ADVANCED PLANT LEAF DISEASE DETECTION SYSTEM

N V ASHOK[*1], SK FARZANA BEGUM[*2], N. NAGA HARI SAIRAM[*3]
*[*1, 2, 3]Assistant Professor, Dept. of Agricultural Engineering.*
A.M Reddy Memorial College of Engineering and Technology, Andhra Pradesh.

Abstract: **Agriculture is the mainstay of the Indian economy. Almost 70% people depend on it & shares major part of the GDP. Diseases in crops mostly on the leaves affects on the reduction of both quality and quantity of agricultural products. Perception of human eye is not so much stronger so as to observe minute variation in the infected part of leaf. In this paper, we are providing software solution to automatically detect and classify plant leaf diseases. In this we are using image processing techniques to classify diseases & quickly diagnosis can be carried out as per disease. This approach will enhance productivity of crops. It includes several steps viz. image acquisition, image pre-processing, segmentation, features extraction and neural network based classification.**

I. INTRODUCTION

Agriculture has played a key role in the development of human civilization. If there is decrease in agro products, total economy will get affected. Therefore judicious management of all input resources such as soil, seed, water, fertilizers etc. is essential for sustainability. As diseases are inevitable, detecting them plays major role. One can refer incident that occurred in 2007, Georgia (USA), it is estimated that approximately 539 USD was the loss incurred due to plant diseases as well as controlling them. The naked eye observation of farmers followed by chemical test is the main way of detection and classification of agricultural plant diseases. In developing countries, farming land can be much larger and farmers cannot observe each and every plant, every day. Farmers are unaware of non-native diseases. Consultation of experts for this might be time consuming & costly. Also unnecessary use of pesticides might be dangerous for natural resources such as water, soil, air, food chain etc. as well as it is expected that there need to be less contamination of food products with pesticides.

There are two main characteristics of plant disease detection machine-learning methods that must be achieved, they are: speed and accuracy [1]. There is need for developing technique such as automatic plant disease detection and classification using leaf image processing techniques. This will prove useful technique for farmers and will alert them at the right time before spreading of the disease over large area. Solution is composed of four main phases; in the first phase we create a color transformation structure for the RGB leaf image and then, we apply color space transformation for the color transformation structure. Then image is segmented using the K-means clustering technique. In the second phase, unnecessary part (green area) within leaf area is removed. In third phase we calculate the texture features for the segmented infected object. Finally, in the fourth phase the extracted features are passed through a pre-trained neural network [1].
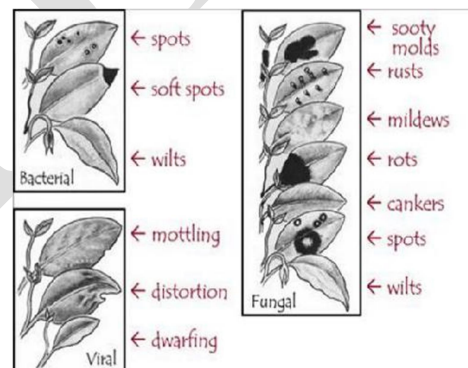

Figure 1. Types of Diseases [3]

N V Ashok, Sk Farzana Begum, N. Naga Hari Sairam *et. al.,* /International Journal of Engineering & Science Research

Basically crop leaf diseases are broadly classified into bacterial, viral, fungal. These are further categorized that are shown above in Figure 1. Some of them are early scorch, cottony mold, tiny whiteness, late scorch etc.

Bacterial type diseases are characterized by tiny pale green spots which soon come into view as water-soaked. Among all types of diseases viral type diseases are difficult to diagnose and control once they starts spreading. Leaves might have wrinkled, curled pattern and growth may be stunted. Fungi are identified primarily from their morphology, with emphasis placed on their reproductive structures.

II. PROPOSED APPROACH

First the images of various leaves are acquired using high resolution camera so as to get the better results & efficiency. Then image processing techniques are applied to these images to extract useful features which will be required for further analysis.

The basic steps of the system are summarized as:

---

1. RGB image acquisition
2. Create color transformation structure & convert color values from RGB to the space specified in that structure.
3. Apply K means clustering for image segmentation
4. Masking of green pixels
5. Remove the masked cells inside the boundaries of the infected cluster.
6. Convert the infected cluster from RGB to HSI
7. SGDM matrix generation for H and S
8. Calling GLCM function to calculate the features
9. Computation of texture statistics
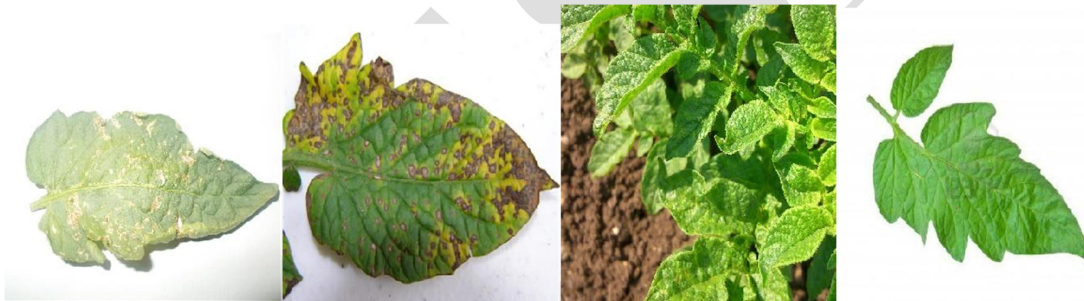10. Configure neural network for recognition

---

Figure 2. Sample images from our dataset.

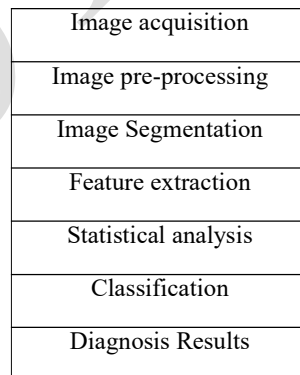| Image acquisition |
| Image pre-processing |
| Image Segmentation |
| Feature extraction |
| Statistical analysis |
| Classification |
| Diagnosis Results |

Figure 3. Block Diagram of proposed approach

Figure 3 shows the basic block diagram of the proposed system. Step by step explanation of the system is as follows:

A. Image Pre-processing

Noise gets added during acquisition of leaf images. So we use different types of filtering techniques to remove noise. We create device independent color space transformation structure. Thus we create the color transformation structure that defines the color space conversion. The next step is that we apply device-independent color space transformation, which converts the color values in the image to color space specified in the color transformation structure. The color transformation structure specifies various parameters of transformation. A device independent color space is the one where the resultant color

depends on the equipment used to produce it. For example the color produced using pixel with a given RGB values will be altered as brightness and contrast on display device used. Thus the RGB system is a color space that

is dependent. To improve the precision of the disease detection and classification process, a device independent color space is required. In device independent color space, the coordinates used to specify the color will produce the same color regardless of the device used to take the pictures. CIE L*a*b is a device independent color space in which a & b components carry color information [1].

B. Image segmentation (k-means clustering)

Image segmentation is the process used to simplify the representation of an image into something that is more meaningful and easier to analyse. K-means clustering is a partitioning method. The function 'kmeans' partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. Unlike hierarchical clustering, k-means clustering operates on actual observations (rather than the larger set of dissimilarity measures), and creates a single level of clusters. The distinctions mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data. K-means treats each observation in your data as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible [2].

C. Masking green pixels

In this step, we identify the mostly green colored pixels. After that, based on specified threshold value that is computed for these pixels, the mostly green pixels are masked as follows: if the green component of the pixel intensity is less than the pre computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that the green colored pixels mostly represent the healthy areas of the leaf and they do not add any valuable weight to disease identification and furthermore this significantly reduces the processing time.

D. Removing the masked cells

The pixels with zeros red, green, blue components as well as pixels on the boundaries of infected cluster are completely removed. This is helpful as it gives more accurate disease classification and significantly reduces the processing time. Infected cluster is converted from RGB to HSI color format [2].

E. GLCM methodology

Gray level Co-occurrence matrix (GLCM) is generated for each pixel map for H & S images of infected cluster.

1. The graycomatrix function creates a gray level co-occurrence matrix by calculating how frequently a pixel with the particular intensity value i occurs in a specified spatial relationship to a pixel with the value j.
2. By default this spatial relationship is the pixel of interest and its immediate right pixel.
3. However we can specify some other spatial relationship between twos. To create multiple GLCMs, specify an array of offsets to the graycomatrix function. These offsets define pixel relationships of varying direction and distance. Directions can be horizontal, vertical, along two diagonals.
4. Calculating statistics from GLCM matrix also known as SGDM [2].

F. Texture analysis (features computation)

1. Contrast: Returns a measure of the intensity contrast between a pixel and its neighbour over the whole image. Range = [0 (size (SGDM, 1)-1) ^2] Contrast is 0 for a constant image. C(i,j) is pixel value of location (i,j).

2. Energy= Returns the sum of squared elements in the SGDM. Range = [0 1]. Energy is 1 for a constant image.

3. $= \sum_{i=0}^{\, } \sigma^{-1}(\,.\,)^2$

Homogeneity= Returns a value that measures the closeness of the distribution of elements in the SGDM to the SGDM diagonal. Range = [0 1] Homogeneity is 1 for a diagonal SGDM.

G. Neural network based classification

The extracted features are given as inputs to pre trained neural network for automatic classification of diseases. BPNN, SVM, Radial basis functions, K-nearest neighbours are some well-known neural networks. 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

1. The training feature set which are used to train the NN model.
2. The testing features sets are used to verify the accuracy of the trained NN model [4].

Before the data can be fed to the ANN model, the proper network design must be set up, including type of the network and method of training. This was followed by the optimal parameter selection phase. However, this phase was carried out

N V Ashok, Sk Farzana Begum, N. Naga Hari Sairam *et. al.,* ∕International Journal of Engineering & Science Research

simultaneously with the network training phase, in which the network was trained using the feed-forward back propagation network (BPNN). In the training phase, connection weights were always updated until they reached the defined iteration number or acceptable 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 [8].

III. EXPERIMENTAL RESULTS AND OBSERVATIONS

As discussed in the introduction, solution is composed of four phases. In this paper, we have tabulated results up to three phases. Graphical user interface (GUI) is built for the system to carry out different image processing steps. Dataset folder is created consisting of various images of leafs under test. The crops we have selected for experimental purpose are tomato and potato from Pune district and cotton from Marathwada region of Maharashtra.

In phase 1, input image is selected from dataset folder as shown in Figure 4. Color transformation structure is created from RGB to CIELAB color space. Then color based segmentation using k-means clustering is applied to get the infected region of interest. The choice of selection of number of clusters is user dependent. Generally best results come up with either 3 or 4. The sample results of clustering on Septoria Leaf Spot of Tomato (fungal) are shown in next Matlab figures. We can see that Figure 5, 6 & 8 represent the intact (healthy) parts of leaf whereas Figure 7 consists of diseased portion of leaf. This is the region of interest for further processing.

In phase 2, infected cluster is selected. Green pixels are masked based on threshold value set. Also Pixels on the boundaries are also removed as they both do not contribute to disease identification process. Then resultant image is converted from RGB to HSI color format.
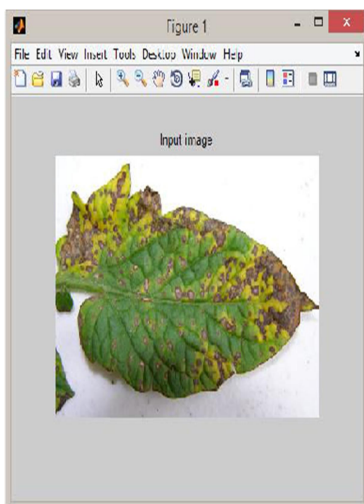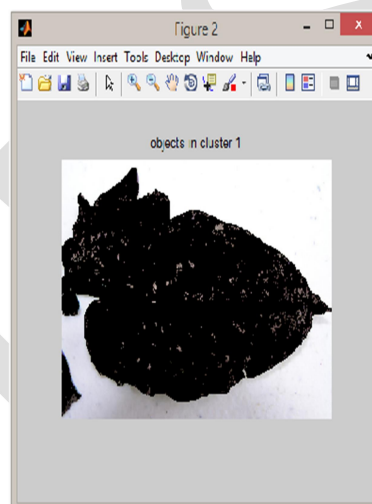

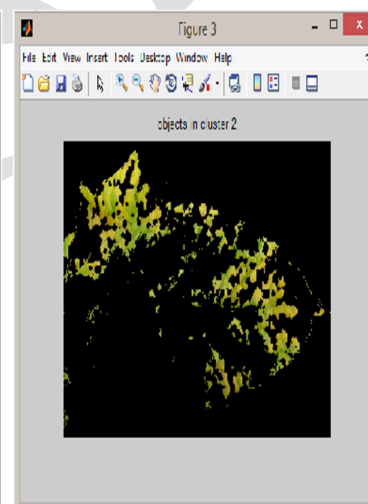Figure 4. Input image   Figure 5. Objects in cluster 1   Figure 6. Objects in cluster 2
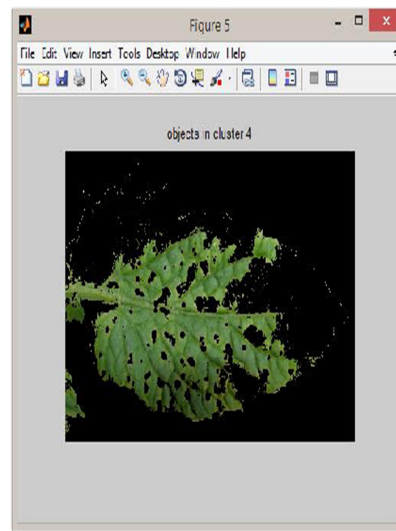
N V Ashok, Sk Farzana Begum, N. Naga Hari Sairam *et. al.,* / International Journal of Engineering & Science Research
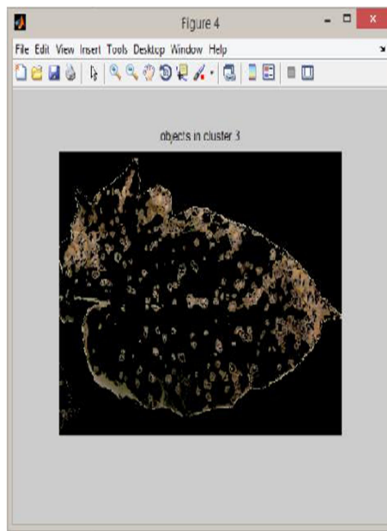
Figure 7. Objects in cluster 3 (infected)

Figure 8. Objects in cluster 4 (green area)

In phase 3, feature extraction & statistical analysis on modified infected cluster obtained in previous phase is carried out. In practice, analysis can be done on five different models also known as color features. Reason behind this selection is the variation in accuracy that we get during the neural network based classification. These models are listed in Table I:

Table I. Color Feature Models

Model	Color features
M1	H
M2	S
M3	I
M4	HS
M5	HSI

Here we have considered Model M1 & M2. The GLCM properties for Hue (H) & saturation (S) images are tabulated in following Table II.

Table II. Features set for Septoria spot on Tomato (fungal)

GLCM Properties	'H'	'S'
Contrast	0.1061	1.7162
Correlation	0.7283	0.7666
Energy	0.6949	0.6863
Homogeneity	0.9592	0.9276

Similarly we have tested for healthy tomato leaf from our dataset that is shown in Figure 2 (4th leaf) & get the features as given in Table III.

Table III. Features set for Healthy Tomato leaf

GLCM Properties	'H'	'S'
Contrast	0.0081	0.1883
Correlation	0.8141	0.8801
Energy	0.9702	0.9562
Homogeneity	0.9971	0.9946

Likewise we can get such feature values for other types of diseases. After analyzing these feature sets, decision can be made by comparing four properties of healthy leaf with that of infected leaf whether it is healthy or infected. Another technique can be implemented for detection is Histogram matching. Thus up to this stage, only detection of plant leaf disease is made (healthy or not). Next task is to classify them into specific type's e.g. early scorch, cottony mold, septoria spot, brown spot etc.

In phase 4, neural network based classification is to be carried out. Feed-forward Back propagation neural network is preferred as it has best performance being a function of Mean Square Error (MSE) with number of iterations be 10,000 and maximum allowed error is 10-5. First, the network is trained using training feature sets in order to get target output. After that it is tested for different types of diseases & accuracy is computed. It is experimentally found that using HS color feature (M4) we can get highest classification accuracy of 94-96% compared to other models.

To measure the severity of disease, grading can be assigned based on affected pixel area. Image is binary thresholded so as to measure number of black and white pixels. As per the percentage, disease grades 0, 1, 2 etc. can be assigned. This can be calculated by using next equation.

When disease is classified, information related to it such as causes, symptoms and cure techniques is to be displayed. For that purpose we have prepared Diagnosis database with the help of GUI. One such sample is shown in Figure 9.
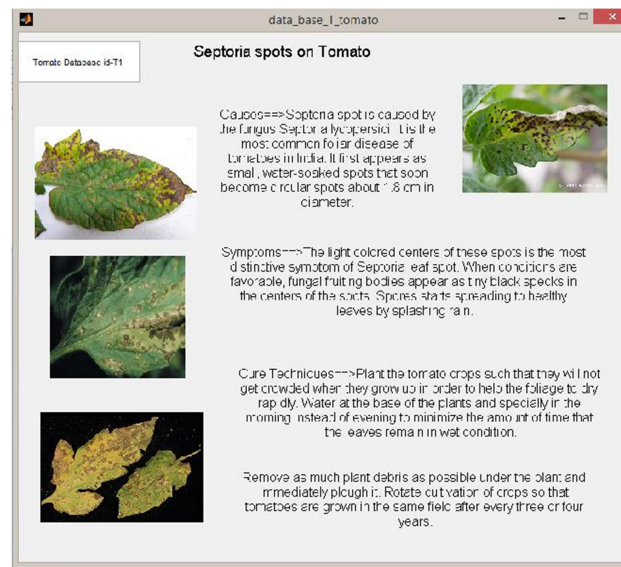
Figure 9. Diagnosis Database for Septoria spot Disease (Tomato)

CONCLUSION

The study reviews and summarizes image processing techniques for several plant species that have been used for recognizing plant diseases. The major techniques used are K-means clustering, GLCM and BPNN. Some of the challenges in these techniques are optimization of the technique for a specific plant, effect of the background noise in the acquired image and automation technique for a continuous automated monitoring of plant leaf diseases under real world field conditions. The proposed approach is a valuable approach, which can significantly support an accurate detection of leaf diseases in a little computational effort. Further future work can be extended by developing better segmentation technique; selecting better feature extraction and classification algorithms and NNs in order to increase the recognition rate of final classification process. Also by computing severity and amount of disease present on the crop, only necessary and sufficient amount of pesticides can be used making agriculture production system economically efficient. So there is a scope of improvement.

REFERENCES

[1] Smita Naikwadi, NiketAmoda, "Advances in image processing for detection of plant diseases", International journal of application or innovation in engineering and managrment(IJAIEM) Volume 2,Issue11,November 2013

[2] www.mathworks.in

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

[4] Arti N. Rathod, Bhavesh A. Tanawala, Vatsal H. Shah, International Journal of Advance Engineer ing and Research Development (IJAERD) Volume 1,Issue 6,June 2014, e-ISSN: 2348 - 4470

[5] Simona E. Grigorescu, Nicolai Petkov, and Peter Kruizinga, Comparison of Texture Features Based on Gabor Filters ,IEEE transaction on Image processing vol 11,No.10,pp1160-1167, oct 2002

[6] Book: Gonzalez and Woods,"Digital image processing using MATLAB",Pearson Education

[7] S. Arivazhagan, R. Newlin Shebiah* , S. Ananthi, S. Vishnu Varthin, Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features

[8] 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 Applications.

[9] Arti N. Rathod, Bhavesh A. Tanawala,Vatsal H. ShahLeaf disease detection using image processing and neural network, International Journal of Advance Engineer ing and Research Development (IJAERD),Volume 1,Issue 6,June 2014.

[10] Mr. Hrishikesh P. Kanjalkar, Prof. S.S.Lokhande, Detection and Classification of Plant Leaf Diseases using ANN, International Journal of Scientific & Engineering Research, Volume 4, Issue 8, August-2013

[11] Book: Principles of Soft Computing, By S.N Sivanandam and S.N.Deepa

[12] Book: Neural Networks, A classroom Approach by Satish Kumar, Tata McGraw Hill, publications.

[13] www.pdkv.ac.in

[14] www.ikisan.com