vol/issue: 31(1), pp. 63-68, 2006.

[3] P. Vincelli and D. E. Hershman, "Assessing Foliar Diseases of Corn, Soybeans, and Wheat," *Principles and Practices PPFS-MISC-06*, 2011.

[4] A. Meunkaewjinda, et al., "Grape leaf disease detection from color imagery using hybrid intelligence system," *IEEE Proceedings of ECTI-CON*, pp. 513-516, 2008.

[5] A. Camargo and J. S. Smith, "An image-processing based algorithm to automatically identify plant disease visual symptoms," *Research Paper: AE—Automation and Emerging Technologies, Biosystem Engineering*, vol. I02, pp. 9-21, 2009.

[6] D. G. Kim, et al., "Classification of grapefruit peel diseases using color texture feature analysis," *International Journal on Agriculture and Biological Engineering*, vol/issue: 2(3), Sep 2009.

[7] D. Al Bashish, et al., "A framework for Detection and Classification of plant leaf and stem Diseases," *IEEE International conference on signal and Image Procesing*, pp. 113-118, 2010.

[8] S. B. Patil and S. K. Bodhe, "Leaf Disease Severity Measurement Using Image Processing," *International Journal of Engineering and Technology*, vol/issue: 3(5), pp. 297-301, 2011.

[9] J. K. Patil and R. Kumar, "Advances in image processing for detection of plant diseases," *Journal of Advanced Bioinformatics Applications and Research*, vol/issue: 2(2), pp. 135-141, Jun 2011.

[10] Y. Zhang and Lenan, "Wu Classification of Fruits Using Computer Vision and a Multiclass Support Vector Machine," *Open Acess Sensors*, vol. 12, pp. 12489-12505, 2012.

[11] S. Phadikar, et al., "Classification of Rice Leaf Diseases Based on Morphological Changes," *International Journal of Information and Electronics Engineering*, vol/issue: 2(3), May 2012.

[12] V. Singh, et al., "Detection of unhealthy region of plant leaves using Image Processing and Genetic Algorithm," *IEEE International Conference on Advances in Computer Engineering and Applications (ICACEA)*, pp. 1028-1032, 2015.

[13] J. Lu, et al., "An in-field automatic wheat disease diagnosis system," *Computers and Electronics in Agriculture,* vol. 142, pp. 369-379, 2017.

[14] T. Munisami, et al., "Plant leaf recognition using shape features and color histogram with K-nearest neighbour classsifier," *Procedia Computer Science*, vol. 58, pp. 740-747, 2015.

[15] S. Kaur, et al., "Plants Disease Identification and Classification through Leaf Images: A Survey," *Archives of Computational Methods in Engineering*, 2018.

[16] D. Arthur and S. Vassilvitskii, "k-means++: The Advantages of Careful Seeding," pp. 1-11.

[17] E. K. Ratnasari, et al., "Sugarcane Leaf Disease Detection and Severity Estimation Based On Segmented Spots Image," *IEEE International Conference on Information, Communication Technology and System*, pp. 93-98, 2014.

[18] S. B. Jadhav and S. B. Patil, "Grading of Soybean Leaf Disease Based on Segmented Image Using K-means Clustering," *IAES International Journal of Artificial Intelligence (IJ-AI),* vol/issue: 5(1), pp. 13-21, Mar 2016.

[19] A. Zisserman, "Supervised Learning Support Vector Machines," www.inf.unibz.it/~mkacimi/svm.pdf.

[20] C. Bishop, "Pattern Recognition and Machine Learning," Springer 2006.

BIOGRAPHIES OF AUTHORS

Sachin. B. Jadhav has pursuing Ph. D from VTU Belagavi and awarded M.E in Electronics and B. E. in Electronics & Telecommunication from Shivaji University Kolhapur. Currently he is working as an Assistant Professor in Department of Electronics & Telecommunication Engineering at Bharati vidyapeeth College of Engineering, Kolhapur. He has keen interest in the field of Image processing.

Vishwanath R. Udupi has Awarded Ph.D, Graduation and Post-graduation in Electronics Engineering. His field of study is Image processing, Cryptography, Neural Networks, Fuzzy Logic, Bio-Medical Application. His keen interest in application of image processing. He had worked as Professor, Assistant professor and Lecturer under Visveswaraya Technological University, Belagavi as well as Shivaji University Kolhapur. His total education experience is about more than 31 years. Currently working as Principal at Maratha Mandal Engineering College, Belagavi. He has written several technical papers in reputed international Journal and Conferences. He is life member of LM ISTE.

Sanjay Bapuso Patil has Awarded Ph.D in subject of Electronics & Telecommunication, Graduation and Post-graduation in Electronics Engineering. His field of study is Digital Image Processing. His keen interest in application of image processing for ―Precision Farming‖. He had worked as Professor, Assistant professor and Lecturer under Pune as well as Shivaji University Kolhapur. His total education as well as industrial experience is about more than 24 years. Currently working as Principal at Shri. Chattrapati Shivajiraje College of Engineering, Bhor, Pune, India,. He has written several technical papers in reputed international Journal and Conferences. He is life member of ISTE.