Leaf Disease Detection and Classification based on Machine Learning

Sandeep Kumar
Department of ECE
Sreyas Institute of Engineering & Technology,
Hyderabad, India
drsandeep@sreyas.ac.in

KMVV Prasad
Department of ECE
Sreyas Institute of Engineering & Technology,
Hyderabad, India
mvvprasad@sreyas.ac.in

A. Srilekha
Department of ECE
Sreyas Institute of Engineering & Technology,
Hyderabad, India
srilekha.aruna@gmail.com

T.Suman
Department of ECE
Sreyas Institute of Engineering & Technology,
Hyderabad, India
suman.t@sreyas.ac.in

B. Pranav Rao
Department of ECE
Sreyas Institute of Engineering & Technology,
Hyderabad, India
pranavrao01@gmail.com

J.Naga Vamshi Krishna
Department of ECE
Sreyas Institute of Engineering & Technology,
Hyderabad, India
vamshisharma6411@gmail.com

***Abstract*—Detection of diseases in plants is a significant task that has to be done in agriculture. This is something on which the economy profoundly depends. Infection discovery in plants is a significant job in the agribusiness field, as having diseases in plants is very common. To recognize the diseases in leaves, a continuous observation of the plants is required. This observation or continuous monitoring of the plants takes a lot of human effort and it is time-consuming too. To make it simply some sort of programmed strategy is required to observe the plants. Program based identification of diseases in plants makes easier to detect the damaged leaves and reduces human efforts and time-saving. The proposed algorithm distinguishing sickness in plants and classify them more accurately as compared to existing techniques.**

Keywords— Leaf, Diseases, K-means algorithm, GLCM, Support Vector Machine (SVM).

I. INTRODUCTION

Agriculture is the spine of our nation. Our country is renowned for agriculture [1-2]. In India, people are most likely interested in agriculture. It plays a very important role in the Indian economy. Around 70% of rural areas depend on agriculture. A total of 17% is paid to GDP and gives the people employment opportunities for 60% of the total population [3-5]. Hence plant disease detection plays a significant role in agriculture. The method of processing the image is an appropriate technique used in the application of agriculture. Plants are most affected by fungus and various bacterial diseases [6]. The automatic system helps to avoid the farmers to frequently visit the fields. Automatic detection systems also are helpful in the case of large crops. From the symptoms that appear on the leaves, this system detects the diseases. Due to the increase in population, the changes in climate affect plants [7-8]. With new technologies, plant disease can easily be predicted. Taking care of leaf needs continuous observation. Especially for the diseases that affect production and post-harvest life [9]. A disease leaf is taken and its image is processed. We aim to detect diseases namely Alternaria alternate, Anthracnose, bacterial blight, and Cercospora leaf spot [10-13]. The area infected, accuracy, and percentages of infection are measured [14-15].

II. LITERATURE SURVEY

Abirami Devaraj [6] mentions the usage of a K-means clustering algorithm in the segmentation of the image which partitions the image into k clusters. That one part of the cluster contains an image with the majority of the unhealthy part. He classified diseases by using Random Forest Classifier which is time-consuming and very complex. Suja Radha [3] proposed a method to calculate the affected region in the leaf. For segmentation, K-means clustering is the method used. She used Support Vector Machine (SVM) to classify the diseases which have high dimensional input space compared to that of remaining classifiers. Latha S et al. [5] proposed a novel methods that are used for disease detection. He discussed the Otsu Threshold algorithm and K-means algorithm for the segmentation of the images. The Color co-occurrence method and Leaf color extraction using H and B components are discussed for the feature extraction. And he compared ANN and BPNN classifiers to classify the diseases. P. Kaur et al. [4] discussed a method for spotting diseases which will show in the leaves of cucumber plants. Segmentation of healthy and infected areas is accomplished utilizing a statistic pattern recognition approach. Inside that features like color, shape, and texture will be extracted. Those features are supplied to the SVM, which performs the utmost classification. He concluded saying, results obtained from the SVM is so good when compared to those we got from neural networks.

III. PROPOSED METHODOLOGY

In this proposed algorithm, we aimed to detect the diseases in plants by considering leaves. In our proposed methodology, we are providing the type of disease that the leaf is effected by, from the images, and shows the diseased region of the leaf by the means of image processing technique. The result will be provided within less time and the percentage of the area affected and also the accuracy. Samples of images are

978-1-7281-7213-2/20/$31.00 ©2020 IEEE

collected that comprised of different plant diseases like Alternaria Alternate, Bacterial Blight, Cercospora leaf spot.

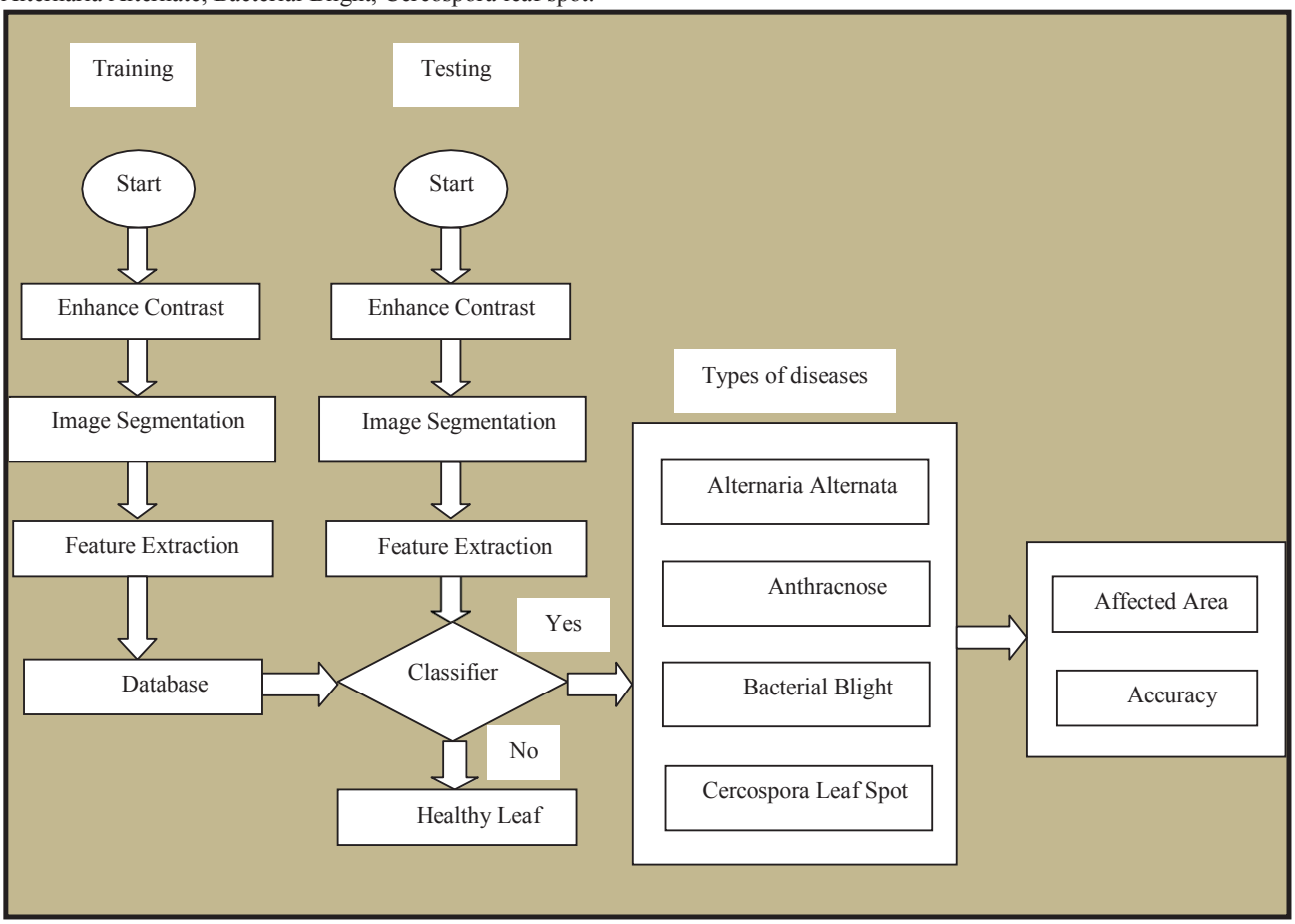

Fig. 1. Flow Chart of Proposed Work

Anthracnose, and Healthy leaves. Different images of the leaves are collected for every disease that was classified into database images and input images [1-2]. The allocation of the image depends upon the shape and texture oriented features. The leaf samples display the disease detection using color-based segmentation model. In this paper, we have discussed the plant disease prediction using the K-means clustering algorithm for segmentation. GLCM is for the extraction of features. The different types are diseases are classified by using the Support Vector Machine (SVM) classifier method.

A. Enhance Contrast

The enhancement of the image involves changes in brightness and contrast. We need to adjust the image intensity value by the color map and find the limits also, to contrast stretch the image. Contrast adjustment is done by scaling all pixels of the image by a constant (K) [6], [8-9]. Conversion of RGB to Grayscale can be done in two ways. These methods have both advantages and disadvantages. They are as follows:
1. Average method: It is an easy method that considers the average of three colors RGB. 2. Weighted method: In this method it calculates grayscale image by using Grayscale = 0.3R + 0.59G + 0.11B

B. Image Segmentation

Image Segmentation is a technique implemented for converting the digital image into several segments that help in easy analysis. RGB is the most widely used color representation method in computer graphics. Each primary color may take an intensity value [10]. If the obtained three components are the same then it is undefined. The pixel scope of RGB is [0, 255]. By calculating the components Hue, Saturation, Intensity pixel range can be converted.

C. K-Means Clustering

It is used to segregate the objects based on the leaf's features into groups. K-means technique separates the obtained image into clusters. After the formation of clusters, one part of it contains the majority of the affected region. K-means cluster classifies the images into k number of categories [11]. The distance between instance and cluster center is calculated. This can be done by using the Euclidean distance metric.

$$Dist_{XY} = \sqrt{\sum_{k=1}^{m} (X_{ik} - X_{jk})^2}$$

D. *Feature Extraction (Gray Level Co-occurrence Matrix (GLCM))*

Feature extraction is mainly used to extract important information that represents diverse classes. Through segmentation, we have separated the diseased area from the image. Various features are extracted from the segmented image in feature extraction. It's an old statistical method that uses feature extracting and texture classifying methods. This method gives the relation between pairs of pixels in an image. The texture features are given by GLCMs [16-17]. Example contrast, correlation, energy, entropy, and homogeneity. In recent times, instead of using GLCM individually, it is used in combination with other methods. Few implementations of the GLCM, are one-dimensional GLCM and second-order statistical GLCM. These are additionally applied to various color space for the color co-occurrence matrix.

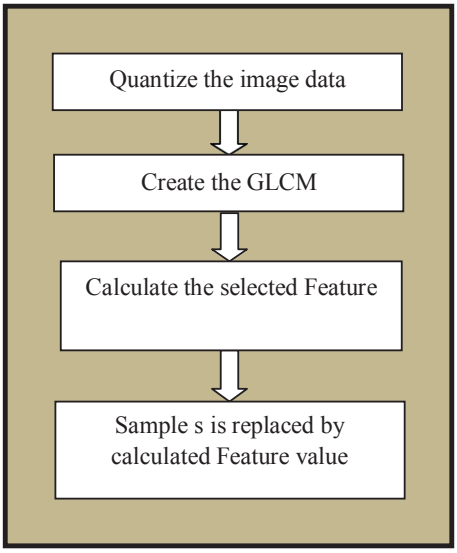

Fig. 2. Flow Chart of GLCM

E. *SVM Classifier*

A support vector machine (SVM) is a supervised learning technique that categorizes data. The hyperplane is the line that splits the data based on the distance between support vectors. Support vectors are extreme points of the dataset. Based on the hyperplane, it is easy to classify the new data and assign them to it [16-19]. It is most generally utilized in recognizing facial expressions, speech recognition, and texture classification, etc. The samples that are closer to the margin are selected which determines the hyperplane is known as support vectors. SVM will be mainly based upon increasing the distance from the separating hyperplane to the nearest example. In fundamental SVM only binary classification is upheld, however, an extension, a multiclass classification case will be possible. Two types of SVM classifiers, namely: 1. Linear SVM Classifier 2. Non-Linear SVM Classifier The main advantage of SVM is, it is vigorous in any event.

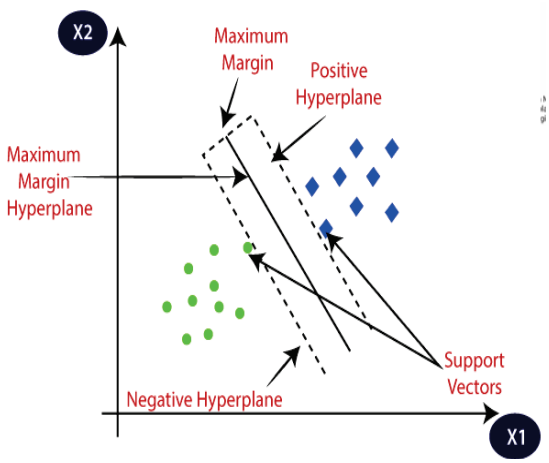

Fig. 3. SVM Classification

IV. RESULTS

When the color segmentation is done, we will get three clusters. These clusters contain the disease spots. The diseased area cluster object is selected to find the percentage of disease area. Let the diseased area be (X). The total number of pixels is taken for finding out the whole area of the leaf. Let it be taken as (Y). The infected area (%) is calculated by using the formula:

Affected region = (X /Y) × 100.

Accuracy is determining correctly classified diseases. We may have any number of images of the leaves that comprises of different diseases. Hereby using classifier we are classifying the various diseases. Through classification, we may get correct outputs. The rate of getting the correct output is accurate. Accuracy can be calculated by:

Accuracy (%) = 100* ((No. of correctly Classified) / (Total no of leaves in Datasets))

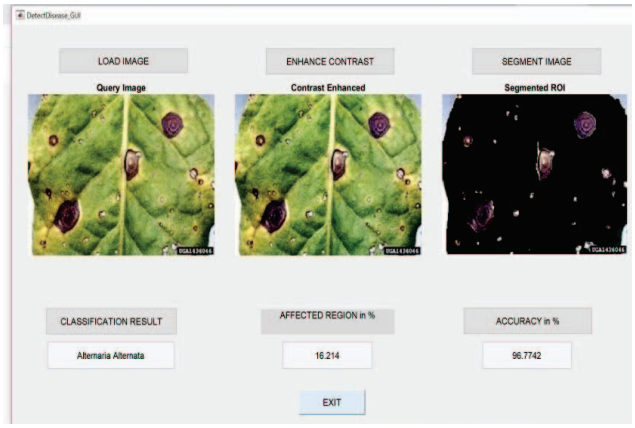

Fig. 4. Flow Chart of GLCM

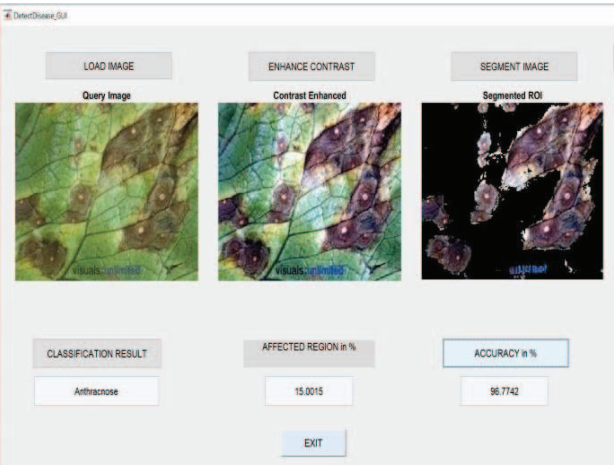

Fig. 5. Flow Chart of GLCM

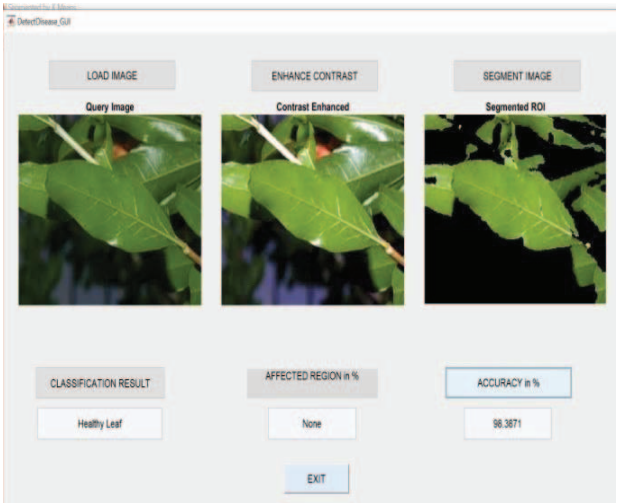

Fig. 6. Flow Chart of GLCM

TABLE 1: Accuracy of Proposed Work

S.No	Disease	Affected Region (%)	Accuracy in %
1	Alternaria Alternate	16.214	96.7102
2	Bacterial Blight	15.0142	95.8042
3	Anthracnose	15.0015	96.0042
4	Cercospora Leaf Spot	15.6087	95.1772
5	Healthy leaf	None	98.3871

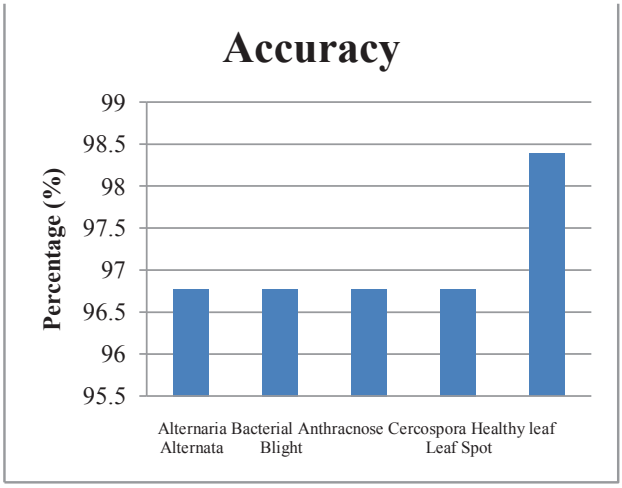

Fig. 7. Accuracy on Proposed Work

V. Conclusion

In our proposed work, we have used various images for detecting leaf diseases. We have used segmentation technique like k-means clustering, for extracting various features Gray Level Co-occurrence Matrix (GLCM) is used and Support Vector Machine (SVM) classifier to classify different types of diseases. This process helps us to find the different diseases in leaves precisely. The dataset consists of the different leaf images affected by various diseases like Cercospora leaf spot, bacterial blight, Anthracnose, and Alternaria alternata. The results show the area affected and the percentage that is affected with great accuracy.

ACKNOWLEDGMENT

We are very thankful for the center of excellence in the Sreyas Institute of Engineering & Technology for fruitfully helping this paper.

REFERENCES

[1] 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.

[2] 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.

[3] Suja Radha, Leaf Disease Detection using Image processing, Journal of Chemical and Pharmaceutical Sciences, March 2017.

[4] Prabhjeet Kaur, Sanjay Singla, Sukhdeep Singh, Detection and classification of leaf diseases using an integrated approach of support vector machine and particle swarm optimization, International Journal of Advanced and Applied Sciences, March 2017.

[5] Saradhambal.G, Dhivya.R, Latha.S, R.Rajesh, Plant disease detection, and its solution using image classification, International Journal of Pure and Applied Mathematics, Volume 119 No. 14 2018, 879-884.

[6] Abirami Devaraj, Karunya Rathan, Sarvepalli Jaahnavi, and K Indira, Identification of Plant Disease using Image Processing Technique, International Conference on Communication and Signal Processing, April 4-6, 2019.

[7] Vishal Mani Tiwari&Tarun Gupta "Plant Leaf Disease Analysis using Image Processing Technique with Modified SVM-CS Classifier" Research Gate2017.

[8] B.V. Ramana Reddy, A. Suresh, M. Radhika Mani, and V.Vijaya 50 Kumar, "Classification of Textures Based on Features Extracted from Pre-processing Images on Random Windows", International Journal of Advanced Science and Technology, Volume 9, pp 9 – 18, August2009.

[9] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. AL Rahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011.

[10] Gavhale, K.R., Gawande, U., „An Overview of the Research on Plant Leaves Disease Detection using Image Processing Techniques", IOSR Journal of Computer Engineering, Volume 16, Issue 1, pp. 10-16,2014.

[11] 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.

[12] Sonal P. Patil, Ms. Rupali S.Zambre"Classification of Cotton Leaf Spot Disease Using Support Vector Machine," International Journal of Engineering Research and Applications, ISSN: 2248-9622, Vol. 4, Issue 5, May 2014, pp.92-97.

[13] Ijai Singh, A. K Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques, Volume 4, Issue 1, March 2017, Pages 41-49.

[14] Gaurav Kushinagar, A. N Thakre, " Plant Disease Detection in Image Processing using MATLAB," in International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 6 Issue 4.

[15] Kumar S., Sukhwinder Singh and Jagdish Kumar "Live Detection Of Face Using Machine Learning with Multi-Feature Method" in Wireless Personal Communication Springer Journal (SCI) DOI: 10.1007/s11277-018-5913-0. 4.

[16] Kumar S., Sukhwinder Singh and Jagdish Kumar "Automatic Live Facial Expression Detection Using Genetic Algorithm with Haar Wavelet Features and SVM" in Wireless Personal Communication Springer Journal (SCI) DOI: 10.1007/s11277-018-5923-y.

[17] Kumar S, Sukhwinder Singh, and Jagdish Kumar, "A Multiple Face Detection Using Hybrid features with SVM Classifier" in the Springer Nature on Data Communication and Networks with ISBN: 978-981-13-2254-9.

[18] Kumar S, Sukhwinder Singh, and Jagdish Kumar, "Gender Classification Using Machine Learning with Multi-Feature Method" in IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, USA, January 7th-9th, 2019.

[19] Kumar S., and Sanjay Sharma. "Image Compression Based On Improved Spiht and Region Of Interest." PhD diss., 2011.