Diagnosis of Grape Leaf Diseases Using K-Means Clustering and Neural Network

S.Sankareswari,
Dept of Computer Science and Engineering,
National Engineering College,
Kovilpatti, Tamil Nadu, India
sivi.ranjani22@gmail.com

G.Sivakamasundari,
Assistant Professor,
Dept of Computer Science and Engineering,
National Engineering College,
Kovilpatti, Tamil Nadu, India
gssundari1976@gmail.com

Abstract— Plant diseases cause significant damage and economic losses in crop. Subsequently, reduction in plant diseases by early diagnosis results in substantial improvement in quality of the product. The goal of proposed work is to diagnose the disease in grape leaf using image processing. First the color transformation of grape leaf from RGB to LAB color space is done. Then segmentation is done using k-Means clustering. Useful i.e. diseased segment is identified and texture features are extracted from GLCM matrix of the diseases segment. At last, the classification is done using Feed Forward Neural Network (FFNN).

Keywords: K-means, GLCM, Feed Forward Neural Network

I INTRODUCTION

Plant diseases cause significant reduction in both quality and quantity of agricultural products. 70% of the Indian population depends on agriculture. Farmers have wide range of diversity to select suitable Leaf crops. But the cultivation of these crops for optimum yield and quality produce is highly technical. It can be improved by the aid of technological support. Many authors have worked on the development methods for the automatic detection and classification of leaf diseases. The philosophy behind precision agriculture is not only including a direct economical optimization of agricultural production, it also stands for a reduction of harmful outputs into environment and non-target organisms.

This an automatic diseases identification system is proposed. Depending on the applications, many systems have been proposed to solve or at least to reduce the problems, by making use of image processing, pattern recognition and some automatic classification tools. In the next section this paper tries to present those proposed systems in meaningful way.

II LITERATURE SURVEY

[1].Paul Arora, N.Ghaiwat, March 2014.The purpose of Agriculture is not only to feed ever growing population but it's an important source of energy and a solution to solve the problem of global warming. Plant diseases are extremely significant, as that can adversely affect both quality and quantity of crops in agriculture production. Plant disease diagnosis is very essential in earlier stage in order to cure and control them. Generally the naked eye method is used to identify the diseases. In this method experts are involved who have the ability to detect the changes in leaf color. This method involves lots of efforts, takes long time and also not practical for the large fields. Many times different experts identify the same disease as the different disease. This method is expensive as it requires continuous monitoring of experts. Depending on the applications, many systems have been proposed to solve or at least to reduce the problems, by making use of image processing, pattern recognition and some automatic classification tools. In the next section this paper tries to present those proposed systems in meaningful way. [2]S.Mohana Valli, S.Raju, Vol 2, Issue No 2, May 2011.This Project searches a very huge image databases. Botanists are usually brought to use large collections of plants images. They need automatic tools to assist them in their work. This paper presents a plant retrieval system which takes as input the image of a plant and returns the most similar images from a database. The system is intended to be used as an e-commerce service where users can send images of their house plants (which they often do not know by name) to find their Latin names and care instructions. The problem involves identification of the matching plant, as well as retrieval of related varieties which may be also of interest to the user.[3] Malvika Ranjan, Neha Joshi, Volume 3, Issuse No 3, June 2015.Plant disease diagnosis is an art as well as science. The diagnosis process (i.e. recognition of symptoms and signs), is inherently visual and requires intuitive judgment as well as the

use of scientific methods. The work begins with capturing the images. Color feature like HSV features are extracted from the result of segmentation and Artificial neural network (ANN) is then trained by choosing the feature values that could distinguish the healthy and diseased samples appropriately. Experimental results showed that classification performance by ANN taking feature set is better with an accuracy of 80%. The present work proposes a methodology for detecting cotton leaf diseases early and accurately, using diverse image processing techniques and artificial neural network (ANN). [4]H.Kulkarni,K.R.Ashwin Patil ,Vol 2,Issue No 5,Oct 2012.Agriculture is the mother of all cultures. It has played a key role in the development of human civilization. Agricultural practices such as irrigation, crop rotation, fertilizers, and pesticides were developed long ago, but have made great strides in the past century. By the early 19th century, agricultural techniques had so improved that yield per land unit was many times that seen in the middle ages. Therefore, judicious management of all the inputs is essential for the sustainability of a complex system. The focus on enhancing the productivity, without considering the ecological impacts has resulted into environmental degradation. Without any adverse consequences, enhancement of the productivity can be done in a sustainable manner. Plants exist everywhere we live, as well as places without us. Many of them carry significant information for the development of human society. As diseases of the plants are inevitable, detecting disease plays a major role in the field of Agriculture. Plant disease is one of the crucial causes that reduces quantity and degrades quality of the agricultural products.[5] S.Arivazhagan, R.Newin Shabiah, Volume 15, Issuse No 1, March 2013.Plant diseases have turned into a dilemma as it can cause significant reduction in both quality and quantity of agricultural products. Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. The proposed system is a software solution for automatic detection and classification of plant leaf diseases. The developed processing scheme consists of four main steps, first a color transformation structure for the input RGB image is created, then the green pixels are masked and removed using specific threshold value followed by segmentation process, the texture statistics are computed for the useful segments, finally the extracted features are passed through the classifier. The proposed algorithm's efficiency can successfully detect and classify the examined diseases with an accuracy of 94%. Experimental results on a database of about 500 plant leaves confirm the robustness of the proposed approach.

III Proposed System

The proposed system automates the detection of disease in lemon leaf in order to find out the diseases with ease so that farmers need not check manually and randomly. Image pre-processing is a task of leaf image which is in RGB color format is converted to a Lab. Image based on segmentation is adapted and the binary image is obtained and it can be performed by using K- Means Clustering.The image analysis focuses on the leaf texture Feature using GLCM. Classification is done using Feed Forward Neural Network algorithm.Image are classified by comparing each image with a particular threshold range. if the disease is detected the name of the disease.This works reduce the man force.

IV System Design

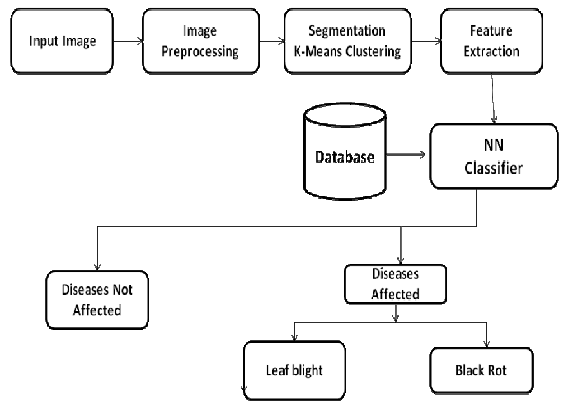

Figure 1

A.Image Conversion

Sample grape leaf images are taken from plant village website. In image preprocessing, RGB image is converted into LAB image. LAB is used for color spacing in defected leaf like Luminance at different position and Brightness of different spots on leaf. A and B are chromatic components.

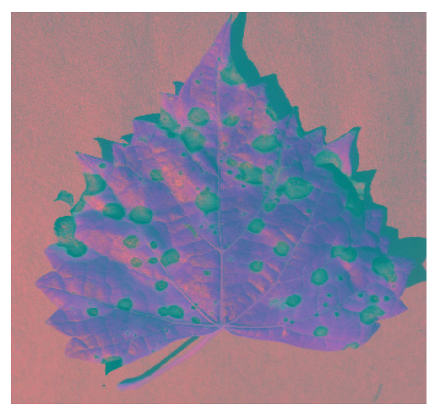

Figure 2: LAB image

B. Segmentation

The segmentation based on K-means technique is a partition clustering technique used to partition n number of observations into k clusters .In this technique, k is the number of clusters in the segmented image and colors present in an image are used for the clustering. The main advantage of segmentation based K-means clustering technique is that it works on local information and global information of image. K-means clustering algorithm is easy to implement, fast and flexible than others. From segmented image extracts the features of image such as texture feature.

K-means clustering algorithm:

- The **k-means algorithm** is an algorithm to cluster n objects based on attributes into k partitions, where $k < n$.
- It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data.
- It assumes that the object attributes form a vector space.
- An algorithm for partitioning (or clustering) N data points into K disjoint subsets S_j containing data points so so as to minimize the sum-of-squares criterion

$$\mathbf{j} = \sum_{j=1}^{K} \sum_{n \in S_j} |x_n - \mu_j|^2$$

- where x_n is a vector representing the n^{th} data point and u_j is the geometric centroid of the data points in S_j.
- Simply speaking k-means clustering is an algorithm to classify or to group the objects based on attributes/features into K number of group.
- K is positive integer number.
- The grouping is done by minimizing the sum of squares of distances between data and the corresponding cluster centroid.

C Feature Extraction

In 1973, Haralick introduced the co-occurrence matrix and texture features which are the most popular second order statistical features today. Haralick proposed two steps for texture feature extraction. First step is computing the co-occurrence matrix and the second step is calculating texture feature based on the co-occurrence matrix. This technique is useful in wide range of image analysis applications from biomedical to remote sensing techniques.

Contrast

Returns a measure of the intensity contrast between a pixel and its neighbor over the whole image. Range = [0 (size (GLCM, 1)-1) ^2] Contrast is 0 for a constant image.

$$\textbf{Contrast} = \sum_{i,j=0}^{N-1} (i , j) (i , j)^2$$

Energy

Returns the sum of squared elements in the GLCM. Range = [0 1] Energy is 1 for a constant image.

$$\textbf{Energy} = \sum_{i,j=0}^{N-1} (i,j)^2$$

Homogeneity

Returns a value that measures the closeness of the distribution a of element in the GLCM to the GLCM diagonal. Range = [0 1] Homogeneity is 1 for a diagonal GLCM.

$$\textbf{Homogeneity} = \sum^{N-1} C (i,j)/(1+(i -j)^2$$

Correlation:

Returns a measure of how correlated a pixel is to its neighbour over the whole image. Range = [-1 1] Correlation is 1 or -1 for a perfectly positively or negatively correlated image. Correlation is Nan for a constant.

$$\textbf{Correlation} = \sum i,ji-\mu i)(j-\mu j4 \quad (i,j) / \sigma i\sigma ji,j$$

Inverse Difference Moment

Inverse Difference Moment (IDM) is a measure of image texture as Defined.IDM is usually called homogeneity that measures the local homogeneity of an image. IDM feature obtains the measures of the closeness of the distribution of the GLCM elements to the GLCM diagonal. IDM has a range of values so as to determine whether the image is textured or non-textured.

$$\textbf{IDM} = \sum_{I-0}^{N-1} \sum_{j-0}^{N-1} \frac{1}{1+(i-j)^2} M(i, j)$$

D Classification

Neural networks are used in the automatic detection of leaves diseases. Neural network is chosen as a classification tool due to its well-known technique as a successful classifier for many real applications. The training and validation processes are among the important steps in developing an accurate process model using NNs. The dataset for training and validation processes consists of two parts; the training feature set which are used to train the NN model; while a testing features sets are used to verify the accuracy of the trained NN model. Before the data can be fed to the ANN model, the proper network design must be set up, including type of the network and method of training. This was followed by the optimal parameter selection phase. However, this phase was carried out simultaneously with the network training phase, in which the network was trained using the feed-forward network. In the training phase, connection weights were always updated until they reached the defined iteration number or acceptable error. Hence, the capability of ANN model to respond accurately was assured using the Mean

Square Error (MSE) criterion to emphasis the model validity between the target and the network output.

V. Experimental Results

The data consisted of 25 images of block rot (class 1) and 10 images of leaf spot (class 2). MATLAB 7.1 Neural network pattern recognition tool was used for training. In 35, 29 images are used for training and 2 each for testing and validation. Data was given in two files namely input and target. Input file with 7rows representing texture features and 35 columns representing sample images.

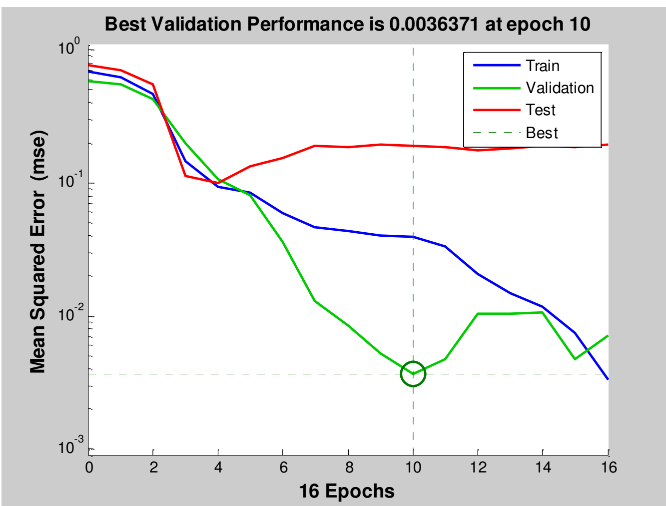

Figure 3 Validation results after training

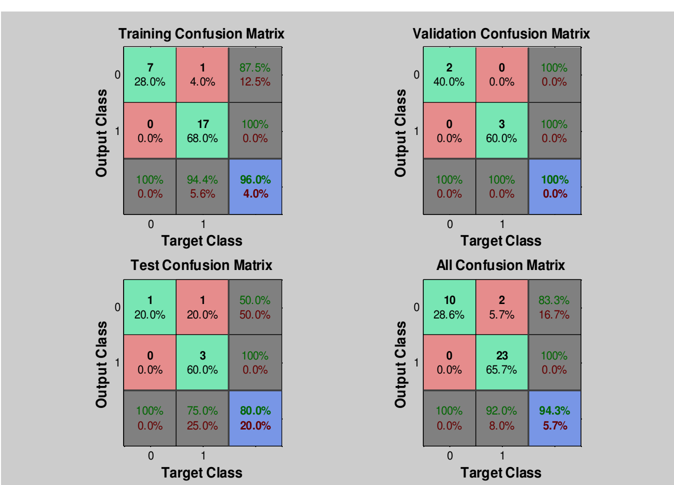

Figure 4 Sensitivity and specificity of system and confusion matrix

Target file consisted two rows with binary values. [0 1] for block rot and [1 0] for healthy. It gives good validation results as shown in figure 3. Confusion matrix in figure 4 shows sensitivity (True positive) and specificity (False positive) rate of the system. During training it gave 94.3% correct results which shows the system will work almost accurately.

The model that used hue features gives accurate results reaching a perfect 94.3%.

VI Conclusion and Future Work

Study involved collecting leaf samples from different regions.

Work was carried out to investigate the use of computer Vision for classifying grape leaf diseases. Two classes Of grape leaves, i.e., block rot and leaf spot were considered in the experiments. orithms based on image-processing techniques, feature extraction and classification

VII. References

1) Detection and Classification of Plant Leaf Diseases Using Image processing Techniques, Department of Electronics and Telecommunication Engineering, Wagholi, Pune, Publication Mar, 2014.

2) Applying image processing technique to detect plant diseases, department of Computer Science and Engineering, Gogte Institute Of Technology, Belgaum, Karnataka, India Publication Sep-Oct. 2012

3)Content Based Leaf Image Retrieval Using Shape, Color and Texture Features, Department of Electronics And Communication Engineering, Thiagarajar College Of Engineering,Madurai,Tamil Nadu, India, May 2011

4)Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features, Department of Electronics and Communication Engineering, Mepco

Schlenk Engineering College, Sivakasi Tamilnadu, 626 005, India, March, 2013

5) Detection and Classification of Leaf Disease Using Artificial Neural Network, SAE, Kondhwa, Dept. of E&TC. Engineering, Pune University, Pune, India, June 2015

6) Leaf Disease Grading by Machine Vision and Fuzzy Logic, Int. J. Comp. Tech. June 2011.

7) Grey level co occurrence matrices: generalization and some new features, International Journal of Computer Science, Engineering and Information Technology, 2012.