International Journal of Advance Engineering and Research Development (IJAERD)
Volume 1,Issue 6,June 2014, e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

LEAF DISEASE DETECTION USING IMAGE PROCESSING AND NEURAL NETWORK

Arti N. Rathod[1],Bhavesh A. Tanawala[2], Vatsal H. Shah[3]

[1] Computer Engineering Department, Birla Vishvakarma Mahavidyalaya Vallabh Vidhyanagar, er.artirathod@gmail.com
[2] Computer Engineering Department, Birla Vishvakarma Mahavidyalaya Vallabh Vidhyanagar,bhavesh.tanawala@bvmengineering.ac.in
[3] Information Technology Department, Birla Vishvakarma Mahavidyalaya Vallabh Vidhyanagar, vatsal.shah@bvmengineering.ac.in

Abstract— In agriculture research of automatic leaf disease detection is essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect symptoms of disease as soon as they appear on plant leaves. There are the main steps for disease detection of Image Acquisition, Image Preprocessing, Image Segmentation, Feature Extraction and Statistical Analysis. This proposed work is in first image filtering using median filter and convert the RGB image to CIELAB color component, in second step image segmented using the k-medoid technique, in next step masking green-pixels & Remove of masked green pixels, after in next step calculate the Texture features Statistics, in last this features passed in neural network. The Neural Network classification performs well and could successfully detect and classify the tested disease.

Keyword— Leaf disease, Image processing, CIELAB color model, SGDM Matrix, Color Co-occurrence Method, k-medoids, Neural Network.

I. INTRODUCTION

India is a cultivated country. Sharecroppers (Farmers) have huge range of diversity to select suitable crops. Research work develops the advance computing environment to identify the diseases using infected images of various leaves. Images of leaves are taken from digital camera, smart phones and processed using image growing, then the part of the leaf sport has been used for the classifying purpose of the train and test of disease. The technique evolved into the system is both Image processing techniques and advance computing techniques to identify disease from leaves.

A. Image Analysis Can Be useful For The Following Purposes:

1. To identify diseased leaf, stem, fruit.
2. To measure affected area by disease.
3. To find the boundaries of the affected area.
4. To find out the color of the affected area.
5. To determine size & shape of leaf.
6. To identify the Object correctly.
   Etc.

Disease management is a challenging task. Generally diseases are seen on the leaves or stems of the plant. Precise quantification of these visually experimental diseases, pests, traits has not studied yet because of the complexity of visual patterns. Hence there has been growing demand for more specific and sophisticated image pattern understanding.

B. Different Types Of Leaf Spot Diseases:

– Bacterial
– Fungal
– Viral

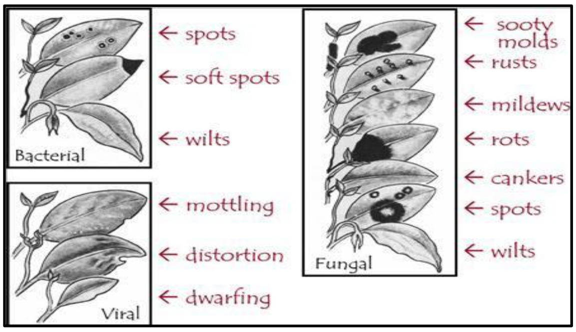

Fig. 1. Various types of diseases [11]

Most leaf diseases are caused by fungi, bacteria and viruses. Fungi are identified primarily from their morphology, with emphasis placed on their reproductive structures. Bacteria are measured more primitive than fungi and generally have simpler life cycles. With few exceptions, bacteria exist as single cells and increase in numbers by dividing into two cells during a process called binary fission viruses are extremely tiny particles consisting of protein and genetic material with no associated protein [9].

In biological science, sometimes thousands of images are generated in a single experiment. There images can be required for further studies like classifying lesion, scoring quantitative traits, calculating area eaten by insects, etc. Almost all of these tasks are processed manually or with distinct software packages. It is not only incredible amount of work but also suffers from two major issues: excessive processing time and subjectiveness rising from different individuals. Hence to conduct high throughput experiments, plant biologist need efficient computer software to automatically extract and analyze significant content. Here image processing plays important role [1]. This paper provides a survey to study in different image processing techniques used for studding leaf diseases.

II. LITERATURE REVIEW

Some papers are describing to detecting leaf disease using various methods suggesting the various implementation ways as illustrated and discussed here.

[2] In this paper consists of two phases to identify the affected part of the disease. Initially Edge detection based Image segmentation is done, and finally image analysis and classification of diseases. This work the input images using the RGB pixel counting values features used and identify disease wise and next using homogenization techniques Sobel and Canny using edge detection to identify the affected parts of the leaf spot to recognize the diseases boundary is white lighting and then result is recognition of the diseases as output.[3] in this paper detection of leaf diseases has been used method is threefold: 1) identifying the infected object based upon k-means clustering; 2) extracting the features set of the infected objects using color co-occurrence methodology for texture analysis; 3) detecting and classifying the type of disease using NNs, moreover, the presented scheme classifies the plant leaves into infected and not-infected classes. [4] In this paper a comparison of the effect of CIELAB, HSI and YCbCr color space in the process of disease spot detection is done. All these color models are compared and finally ‘A’ component of CIELAB color model is used. [5] In this paper Support vector machines are a set of related supervised learning method used for classification and regression. The detection accuracy is improved by SVM classifier. [6] The process of image segmentation was analyzed and leaf region was segmented by using Otsu method. In the HSI color system, H component was chosen to segment disease spot to reduce the disturbance of illumination changes and the vein. Then disease spot regions were segmented by using Sobel operator to examine disease spot edges. Finally plant diseases are graded by calculating the quotient of disease spot and leaf areas. [7] This paper wills two techniques for feature extraction and

comparison of two techniques. Otsu Threshold: thresholding creates binary image from gray level ones by turning all pixels below some threshold to zero and all pixels about that threshold to one. K-Means clustering is an unsupervised learning task where one seeks to identify a finite set of categories termed clusters to describe the data. [8] This paper describes the segmentation consist in image conversion to HSV color space and fuzzy c-means clustering in hue-saturation space to distinguish several pixel classes. These classes are then merged at the interactive stage into two final classes, where one of them determines the searched diseased areas. [13] This paper presents a color image segmentation method which divides colour space into clusters. Here also compare the efficiency of available algorithm for segmentation of gray as well as noisy images. Unlike k-means algorithm, k-medoids is not sensitive to dirty data and abnormal data. The k-medoids algorithm is suitable for noisy images. It is seen that, the segmented images are highly dependent on the number of segments or centroids.

III. OBJECTIVES OF RESEARCH

The major objectives of this research are follows:

1. To collect image data sets of various common leaves diseases.
2. To identifying infected area based upon clustering algorithm.
3. To evaluate the color co-occurrence method, for disease detection in leaves.
4. To develop NNs strategies classification of leaves based on the features obtained from the color co-occurrence method.

The objective behind this research work is to progress the performance of disease detection technique. The image data of the leaves selected for this study would be collected. Algorithms based on image processing techniques for image preprocessing, image segmentation, feature extraction, statistical analysis and classification would be designed. Comparison is also presented. Thus, the proposed algorithm was tested on five diseases are: Early scorch, cottony mold, and late scorch, brown spot and bacterial-fungal.

IV. THEORETICAL BACKGROUND

A. The Basic Approach Procedure

The overall concept of this work for any vision related algorithm of image classification is almost the same. First, the digital images are acquired from the plants using a digital camera. Then image processing techniques are applied to the acquired images to take out useful features that are essential for further analysis. After that, several analytical selective techniques are used to classify the images according to the specific problem. Figure 3 show the basic procedure of the proposed detection algorithm in this research.

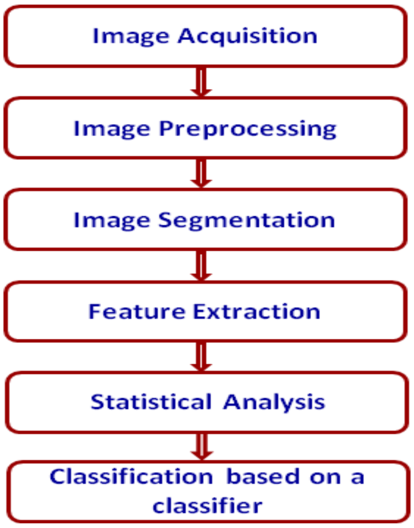

Fig. 2. The basic procedure based disease detection solution

B. Image Acquisition and Preprocessing:

First, the images of various leaves are going to acquire by a digital camera. The digital images are acquired from the location referring different sites. Other way is collect the images from agriculture research units. The images which have input are always not satisfactory regardless of what image acquisition devices are adopted. If there are noises in the image, the region of curiosity in the image is not clear or other objects' interference exists in the image and so on. Different pre-processing methods should be selected for different image applications. There are three steps included in preprocessing phase: clipping, smoothing and enhancement [7]. Noises which may be brought from the process of image collection and lots of information which may be easily leaded from the operating and saving to the image would make the quality of image dropped, thereby affects following of disease. So, the image with low quality must be smoothed by filter [7]. Different kinds of noises exist in an image and variety of noise reduction techniques is available to perform de-noising. Selection of the de-noising algorithm depends on the application. Median filter performs better with salt and pepper noise by choosing appropriate threshold. An image has salt and pepper noise will have dark pixels (black dots or pepper) in bright region and bright pixels (white dots or salt) in dark region. An effective method to remove this type of noise involves the use of median filter [5]. Median filter is non-linear filter. The median filter is to find the median value by across the window, replacing each entry in the window with the median value of the pixel [14]. It's best in removing salt and pepper noise and impulse noise. Median filter erases black dots called the pepper and fills in the white holes in the image, called salt. It simply replaces each pixel values by the median of the intensity level in the neighborhood of the pixel [6].

C. Image Segmentation (region of interest):

The image will be segmented into different parts according to the region of interest. Image segmentation is to divide the image into same meaningful regions. Simply to say, image segmentation means to separate the object from background for following processing in an image [7]. In this step the images are segmented using k-medoids clustering methods. K-medoids clustering is partitioning based clustering method. K-medoids or PAM (Partition around medoids): Every cluster is represented by one of the objects in the cluster. K-medoids more robust than k- means in the presence of noise and outlines; because a medoids is less influenced by outlines or other extreme values than a mean.

K-medoids Algorithm [12]**:**

Input: 'k', the number of cluster to be partitioned; 'n', the number of objects

Output: A set of 'k' cluster that minimize the sum of the dissimilarities of all the objects to their nearest medoids.

Steps:

I. Arbitrarily choose 'k' objects as the initial medoids;

II. Repeat,
   a. Assign each remaining object to the cluster with the nearest medoids;
   b. Randomly select a non-medoids object;
   c. Compute the total cost of swapping old medoids object with newly selected non-medoids object.
   d. If the total cost of swapping is than zero, than perform that swap operation to form the new set of k-medoids.

III. Until no change

Calculate distance so as to associate each data object to its nearest medoid. Cost is calculated using Manhattan distance. Costs to the nearest medoid are shown bold in the table [17]. The Manhattan distance D between two vectors X and Y is D- sum (abs(x-y)). Where cost between any two points is round using formula $cost(x,c) = \sum_{i=1}^{d} |x_i - c_i|$ where x is any data object, c is the medoid, and d is the dimension [12]. K-means and k-medoids – both the methods find out clusters from the image. K-means drawback is sensitivity to noisy data outlines. Compared to this, k-medoids is not sensitive to noisy data, outlines and effective for gray scale too.

An example of the output of K-medoids clustering for a leaf infected with *early scorch* disease is shown in Figure 3. It is observed from Figure 3 that cluster 3 contains infected object of early scorch disease. Furthermore, clusters 2 and 4 contain the intact parts of leaf, although they are distinct from each other. However cluster 1 represents the black background of the leaf which can be discarded primarily. Finally, the image in facilitates the segmentation procedure followed in K-medoids algorithm.

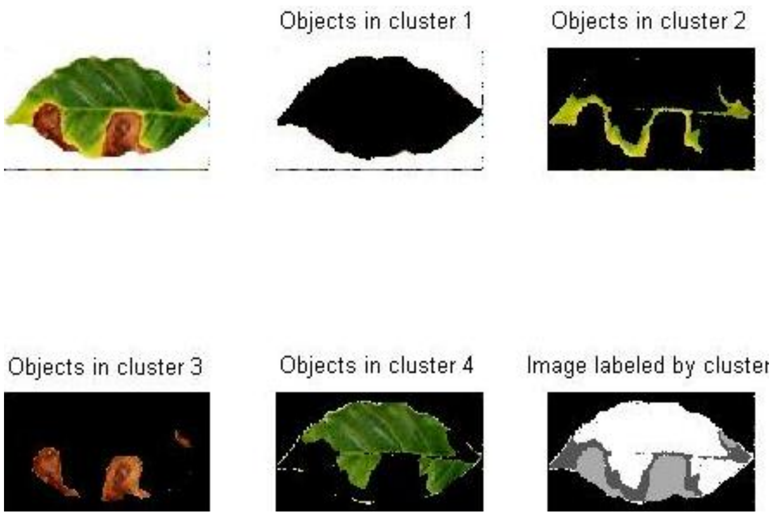

Fig. 3. Output of K-Medoids clustering for a leaf that is infected with early scorch disease.

D. Extraction Of Features And Statistics Analysis:

The input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a compact representation set of features. The input data Transform into the set of features is called features extraction. If the features extraction is carefully chosen to so expect that the features set will extract the relevant information from the input data in order to perform the desired task [7]. The color co-occurrence texture analysis method is developed by the Spatial Gray-level Dependence Matrices (SGDM). The gray Level Co-occurrence Methodology (GLCM) is a statistical way to describe shape by statistically sampling the way certain gray level occurs in relative to other gray levels.These matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance orientation from any pixel given that pixel has a second particular gray level. For a position operator p, we can define a matrix Pij that counts the number of times a pixel with gray level I occurs at position p from a pixel with gray-level j. The SGDMs are represented by the function P (i, j, d, Θ) where I represent the gray level of the location (x,y) at an orientation angle of Θ. The reference Pixel at image position (x, y) is shown as am matrix. All the neighbor from 1 to 8 are numbered in a clockwise direction Neighbors 1 and 5 are located on the same plant at a distance of l and an orientation of 0 degree. An example image matrix and its SGDM are already given in the three equations above. In this research, a one pixel offset distance and a zero degree orientation angle was used [3].After the transformation, we calculated the feature set for H and S, we dropped (I) since it does not give extra information. However, we use GLCM function in Matlab to create gray-level co-occurrence matrix; the number of gray levels is set to 8, and symmetric value is set to “true”, and finally, offset is given a “0” value [3]. Properties of Spatial Gray-level Dependence Matrices (SGDM) like Contrast, Energy, Local homogeneity, and correlation are compound for the Hue content of the images as given following Eqns.

Table.1 Properties of SGDM [18].

Property	**Description**	**Formula**
Contrast	Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image. Range = [0 (size(GLCM,1)-1)^2] Contrast is 0 for a constant image.	$\sum_{i,j} \lvert i-j \rvert^2 P(i,j)$
Correlation	Returns a measure of how correlated a pixel is to its neighbor over the whole image. Range = [-1 1] Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is NaN for a constant.	$\sum_{i,j} \frac{(i-\mu i)(j-\mu j)P(i,j}{\sigma_i \sigma_j}$
Energy	Returns the sum of squared elements in the GLCM. Range = [0 1] Energy is 1 for a constant image.	$\sum_{i,j} P(i,j)^2$
Homogeneity	Returns a value that measures the closeness of the distribution a\of element in the GLCM to the GLCM diagonal. Range = [0 1] Homogeneity is 1 for a diagonal GLCM.	$\sum_{i,j} \frac{P(i,j)}{1+\lvert i-j \rvert}$

E. Classification based on classifier:

In this paper, neural networks are used in the automatic detection of leaves disease. Neural network is chosen as a classification tool due to its well known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using NNs. The dataset for training and validation processes consists of two parts; the training features set which are used to train the NN model; whilst a testing features sets are used to verify the accuracy of the trained using the feed-forward back propagation network. In the training part, connection weights were always updated until they reached the defined iteration number or suitable error. Hence, the capability of ANN model to respond accurately was assured using the Mean Square Error (MSE) criterion to emphasis the model validity between the target and the network output.

V. BASIC STEPS OF THE PROPOSED ALGORITHM

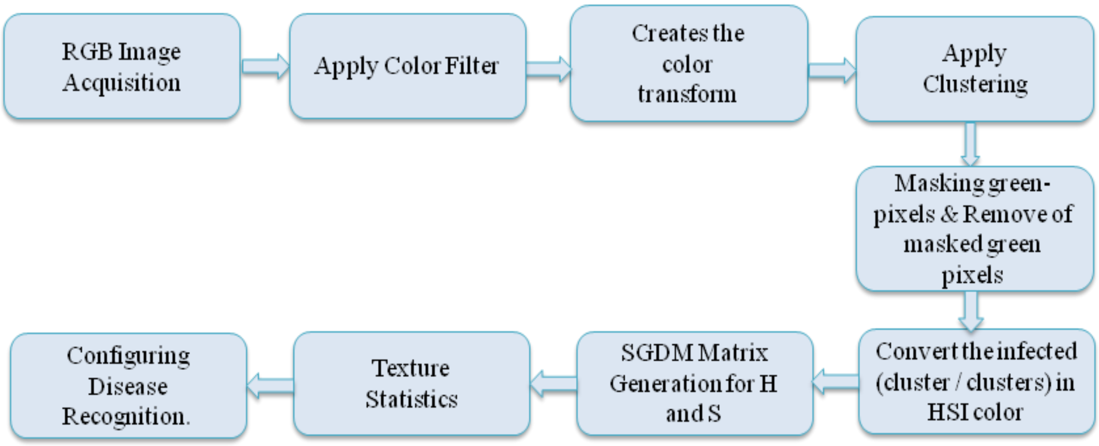

Fig. 4. Disease detection System Architecture.

Algorithm:

1. RGB image acquisition.
2. Apply Color Median Filtering by Median filter.
3. Create the color transformation structure RGB to CIELAB Color structure.
4. Apply K-medoids clustering.
5. Masking green-pixels & Remove of masked green pixels.
6. Convert the infected (cluster / clusters) in HSI Color.
7. SGDM Matrix Generation for H and S.
8. The GLCM function to calculate the Texture features Statistics.
9. Configuring Neural Networks for Recognition.

VI. EXPERIMENTAL RESULTS AND OBSERVATIONS

A. Input Data Preparation and Experimental Settings:

In this experiment, two main dataset were generated, namely: (i) Training texture feature data, and (ii) Testing texture feature data. Each row had a unique number (1 to 6) which represented the class (i.e., the disease) of the particular row of data. 1st, represented early scorch disease infected leaf. 2nd, represented Cottony mold disease infected leaf. 3rd, represented Late scorch disease infected leaf. 4th, represented Brown spot disease infected leaf. 5th, represented Bacterial-Fungal disease infected leaf, and 6th, represented normal leaf. Then, a software program was written in MATLAB that would take in **.mat** files representing the training and testing data, train the classifier using the "train dataset", and then use the "test dataset" to perform the classification task on the test data. Consequently, a Matlab routine would load all the data files (training and testing data files) and make modifications to the data according to the proposed model chosen [3].

The architecture of the network used in this study was as follows. A set of 10 hidden layers in the neural network was used with the number of inputs to the neural network (i.e. the number of neurons) is equal to the number of texture features listed above. The number of output is 6 which is the number of classes representing the 5 diseases studied along with the case of normal (uninfected) leaf. Those diseases are early scorch, cottony mold, late scorch, brown spot, Bacterial-Fungal. The neural network used is the feed forward back propagation with the performance function being the

Mean Square Error (MSE) and the number of iterations was 10000 and the maximum allowed error was 10-5.

B. Experimental Results:

Table 2 shows the results for NN classification strategy for testing. These results were obtained using a NN classifier for different five diseases. In Color feature HS get the highest overall classification accuracy, in which it achieved an overall accuracy of 96% compared to the 94% accuracy achieved in [3]. Also, Figure 5 shows a graph that representing the percentage classification of various disease of all the color features models shown in Table 2.

Model	Color Features	Early scorch	Cottony mold	Late scorch	Brown spot	Bacterial-Fungal	Normal	Overall average
M1	H	92	92	89	90	93	98	92.33
M2	S	93	90	85	89	85	98	90
M3	I	92	89	87	88	86	99	90.16
M4	HS	95	96	89	96	95	100	95.17
M5	HSI	87	85	80	79	84	99	85.67

Table.2 Percentage classification of various diseases

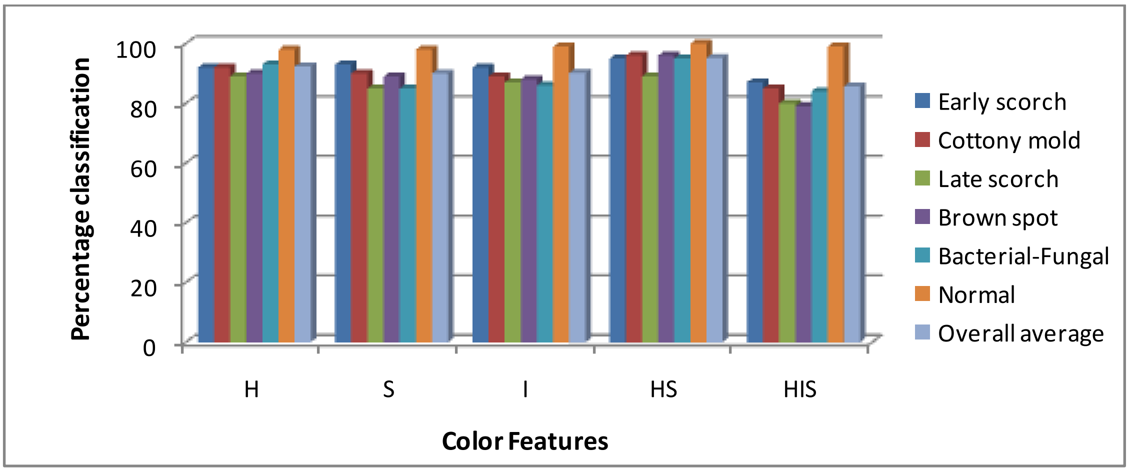

Fig. 5 Percentage classification of various diseases

It can be implied from Table 2 that Model M4 which has used only the H and S components in computing the texture features, has emerged as the best model among the various models. The numbers of leaf samples that were classified into each of the five tested categories using HS model with specific threshold value are shown in Table 3 and Figure 6. It is observed from Table-3 that only two samples from brown spot leaves were misclassified. Similarly, in the case of bacterial-fungal images, only three test images from the class were misclassified.

From species	Early scorch	Cottony mold	Late scorch	Brown spot	Bacterial-Fungal	Normal	**Accuracy**
Early scorch	25	0	0	0	0	1	100
Cottony mold	0	24	0	1	0	1	96
Late scorch	0	0	25	0	0	0	100
Brown spot	0	0	0	24	1	0	96
Bacterial-Fungal	0	1	1	0	23	0	92
Normal	0	0	0	1	2	23	92

Average	96

Table 3: Classification results per class for neural network.

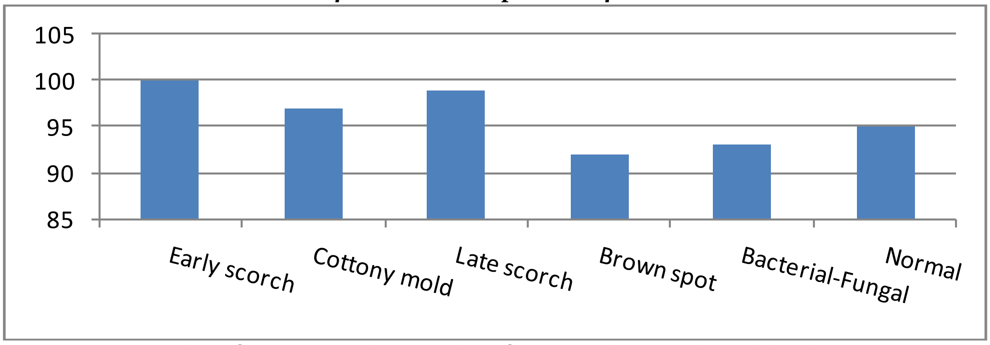

Fig. 6 Classification results per class for neural network with back propagation.

VII. CONCLUSION AND FUTURE WORK

There is main characteristics of disease detection is system can identify the affected part of a leaf spot by using the image processing technique. For filtering Median filter performs better with salt and pepper noise. In Color model CIELAB color model is accurately detected disease and results are not affected by background, type of leaf, type of disease spot and camera flash. The k-medoids algorithm is working for gray scale images and better performs for large databases not sensitive to noisy data and outliers, too. K-medoids performs reasonably better than the K-Means algorithm. The applications of Median filter, CIELAB color model, and clustering and texture analysis have been formulated for clustering and classification of diseases that affect on plant leaves. Neural networks in order to increase the recognition rate of final classification process. The experimental results indicate that the proposed approach is a valuable approach, which can significantly support an accurate detection of leaf diseases in a little computational effort.

Recognizing the disease is mainly the purpose of the proposed approach. For future research, Work can be extended for development of hybrid algorithms such as other clustering method and NNs in order to improve the recognition rate of the final classification process. Further to needed to compute amount of disease present on leaf.

REFERENCES

[1] Jayamala K. Patil, Raj Kumar, "Advances In Image Processing For Detection of Plant Diseases", JABAR, 2011, 2(2), 135-141.

[2] P.Revathi, M.Hemalatha, "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques", ISBN, 2012, 169-173, IEEE.

[3] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", IJCA, 2011, 17(1), 31-38, IEEE-2010.

[4] Piyush Chaudhary, Anand K. Chaudhari, Dr. A. N. Cheeran and Sharda Godara, "Color Transform Based Approach for Disease Spot Detection on Plant Leaf", IJCST, 2012, 3(6), 65-70.

[5] S. Arivazhagan, R. Newlin Shebiah, S. Ananthi, S. Vishnu Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", CIGR, 2013, 15(1), 211-217.

[6] Chanchal Srivastava, Saurabh Kumar Mishra, Pallavi Asthana, G. R. Mishra, O.P. Singh, "Performance Comparison Of Various Filters And Wavelet Transform For Image De-Noising", IOSR-JCE, 2013, 10(1), 55-63.

[7] Mrunalini R. Badnakhe, Prashant R. Deshmukh, "Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering", IJARCSSE, 2012, 2(3), 449-452.

[8] Mr. Salem Saleh Al-amri, Dr. N.V. Kalyankar and Dr. Khamitkar S.D, "A Comparative Study of Removal Noise from Remote Sensing Image", IJCSI, 2010, 7(1).

[9] Prof. Sanjay B. Dhaygude, Mr.Nitin P.Kumbhar, "Agricultural plant Leaf Disease Detection Using Image Processing" IJAREEIE, 2013, 2(1), 599-602.

[10] Dheed Al Hiary, S. Bani-Ahmad, M. Braik, "A Framework for Detection and Classification of Plant Leaf and Stem Diseases", IEEE, 2010, 113-118.

[11] S. Ananthi, S. Vishnu Varthini, "Detection and classification of plant leaf diseases", IJREAS, 2012, 2(2), 763-773.

[12] Shalini S Singh, N C Chauhan, "K-means v/s K-medoids: A Comparative Study", National Conference on Recent Trends in Engineering & Technology, 13-14 May 2011.

[13] Amit Yerpude, Dr. Sipi Dubey, "Colour image segmentation using K – Medoids", IJCTA, 2012, 3(1), 152-154.

[14] Govindaraj.V, Sengottaiyan.G, "Survey of Image Denoising using Different Filters", IJSETR, 2013, 2(2), 344-351.

[15] Bindu Tiger, Toran Verma, "Identification and Classification of Normal and Infected Apples using Neural Network", IJSR, 2013, 2(6), 160-163.

[16] Alham F. Aji, Qorib Munajat, Ardhi P. Pratama, Hafizh Kalamullah, Aprinaldi, Jodi Setiyawan, and Aniati M. Arymurthy "Detection of Palm Oil Leaf Disease with Image Processing and Neural Network Classification on Mobile Device", 2013, 5(3), 528-532.

[17] K-medoids - Wikipedia, the free encyclopedia, http://en.wikipedia.org/wiki/K-medians_clustering

[18] httt://www.mathworks.com/

[19] "Digital Image Processing" By Rafael C. Gonzalez Richard E. Woods 2nd Edition.

[20] "Introduction to MATLAB & SIMULINK A PROJECT APPROACH" By O. Beucher and M. Weeks, 3rd Edition

