Disease Identification in Plants Using K-means Clustering and PCA Based GLCM using Classification with ANN

[1]Ms.S.Kousalya ,[2]J.Devika ,[3]S.Karthiga, [4]S.Roopa Sri

[1]Assistant professor.,[2,3,4] PG Scholars,Sri. SRNM College, TamilNadu. India

[1]skousalya.07@gmail.com, [2]devi4844@gmail.com ,[3]karthigaselvaraj1596@gmail.com, [4]srirubabca1995@gmail.com.

***Abstract*; Image processing is the information from the images. The plant disease detection is the technique which is applied to detect disease from the input images. In this work, the technique is applied which is based on textural feature extraction, segmentation, and classification. The GLCM algorithm is applied which extracts textural features from the image. The k-mean clustering algorithm is used for the segmentation of input images. The SVM classifier is applied to the existing algorithm which will classify the input two classes. To improve performance of existing algorithm the SVM classifier is replaced with ANN classification. This leads to an improved accuracy of disease detection, moreover, classifying the data into multiple classes.**

KEY TERMS: textural feature, GLCM, segmentation, ANN classification

I INTRODUCTION

Agriculture is one of the most important sources for human sustenance on Earth. Not only does it provides the much necessary food for human existence and consumption but also plays a major vital role in the economy of the country. A disease is caused by the pathogen which is an agent causing disease. In general, there are two types of factors which can bring death and destruction to plants; biotic and abiotic agents. Biotic agent including insects, bacteria, fungi, and viruses. Abiotic agents include extremes of temperature, excess moisture, poor light, insufficient nutrients, and poor soil pH and air pollutants. In this paper, we have used ANN classifier to identify the pest disease in the plants. Image acquisition devices are used to acquire images of plantations at regular intervals. These images are subjected to preprocessing. The pre-processed leaf images are then segmented using K-means clustering method. Then, the color features(mean, skewness), texture features such as energy, entropy, correlation, contrast, edges are extracted from diseased leaf image using gray level co-occurrence matrix (GLCM) in the texture & compared with normal leaves image. The Artificial neural networks (ANN) classifier is used to classify the pest Disease in plants. In this changing environment, appropriate and timely disease identification including early prevention has never been more important. There are numerous ways to sense plant pathologies. Some diseases do not have any visible symptoms, or the effect becomes noticeable too late to act, and in those situations, a sophisticated analysis is obligatory. However, most diseases generate some kind of manifestation in the visible spectrum, so the naked eye examination of a trained professional is the prime technique adopted in practice for plant disease detection. In order to achieve accurate plant disease diagnostics, a plant pathologist should possess good observation skills so that one can identify characteristic symptoms [8]. Variations in symptoms indicated by diseased plants may lead to an improper diagnosis since amateur gardeners and hobbyists could have more difficulties determining it than a professional plant pathologist. An automated system designed to help identify plant diseases by the plant's appearance and visual symptoms could be of great help to amateurs in the gardening process and also trained professionals as a verification system in disease diagnostics. Advances in computer vision present an opportunity to expand and enhance the

practice of precise plant protection and extend the market of computer vision applications in the field of precision agriculture. Exploiting common digital image processing techniques such as color analysis and thresholding [9] were used with the aim of detection and classification of plant diseases. Various different approaches are currently used for detecting plant diseases and most common are artificial neural networks (ANNs) [10] and Support Vector Machines (SVMs) [11]. They are combined with different methods of image preprocessing in favor of better feature extraction. In machine learning and cognitive science, ANN is an information-processing paradigm that was inspired by the way biological nervous systems, such as the brain, process information. The brain is composed of a large number of highly interconnected neurons working together to solve specific problems.

The affected leaves are classified as diseases;

a) Bacterial disease: e.g. Bacterial Blight, Crown Gall,LintDegradation.

b) Fungal diseases: e.g. Anthracnose, Leaf Spot.

c) Viral disease: e.g. Leaf Curl, Leaf Crumple, Leaf Roll.

II PROPOSED SYSTEM

In this work concentrated on identification of Leaf Spot disease and Leaf Miner from the photographic signs and classify them using image processing techniques. The proposed framework has been implemented in three steps. First, image segmentation is performed using K means clustering to identify the infected area. In the next step, leaf features are extracted from segmented regions using feature extraction techniques such as GLCM. These features are used for classification into infected or non-infected leaf type. As third step these features are given to the classifier to classify the disease in the plants. This work uses ANN classifier to obtain effective results.

III METHODS AND MATERIALS

- Disease identification in plants presented in this paper uses k-means clustering.
- As a tested, investigate a collection of plant pests, which were obtained from a digital camera.
- Algorithm can be tested are more than ten affected images as shown in figure 2.

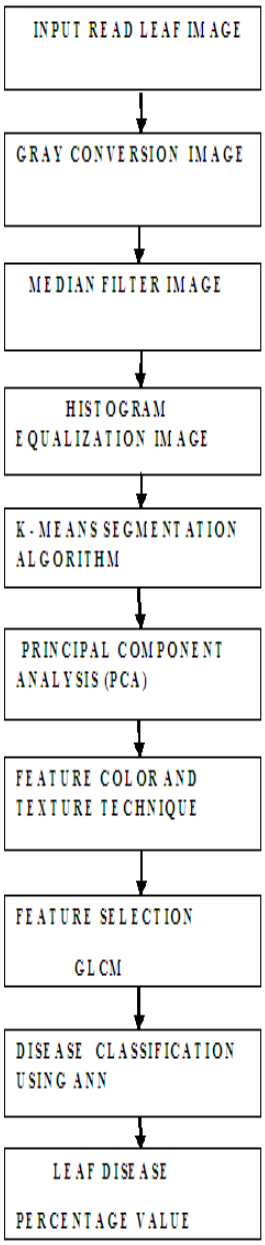


Figure 1: Data flow diagram.

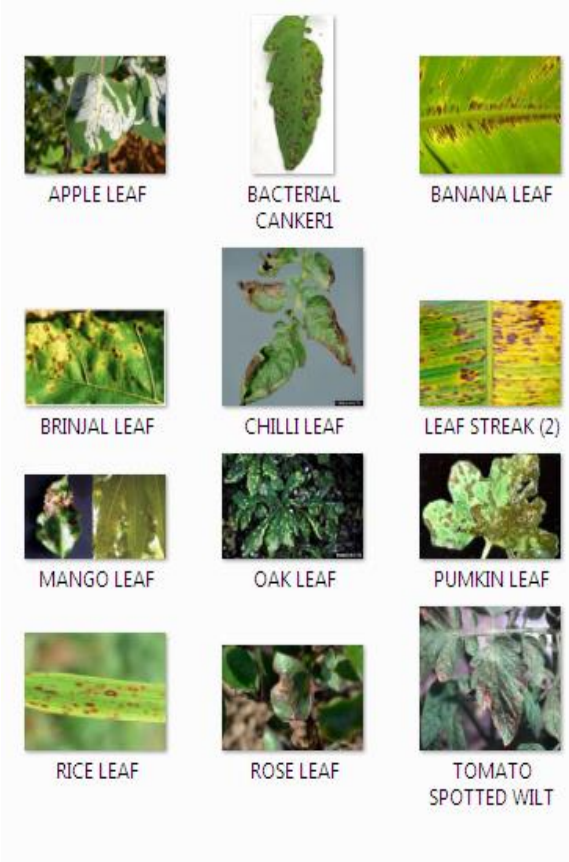


Figure 2: Affected images.

The common approach used by the farmers to identify the plant diseases is by consulting the experts which very tedious, consumes more time and money. In addition, there are some diseases which cannot be identified by the naked eyes. Hence in this context, a fast, reliable and automatic method is required to accurately identify the plant diseases. This paper provides the review of image processing techniques applied for identifying plant diseases detection. The figure represents the general image processing steps adopted for plant disease detection using image processing.

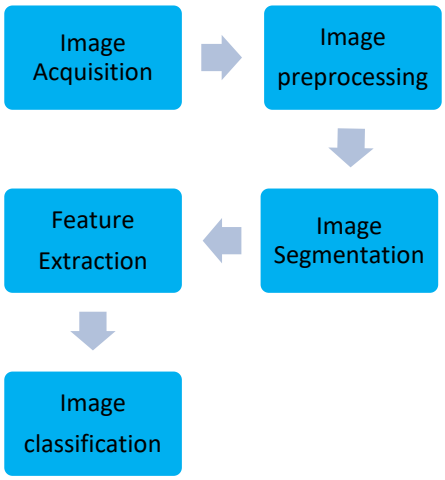
Image Acquisition: The digital images are acquired using a digital mobile camera or digital camera and given as input to the identification system. This is the image in which the leaf disease has to be identified by the system.

Image Preprocessing: It is the technique for improving picture quality prior to computational processing and also used to remove the low-frequency noise, reflections and masking portions of the images.

Image Segmentation: Segmentation is the process of partitioning the digital image into multiple parts/segments. It helps in simplifying the image into more meaningful and easier to analyze.

Feature extraction: After the segmentation process, various features are extracted from the infected region. The features which can be used in plant diseases detection are color, texture, and shape etc.

Image Classification: It is most important part in the digital analysis. Classification can be executed on spectral features like density, texture etc and then divides the features space into many classes using different machine learning algorithms.



Result:

- The main purpose of the RGB color model is for the sensing, representation, and

display of images in electronic systems, such as televisions and computers, though it has also been used in conventional photography.

- Before the electronic age, the RGB color model already had a solid theory behind it, based in human perception of colors.

Figure 3:Input image.

- The affected leaf images were converted from RGB color format to gray scale images.
- Segmentation refers to the process of clustering the pixels with certain properties into salient regions and these regions correspond to different faces, things or natural parts of the things.

Figure 4: Median filter

- In this paper, k-means segmentation technique are used to fragment goal areas.
- Target regions are those areas in the image that represented visual symptoms of a fungal disease.

Figure 5: K-means clustering.

Figure 6: k-means Segmentation.

- As the paddy leaf disease consists of several types of disease blast, brown spot and narrow brown spot that had different lesion shape and lesion color.
- Shape is one of the important parameter of the image. Breadth and length of the image are significant characteristic to describe the shape
- A simple approach is to measure the breadth and height of the image is to measure the count of the object pixel.

Color plays an important role in image processing. Digital image processing produce quantitative color measurement that are useful for the work of inquiring the lesion for early diagnosis. The pixel in the color images are commonly represented in RGB format, where RGB are RED GREEN BLUE values respectively from the color images capturing device, Grey-Level Co-occurrence Matrix (GLCM). Grey-Level Co-occurrence Matrix texture measurements have been the workhorse of image texture 'Correlation' Returns a measure of how to correlate a pixel to its neighbour over the whole image. The GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image.

Figure 7:Otsu's thresholding.

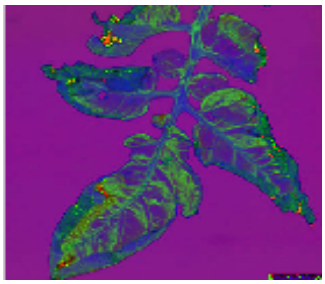


Figure 8:HIS

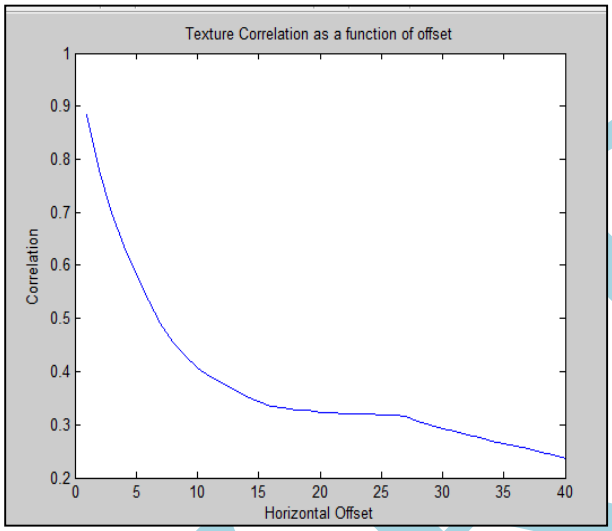


Figure 9: Textural feature.

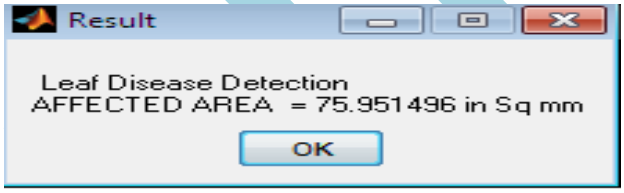


Figure 10: Result for affected area.

IV KMEANS SEGMENTATION

k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

K-MeansClusteringAlgorithm:

Clustering is a method to divide a set of data into a specific number of groups. It's one of the popular methods is k-means clustering. In k-means clustering, it partitions a collection of data into a k number group of data. It classifies a given set of data into k number of disjoint cluster. K-means algorithm consists of two separate phases.In the first phase it calculates the k centroid and in the second phase, it takes each point to the cluster which has the nearest centroid from the respective data point. There are different methods to define the distance of the nearest centroid and one of the most used methods is Euclidean distance. Once the grouping is done it recalculate the new centroid of each cluster and based on that centroid, a new Euclidean distance is calculated between each center and each data point and assigns the points in the cluster which have minimum Euclidean distance. Each cluster in the partition is defined by its member objects and by its centroid. The centroid for each cluster is the point to which the sum of distances from all the objects in that cluster is minimized. So K-means is an iterative algorithm in which it minimizes the sum of distances from each object to its cluster centroid, over all clusters. Let us consider an image with a resolution of x×y and the image has to be cluster into k number of cluster. Let p(x, y) be an input pixel to be cluster and ck be the cluster centers. An algorithm for k-means13 clustering is following as:

- Initialize the number of cluster k and centre.
- For each pixel of an image, calculate the Euclidean distance d, between the center and each pixel of an image using the relation given below.
- Assign all the pixels to the nearest centre based on distance d.
- After all, pixels have been assigned, recalculate the new position of the centre.

using the relation given below.

$$c_k = \frac{1}{k} \sum_{y \in c_k} \sum_{x \in c_k} p(x, y)$$

- Repeat the process until it satisfies the tolerance or error value.
- Reshape the cluster pixels into the image. Although k-means has the great advantage of being easy to implement, it has some drawbacks. The quality of the final clustering results depends on the arbitrary selection of initial centroid. So if the initial centroid is randomly chosen, it will get the different result for different initial centers.So the initial center will be carefully chosen so that we get our desire segmentation.And also computational complexity is another term which we need to consider while designing the K-means clustering. It relies on the number of data elements, number of clusters and number of iteration.

V CONCLUSION

In this paper, the proposed system work discrimination between healthy plants and diseased plants by using an ANN technique. In this paper, respectively, the applications of K-means clustering have been formulated for clustering and classification of diseases that affect on plant leaves. The whole process of leaf classification can be implemented In this study, extracted using the Gray-Level Co-occurrence Matrix (GLCM) and the Principal Component Analysis (PCA) algorithms, using leaf detection, feature extraction and GLCM features are rendered to an ANN classifier for purpose of classification.

References

[1] Ananthi, S. Vishnu Varthini, Detection, And Classification Of Plant Leaf Diseases International Journal of Research in Engineering & Applied Sciences, Volume 2, Issue 2 (February 2012) ISSN: 2249-3905

[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. Applications (0975-8887), Volume 17-No.1.March2001.

[3] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad , (2010)A Framework for Detection and Classification of Plant Leaf and Stem Diseases, International Conference on Signal and Image Processing pp 113-118.

[4] Dae Gwan Kim, Thomas F. Burks, Jianwei Qin, Duke M. Bulanon, Classification of grapefruit pee diseases using color texture feature analysis, International Journal Agriculture and Biological Engineering, Vol:2,No:3,September 2009. Pp 41-50.

[5] Sabine D. Bauer , Filip Korc, Wolfgang Forstner, The Potential of Automatic Methods of Classification to identify Leaf diseases from Multispectral images, Published online: 26 January 2011,Springer Science+Business Media, LLC 2011., Precision Agric (2011 12:361–377, DOI 10.1007/s11119-011-921.

[6] Wang Jue,Wang shitong,Image Thresholding Using Parzen Window Estimation .Jurnal of applied sciences8(5):772-779,2008,ISSN 1812- 5654,Asian Network for Scientif Information,2008

[7] Muhammad Hameed Siddiqi1, Suziah Sulaiman, Ibrahima Faye and Irshad Ahmad, A Real Time Specific & Weed Discrimination System Using Multi-Level Wavelet Decomposition, International Journal of Agriculture Biology, ISSN Print: 1560–8530; ISSN Online1814-9596 ,09– 118/YHP/2009/11–5–559–565

[8] M. B. Riley, M. R. Williamson, and O. Maloy, “Plant disease diagnosis. The Plant HealthInstruction”2002.

[9] J. G. Arnal Barbedo, “Digital processing techniques for detecting, quantifying and classifying plant diseases,” SpringerPlus, vol. 2, article 660, pp. 1–12, 2013. View at Publisher · View at Google Scholar · View at Scopus

[10] H. Cartwright, Ed., Artificial Neural Networks, Humana Press, 2015.

[11] I. Steinwart and A. Christmann, Support Vector Machines, Springer Science & Business Media, New York, NY, USA, 2008. View at MathSciNet.

