Plant Leaf Detection and Disease Identification System

Somesh Kumar, Vishakha Rastogi, Zainab Azeem, Rishabh Kumar Sharma, Siddhant Mishra
Department of Computer Science & Engineering
Moradabad Institute of Technology, Moradabad, U.P, India

Abstract **- In an Agricultural Country like India,the Agricultural productivity plays a vital role in boosting up the economy. Many a times, it happens that the farmers and the cultivators are not able to take proper care of the crop and it leads to some serious effects on the plants like plant diseases. It requires a proper plant pathologist's training in order to diagnose the disease. Since, it is an overhead for the farmers to take their plant to the laboratories. Thus, it will be beneficial if the technological measures are applied in this area. The most common diseases like Black Spot, Mildew and Canker are easy to be detected but the not so common ones, if not detected in due time can alter plant growth, reproduction and their survivorship. This paper is all about designing a project that will help in recognizing the disease a plant leaf is suffering from. The system uses the image processing techniques like resizing and enhancement, followed by image segmentation to keep only the meaningful parts and discard the other partitions. Then, the features from the segmented image are extracted which is then fed to the supervised learning model called, multi SVM for the classification purpose. The recommended system has been implemented utilizing MATLAB and has an easy to use User Interface, the GUI. After the experiments we found that, it is able to classify the diseased and healthy leaves with an accuracy of 98.39%.**

Index Terms - Image Pre-Processing, Segmentation, K-Means, Features Extraction, SVM

I. INTRODUCTION

Agricultural field plays a crucial role in everyone's life. Why does a man earn? The answers to this can be many, and one of the important answer is just to have enriched quality food. As we know that, diseases in plants are quite natural, but if proper care is not given then it can influence the quality of the crop. And, it can result in lesser productivity, also the people will not buy it. So, inorder to maintain a balance between quantity and quality, one has to keep an eye on the plants. So that, if the leaves get infected then it could be discovered that how much part is infected and what measures could be taken accordingly. Also, it is very important to identify the correct disease, before it's too late, else the wrong identification and providing wrong remedies like chemical and fertilizers can lead to financial loss, shortened harvest, and productivity of the food.

It requires proper plant pathologist's training or experts and laboratories to identify the plant disease by looking at the infected part, color, veins and other parts of the leaves . To design a programmed approach for the purpose of plant leaf and disease identification quite an analogous process is used. After considering all the aspects to assess important features of the plant, we have proposed a methodology that not only helps in identifying whether a leaf is healthy or diseased but also tell us the percentage of diseased region. And the percentage of infected part, can help us in providing proper remedy to cure the unhealthy plant. The main purpose of our project is to help farmers in proper yielding of their crop. The system is designed in an easy to use fashion and the cultivators of the crop, the farmers will be able to afford it easily.

This paper presents a framework that consists of four major steps-

1) Image Preprocessing, where the image is resized and enhanced to remove the noise and improve its quality.

2) Image Segmentation, using the popularly used clustering algorithm (K-Means), utilizing the Euclidean Distance method to calculate the distance in between the cluster data points.

3) Feature Extraction, using co-occurrence matrix (GLCM).

4) Classification with the help of supervised learning model, the multi SVM Classifier with 500 epochs.

Following the process discussed above, the system is good enough to attain the results with a precision of 98.39% in identifying the plant disease. Thus, helping the farmers with limited knowledge in proper cultivation by providing the correct soil minerals and fertilizers or pesticides accordingly.

II. METHODOLOGY

The methodology used in designing the proposed framework is described in fig.1 in the form of a flowchart. It is implemented using the Machine learning techniques like image processing involving resizing and enhancing of image, segmentation of preprocessed image using K-Means algorithm, feature extraction using gray level co-occurrence matrix and finally using a multi-class SVM classifier the type of disease from which a plant leaf is infected is classified.

It helps us to create a leaf disease detection system that uses the image processing techniques like resizing and enhancement, followed by image segmentation to keep only the meaningful parts and discard the other partitions. Then, the features from the segmented image are extracted which is then fed to the supervised learning model called, multiclass SVM for the classification purpose.

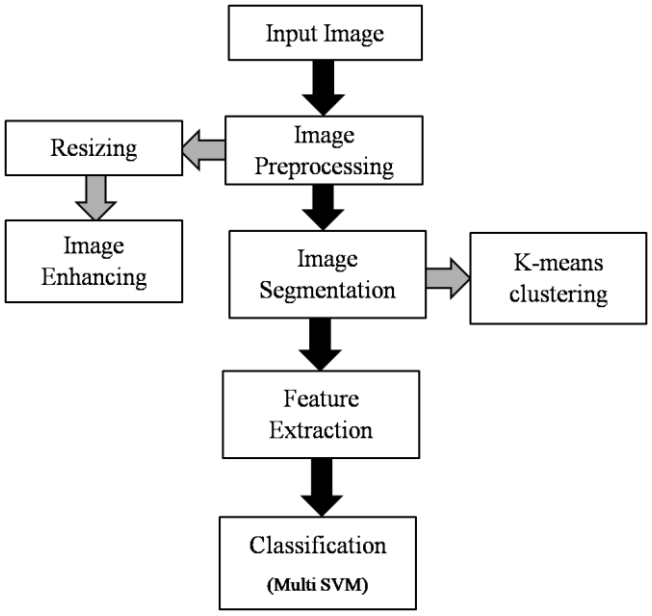

Fig. 1 Flowchart of proposed methodology

Each step involved in the proposed methodology is discussed in detail as follows:

A. Image Pre-processing

The images in the dataset need to be cleaned to improve the quality and remove the unwanted noise. Image pre-processing is the initial task done before designing the model. This process may also decrease the training time of the model, thus resulting in increased model inference speed. The proposed framework follows three steps of image preprocessing, i.e., image cropping, image enhancement and image converting. The image is firstly cropped on the leaf area with disease, and then it is converted to gray levels. Then, to increase the contrast, image enhancement is performed.

1) Image Resizing: Here image is resized and it is done by changing its original size value.

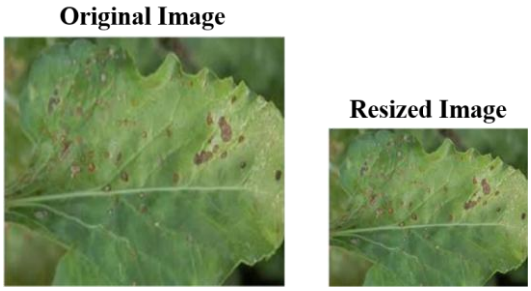

Fig. 2 Image resizing

2) Image enhancement: The resized image is given as the input, to which smoothing and filtering is applied. The sharp edges are removed and the enhanced version of the image is obtained as the output in this step.

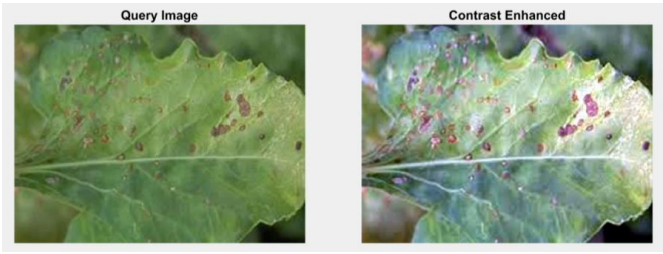

Fig. 3 Image enhancement

3) Image Color Conversion: The enhanced image which is in RGB form is then converted into HSI image.

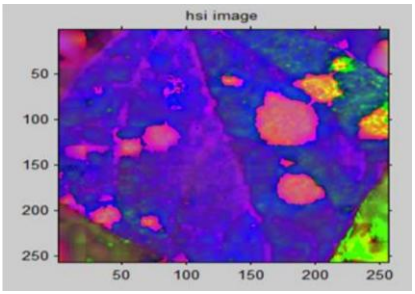

Fig. 4 Image color conversion

B. Image Segmentation

Not all part of the image is useful in the detection process. So, to keep only the meaningful part of the image for future purpose, it is segmented in several partitions as per the concerned regions. This detects the division of the same and meaningful regions. For this purpose, the pixels of same type are identified and grouped together. Image segmentation is important as it helps in the detecting the objects and edges of the image. In the proposed project, clustering algorithm called K- means is applied for segmentation purpose. This algorithm classifies the objects into K classes according to set of attributes. The Mathematical support used with K-Means is the Euclidean distance among objects and related cluster.

1) K-Means Clustering Algorithm: This Algorithm is used for the implementation to help identify the clusters of infected portions on the plant's leaf. Since, no prior knowledge about such patterns or clusters was available, therefore, the K-Means algorithm is selected. The algorithm is discussed stepwise, as follow:

i) The data is to be clustered in K groups, where K is predefined. (the value of K can be varied as per the necessity)
ii) Arbitrarily, K data points are selected as the centers of the clusters.
iii) Data points are allocated to their close by cluster using the distance function.
iv) Clusters are reformed and the new centroid for each cluster is evaluated depending on the Euclidean distance.
v) Step 3 and 4 are repeated until, not much change in the centroids is noticed.
vi) Return those centroids and stop the algorithm.

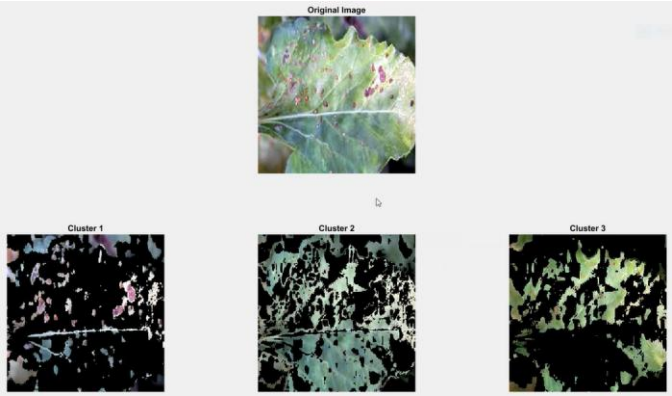

Fig. 5 Clusters formed

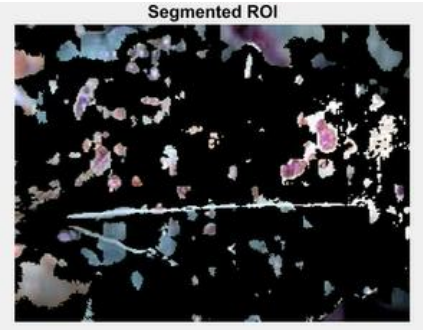

Fig. 6 Segmented region of interest

2) Euclidean Distance Function: K-Means algorithm, is an effective method, but the no. of clusters to be defined beforehand. Also, it requires the mathematical support for the clustering process. Here, in the proposed methodology, the proximity standard considered is Euclidean distance. And, it is calculated with the help of the equation given below:

$$D(x,y) = \sqrt{\sum |x_i - y_i|^2} \quad (1)$$

C. Feature Extraction

Features are the characteristics or attributes of the data objects. For a plant leaf and disease detection, features can be color, texture, morphology and structure. Here, one of the earliest technique of extracting the features, Co-occurrence matrix method is used. A grey-level co-occurrence matrix shows different distributions of grayscale values or colors at a given offset. A GLCM is also called, co-occurrence distribution.

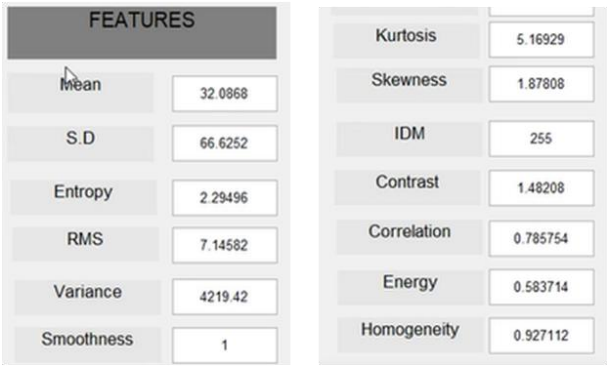

Fig. 7 Features extracted

D. Classification

A multi-class 500 epochs SVM (Support Vector Machine) is implemented for the classification purpose. In machine learning, it comes under the supervised learning model. SVMs work in a very efficient manner for classification and regression analysis. In order to produce an output, SVM has to be associated with a learning algorithm. Experiments and studies show that, SVM is better at classifications and regressions as compared to other machine learning models.

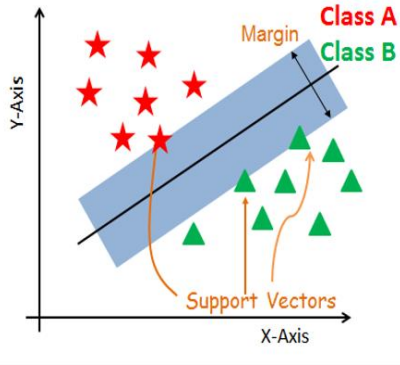

Fig. 8 SVM classifier

III. RESULTS

The results of the plant leaf disease detection are as follows:

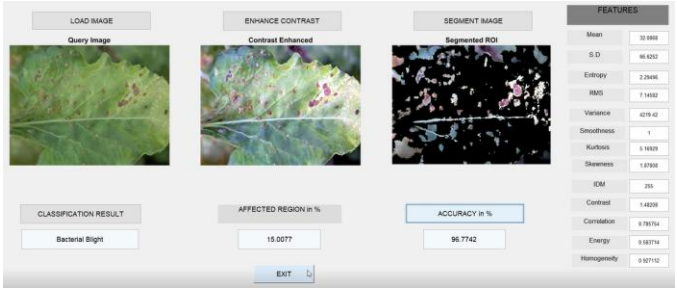

Fig. 9 Results of leaf disease identification

IV. PERFORMANCE OF DIFFERENT TECHNIQUES IN PLANT DISEASE IDENTIFICATION

TABLE I
PERFORMANCE OF DIFFERENT TECHNIQUES

Classifier	Accuracy (in %)
Convolutional Neural Network	83.5
K-Means	93.6
Bayes Network	93
Proposed Method	98.39

V. CONCLUSION

This paper has presented an automated framework for the plant leaf detection and disease identification. The approach follows four steps, i.e. pre-processing, segmentation, extracting the features and classification of the image. We found that the presented framework attained the results with the correctness of 98.39% for the disease detection purpose, which is better than other techniques already designed. With the correct identification of the disease and the percentage of the infected part, the work of the farmers is made easy. They can use desired quantity of the pesticides accordingly. Since, the disease is identified correctly, this will result in controlled pests and increased harvest.

In future, we can expand this project by adding the treatment for the detected disease. Also, we can release a web version of the project. We can also try to improve the performance of the proposed approach by combining different techniques of segmentation or classification. With this concept, it will be easy to identify disease and rectify the problem with less cost. We can also make an android app where the user can directly click the picture of the leaf using the mobile camera. Also, we can try to combine this approach with deep learning algorithm like CNN and can check if we can get better accuracy.

REFERENCES

[1] Barbedo, Jayme Garcia Arnal, A new automatic method for disease symptom segmentation in digital photographs of plant leaves, European Journal of Plant Pathology, 2016, 1-16.

[2] Bhong, Vijay S and Pawar B.V, Study and Analysis of Cotton Leaf Disease Detection Using Image Processing, International Journal of Advanced Research in Science, Engineering and Technology, 3 (2), 2016.

[3] Lumb, Manisha, and Poonam Sethi, Texture Feature Extraction of RGB, HSV, YIQ and Dithered Images using GLCM, Wavelet Decomposition Techniques, International Journal of Computer Applications, 68 (11), 2013.

[4] Reena Tijare, Pawan Khade, Rashmi Jain, The Survey of Disease Identification of Cotton Leaf, International Journal of Innovative Research in Computer and Communication Engineering, 2015.

[5] Sasirekha N, Swetha N, An Identification of Variety of Leaf Diseases Using Various Data Mining Techniques, International Journal of Advanced Research in Computer and Communication Engineering, 4 (10), 2015.

[6] Zhang S. W, Shang Y.J, and Wang L, Plant Disease Recognition Based on Plant Leaf Image, Journal of Animal and Plant Sciences, 25 (1), 2015, 42-45.

[7] Khirade SD, Patil AB (2015) Plant Disease Detection Using Image Processing. Int Conf on Comp Communication Control and Automation IEEE 153: 768-771

[8] Surya Prabha D, Satheesh Kumar J (2014) Study on Banana Leaf Disease Identification Using Image Processing Methods. Int Jour of Res in Comp Sc and Inf Tech IJRCSIT 2(A): 89-94

[9] Weizheng S, Yachun W, Zhanliang C, Hongda3 W (2008) Grading Method of Leaf Spot Disease Based on Image Processing. Int Conf on Comp Sc and Soft Eng, IEEE 491-494 [4] Phadikar S, Sil J (2008) Rice Disease Identification using Pattern Recognition Techniques, Proce of 11th Int Conf on Comp and Inf Tech, ICCIT, Khulna, Bangladesh, IEEE 420-423

[10] Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. 7 (2016). Article: 1419