A Machine Learning System For Recognition Of Vegetable Plant And Classification Of Abnormality Using Leaf Texture Analysis

Sandhya Tippannavar, Dr.Shridevi Soma

Abstract: India being an agriculture major country, crop productivity plays a significant role towards the overall Gross Development Product (GDP) of the country. Vegetables are grown throughout the year under particular climatic condition and cultivation period in India. The vegetables may be affected by bacteria, viruses or insects. It is important to monitor the crops to control the spreading of disease, hence a novel machine learning based technique is proposed for vegetable leaf identification and abnormality detection based on leaf texture analysis. In the proposed method 500 leaf images of 6 different types of vegetables are used for experimentation. Median and tri-lateral filters are used for image enhancement. The leaf part is segmented from the background by threshold and morphological operation, texture and color features are extracted by Fractal feature and Color correlogram respectively. Two-level Classification such as vegetable identification and disease, disorder or normal leaf identification is done using machine learning techniques such as k-nearest neighbor and Probabilistic Neural Network. The accuracy of the vegetable identification and abnormality detection using k-nearest neighbor (k-NN) is 86.39% and 75.04% respectively and Probabilistic Neural Network (PNN) is 75.70% and 71.24% respectively.

Keywords: Fractal features, Color correlogram, k-nearest neighbor (k-NN), Probabilistic Neural Network (PNN).

1. INTRODUCTION

India is an agricultural land, 70% of the population is dependent on agriculture. Vegetables are grown as a major horticultural crop in India throughout the year. Monitoring the pest and disease affecting the growth is very important and they may lead to loss of the farmers. Digital image processing techniques are best suitable to detect disorder type, disease type and different stages of diseases. Such system is used to control spreading of leaf diseases and increase the productivity.

Previous Literature survey [1,2,3,4,5,6,7,8,9, 13, 18,19,20] were based on detection of diseases in plant using leaf images and few [10,11,12,14,15,16,17] in early identification of the diseases so as to control diseases and increase production of the crop. [4] Haiguang Wang et. al. have proposed disease identification of wheat and grape based on image processing techniques where accuracy of 6 classifiers namely PCA, ANN, BPN, RBF, GRNN and PNN are compared. Their experiment result showed the higher accuracy using PCA. [5] Anand.H .Kulkarni et. al. have proposed image Processing Technique to classify and detect pomegranate diseases using ANN with 91% accuracy. [7] Tuan Norjihan et. al. have proposed Classification of watermelon leaf diseases based on color feature extraction and ANN. [8] Jagadeesh D. Pujari et. al. have proposed classification of Fungal Disease Symptoms affected on Cereals SVM and ANN. Best result obtained using SVM.

- *Sandhya Tippannavar is currently pursuing masters degree program in computer science and engineering at PDA college of engineering. E-mail: stippannavar@gmail.com*
- *Dr. Shridevi Soma is Professor and PG Co-ordinator at PDA college of engineering.*

[10] Ms. Kiran R. Gavhale et. al. have proposed detection of unhealthy region of citrus leaf using SVM classifier. [16] Aakanksha Rastogi et. al. have proposed automatic leaf disease detection and grading by computer vision technology using ANN. Total leaf area, diseased area along with percentage of infection is calculated and grading is done using fuzzy logic toolbox.

From the work carried out in the past it is observed that most of the researchers concentrated on detection and recognition of type of plant and diseases and few of them on particular type of plant. The novelty of proposed system addresses the problem of differentiating disease and disorder and the amount of the disease spread with its stages of plant leaf using texture features and color features.

2. VEGETABLE DISEASE AND DISORDER DESCRIPTION

Vegetables such as brinjal, broad beans, cucumber, ridge guard, spinach and tomato leaf images are considered for conduction of the experiment. Disease of different stages for example Leaf_minor_initial and Leaf_minor_final and disorders are considered for detailed analysis. Disease and disorder seem to be synonymous but they are different from each other.

A. Disease

Disease in plants is defined as abnormal growth or dysfunction of plant as a result of some disturbance in that plant life-cycle. Disease in plants may be caused by pathogens (infectious organisms) such as bacteria, fungi, viruses, viroid, nematodes etc. or environmental

conditions (physiological factors). Control of the plant

disease is crucial step towards the production of yield which involves identification and prevention of the disease.

Figure-1 : Example of diseased leaf

A. Disorder

Disorder in plants is caused due to variety of environmental conditions. Unsatisfactory environmental conditions stress plants and cause abnormal growth or disease-like symptoms. The reasons for cause of disorder are air pollution, lack of sunlight, water or soil nutrients, Toxic chemicals, extreme variations in temperature etc.

Figure-2 : Example of Disordered leaf

3. METHODOLOGY

The proposed methodology involves the following stages as shown in Figure-3, to identify the leaf among six different types of vegetable leaves and detect the type of abnormality.

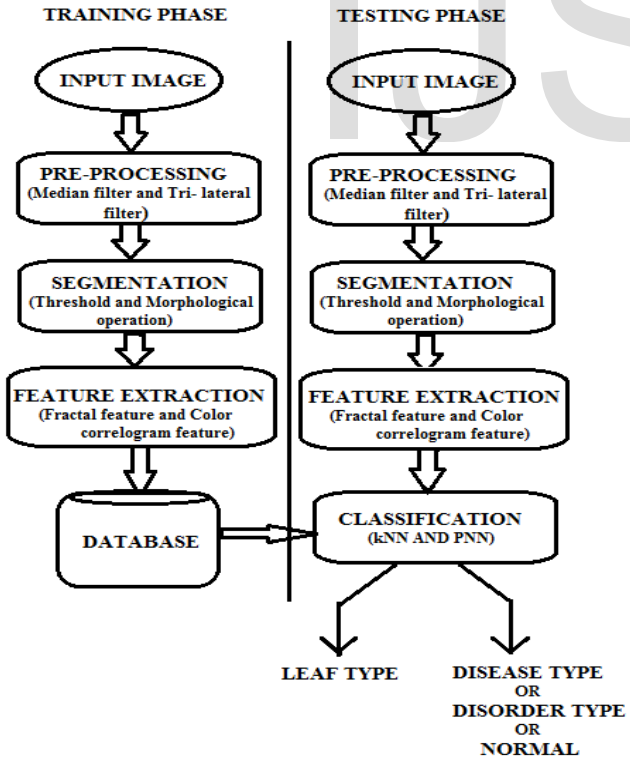


Figure-3 : System Design

A. IMAGE ACQUISITION

The photo graphic images of the diseased, disordered and normal vegetables such as Brinjal, Broad Beans, Cucumber, Ridge Guard, Spinach and Tomato are acquired from the fields by collecting the leaves. The leaves are kept on a white sheet and taken photos under the spot light setup by digital camera Nikon D7000(16MP). The images are in .JPEG format. The main reason for JPEG format is, it compresses the image while storing and it converts he image into bitmap image.

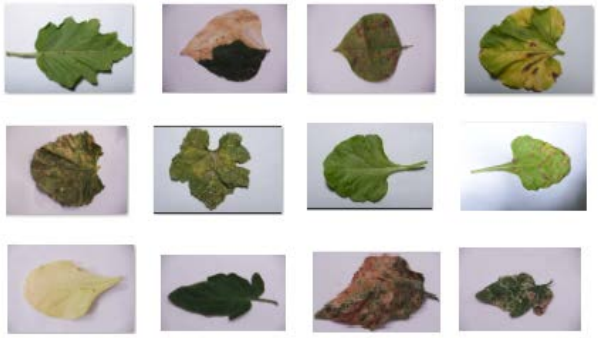


Figure-4 : Database Collection of vegetable leaves

B. IMAGE PRE-PROCESSING

Pre-processing is common procedure for the images which consist of some sort of impurities or noise in it. The aim of pre-processing is improving the data of an image that inhibit unwanted distortions or increase some image features

i. Median Filtering

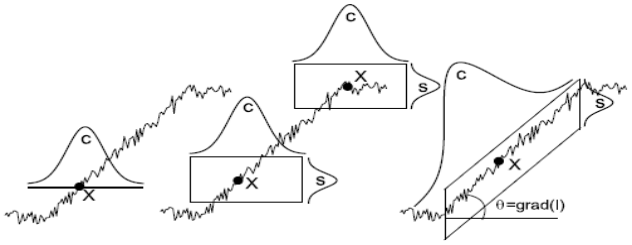
The median filter is used to remove noise and it is type of nonlinear digital filtering technique. It is low pass filter which blocks the high frequency content of the image (mainly boundaries) and allow the low frequency content to pass through it. In median filtering, the neighboring pixels are ranked according to brightness (intensity) and the median value becomes the new value for the central pixel. The median is calculated by first sorting all the pixel values from the surrounding neighborhood into ascending order and then replacing the pixel being considered with the middle pixel value.

$$P_{med}(m,,n) = median\{P(m-k,n-1) \mid -\frac{N_{med}-1}{2} \le k,$$

$$l \le \frac{N_{med}-1}{2} \wedge 1 \le m-k \le m \wedge 1 \le n-1 \le N \}$$

ii. Tri-Lateral Filtering

Bilateral filter is non-linear edge preserving and noise reducing smoothing filter for images. Intensity value at each pixel in an image is replaced by weighted average of intensity value from near-by pixels. The weight calculation can be based on Gaussian distribution. Trilateral filtering is a modification over bilateral filter and can be used for Gaussian noise as well as impulse noise. The 'winsize' should be odd.



(a) (b) (c)

Figure-5 : (a) Unilateral filter window, (a) Bilateral filter window, (c) Trilateral filter window.

C. SEGMENTATION

i. Threshold

It is the simplest method of segmentation. Red, green and blue values will never be 255 on a white background. The value of red and blue color will be less compared to green because leaves are green in color so an intensity value is selected (here the value is 130) and compared with all the color channels and assigned true (1) if the value is satisfied else false (0). This is known as masking which results in masked image.

ii. Morphological Operations

Mask image generated from threshold will have anamolies, some white pixel will be less so morphological filtering is performed. Closed area is removed by erosion which also removes actual area from the edge, hence dilation operation is performed such that most of the areas will be covered. Dilation is taken wider angle to include most of the leaf part. Dilation followed by erosion is known as opening operation. Mask image is subtracted from the original image to get the segmented part i.e., leaf part.

D. FEATURE EXTRACTION

i. Fractal Features

Fractal is described as a family of non-differentiable functions that are infinite in length. Fractal is combination of both shape and texture features. Leaf is a combination of texture patterns hence fractal features can be found. A MATLAB tool, Segmentation-base d Fractal Texture Analysis (SFTA) is used to calculate fractal feature. The path is added and fractal dimension is calculated for RGB color channels using 4 levels of iteration. The Fractal features are 63 in number.

0.0001	0.0191	0.2001
0.0001	0.0193	0.2001
0.0001	0.0063	0.8248

• • • • and so on upto 63

Figure-6 : Fractal feature values

ii. Color Correlogram

Color correlogram is used to extract the color features. Color auto correlogram generates vectors for input image and it is correlation statistics. Color auto correlogram can be stored as table index by color 'i' where 'd'th entry shows the probability of finding a pixel 'i' from the same pixel at distance 'd'. The number of colors repeated and its neighbourhood pixel is calculated. The Color features are 128 in number

0.9960	0.2214	0.4919
0.4791	0.6460	0.2936
0.3811	0.7220	0.3255
0.6187	0.3671	0.5871
0.3060	0.7977	0.1660
0.6455	0.3202	0.4385
0.2544	0.3365	0.2806
0.0755	0.4790	• • • •

and so on upto 128

Figure-7 : Color Feature values

E. IMAGE CLASSIFICATION

The fractal features and color correlogram features extracted from the leaves are combined and used for training. The dataset consists of 500 leaf images out of which 250 images are trained and 250 images are used for tesing.

1. K Nearest Neighbour (k-NN)

1.1 Training

Database training folder consists of 6 leaf type folder which further consists of two folders, abnormal and normal. Abnormal folder consists of one or more disease and disorder folders while normal folder does not have any subfolders. The features are trained in the similar way as mentioned above. It will initially collect "CT" (class type) information then loop through abnormal folder and all the categories in it to collect "CD" (class disease type) information and further to normal which does not have any folders. All 191(Fractal and Color) features are stored in 'F'.

1.2 Testing

Test images are pre-processed using median filter, segmented using threshold and morphological operation, features are extracted using fractal and color correlogram

methods and tested. 250 images are used for testing. The *k* nearest neighbour algorithm (*k*-NN) is a non-parametric method used for classification and regression. The features extracted from the test image are stored in 'f' and these features are tested using the stored features 'F' obtained from trained images for class type as well as disease type. Two-level Classification is done, first result is "Leaf type" and

second result is either "disease, disorder or normal", the result is displayed using message box.

2. Probabilistic Neural Network (PNN)

2.1 Training

Among the various types of neural networks, probabilistic neural network (PNN) is used because it is much faster to train and results are accurate. The Dataset is looped in the same manner as mentioned in k-NN and extracted features are stored. Transpose of the feature vector and the vectorised version of the classes extracted are provided to neural network for training. Out of all the feature values some of the feature values are "NaN" (Not-a-Number obtained when divided by 0). As Neural network will not work for "NaN" data hence all NaN value in the feature vector are made 0. Probabilistic neural network is used to classify leaf type at the first level and for each leaf type the classification of disease and disorder are carried at the second level.

2.2 Testing

A probabilistic neural network (PNN) is a feed forward neural network, which is widely used in classification and pattern recognition. An input image is selected for testing and its features are extracted. The class type and class disease type stored in a variable during training is loaded and the features extracted from test image is compared with the features of the trained images in order to classify given test image for leaf type at first level and disease or disorder at the second level.

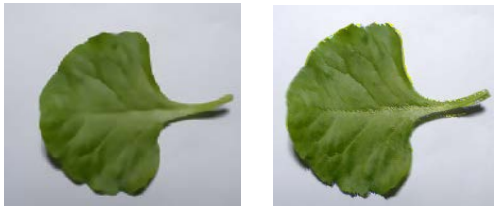
3. Hist Based Testing

Hist based testing involves calculation of nitrogen content in the leaf image. The boundaries of the masked image is found and nitrogen percent is calculating for all the three color channels and it is categorized as either normal or abnormal and the result is displayed with message box.

4. RESULTS AND DISCUSSION

The techniques are implemented using MATLAB. Total of 500 images are collected to create database, out of which 250 images are used for training and 250 images are used for testing. The results of each stage of spinach are as shown below,

A. Pre-Processing



(a) (b)

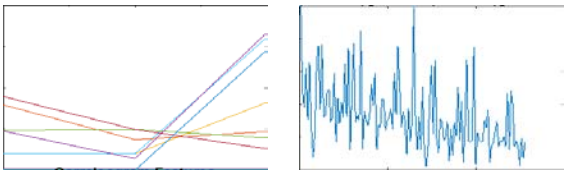
Figure-8 : Pre-processed images (a) Median Filter (b) Tri-lateral Filter.

B. Segmentation

(a) **(b)** **(c)**

Figure-9 : (a) Thresholded image (b) Morphological image (c) Segmented image

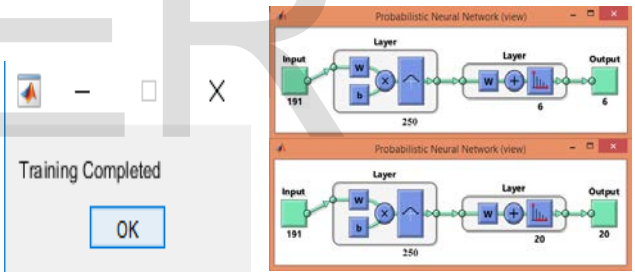
C. Feature Extraction



(a) *(b)*

Figure-10 : (a) Fractal Feature (b) Color correlogram

D. Training



(a) (b)

Figure-11 : (a) Training using features (b) Training using Probabilistic neural network (PNN)

E. Testing

Testing using K-nearest neighbor (k-NN), Probabilistic neural network (PNN) and Hist based testing. The result of testing is given through a message box with name of the leaf followed by diagnosis. Here is the sample result of spinach normal, diseased and disordered with the message box.

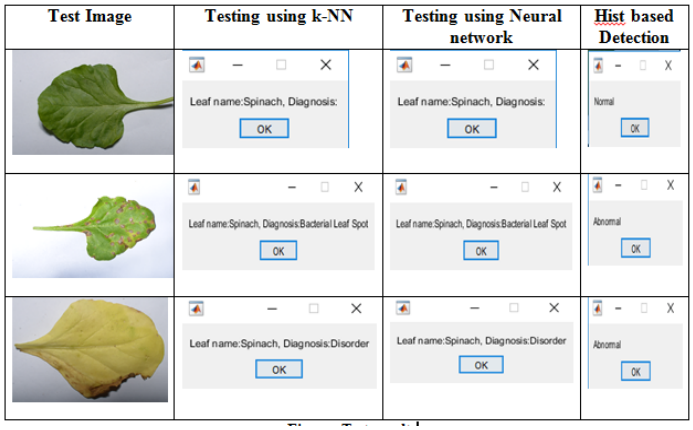


Figure-12 : Test Results of Spinach leaf

F. Accuracy Calculation

The Leaf types are six in number which are considered as class1 to class 6. Hence the accuracy for leaf type is calculated as "leaf_accuracy" and further for disease, disorder and normal type the accuracy is calculated as "accuracy" for both the classifiers using the formula as shown below,

$$\text{Percentage accuracy (\%)} = \frac{\text{Correctly Recognized Image Samples}}{\text{Total Number of Test Image Samples}} * 100.$$

(a) The accuracy calculation result of k-NN is as shown in Figure-13. The accuracy of all six classes of leaf identification is 86.39% and that of disease, disorder and normal is 75.04%.

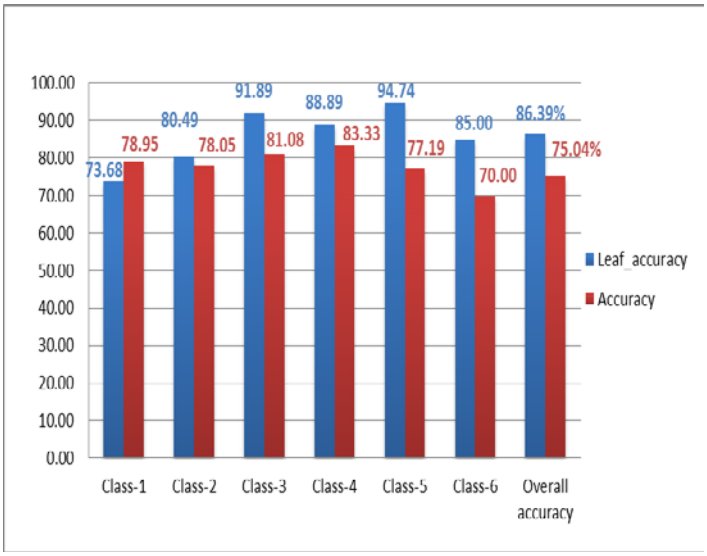


Figure-13 : Accuracy of kNN

(b) The accuracy calculation result of PNN is as shown in Figure-14. The accuracy of all six classes of leaf identification is 75.70% and that of disease, disorder and normal is 71.24%.

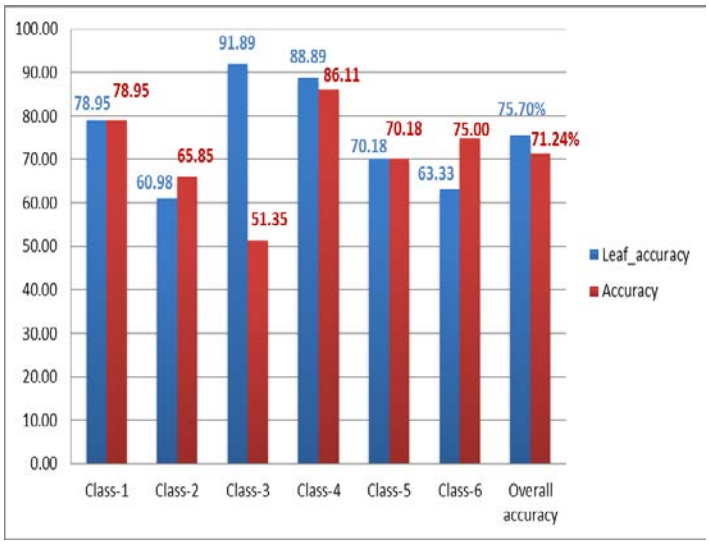


Figure-14 : Accuracy of kNN

(c) Overall Accuracy

The overall accuracy of k-NN and PNN is as shown in Figure-15. The performance of k-NN is better compared to PNN.

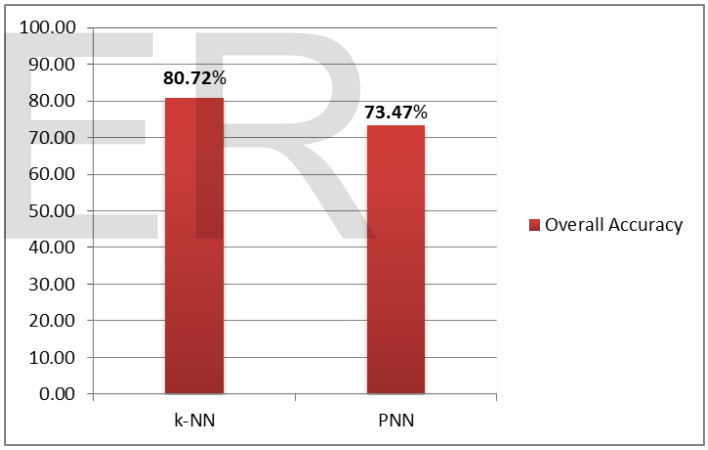


Figure-15 : Overall accuracy of classifiers

5. CONCLUSION

A two-level classification system to classify leaf images of various diseased and disordered is developed. The leaf images are pre-processed and segmented to enhance the images and extract the region of interest part that is leaf part from the background. Texture and color features are extracted using Fractal and color correlogram techniques. These features are trained. The images are tested using k-nearest neighbor and probabilistic neural network. From the experiment it is observed that k-NN classifier performs better than PNN with an overall accuracy of 80.72%.

6. FUTURE SCOPE

Dataset collection may involve diseases of different season and on different land. The proposed system can be further improved by incorporating other high level features such as shape features (Zernike moments, higher order statistics, counterlet transforms etc).

References

[1] Min Zhan, Qinggang Meng , "Citrus canker detection based on leaf images Analysis", IEEE, 978-1-4244- 7618, 2010.

[2] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of Computer Application(0975 – 8887) Volume 17, March 2011.

[3] Song Kai, liu zhikun, Su hang, Guo chunhong ,"A Research of maize disease image recognition of Corn Based on BP Networks", Third International Conference on Measuring Technology and Mechatronics Automation, 978-0-7695-4296, 2011 lEEE.

[4] Haiguang Wang, Guanlin Li, Zhanhong Ma, Xiaolong Li, "Image Recognition of Plant Diseases Based on Principal Component Analysis and Neural Networks", 8th International Conference on Natural Computation (ICNC), 978-1-4577-2133, 2012 IEEE.

[5] Anand.H.Kulkarni, Ashwin Patil R. K, " Applying image processing technique to detect plant diseases", International Journal of Modern Engineering Research (IJMER Vol.2, Issue.5, 2012 pp-3661-3664.

[6] P. Revathi , M. Hemalatha, "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques", International Conference on Emerging Trends in Science, Engineering and Technology ISBN :, 978-1-4673-5144-7, 2012 IEEE.

[7] Tuan Norjihan Tuan Yaakub, Puteri Nor Ashikin Megat Yunus, "Classification of Watermelon Leaf Diseases Using Neural Network Analysis", 978-1-4673-5968-9, IEEE 2013.

[8] Jagadeesh D. Pujari, Rajesh Yakkundimath, "Classification of Fungal Disease Symptoms affected on Cereals using Color Texture Features", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.6, 2013.

[9] Monika Jhuria, Ashwani Kumar, Rushikesh Borse, "Image Processing for Smart Farming: Detection of disease", 978-1-4673-6101-9, IEEE 2013.

[10] Ms. Kiran R. Gavhale,Prof. Ujwalla Gawande, Mr. Kamal O. Hajari, "Unhealthy Region of Citrus Leaf Detection Using Image Processing Techniques", International Conference for Convergence of Technology, 978-1-4799-3759, IEEE 2014.

[11] Jagadesh D. pujari, Rajesh Yakkundimath, Abdulmunaf S. Byadgi, "Image Processing Based Detection of Fungal Diseases in Plants", International conference on information and communication technologies(ICICT), Elsevier 2014.

[12] Sachin D. Khirade, A. B. Patil, "Plant Disease Detection Using Image Processing", International Conference on Computing Communication Control and Automation, 978-1-4799-6892 , IEEE 2015.

[13] Bhushan R. Adsule, Jaya M. Bhattad, "Leaves Classification Using SVM and Neural Network for Disease Detection", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 6, IJIRCCE 2015.

[14] S. W. Zhang, Y. J. Shang and L. Wang, "PLANT DISEASE RECOGNITION BASED ON PLANT LEAF IMAGE", The Journal of Animal & Plant Sciences, 2015

[15]Ramakrishnan. M and Sahaya Anselin Nisha. A, "Groundnut Leaf Disease Detection and Classification by using Back Propagation Algorithm", 978-1-4 799-8081, IEEE 2015.

[16] Aakanksha Rastogi, Ritika Arora, Shanu Sharma, "Leaf Disease Detection and Grading using Computer Vision Technology & Fuzzy Logic", 2nd International Conference on Signal Processing and Integrated Networks (SPIN), 978-1-4799-5991, IEEE 2015.

[17] P. R. Rothe, R. V. Kshirsagar, " Cotton Leaf Disease Identification using Pattern Recognition Techniques", International Conference on Pervasive Computing (ICPC), -1-4799-6272, IEEE 2015.

[18] Mrunmayee Dhakate, Ingole A. B, "Diagnosis of Pomegranate Plant Diseases using Neural Network", IEEE 2015.

[19] Anand R,Veni S,Aravinth J, "An Application of image processing techniques for Detection of Diseases on Brinjal Leaves Using K-Means Clustering Method", Fifth International Conference On Recent Trends In Information Technology, IEEE 2016.

[20] Pranjali B. Padol, Anjali A. Yadav, "SVM Classifier Based Grape Leaf Disease Detection", Conference on Advances in Signal Processing (CASP), IEEE 2016.