

A Comparative Study of Algorithms used for Detection and Classification of Plant Diseases

Roshni C.R¹, Dr. M. Safish Mary²

¹M.Phil Scholar, Department of Computer Science, St. Xavier's College (Autonomous), Tirunelveli, India

²Assistant Professor, Department of Computer Science, St. Xavier's College (Autonomous), Tirunelveli, India

Abstract: *Our Country's economy prospect lies mainly in agricultural sector. Although there is much advancement in technology, still chances of predicting the diseases in plants are vague. In this paper, a technical solution for the farmers to detect and diagnose the right disease affecting the plants is discussed. The Content Based Image Retrieval (CBIR) technique is used to retrieve the images of diseased plant from the training dataset based on a query image. The images thus retrieved are segmented using Hierarchical Clustering which produces cluster of diseased plant images. The clusters are then classified using Support Vector Machine (SVM) classifier based on the features extracted from clusters which verifies correct type of disease affecting the plant set.*

Keywords: CBIR, Hierarchical Clustering, Segmentation, Feature Extraction, SVM.

1. Introduction

Agriculture is the backbone of our country. Many farmers are still struggling hard in detecting diseases which are affecting plants, vegetation crops and cereals. Though experts are available to detect the diseases, still prediction by naked vision may not be correct every time. So it will be effective to have an automated expert system. Even though research is done periodically to predict the disease in plants, accuracy in finding the right disease and diagnosis are still not perfect. In this paper, accuracy is the main performance measure to be concentrated in detecting the right type of disease in plants.

Dataset required for this work is collected from Online Data Repository [1]. The diseased leaves of Tomato plant are taken as the dataset here. Fungal, bacterial and viral diseases that affect the Tomato plant are studied. This paper concentrates on the diseases like Septoria Leaf Spot, Fusarium, Verticillium Wilt, Mosaic Virus and Bacterial Spot. These diseases are diagnosed by experts with naked eye observation which is time consuming. Nowadays there are relatively few expert level consultants available and it is very expensive to get their consultancy. Since the plant diseases degrade the quality and quantity of the plant, it is mandatory to use the technology for automatic detection and diagnosis of diseases.

The purpose of this paper is to discuss various techniques for image retrieval, image segmentation and classification. The paper is organized into the following sections. Section 2 describes the disease detection methodology. In Section 3 a review of the existing literature is given in detail. Section 4 covers various algorithms for disease detection in plants and a comparative study of these algorithms performance is tabulated. Finally section 5 covers conclusion made from the comparative study.

2. Disease Detection Methodology

In this section, the basic idea of detection and diagnosis of the plant disease is discussed. In Figure 1, the proposed methodology for detection and classification of plant diseases is given.

2.1 Image Acquisition

Images of normal and infected plant leaves are taken from the online repository and from nurseries nearby and the Image Dataset is created.

2.2 Image Preprocessing

To remove noise from an image or for any unwanted object removal, there are various techniques available. Among the various filtering techniques used, Median Filter is found to produce better noise removal results in plant leaves.

2.3 CBIR for Image Retrieval

Content Based Image Retrieval (CBIR) system mainly uses color, shape, texture as their basic features. An image containing these basic features is set as the query image. The query image is then compared with the image in the dataset and the images whose features match exactly with that of the query image are retrieved. The obtained images are then stored in a separate database for further proceedings. These images are used as the training dataset which is the best suited diseased image set to be used for disease prediction and diagnosis. The process of image retrieval done using CBIR technique [2] is explained as follows:

2.3.1 Color Retrieval

When an image is used for comparison, it is first preprocessed and color histogram is derived from the image. The histogram shows the proportion of colors in each pixel of the image. Images whose color matches closely with that of the query image are considered for retrieval.

2.3.2 Texture Retrieval

Texture similarity is done by extracting the texture defining values from the query image. The values are calculated on the basis of scale, degree of contrast and directionality periodically for the texture analysis.

2.3.3 Shape Retrieval

The global features of shape like aspect ratio, circularity, moment invariants and local features like sets of successive boundary segments are extracted from the query image. These features are compared with the features obtained from the dataset images and those matching with the query image are retrieved. The images thus obtained from the CBIR techniques are separately stored in a database.

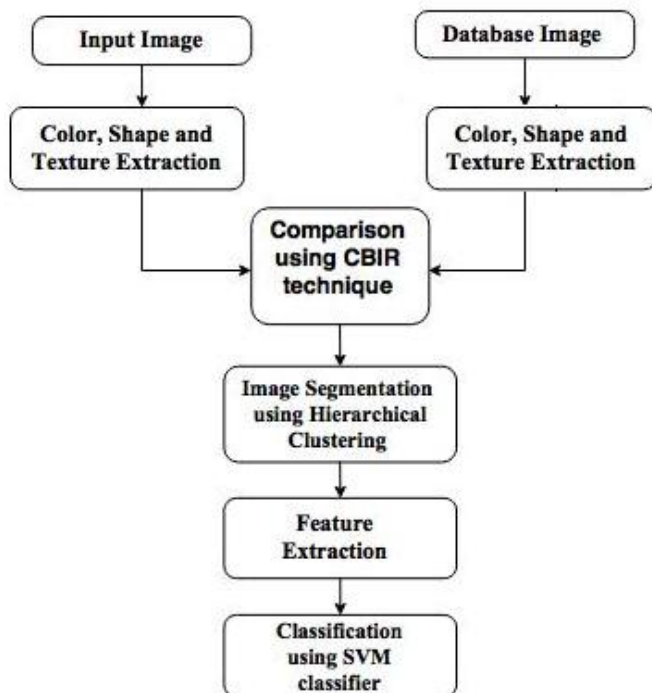


Figure 1: Disease detection and diagnosis methodology

2.4 Image Segmentation

The images in the database are grouped into clusters using Hierarchical Clustering algorithm. Many cluster based on diseased and healthy leaves are produced. Since the segmentation process segments the image into multiple parts, the infected regions are easy to visualize [3].

2.5 Feature Extraction and Classification

There are 13 features that are considered for classification to evaluate the disease affected region in the leaves which are contrast, correlation, energy, homogeneity, mean, standard deviation, entropy, RMS (Root Mean Square), variance, smoothness, kurtosis, skewness, IDM (Inverse Difference Moment). The resultant image thus obtained based on the extracted features are fed into the SVM (Support Vector Machine) classifier which diagnose the right type of disease [3].

3. Literature Review

In [2], the authors presented an idea of mobile phone approach. Image captured using mobile phone is used as the query image whose features are compared with that of the database image using CBIR technique. Images are retrieved mainly based on three features like color, texture and shape. The resultant images are segmented using k-means clustering and Euclidean Distance measure is calculated to find the nearest matching image to the query image and the result is returned to the user via android phones.

In [3], the authors presented the prediction of leaf diseases based on Hierarchical clustering, k-means clustering and Fuzzy C- Mean (FCM) clustering. The clustering techniques produced many diseased and undiseased clusters. In that Hierarchical clustering produced better accuracy in terms of performance. Then the features are extracted for disease classification using SVM classifier.

In [4], the authors had taken the approach of CBIR Technique, which is used for content based retrieval of images. Both the query image and database image are extracted based on color, shape and texture and are stored in a separate database which has undergone similarity measurement using Euclidean distance. The output images are retrieved based on the threshold value.

In [5], the authors discussed the segmentation of plant leaf images using k-means clustering technique to find the affected regions. Once the affected regions are found, the color and texture features are extracted from it using the color co-occurrence matrix and special gray level dependency matrix after being converted from RGB to HIS pixel value to contribute as discriminating feature for classification. Then the classifiers ANN and SVM are used for disease classification. The extracted features are used to train SVM and ANN. The test images are then used to recognize and classify the image samples. The SVM achieves a significant improvement in classification accuracy over ANN and proved to be a powerful tool for automatic classification of fungal affected symptoms.

In [6], Revathi and Hemalatha discussed about the cotton leaf plants with six diseases. After the features are extracted the skew divergence color variance is calculated by color histogram and color descriptor. The skew divergence shape feature is calculated by the Sobel and Canny edge detection method and finally texture is calculated by Gober filter and texture descriptor. Then feature selection is done by PSO (Particle Swarm Optimization) technique to analyze the best matching image of affected leaf. These features result in optimal solution for disease detection. The classification is done using CIG-DFNN (Cross Information Gain - Deep Forward Neural Network). In its first phase, the parameters are initialized and training data is fed to the input layer and is processed in the hidden layer producing the classes as output. In the second phase, the parameter value is fine-tuned with gradient descent and back propagation method. In DFNN, CIG rate is performed to diminish the error rate of hidden and input vector layer. Its advantages are accurate

classification of diseases by increasing the accuracy rate and diminish the error rate of layers.

In [7], the authors proposed a pre-processing technique to classify cotton, orange and lemon leaves diseases. They preprocessed the images using median filtering. The grayscale images are used for histogram distribution. In the histogram, segmentation is done using global thresholding by converting grayscale image into binary image. In Global Thresholding the value is selected by separating leaf image peak intensity into two classes namely background and foreground while minimizing the overlap between clusters. Classification is done by SVM and Neural Network Pattern Recognition. In SVM the accuracy of normal and abnormal leaves are tested and in Neural Network Pattern Recognition, with 250 neurons in hidden layer is proposed. The plots of confusion matrix for each training, validation, testing and overall performance are developed. It consists of TPR (True Positive Rate), TNR (True Negative Rate) and overall accuracy rate.

In [8], Annesha and Joydeep presented a new multi-feature image clustering technique which helps to classify large volumes of data with high accuracy. Images are retrieved based on feature extraction methods. Here there are three approaches as color extraction based on color moments and histogram and edge detection method. The three ID array created is merged into Single ID array. This is done for all images. Then k-means clustering is used to cluster data objects which produce four clusters. To evaluate the clusters, Euclidean distance metrics is used.

In [9], the author dealt with the classification of leaf diseases using SVM classifier. The diseased images are preprocessed with the Gaussian filter and are segmented using k-means clustering which produced three clusters. Euclidean distance measure is calculated. Nine texture and color features are calculated for all segmented parts of single leaf image. The feature values are given to train SVM classifier which classifies the diseases.

4. Comparative Study

Table 1 shows the overview of all methods that are used in different papers, together with the various methods they used for the detection and classification of diseases found in plants, the accuracy obtained in using each method. The advantages of using these methods as discussed in the referred papers are also tabulated.

The graph in figure 2 shows the comparative results of the widely used algorithms for plant disease detection in terms of the accuracy achieved in the related work carried out by the researchers. From the graph it is very clear that hierarchical clustering and SVM classifier produce better accuracy in classifying plant disease detection and diagnosis. Hence a combination of these two approaches will surely improve the plant disease detection technique.

Table 1: A comparative study of algorithms and their advantages

Paper Reference	Algorithm Used	Accuracy Rate (%)	Advantages
[3]	K-Means Fuzzy C Means Hierarchical	75.86 80.05 92.72	Valuable Less Effort Accuracy
[4]	CBIR Thresholding	-	Faster Output
[5]	SVM ANN	83.83 77.75	Comparison results shows SVM is better
[6]	Feature Selection with PSO CIF-DFNN	95	Accurate Diminishes Error Rate High Performance
[7]	SVM Neural Network Pattern Recognition	70.21 96.27	Larger datasets are used More Features are extracted
[8]	K-Means	90.50	Complexity Decreases
[9]	K-Means SVM	88.89	Normal & abnormal leaves are studied Confusion matrix is plotted

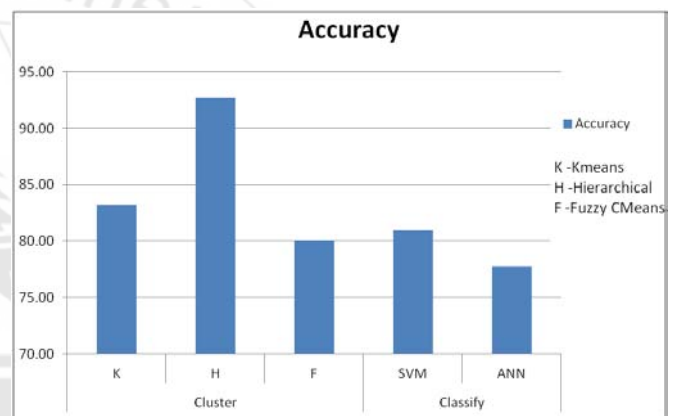


Figure 2: Comparison of most frequently used algorithm's overall accuracy rate in percentage

5. Conclusion

The CBIR technique used to retrieve diseased leaf images yields better result in image extraction. From the comparative study, it is clear that there are a number of segmentation and classification algorithms used by researchers for plant disease detection and diagnosis. So this paper proposes a combination of clustering and classification algorithms for disease detection and diagnosis in plants which will result in better accuracy. It is proposed to retrieve diseased images by implementing CBIR technique and segmenting the retrieved diseased leaf image by hierarchical clustering technique to predict the disease present in the leaf and SVM classifier to diagnose the right type of disease to produce more accurate results.

References

- [1] <https://www.plantvillage.org/>
- [2] A.S Deokar, Akshay Pophale, Swapnil Patil, Prajakta Nazarkar, Sukanya Mungase, "Plant Disease Identification using Content Based Image Retrieval Techniques Based on Android System", International

- Advanced Research Journal in Science, Engineering and Technology Vol. 3, Issue 2, February 2016
- [3] N. Swetha, Mrs. Dr. N. Sasirekha, "Prediction of Leaf Disease using Segmentation with Hierarchical Clustering", International Conference on Recent Innovation in Science, Technology and Management (ICRISTM-16) at Indian Federation of United Nations Associations (IFUNA), New Delhi, India, ISBN: 978-81-932712-3-0. 12th June 2016
- [4] Parash Mehta, Lavanya Saoji, Ankit Dodrajka, Tushar Chug, "Detection of Diseases on leaves and its possible diagnosis Using CBIR technique", International Education and Research Journal (IERJ) E-ISSN No: 2454-9916 Volume: 2 Issue: 2 Feb 2016
- [5] Jagadeesh D. Pujari, Rajesh Yakkundimath and Abdulmunaf S. Byadgi, "Classification of Fungal Disease Symptoms affected on Cereals using Color Texture Features", ICICES2014 - S. A. Engineering College, Chennai, TamilNadu, India J. Gerald's, "Sega Ends Production of Dreamcast," vnunet.com, para. 2, Jan. 31, 2001
- [6] P.Revathi and M.Hemalatha, "Identification of Cotton Diseases Based on Cross Information Gain Deep Forward Neural Network Classifier with PSO Feature Selection", International Journal of Engineering and Technology (IJET), Department of Computer Science Karpagam University
- [7] 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
- [8] Annesha Malakar, Joydeep Mukherjee, "Image Clustering using Color Moments, Histogram, Edge and K-means Clustering", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064
- [9] Pranjali B. Padol, Prof. Anjali A.Yadav, "SVM Classifier Based Grape Leaf Disease Detection", Conference on Advances in Signal Processing (CASP) Cummins College of Engineering for Women, Pune. Jun 9-11, 2016
- [10] Zalak R. Barot, Narendrasinh Limbad, "An Approach for Detection and Classification of Fruit Disease: A Survey", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14

Author Profile



Roshni CR is presently doing her M. Phil degree in St. Xavier's College. She finished her BSc in Computer Science from Rose Mary College and MCA degree from Manonmaniam Sundaranar University in the year 2011 and 2014 respectively. She is pursuing her thesis work in image segmentation and classification.



M. Safish Mary received her Ph. D degree in Computer Applications and M. Phil degree in Computer Science from Manonmaniam Sundaranar University, Tirunelveli, India. She is presently working as Assistant professor at St. Xavier's College (Autonomous), Palayamkottai, India. She has a vast teaching experience of about 18 years and research experience of about 10