International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-4 Number-2 Issue-15 June-2014

Disease Detection of Cotton Leaves Using Advanced Image Processing

Vivek Chaudhari[1], C. Y. Patil[2]

Abstract

In this research, identification and classification of cotton diseases is done. The pattern of disease is important part where some features like the colour of actual infected image are extracted from image. There are so many diseases occurred on cotton leaf so the leaf color is different for different diseases. This paper uses k-mean clustering with Discrete Wavelet Transform for efficient plant leaf image segmentation and classification between normal & diseased images using neural network technique. Segmentation is basic pre-processing task in image processing applications and it is required to extract diseased plant leaf from normal plant leaf image and image background. Image segmentation is necessary to detect objects and borderlines in images. Even though different methods are already proposed, it is still hard to accurately segment a random image by one specific method. In last years, additional attention has been given to merge segmentation algorithm and feature extraction algorithm to enhance segmentation results.

Keywords

Pre-processing, Segmentation, Wavelet transform, k-mean clustering, Neural network.

1. Introduction

In whole world, India accounts approximately 25 percent of cotton land. Maharashtra is main cotton growing state in India. Cotton diseases are main problems for decreasing production of cotton. In many advance techniques for agriculture and medical field, image processing is used for multi-dimensional image analysis and applications. Image processing can be used in agriculture for detecting diseased leaf, stem and fruit, size and shape of affected area. Image processing is used in agriculture field for detecting veins, color and texture of plant leaf image. In cotton plants approximately 90 percent of diseases are on cotton leaves.

Hence not whole cotton plant but cotton leaves are mainly suffered from diseases like Alternaria leaf spot, fungus, foliar leaf spot etc. Disease management is a challenging task. Many diseases show visual symptoms of disease but it is quite difficult to precisely quantifying visual patterns. Hence there are many digital imaging techniques to identify such diseases precisely. As human eye cannot differentiate minute variation in color symptoms, various identifying techniques are used. This work represents agriculture atmosphere to sustain the farmers to identify the cotton diseases and get pest recommendation. Cotton foliar diseases recognition and automatic classification using wavelet transform has been used for feature extraction while Support Vector Machine has been used for classification. The work shows that self-organizing feature map combined with a back-propagation neural network can be used for features extraction and those features are used to recognize the color of the image [1]. The modified self-organizing feature map is used to segmentation of cotton leaf disease with genetic algorithms for optimization and support vector machines for classification [2].

The segmentation using clustering approach had many limitations when calculating the number of clusters that are available [3]. Algorithm of leaf recognition using features recognition system is implemented. Plant leaf recognition is done using a Probabilistic Neural Network (PNN). The features are extracted and processed by PCA to reduce those features to input them to PNN [4]. A new Image recognition method based on Wavelet Transform (DWT) and Singular Value Decomposition (SVD) that is capable of retrieving most of the images similar to the target image [5]. In this paper, a proposed approach that makes use of wavelets and K means clustering is applied for leaf images. Through wavelets high pass image is extracted and to enhance edge details further decomposition is applied, which provides fine enhanced edge details with wavelet features like energy and entropy. The wavelet features and the k means are combined in our method to give better accuracy results. Also PCA analysis is

Manuscript received June 06, 2014.
Vivek Chaudhari, Department of Instrumentation and Control, College Of Engineering, Pune, India.
C. Y. Patil, Department of Instrumentation and Control, College Of Engineering, Pune, India.

used for feature reduction which reduces time required for processing. Further these diseases are classified using back propagation neural network. This method reduces processing time as well as accuracy is increased.

2. Methodology of Work

After looking at various methods used in literature for detection of diseases on cotton leaves we are using following methodology:

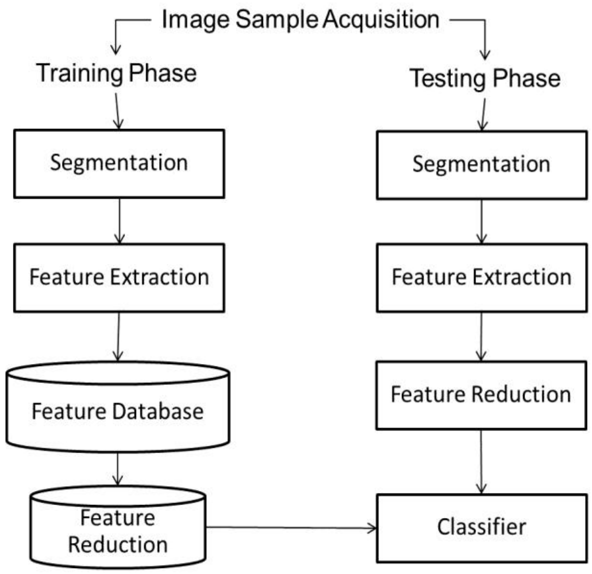

Fig. 1: Methodology for classification of cotton leaves

In this Methodology, images are captured using digital camera and are given to above methodology. Each image is segmented by using clustering technique. From these clustered images features are being extracted using discrete wavelet transform. As extracted features are large to be considered for further classification process, feature reduction is been done using principal component analysis. After feature reduction, images stored in database are classified using back propagation neural network.

A) Image Samples

Various bacterial and fungal affected image samples on cotton leaves such as foliar leaf spot, Alternaria leaf spot, bacterial blight, and cerco spor-leaf spot are considered for detection. Images of leaves in color are acquired using a Sony digital color camera in real field. The leaves used for processing are randomly selected from the cotton field and captured at uncontrolled lighting condition. Before segmentation, image preprocessing is done in which image is resized into 256 x 256 dimensions.

B) Segmentation

Image segmentation is used to determine components of an image into which are more significant and easier to examine. Image segmentation is done using k-means clustering algorithm which is easy to analyze images [12]. Image segmentation is to be done for dividing and extracting image objects even from blur boundary [13]. Clustering needs various image objects which are easily separable from each other to form number of clusters. Hence we are transforming r*g*b color image into L*a*b space.
L*a*b color space is used because it is made up of luminosity layer in 'L' component and two chromaticity layers in 'a' and 'b' components. L*a*b is more organized because all color elements are exists in 'a' and 'b' layers. Here 'a' represents 'red' and 'green'. Similarly 'b' represents 'yellow' and 'blue'. To measure the difference between two colors 'Euclidean distance matrix' is used. K-means classifies the two colors between 'a*b' space. For every pixel in image k-mean determines cluster value.

K-mean Clustering

To study and understand images segmentation is done by using k-means clustering. K-mean was first used by James MacQueen in 1967. A cluster is a group of same items and dissimilar to items which are not same to any other cluster. Clustering is the sorting of items into various groups so that the data in every subset show some common part according to some defined distance measure. An image can be grouped using shapes, textures or any other information that can be taken from the image itself. In K-means, data parts are divided into predefined number of clusters [6]. Firstly the centroids of defined clusters are set randomly. The next step is to take every point belonging to a given data set and link it to the closest centroid. Each pixel is assigned to the cluster based on the closeness of the pixel [7], which is determined by the Euclidian distance measure.

K-means Clustering Steps

K-means is an unsupervised clustering method which classifies the input image into various clusters referring to their distance from each other [8]. Following fig. 2 shows the flow chart of k-means algorithm which is relatively efficient and applicable only when mean is defined.

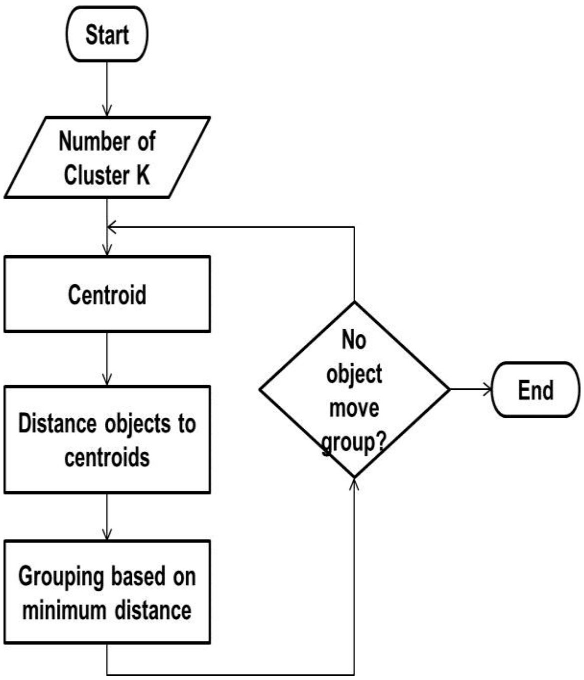

Fig. 2: Flow chart for clustering of cotton leaves

The items are divided around centroids $\mu_i \,\forall i \ldots\ldots k$ which are calculated by minimizing the following objective [8]:

Equation (1)

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

Here k is number of clusters i.e. $S_i, i = 1, 2 \ldots\ldots k$ and μ_i is centroid of all the points $x_j \in S_i$

We have implemented the algorithm as follows [8]:

1) Compute the intensity values
2) Using k unusual intensities initialize the centroids
3) Repeat steps 4 and 5 till the tag of the cluster do not change further
4) Cluster the image points according to distance of their intensity values from the centroid point values [8]

Equation (2)

$$c^{(i)} = \arg\min \| x^{(i)} - \mu_j \|^2$$

5) Compute new centroid for each cluster [8].

Equation (3)

$$\mu_i = \frac{\sum_{i=1}^{m} 1\{c_{(i)} = j\} x^{(i)}}{\sum_{i=1}^{m} 1\{c_{(i)} = j\}}$$

Where k is the number of clusters, *i* repeat over all the intensity values, *j* repeats over all the centroids and μ_i are the centroid intensities.

C) Feature Extraction

After image segmentation, features are extracted from image. Feature to be extracted carries maximum information about the image. Wavelet transform gives proper method for feature extraction because of its following features:

- It compresses most of the signal's energy into some transformation coefficients.
- It has ability to display efficiently low frequency components as well as high frequency transients.

In wavelet transform technique, coefficients are extracted from decomposition vectors. This is done by decomposition vector as input. It uses four coefficients namely approximation coefficients, horizontal, vertical and diagonal coefficients. A parameter for size is calculated which also gives two more functions, matrix rescale and coefficient decimation. The rescale function scales the approximation coefficients, horizontal coefficients, vertical coefficients and diagonal coefficients. The coefficient decimation function selectively extracts the coefficients.

Wavelets in Image Decomposition

The Haar wavelet is the easiest way of compression technique in wavelet group. In 1-D, the following algorithm [9][10] transforms a 2-D vector [*x*(1), *x*(2)]*T* into [*y*(1), *y*(2)]*T* by relation:

Equation (4)

$$\begin{bmatrix} y(1) \\ y(2) \end{bmatrix} = \mathrm{T} \begin{bmatrix} x(1) \\ x(2) \end{bmatrix} \quad where \quad \mathrm{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

T is an orthogonal matrix as its rows are orthogonal to each other. Therefore $\mathrm{T}^{-1} = \mathrm{T}^T$ and it is possible [9] to recover *x* from *y* by relation

Equation (5)

$$\begin{bmatrix} x(1) \\ x(2) \end{bmatrix} = \mathrm{T}^T \begin{bmatrix} y(1) \\ y(2) \end{bmatrix}$$

In 2-dimensions, *x* and *y* become *2 × 2* matrices, at first transform the columns of *x*, by pre-multiplying by T, and then the rows of the result by post-multiplying [9] by T^T to find

Equation (6)
$y = \mathrm{T}\,x\,\mathrm{T}^T$ And in the next step $x = \mathrm{T}^T x\,\mathrm{T}$
To show more precisely what is happening in a that matrix is x of the form [9] is applied
Equation (7)

$$x = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad y = \frac{1}{\sqrt{2}} \begin{bmatrix} a+b+c+d & a-b+c-d \\ a+b-c-d & a-b-c+d \end{bmatrix}$$

The following table shows filtering applies to image:

Table 1: Filtering process applied to 2x2 matrix blocks

Top left : Low-Low Filter	Top right : High-low Filter
Bottom Left: Low-High Filter	Bottom Right: High-High Filter

To apply this filtering to a whole image, the pixels are grouped into *2 × 2* blocks and apply Eq. (6) to each block. In fig. 3, A shows approximation coefficient which is determined by grouping all the top left components in of the *2 × 2* blocks in ***y***. The same goes for the components DL, DC and D. As shown in Fig. 4 the maximum of energy is contained in the upper-left sub-image and the least energy is in the lower-right sub-image.

A	*DL*
DC	*D*

Fig. 3: Wavelet decomposition scheme

Fig. 4: Decomposed image

D) Feature Reduction

We cannot use all the features present in the image as it may end up to the increasing rate of misclassification and increased computation time. Therefore we must choose accurate and useful features only, while discarding remaining features. It is attained by transforming the data set to another set of ordered variables in accordance with their variances. This method is known as "feature reduction". Principal Component Analysis (PCA) is mostly used for feature reduction.

PCA method searches a mapping to find the best representation for issuing of data. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible.

Following are the steps applied for feature reductions [11]:

1. Obtain the mean value of the wavelet feature set
2. Subtract the mean value from original mean value. From these values a new matrix is obtained.
3. Calculate the co-variance matrix of matrix obtained from previous step.
4. Calculate the eigenvectors and eigenvalues of the covariance matrix.
5. Choose components and form a feature vector mainly containing eigenvalues.
6. Determine the new wavelet feature set.

The features obtained from these components are combined to form feature vector which are given as input to the classifiers.

E) Classifier

A back propagation neural network is used as the classifier here. To complete the task of leaf recognition and classification, the multi-layered feed forward network is used with nonlinear sigmoid function [14][15]. Various classes are shown in output by showing various output units in disease classification. The possible output pattern class would be approximately an interpolated version of the output pattern class corresponding to the input learning pattern close to the given test input pattern. This method involved the back propagation learning rule based on the principle of gradient descent along the error surface in the negative direction.

The classification of affected cotton crops procedure:
Start
Step1: Accept the train images on cotton
Step2: Segment the images by clustering

Step3: Extract wavelet based features
Step4: Perform PCA feature reduction
Step5: Train the classifiers with reduced features
Step6: Accept test images and perform Step 2 & 3.
Step7: Recognize and classify the images using classifiers
Stop

3. Experimental Results

After processing the input image segmentation and feature extraction for single image input is done using k-means and DWT algorithm as shown in figure 5 & 6:

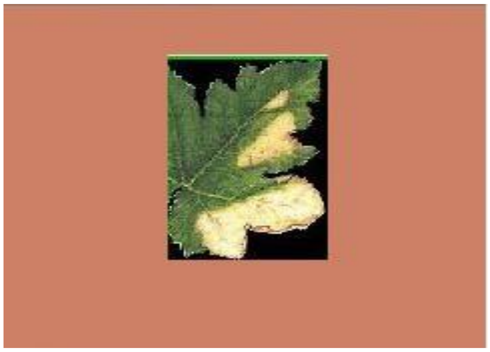

Fig. 5: Input Image

Level 3 decomposition is shown in figure 6 from which we can see that the upper-left corner of image has highest intensity value which shows the processed image. Upper-left corner is combination of approximate values determined from input image.

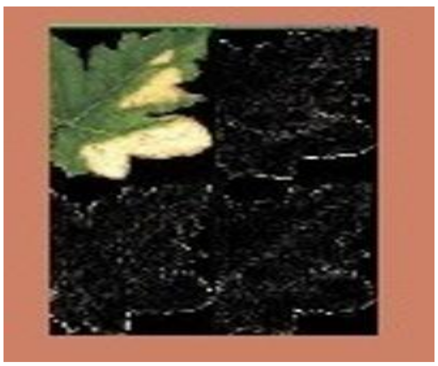

Fig. 6: Level 3 Decomposition DWT Image

Then for each image three predefined clustered are determined with the help of clustering approach. These clusters are shown in fig. 7.

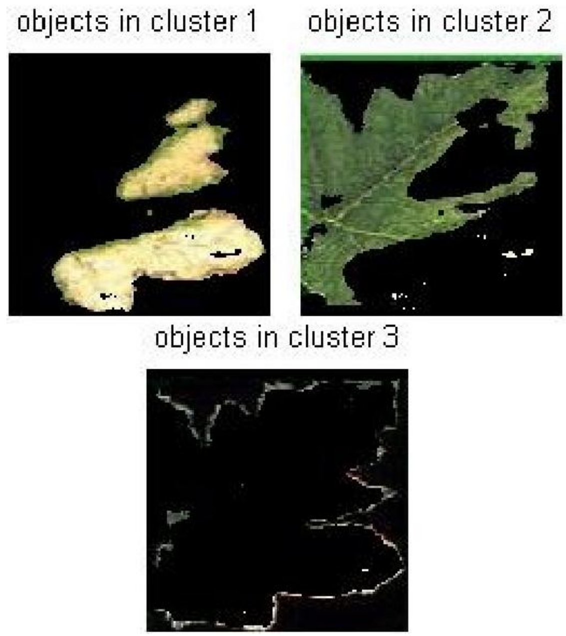

Fig. 7: Clustered Image

By considering the highest cluster value diseased part images are segmented. Segmented image shows the area affected by disease as shown in fig 8. Where in left image is input image and right image is segmented image. Learning rate in GUI can be changed in order to get more accuracy. Accuracy, specificity, sensitivity and type of detected disease are displayed in GUI.

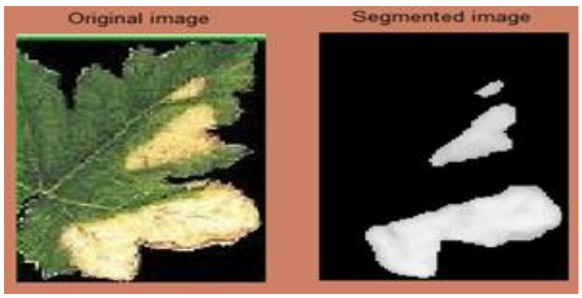

Fig. 8: Segmented Image

Feed forward back-propagation neural network is trained and extracted features are used for classification. Neural network classification output sensitivity and accuracy are displayed in GUI. This work is analyzed with eight types of disease in cotton leaves. The figures show the example for edge detected diseases. Input image and feature extracted image for two diseases are shown in figure 9a and 9b.

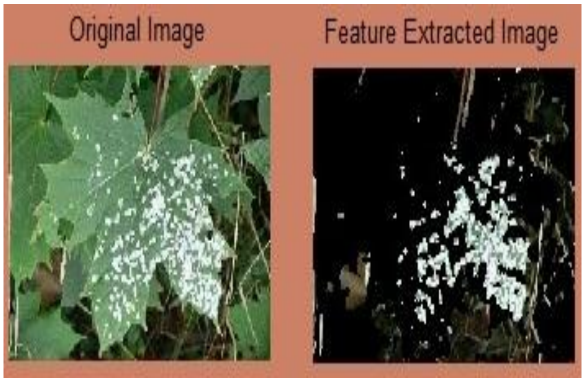

Fig. 9a: Segmented Image

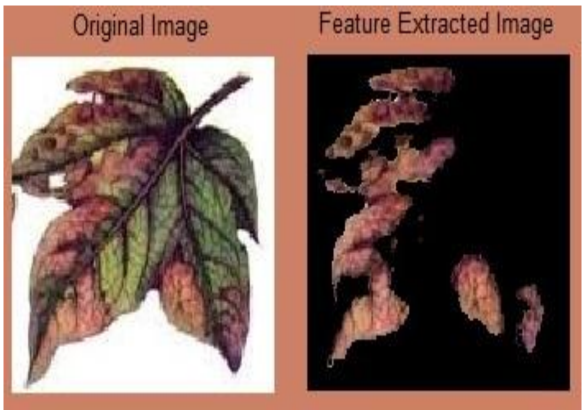

Fig. 9b: Segmented Image

For each single image disease detection accuracy of 97% is attained. Further in following table various classification outputs are shown. In this output accuacy values are taken from literture for SOM + BPNN + PCA + Eigen Vector and RPM & Dis bin CYMK Component methods are taken for comparison. In table 2 number of classified and missclassified images for various eight classes is given.

Table 2: Classification output comparison

Class	Correctly Classified	Missclassified
1	15	0
2	14	1
3	15	0
4	15	0
5	15	0
6	14	1
7	15	0
8	10	5

Table 3: Classification output comparison

Algorithms	Accuracy
SOM+BPNN PCA +Eigen Vector	90
RPM and Dis bin CYMK Components	83
K-mean+DWT+PCA+BPN	97

4. Conclusion

Clustering algorithm is implemented for segmentation purpose. Also DWT is used for feature extraction from leaf image. Further by reducing those features using PCA algorithm, classification of those images is done by neural network analysis. We have got satisfied outputs by combining above method for disease detection and classification as compared to previously implemented algorithms. For single image input we get appropriate detected disease output after implementation of these algorithms. We have got 98% of accuracy in classification on variation of learning rate of neural network.

In future we can development of real time implementation of this algorithm in farm for continuous monitoring and detection of plant diseases. In real time system, we can monitor and give exact solution to avoid various diseases on cotton plant.

Acknowledgment

Author expresses sincere thanks to mentor, Dr. C. Y. Patil for his expert guidance, encouragement and motivation during the development of this project, without which this project would not have developed to its present form. I am thankful to Dr. S. L. Patil for his support and help. I am also very thankful to my friends for helping me to complete my project. The research has not been possible without them.

References

[1] M. V. A. Gulhane and A. A. Gujrar, "Detection of diseases on cotton leaves and its possible diagnosis," International Journal of Image Processing (IJIP), vol. 5, no. 5, pp. 590-598, 2011.

[2] A. A. Gurjar and V. A. Gulhane, "Disease Detection On Cotton Leaves by Eigen feature Regularization and Extraction Technique," International Journal of Electronics, Communication & Soft Computing Science and

Engineering (IJECSCSE), vol. 1, no. 1, pp. 1-4, 2012.

[3] Chandok, Ms. Chinki, Mrs. Soni Chaturvedi and A. A. Khurshid, "An Approach to Image Segmentation using K-means Clustering Algorithm," International Journal of Information Technology (IJIT), vol. 1, no. 1, 2012.

[4] S. G. Wu, Forrest Sheng Bao, Eric You Xu, Yu-Xuan Wang, Yi-Fan Chang and Quio-Liang Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network," in Signal Processing and Information Technology, 2007 IEEE International Symposium on, IEEE, 2007.

[5] M. Malakooti, F. Zahed and Seyyed Mohsen Hashemi, "Image Recognition Method based on Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD)," The Third International Conference on Digital Information Processing and Communications (ICDIPC2013), Society of Digital Information and Wireless Communication, pp. 42-47, 2013.

[6] M. Seli and E. D. Yegenoglu, "Compare various combinations of similarity coefficients and clustering methods for Olea europaea sativa," Scientific Research and Essays, vol. 5, no. 16, pp. 2318-2326, 2010.

[7] T. Kanungo, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman and Angela Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," Pattern Analysis and Machine Intelligence, IEEE Transaction, vol. 24, no. 7, pp. 881-892, 2002.

[8] S. R. Dubey, P. Dixit, Nishant Singh and J. P. Gupta, "Infected fruit part detection using K-means clustering segmentation technique," International Journal of Artificial Intelligence and Interactive Multimedia, vol. 2, no. 2, 2013.

[9] A. K. Jain and Richard C. Dubes, Algorithms For Clustering Data, Pentice Hall Inc, 1988.

[10] N. Vallimmal and S. N. Geethalakshmi, "Leaf image segmentation based on the combination of wavelet transform and k-means clustering," International Journal of Advanced Research in Artificial Intelligence, vol. 1, no. 3, pp. 37-43, 2012.

[11] J. D. Pujari, R. Yakkundimath and S. B. Abdulmunaf, "Automatic Fungal Disease Detection Based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops," International Journal of Image, Graphics & Signal Processing, vol. 6, no. 1, 2013.

[12] D. S. Guru, P. B. Mallikarjun and S. Manunath, "Segmentation and classification of tobacco seedling diseases," in In Proceedings of the Fourth Annual ACM Bangalore Conference, ACM, Banglore, 2011.

[13] P. Revathi and M. Hemlata, "Classification of cotton leaf spot diseases using image processing edge detection techniques," in In Emerging Trends in Science, Engineering and Technology (INCOSET), IEEE, 2012.

[14] S. R. Deokar, P. H. Zope and S. R. Suralkar, "Leaf Recognition Using Feature Point Extraction and Artificial Neural Network," International Journal of Engineering, vol. 2, no. 1, 2013.

[15] M. B. Wadhonkar, P. A. Tijare and S. N. Sawalkar, "Classification of Heart Disease Dataset using Multilayer Feed forward backpropogation Algorithm," International journal of application or innovation in engineering & management, vol. 2, no. 4, pp. 214-220, 2013.

Vivek R Chaudhari was born in Jalgaon, India, in 1991. He received the B.E. degree in instrumentation and control engineering from the University of Pune, Pune, Maharashtra, India, in 2012. He is currently pursuing M.Tech degree in instrumentation and control engineering from University of Pune, Pune, Maharashtra, India, in 2014.

Dr. C.Y. Patil is a Professor at Department of Instrumentation and Control, College Of Engineering, Pune. He has been an Assistant Professor with the Instrumentation and control Engineering Department, University of Pune, India. He has a teaching experience of 22 years. A PhD in Instrumentation, he has 6 international journal, 3 national journal, 3 international conferences and 3 national conference papers to his name. Currently working on the above research, he has previously worked on Design and development of Wireless Sensor Network for Greenhouse Agriculture and Fuzzy based energy efficient &environment friendly Air conditioner. He is a life member of Indian Society of Technical Education (ISTE), Instrument Society of India (ISOI) and Biomedical Society of India.