International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 03 | Mar-2016 www.irjet.net p-ISSN: 2395-0072

# Healthy and Unhealthy Plant Leaf Identification and Classification Using Hierarchical Clustering

**Vyshnavi.G.K.P[1], Sirpa.M.R [2], Chandramoorthy.M[3], Padmapriya.B[4]**

[1234]*M.E/Dept. of CSE*

**Abstract-** *Disease management is a challenging task. Mostly diseases are seen on the leaves or stems of the plant. Precise quantification of these visually observed diseases, pests, traits has not studied yet because of the complexity of visual patterns. Hence there has been increasing demand for more specific and sophisticated image pattern understanding. This work presents a method for identifying plant leaf disease based on color. Agrarians are suffering from the issue rising from different types of plant leaf diseases. Sometimes biologists are also unable to identify the disease that leads to lack of identification of right type of disease. The goal of proposed work to diagnose the disease using image processing and clustering techniques on image of plant leaves disease. First the input image is pre-processed. Then input image of leaves is converted as Red Green Blue (RGB) to Hue Intensity Saturation (HIS) or Lab color space. Then leaf disease segmentation is done using Hierarchical clustering. After segmentation the mostly green color pixels are masked based on specific threshold values. Support Vector Machine (SVM) and Neural Network (NN) is trained for classification.*

**Keywords:** Support Vector Machine, Neural Network, Hue Intensity Saturation, Disease, Agrarians, Clustering, Classification.

## 1. INTRODUCTION

India is an agricultural country where in most of the population depends on agriculture. Agriculture is one of the major domains which decides economy of the nation. Today India ranks second worldwide in farm output. Agriculture is still the largest economic sector and plays a major role in socio-economic development of India. Agriculture provide the livelihood of almost two thirds of the workforce in India. India has over 210 million acres of farm land. Jowar, wheat, maize, are the major cereals. Apple, banana, sapota, grapes, oranges are the most common fruits. Sugarcane, cotton, chilli, groundnuts are the major commercial crops. Crop cultivation depends on rainfall, quality of the soil and climatic conditions and short of any one of these leads to loss of crop. Diseases are major for loss of crop every year and really it is a challenge to control the diseases.

Plant disease diagnosis is an art as well as science. The diagnostic process (i.e., recognition of symptoms and signs), is inherently visual and requires intuitive judgment as well as the use of scientific methods. Plant diseases reduce both quantity and quality of plant products. Diseases are impairment to the normal state of the plant that modifies or interrupts its vital functions such as photosynthesis, transpiration, pollination, fertilization, germination etc. These diseases are caused by pathogens viz., fungi, bacteria and viruses, and due to adverse environmental conditions. Farmers require continuous monitoring of experts which might be prohibitively expensive and time consuming.

This research work is disclosure to automatic detection of disease on all variety of plant leaves.

### 1.1. Types of diseases on plant Leaves

The diseases on the leaves are classified as,

a) Viral disease: e.g. Leaf Curl, Leaf Crumple, Leaf Roll.
b) Fungal diseases: e.g. Anthracnose, Black Spot.
c) Bacterial disease: e.g. Bacterial Blight, Crown Gall, Lint Degradation.

In this research work mainly concentrated on identification of fungal disease on plant leaves.

### 1.2. Black spot disease on Leaves

Black spot, also spelled black spot, common disease of a variety of plants caused by species of Pseudomonas bacteria or by any number of fungus species in the genera *Asterina*, *Asterinella*, *Diplotheca*, *Glomerella*, *Gnomonia*, *Schizothyrium*, *Placosphaeria* and *Stigmea*. Infections occur during damp periods and appear as round to irregular black spots on leaves and sometimes on petioles, stems, and flower parts of susceptible plants. Black spot of roses is a serious widespread disease caused by the fungus *Diplocarpon rosae*. On roseplants, the spots are roundish and up to 1 cm (0.5 inch) in diameter with fringed margins. Leaves on susceptible varieties turn yellow and drop early. Affected plants may defoliate twice in a season, are greatly weakened, produce fewer and inferior blooms, and are subject to canker diseases and winterkill. Large numbers of spores are formed in speck-size fruiting structures (acervuli) and disseminated by splashing rain, dew, overhead sprinkling, and gardeners working among wet plants. The spores germinate and penetrate rose tissue in 9 to 18 hours or longer; new leaf spots appear in 3 to 16 days and spores in 10 to 18 days. The cycle may be repeated throughout the growing season. Black spot may

be controlled by fungicide application, planting resistant varieties, and removing any infected leaves immediately

**Fig 1.Black Spot Disease**

### 1.3. Anthracnose disease on Leaves

Anthracnose disease attacks all plant parts at any growth stage. The symptoms are most visible on leaves. At first, anthracnose generally appears on leaves as small and irregular yellow, brown, dark-brown, or black spots. The spots can expand and merge to cover the whole affected area. The color of the infected part darkens as it ages. The disease can also produce cankers on petioles and on stems that causes severe defoliation. Infected leaf has small, circular spots that may increase in size up to 1.2 cm in diameter. As it ages, the center of an older spot becomes blackish and emits gelatinous pink spore masses.

**Fig 2.Anthracnose disease leaf**

### 1.4. Fungal disease symptoms

Among all plant leaf diseases, those caused by fungus some of them are discussed below and e.g. Late blight caused by the fungus Phytophthora infesters. It first appears on lower, older leaves like water-soaked, gray-green spots. When fungal disease matures, these spots darken and then white fungal growth forms on the undersides. Early blight is caused by the fungus Alternaria solani. It first appears on the lower, older leaves like small brown spots with concentric rings that form a bull's eye pattern. When disease matures, it spreads outward on the leaf surface causing it to turn yellow. In downy mildew yellow to white patches on the upper surfaces of older leaves occurs. These areas are covered with white to greyish on the undersides.

## 2. RELATED WORK

In [1] proposed their research in "Application of Data Mining A Survey Paper" Data mining is a powerful and a new field having various techniques. It converts the raw data into useful information in various research fields. It helps in finding the patterns to decide future trends in medical field. This paper [2] presents an approach for careful detection of diseases, diagnosis and timely handling to prevent the crops from heavy losses. The diseases on the cotton are critical issue which makes the sharp decrease in the production of cotton. This paper provides a method to detect cotton leaves diseases using image processing technique. Firstly, K-means clustering algorithm is used for segmentation which classifies objects based on set of features into K no. of classes where feature extraction is color feature variance used for matching the train image features from database images and finally recognition is performed using Neural-network. Finally, neural network is used for recognizer where, initialization the images from the database that are highly correlated to the test image, which is given by user. It is used to analyze the cotton diseases which will be useful to farmers. Their research in "Classification Rules and Genetic Algorithm in Data Mining" Databases today are ranging in size into the Tera Bytes. It is an information extraction activity whose goal is to discover hidden facts contained in databases. Typical applications include market segmentation, customer profiling, fraud detection, evaluation of retail promotions, and credit risk analysis. Major Data Mining Tasks and processes include Classification, Clustering, Associations, Visualization, Summarization, Deviation Detection, Estimation, and Link Analysis etc. There are different approaches and techniques used for also known as data mining models and algorithms. Data mining algorithms task is discovering knowledge from massive data sets. In this paper, we are focusing on Classification process in Data Mining. In following "Agricultural Crops Classification Models Based on PCA-GA Implementation in Data Mining" Extraction of knowledge in agricultural data is a challenging task, from discovering patterns and relationships and interpretation. In order to obtain potentially interesting patterns and relationships from this data, it is therefore essential that a methodology be developed and take advantage of the sets of existing methods and tools available for data mining and knowledge discovery in databases. Data mining is relatively a new approach in the field of agriculture. Accurate information in characterizing crops depends on climatic, geographical, biological and other factors. These are very important inputs to generate characterization and prediction models in data mining. In this study, an efficient data mining methodology based on PCA-GA is explored, presented and implemented to characterize agricultural crops. The experimental results show improved classification rates and generated characterization models for agricultural crops. The domain model outcome may have benefits, to agricultural researchers and farmers. These generated classification models can also be utilized and readily incorporated into a decision support system.

## 3. EXISTING SYSTEM

In existing system k-means clustering algorithm, called the filtering algorithm is used. This algorithm is easy to

implement and only requires that a kd-tree be built once for the given data points. Efficiency is achieved because the data points do not vary throughout the computation and hence, this data structure does not need to be recomputed at each stage. A natural question is whether the filtering algorithm can be improved. The most obvious source of inefficiency in the algorithm is that it passes no information from one stage to the next. But this algorithm is quite complex and does not provide significantly faster running time in practice. Firstly, the RGB images of leaves are acquired. Then RGB images are converted into Hue Saturation Value (HSV) color space representation. RGB is an ideal for color generation. Hue is a color attribute that describes pure color as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and Value means amplitude of light. After the transformation process, the Hue component is taken for further analysis. Saturation and Value are dropped since it does not give extra information. 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. The main drawback of the system is this algorithm is quite complex. It does not provide significantly faster running time in practice. It takes more number of iterations.

## 4. PROPOSED SYSTEM AND OVERVIEW OF THE METHOD

To diagnose the disease using image processing and clustering techniques on image of plant leaves disease. First the input image is preprocessed. Already image segmentation of diseased leaves are done in K-mean clustering. But K-mean clustering performed more iterations for segmentation. The proposed system involves Hierarchical clustering for segmentation takes less iterations and are efficient. It consumes less time to attempt disease identification. After segmentation the mostly green color pixels are masked based on specific threshold values. Support Vector Machine is trained for classification. Hierarchical clustering algorithm takes less number of iterations. It takes less running time and efficient for segmentation. Support Vector Machine is done for classification and is useful for identifying right type of disease.

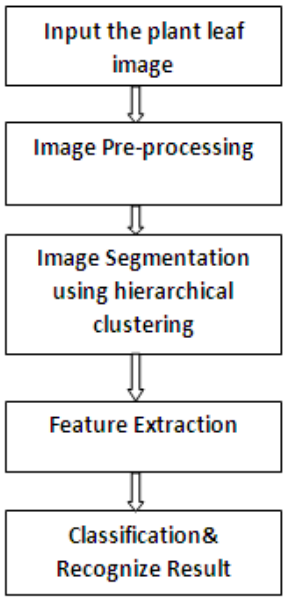


**Fig 3**. System Overview Architecture

### 4.1. Input the plant leaf image

**Figure 4.** Affected leaf

### 4.2. Image pre-processing

Image Pre-processing Noise gets added during acquisition of leaf images. So different types of filtering techniques are used to remove noise .Create device independent color space transformation structure. Thus 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. L*a*b is a device independent color space in which a & b components carry color information.

### 4.3. Image Segmentation

In this research hierarchical clustering is used for segmentation. Convert Image from RGB Color Space to L*a*b* Color Space. The L*a*b* space consists of a luminosity layer 'L*', chromaticity-layer 'a*' and 'b*'.All of the color information is in the 'a*' and 'b*' layers. Classify the colors in a*b* color space using hierarchical clustering. Since the image has 3 colors create 3 clusters. Measure the distance using Euclidean Distance Metric.

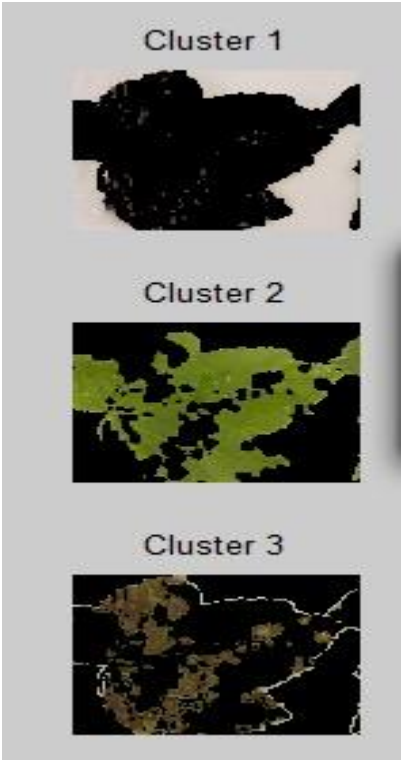


**Fig 5**. Image Segmentation

### 4.4. Feature Extraction

The diseases can be controlled by proper Disease management which is a challenging task. This challenge can be converted to easiest task by using image processing for detecting diseases of leaf, stem, root & fruit. With image processing it is possible to detect the affected area, type of disease &severity of the disease. Mostly diseases are seen on the leaves or stems of the plant. Because of the complexity of visual patterns of the diseases there has been increasing demand for development of more specific and sophisticated image pattern understanding algorithms which can be used for studies like classifying lesion, scoring quantitative traits, calculating area eaten by insects, etc. Now a days almost all of these tasks are processed annually or with distinct software packages. It is not only tremendous amount of work but also suffers from two major issues:

i) Excessive processing time and

ii) Subjectiveness rising from different individuals.

Hence to conduct high throughput experiments, plant biologist need efficient computer software to automatically extract and analyze significant features. As far as the leaf of the plant is considered, the significant features can be obtained by,

1. Color of the leaf
2. Texture of the leaf
3. Shape of the leaf

Color is one of the most widely used features. Color features can be obtained by various methods like Color histogram, Color correlogram, Color Moment, Color structure descriptor. The Color moment method has the lowest feature vector dimension and lower computational complexity. Hence it can be considered as suitable parameter to generate feature vectors which can be further used for classification purpose or for image retrieval.

### 4.5. Support Vector Machine

Support vector machine (SVM) is a non-linear classifier. The idea behind the method is to non-linearly map the input data to some high dimensional space, where the data can be linearly separated, thus providing great classification performance. Support Vector Machine is a machine learning tool and has emerged as a powerful technique for learning from data and in particular for solving binary classification problems. The main concepts of SVM are to first transform input data into a higher dimensional space by means of a kernel function and then construct an OSH (Optimal Separating Hyper Plane) between the two classes in the transformed space. For plant leaf classification it will transform feature vector extracted from leaf's contour. SVM finds the OSH by maximizing the margin between the classes. The SVM estimates a function for classifying data into two classes. Using a nonlinear transformation that depends on a regularization parameter, the input vectors are placed into a high-dimensional feature space, where a linear separation is employed.

### 4.6. Classification

A classification problem deals with associating a given input pattern with one of the distinct classes. Patterns are specified by a number of features (representing some measurements made on the objects that are being classified) so it is natural to think of them as d-dimensional vectors, where d is the number of different features. This representation gives rise to a concept of feature space. Patterns are points in this d-dimensional space and classes are sub-spaces. A classifier assigns one class to each point of the input space. The problem of classification basically establishes a transformation between the features and the classes. The optimal classifier is the one expected to produce the least number of misclassifications. In this research work SVM classification technique is used. SVM classifier is used

to identify the right type of disease affected on input leaf image.

## 5. CONCLUSION

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. Here 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.

## REFERENCES

**[1]** Aarti Sharma, Rahul Sharma,Vivek Kr. Sharma,Vishal Shrivatava "Application of Data Mining – A Survey Paper" Published in: International Journal of Computer Science and Information Technologies (IJCSIT) 2014, Vol. 5 (2) PP:2023-2025, Location: Jaipur Publisher:IEEE

**[2].** Pawan P. Warne, Dr. S. R. Ganorkar "Detection of Diseases on Cotton Leaves Using K-Mean Clustering Method" Volume: 02 Issue: 04 | July-2015.

**[3]** Ajayi Adebowale, Idowu S.A, Anyaehie Amarachi.A "Comparative Study of Selected Data Mining Algorithms Used For Intrusion Detection" International Journal of Soft Computing and Engineering (IJSCE) Volume-3, July 2013 Issue-3 PP:2231-2307, Location: Chennai .

**[4]** Amit Kumar Singh, Rubiya.A, B.Senthil Raja introduced their research in "Classification Of Rice Disease Using Digital Image Processing And Svm Classifier" International Journal of Electrical and Electronics Engineers PP ( IJEEE), Volume 07, 01, Jan- June 2015: 2321-2055 Location: Chennai.

**[5]** Bhushan R. Adsule, Jaya M. Bhattad "Leaves Classification Using SVM and Neural Network for Disease Detection" International Journal of Innovative Research in Computer and Communication Engineering IJIRCCE Vol. 3, Issue 6, June 2015 Location: Nagpur, Maharashtra

**[6]** Geraldin B. Dela Cruz, D. Gerardo, and Bartolome T. Tanguilig "Agricultural Crops Classification Models Based on PCA-GA Implementation in Data Mining" vol 05, 2014, PP.20ka65-2073.

**[7]** Hiary, Bani-Ahmad, Reyalat, Braik and ALRahamneh "Fast and Accurate Detection and Classification of Plant Diseases" International Journal of Computer Applications (0975 – 8887) Volume 17– No.1, March 2011 Publisher:IEEE

**[8]** Jagadeesh D. Pujari, Rajesh.Yakkundimath, Abdulmunaf "Automatic Fungal Disease Detection based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops" I.J. Image, Graphics and Signal Processing, : 2014.01.04 DOI: 10.5815 PP: 24-31 Location: Dharwar, INDIA Publisher: IEEE.

**[9]** Jagadeesh D. Pujari1, Rajesh Yakkundimath and Abdulmunaf S. Byadgi "Classification of Fungal Disease Symptoms affected on Cereals using Color Texture Features" International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.6, 2013, PP:6 Location: Dharwa

## BIOGRAPHIES

Vyshnavi.GKP[1] received the B.E. degree at the stream of Computer Science and Engineering from the Vivekanandha Institute of Engineering and Technology for Women, Tiruchengode, Anna University, Chennai, India, in 2014. Currently doing M.E. at the stream of Computer Science and Engineering in Institute of Road and Transport Technology, Erode, Anna University, Chennai, India. Her research interest includes in "Healthy and Unhealthy Plant Leaf Identification and Classification Using Hierarchical Clustering".

Sirpa.MR[2] received the B.E. degree at the stream of Computer Science and Engineering from the Excel Engineering College, Sankagiri, Anna University, India, in 2013. Currently doing M.E. at the stream of Computer Science and Engineering in Institute of Road and Transport Technology, Erode, Anna University, Chennai, India. Her research interest includes in "Healthy and Unhealthy Plant Leaf Identification and Classification Using Hierarchical Clustering".

Chandramoorthy.M[3] received the B.E. degree at the stream of Electronics and Communication Engineering from the Thangavelu Engineering College, Chennai, Anna University, India, in 2012. Currently doing M.E. at the stream of Computer Science and Engineering in Institute of Road and Transport Technology, Erode, Anna University, Chennai, India. His research interest includes in "Healthy and Unhealthy Plant Leaf Identification and Classification Using Hierarchical Clustering".

Padmapriya.B[4] received the B.E. degree at the stream of Computer Science and Engineering from the SNS College of Technology, Coimbatore, Anna University, Chennai, India, in 2014. Currently doing M.E. at the stream of Computer Science and Engineering in Institute of Road and Transport Technology, Erode, Anna University, Chennai, India. Her research interest includes in "Healthy and Unhealthy Plant Leaf Identification and Classification Using Hierarchical Clustering".