DATA MINING APPROACH FOR AUTOMATED DIAGNOSIS OF PLANT LEAF DISEASE USING IMAGE RECOGNIZING APPLICATION

S.Jawahar[1], S.Fayas Rahiman[2], L.Nithyan[3], D.Madhankumar[4]

[1]Assistant Professor, Department of Computer Science and Applications, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu.

[2,3,4]Student, Department of Computer Science and Applications, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu.

Mail id: [1]shivamjawahar@gmail.com, [2]fayasrahimans@gmail.com, [3]lakshminithyan@gmail.com, [4]kmadhan181@gmail.com

ABSTRACT

The purpose of the project entitled "Data Mining Approach for Automated Diagnosis of Plant Leaf Disease using Image Recognizing Application" is to identify the type of disease affected in plants. The identification of plant disease is very arduous in the agriculture field. Plant disease detection requires a sizably voluminous quantity of work, knowledge in the plant diseases, and withal require more processing time. The objective of this research is to develop an image recognizing application that identifies the diseases on the plant leaves and provides obviation methods. The application is developed utilizing MATLAB. The main function of this application is to load an image of a leaf into the GUI and the data mining technique is utilized to extract the diseased area and Support Vector Machine (SVM) classifier is utilized to relegate the leaf as an infected leaf or not. If the leaf is infected, then the application retrieves the information about the disease which is affected. This project will be very utilizable for farmers as it automatically identifies the disease affected in plants, thus they can utilize adequate magnitude of pesticides to efficaciously control the pests, in turn, the crop yield will be incremented.

Keywords: Data Mining, MATLAB, Support Vector Machine (SVM), Feature extraction.

1. INTRODUCTION

India is an expeditious developing country and agriculture is the backbone for the country's development in the early stages. Because of industrialization and globalization concepts, the field is facing hurdles. On top of that, the cognizance and the indispensability of the cultivation need to be instilled in the minds of the younger generation. Now a day's technology plays a vital role in all the fields but till today we are utilizing some old methodologies in agriculture. Identifying plant disease erroneously leads to immensely colossal loss of yield, time, money and quality of the product. Identifying the condition of the plant plays a paramount role in prosperous cultivation. The identification of plant disease is very arduous in the agriculture field. Leaf disease detection requires a sizably voluminous amount of work, erudition in the plant diseases, and withal require more processing time. The objective of this research is to make utilization of paramount features and prognostication is done utilizing the Data Mining technique and image recognizing.

Data mining is termed as extracting the pertinent information from the sizably voluminous pool of resources. The advents of data mining technologies have been adopted in the prognostication of plant diseases. Nowadays, technology is widely utilized for plant disease prognostication. The management of perennial leaf requires a close monitoring system especially for the diseases that affect engendering and post-harvest life [2].

The image recognizing follows the steps as pre-processing, segmentation, feature extraction, and classification. In the Classification system database is very paramount that contains predefined sample patterns of an object under consideration that compare with the test object to classify it, congruous class. Image recognizing is a paramount task in sundry fields such as biometry, remote sensing, and biomedical images. In a typical relegation system, the image is captured by a camera and consequently processed. In Supervised relegation, first of all, training took place through the kenned group of pixels. The trained classifier used to relegate other images. The Unsupervised relegation utilizes the properties of the pixels to group them and these groups are kenned as cluster and process are called clustering. The number of clusters is decided by users. When trained pixels are not available the unsupervised relegation is utilized. The various classification methods are: Decision Tree, Artificial Neural Network (ANN) and Support Vector Machine (SVM) [2].

2. RELATED WORKS

There is a plethora of research that has worked in the field of identification of plant diseases including Basavaraj Tigadi1 and Bhavana Sharma [3]. In this research, the researcher used Color Features and Artificial Neural Network for relegation a range of diseases affecting bananas. In this work, the first step is converting the image from RGB to gray and HSV color space then extracts the Histogram of template and color features. The researcher utilizes the color features including Mean and Standard Deviation. So as to utilize these features to engender a cognizance predicate that is utilized later by the classifier for training. utilize the victual-forward back propagation neural network to relegate the banana disease.

Plant disease detection and its solution using image classification by Saradhambal.G, Dhivya.R, Latha.S, R.Rajesh[1] is the research to predict the leaf disease using the K-means clustering algorithm. Otsu classifier and k-mean clustering algorithm are used to cluster the segmented images into different sectors. Afore clustering the images, the RGB color model is transformed into the Lab color model. The advantage of Lab color model is to easily cluster the segmented images. Shape and Textural features are extracted in feature extraction.

Leaf rot disease detection of Betel Vine additionally done by utilizing Color analysis [4]. In the pre-processing, the process of cropping was performed to eliminate the background containing dispensable information in the process of disease detection. The color feature is utilized to distinguish rotted leaf area form salubrious leaf area .the image is converted to three types of color space RGB, HSV and YCbCr next by utilizing color analysis the researcher find the hue of HSV give the best result the next step is utilizing hue thresholding for discriminating leaf rotted part of the rest of the background. Conclusively, convert the affected part to a binary image and calculate the white pixel to find the area of the affected part.

3. PROPOSED SYSTEM

In our proposed system we are providing a solution to recover from the leaf diseases and additionally show the affected part of the leaf by image processing technique. The subsisting system can only identify the type of diseases which affect the leaf. We will provide the accuracy of the disease affected in the leaf .

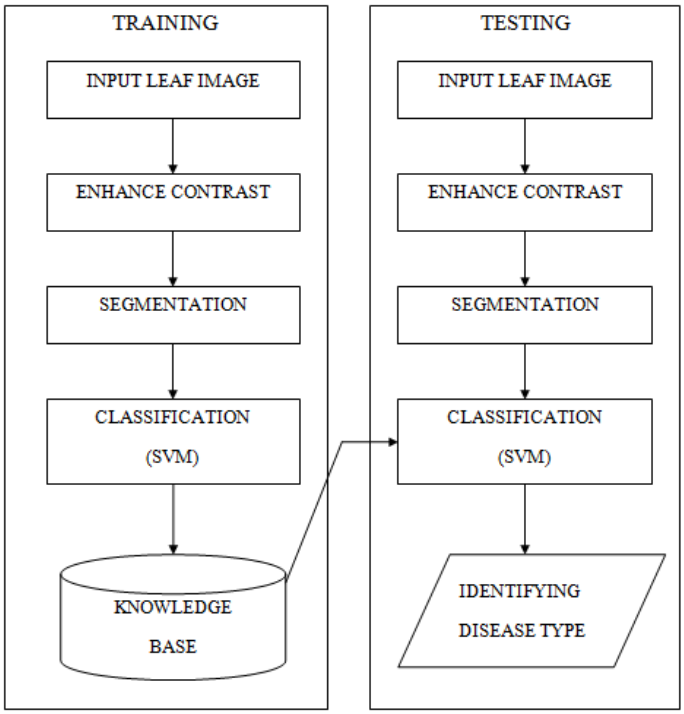

Figure 1: Block diagram of proposed system**.**

The first part of the proposed plant disease detection system is the training process. At this stage, the images of the plants are obtained utilizing a digital camera. Then the image preprocessing techniques are applied to these images. After preprocessing, utilizable image features are extracted utilizing the feature extraction technique that will be utilized as training samples for the support vector machine algorithm (the proposed machine learning algorithm in this system). In the detection phase, the images will be obtained first by capturing them with a digital camera. After that, the image processing techniques referred to in the training phase will be applied and determinately, the case will be relegated as either infected or salubrious through support vector machine (SVM).

4. DATASET

In this study, the images were accumulated from multiple sources. Due to the arduousness of conducting field visits as well because of seasonal conditions, the researcher was coerced to take the pictures from websites, mostly to a group of international universities. The images were amassed for three different diseases such as Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora Leaf Spot, as well as the health status of each of these crops.

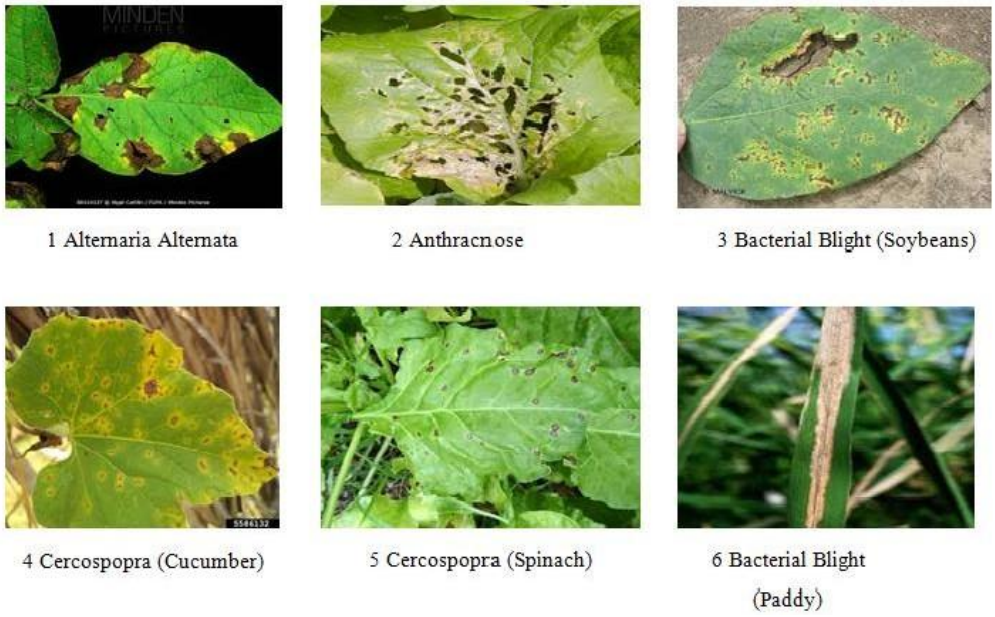

Figure 2: Sample Images from Dataset

Plant Diseases- Fundamentals

In the field of crop engendering, plant disease is a consequential factor that degrades the eminence and quantity of the plants. The mundane approach followed in plant diseases is the relegation and detection model. Both the relegation and detection model is widely studied by the Engineering and IT fields.

Table 1: No. of sample images and estimation of the infected region

Types of diseases	**No. of images**	**Area of infected region (%)**
Alternaria Alternata	22	15.0062
Anthracnose	23	15.0915
Bacterial Blight	6	13.0093
Cercospora leaf spot	9	18.2951

This table shows the types of plant diseases identified by this project. They are Alternaria Alternata, Anthracnose, Bacterial Blight and Cercospopra. The number of images for Alternaria Alternata is 22 and the percentage of the infected area is 15.00062% and for Anthracnose is 15.0915 for 29 images. In 6 images of Bacterial Blight, 13.0093% is an infected area. Out of 9 images of Cercospopra leaf spot, 18.2951% is an infected area.

5. IMAGE ACQUISITION

The process of image processing commences with the acquisition of the image from the environment through the digital camera and stored on the hard disk of the computer and then downloaded to the system for the rest of the operations.

6. ENHANCE CONTRAST

The image that is captured by a digital camera containing about 30% of the infected plant leaf information and the remaining 70% of the rest of the information is not consequential because it represents the background. This background is a nonessential consumption of recollection, and withal in the treatment time in the CPU during the process of retail segmentation In order to gain efficiency in the storage and speed up the processing time. It is paramount we deduct the portion of the image through a process of cropping. In this study, we utilize the command imadjust (I,stretchlim(I)) in Matlab that utilizes the interactive adjusting implement where (I) is the image we require to stretch. We would relish to point out that the process of enhancing must be precise and meticulous not to cut any consequential information from the image because in this type of system the precision is more consequential than the time. Resize all images that will be habituated to a fine-tuned size (300*400) using the command imresize(I4,[300,400]. This fine-tuned size was utilized for all imported images because the precision of the feature extraction process is affected if the images are in different sizes.

7. SEGMENTATION

The K-Means clustering algorithm tries to classify objects supported a collection of features into K number of classes. The classification is finished by minimizing the sum of squares of distances between the objects and also the corresponding cluster or class centroid . In our experiments multiple values of number of clusters are tested. Best results were observed when the amount of clusters was three. So, the image is partitioned into three clusters permanently segmentation result.

Figure 3: Segmented result of Aleternaria Alternata image

8. FEATURE EXTRACTION PROCESS

The image contains an abundance of information, only some of this information can be habituated to distinguish between different situations. so much of the information in the image must be converted to a minimized representation set called the features process that extracts features from the image called feature extraction.

The image has many features such as texture, color, and shape. These features can be utilized as mini-information representing the utilizable information in the image which can be habituated to distinguish between different situations. In this study, a range of texture features and color features were habituated to disunite the affected case from the healthy case, as well as to diagnose the sundry diseases. Shape features were not utilized in this study because the shape of the injury changes perpetually during the stages of disease magnification.

9. TRAINING THE SVM CLASSIFIER

Identification of Patterns utilizing a machine learning approach has two rudimentary stages. In the first stage, the classifier is trained to utilize the training samples to extract the weights. The system then examines the precision of the system utilizing the test samples. In this study, the total dataset was utilized for training the SVM classifier. The settings such as Kernel function: Quadratic, Box constraint level: 4.0. were utilized in SVM to get the best result.

10. RESULTS

A) Alternaria Alternata

It is a fungus which influences the leaf spots over 380 species of plant. It can too influence leaf spots, rots, blight, and other plant components. The fungus Alternaria cucumerina causes Alternaria leaf blight.

This disease is the most prevalent melon, but can additionally affect cucumber, pumpkin, and squash. Alternaria leaf blight does not commonly infect fruit. It can truncate yield and quality through truncated plant vigor and sunscald of exposed fruit.

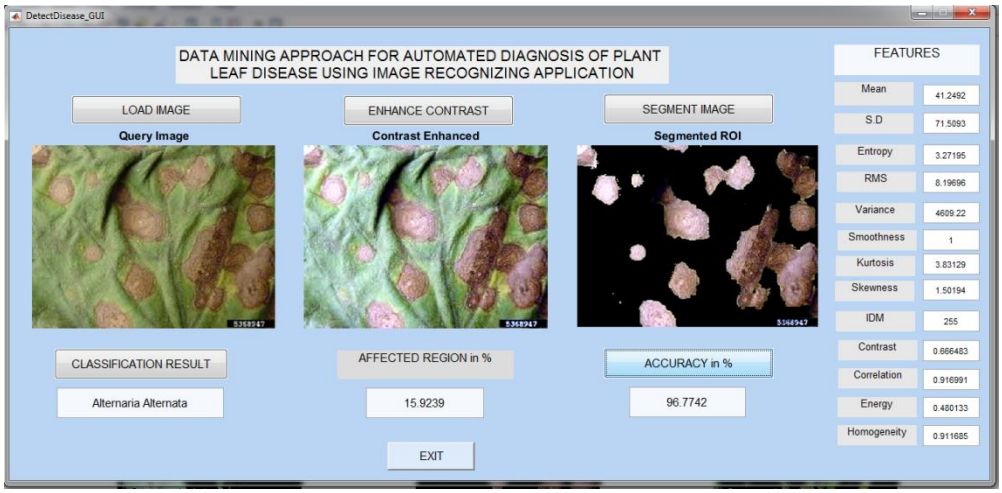

Figure 3: Result for Alternaria Alternata

B) Bacterial Blight

Bacterial Blight is characterized by minuscule, pale green spots or streaks that appeared as dihydrogen monoxide-marinated. The wound will expand then appear as dry dead spots. It may elongate until the full length of the leaf.

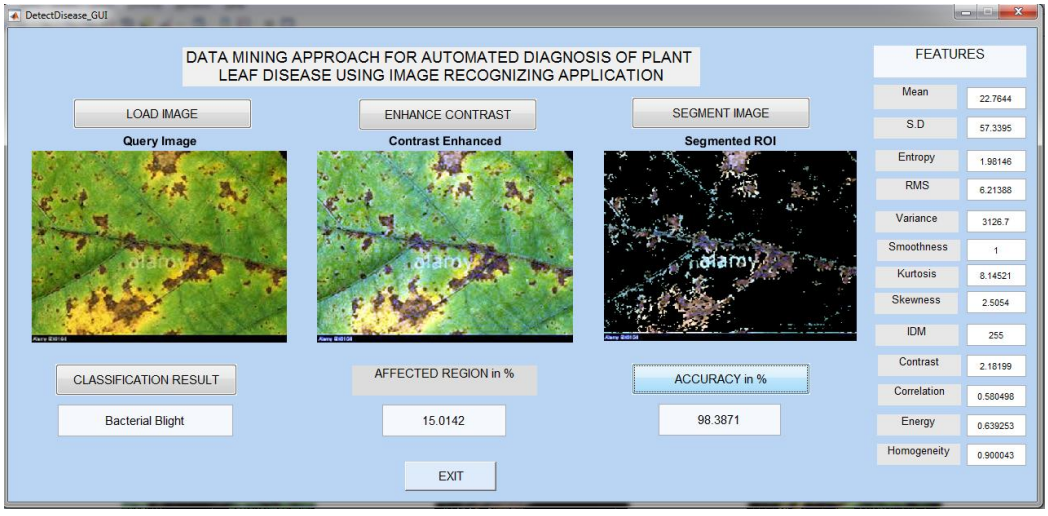

Figure 4: Result for Bacterial Blight

11. CONCLUSION

Data mining technologies have been implemented in the agriculture industry. This project implements an innovative conception to identify the affected crops and provide remedial measures to the agricultural industry. By the utilization of the SVM classifier, the infected region of the leaf is segmented and analyzed. The images are fed into our application for the identification of diseases. It provides a good cull for the agriculture community concretely in remote villages. It acts as an efficient system in terms of abbreviating clustering time and the area of the infected region. Feature extraction technique avails to extract the infected leaf and additionally to relegate the plant diseases. It was concluded that the precision of disease detection is incremented as the number of training samples increases and that the vicissitude in SVM settings additionally affects precision.

REFERENCE

Saradhambal.G , Dhivya.R, Latha.S, R.Rajesh, "Plant Disease Detection And Its Solution Using Image Classification", International Journal of Pure and Applied Mathematics ISSN: 1314-3395

Rama Gaur, Dr. V.S. Chouhan, "Classifiers In Image Processing" International Journal on Future Revolution in Computer Science & Communication Engineering ISSN: 2454-4248

Basavaraj Tigadi, Bhavana Sharma,"Banana Plant Disease Detection and Grading Using Image Processing", International Journal of Engineering Science and Computing IJESC,Volume 6 Issue No. 6

[4] S. Jawahar and Dr. P. Sumathi,” An Efficient K-mer counting and indexing method using Bloom filter for Biological DNA Sequences”, Journal of Web Engineering, Vol.18, Issue.4, May 2019, ISSN: 1540-9589.

[5]S. Jawahar and Dr. P. Sumathi,” A new method for detecting Fuzzy Tandem Repeats (FTR) using Levenshtein Distance for Biological data”, International Journal of Research in Engineering, IT and Social Sciences, Vol.9, Issue.2, February 2019, ISSN: 2250-0588.

[6[S. Jawahar and Dr. P. Sumathi,” Fast and accurate identification of Short Tandem Repeats (STRs) using hash function in DNA sequences”, International journal of Engineering and Technology (UAE), Vol.7, issue.4S2, December 2018, ISSN: 2277-3878. (SCOPUS)

[7] S. Jawahar and Dr. P. Sumathi, “The Risk Of Breast Cancer Associated With Brca1 And Brca2 Gene Genetic Mutation: A Review”, International Journal of Management, IT and Engineering, Vol.8, Issue.8, August 2018, ISSN: 2249-0558.

[8] S. Jawahar and Dr. P. Sumathi, “An Efficient Contiguous Pattern Mining technique to predict mutations in breast cancer for DNA data sequences”, International Journal of Pure and Applied Mathematics, Vol.119, No.15, July 2018, ISSN: 1314-3395. (SCOPUS)

[9] S. Jawahar and Dr. P. Sumathi, “A Survey on Sequential Generator Mining Algorithms”, International Journal of Research in Engineering, IT and Social Sciences, Vol. 7, issue.12, December 2017, ISSN:2250-0588.

[10] Amar Kumar Deya, Manisha Sharmaa, M.R.Meshramb, "Image Processing Based Leaf Rot Disease, Detection of Betel Vine", International Conference on Computational Modeling and Security. 7