Apple Leaf Disease Identification using Support Vector Machine

S.Manimegalai
Department of Computer Science and Engineering
National Engineering College
K.R.Nager,Kovilpatti 628503

G.SivagamaSundari
Assistant Professor
Department of Computer Science and Engineering
National Engineering College
K.R Nager,Kovilpatti 628503

Abstract— **Agriculture is the mainstay of the Indian economy. Almost 70% people depend on it & shares major part of the GDP. Diseases in crops mostly on the leaves affects on the reduction of both quality and quantity of agricultural products. Perception of human eye is not so much stronger so as to observe minute variation in the infected part of leaf. In this paper, we are providing software solution to automatically detect and classify apple leaf diseases. Image processing contains the preprocessing of the plant leaf as color extraction, segmentation and classification of an image. Apple leaf can be classified based on their color and texture feature with help of SVM classifier. The texture feature is extracted from GLCM and the color feature extracted from color moments. This design and implementation of these technologies will greatly aid in selective application, reducing costs and thus leading to improve productivity, as well as improved produce.**

Keywords—HSV Transformation,K-means Clusteing,GLCM, SVM classifier.

I. INTRODUCTION

Agriculture has played a key role in the development of human civilization. If there is decrease in agro products, total economy will get affected. Therefore judicious management of all input resources such as soil, seed, water, fertilizers etc. is essential for sustainability. As diseases are inevitable, detecting them plays major role. One can refer incident that occurred in 2007, Georgia (USA), it is estimated that approximately 539 USD was the loss incurred due to plant diseases as well as controlling them. The naked eye observation of farmers followed by chemical test is the main way of detection and classification of agricultural plant diseases. In developing countries, farming land can be much larger and farmers cannot observe each and every plant, every day. Farmers are unaware of non-native diseases. Consultation of experts for this might be time consuming & costly. Also unnecessary use of pesticides might be dangerous for natural resources such as water, soil, air, food chain etc. as well as it is expected that there need to be less contamination of food products with pesticides. There are two main characteristics of plant disease detection machine-learning methods that must be achieved, they are: speed and accuracy. There is need for developing technique such as automatic plant disease detection and classification using leaf image processing techniques. This will prove useful technique for farmers and will alert them at the right time before spreading of the disease over large area. In the first step we create a color transformation structure for the RGB leaf image and then, we apply color space transformation for the color transformation structure. Then unnecessary part (green area) within leaf area is removed. Then image is segmented using the k-means clustering technique. Next we calculate the texture features and color feature for the segmented infected object. Finally the extracted features are passed through a classification using support vector machine classifier ease of use and the performance of the classifier to calculate precision and recall.

II. PROPOSED SYSTEM

First the images of apple leaves are acquired using high resolution camera so as to get the better results & efficiency. Then image processing techniques are applied to these images to extract useful features which will be required for further analysis.

The basic steps of the system are summarized as

1. Create Color Transformation.
2. Masking and Removing of Green pixels.
3. Apply K-means clustering for further segmentation.
4. Calling GLCM and Color Moments for Feature Extraction.
5. Classify using Support Vector Machine.

2.1 Color Transformation

First, the RGB images of leaves are converted into Hue Saturation Intensity (HSV) color space representation. The purpose of the color space is to facilitate the specification of colors in some standard, generally accepted way. HSV (hue, saturation, value) color model is a popular color model because it is based on human perception. Hue is a color attribute that refers to the dominant color as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and value refers to the amplitude of the light.

2.2 Masking and Removing Green Pixels

In this step, we identify the mostly green colored pixels. After that, based on specified threshold value that is computed for these pixels, the mostly green pixels are masked as, if the green component of the pixel intensity is less than the pre-computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that the green colored pixels mostly represent the healthy areas of the leaf and they do not add any valuable weight to disease identification and furthermore this significantly reduces the processing time.

2.3 Image Segmentation using K-means Clustering

Image segmentation is the process used to simplify the representation of an image into something that is more meaningful and easier to analyze. K-means clustering is a partitioning method. The function 'kmeans' partitions data into k mutually exclusive clusters, and returns the index of the cluster to which it has assigned each observation. Unlike hierarchical clustering, k-means clustering operates on actual observations (rather than the larger set of dissimilarity measures), and creates a single level of clusters. The distinctions mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data. K-means treats each observation in your data as an object having a location in space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible.

2.4 GLCM methodology

Gray level Co-occurrence matrix (GLCM) is generated for each pixel map for H & S images of infected cluster.

- The graycomatrix function creates a gray level co-occurrence matrix by calculating how frequently a pixel with the particular intensity value i occurs in a specified spatial relationship to a pixel with the value j.
- By default this spatial relationship is the pixel of interest and its immediate right pixel.
- However we can specify some other spatial relationship between twos. To create multiple GLCMs, specify an array of offsets to the graycomatrix function. These offsets define pixel relationships of varying direction and distance. Directions can be horizontal, vertical, along two diagonals.
- Calculating statistics from GLCM matrix also known as SGDM

2.4.1 Color Moments methodology

A color image can be represented using three primaries of a color space. Since the RGB space does not correspond to the human way of perceiving the colors and does not separate the luminance component from the chrominance ones, we used the HSV color space in our approach. HSV is an intuitive color space in the sense that each component contributes directly to visual perception, and it is common for image retrieval systems. Hue is used to distinguish colors, whereas saturation
gives a measure of the percentage of white light added to a pure color. Value refers to the perceived light intensity. The important advantages of HSV color space are as follows: good
compatibility with human intuition, separability of chromatic and achromatic components, and possibility of preferring one component to other.The color distribution of pixels in an image contains sufficient information. The mean of pixel colors states the principal color of the image, and the standard deviation of pixel colors can depict the variation of pixel colors. The variation degree of pixel colors in an image is called the color complexity of the image. We can use these two features to represent the global properties of an image.
A color image can be represented using three primaries of a color space. Since the RGB space does not correspond to the human way of perceiving the colors and does not separate the luminance component from the chrominance ones, we used the HSV color space in our approach. HSV is an intuitive color space in the sense that each component contributes directly to visual perception, and it is common for image retrieval systems Hue is used to distinguish colors, whereas saturation gives a measure of the percentage of white light added to a pure color. Value refers to the perceived light intensity.

Mean:
The mean is the average of all numbers and is sometimes called the arithmetic mean. To calculate mean, add together all of the numbers in a set and then divide the sum by the total count of numbers.

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij}$$

Standard Deviation:
The standard deviation (SD, also represented by the Greek letter sigma σ or the Latin letter s) is a measure that is used to quantify the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values.

$$\sigma_i = \sqrt{\left(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - E_i)^2\right)}$$

2.5 SVM Classifier
A support vector machine (SVM) is used to recognize plant disease affecting agriculture/horticulture crops. The study has chosen SVM because of its efficient implementations and performances proved to be excellent for high dimensional problems and small data sets. Viewing training input vector in an *n*-dimensional space, SVM constructs a hyper-plane in the space, which can be used for classification that has the highest distance to the closest training data point of any class (functional margin). To compute the margin, two parallel hyper-planes are constructed, one on every side of the isolating hyper-plane, which are pushed up in opposition to the two data sets. The aim is to determine which class a new data point belongs based on data points associated to one of the two classes. In the case of support vector machines, a data point is computed as a p-dimensional vector (a list of p numbers) and it is meant to know whether such levels can be formed by a (p−1) dimensional hyper-plane. This is called a

linear classifier or maximum margin classifier. The core of SVM Matlab toolbox used in the present work is based on Dr. Lin's Lib SVM version 2.33. It is developed by Junshui Ma, Los Alamos National Lab and Yi Zhao, Electrical Engineering department, Ohio State University.

The notation used to define formally a hyperplane:

$$f(x) = \beta_0 + \beta^{-} x,$$

where β is known as the *weight vector* and β_0 as the *bias*. The optimal hyperplane can be represented in an infinite number of different ways by scaling of β and β_0. As a matter of convention, among all the possible representations of the hyperplane, the one chosen is

$$\beta_0 - \beta' x| - 1$$

where x symbolizes the training examples closest to the hyperplane. In general, the training examples that are closest to the hyperplane are called **support vectors**. This representation is known as the **canonical hyperplane**.

Now, we use the result of geometry that gives the distance between a point x and a hyperplane (β, β_0):

$$\text{distance} = \frac{|\beta_0 + \beta^T x|}{\|\beta\|}.$$

In particular, for the canonical hyperplane, the numerator is equal to one and the distance to the support vectors is

$$\text{distance}_{\text{support vectors}} = \frac{|\beta_0 + \beta^T x|}{\|\beta\|} = \frac{1}{\|\beta\|}.$$

Recall that the margin introduced in the previous section, here denoted as M, is twice the distance to the closest examples:

$$M = \frac{2}{\|\beta\|}$$

Finally, the problem of maximizing M is equivalent to the problem of minimizing a function $L(\beta)$ subject to some constraints. The constraints model the requirement for the hyperplane to classify correctly all the training examples x_i. Formally,

$$\min_{\beta, \beta_0} L(\beta) = \frac{1}{2}\|\beta\|^2 \text{ subject to } y_i(\beta^T x_i + \beta_0) \geq 1 \ \forall i,$$

where y_i represents each of the labels of the training examples.

This is a problem of Lagrangian optimization that can be solved using Lagrange multipliers to obtain the weight vector β and the bias β_0 of the optimal hyperplane.

2.6 Performance Analysis

The performance of individual classifiers with respect to, precision, recall, and average classification accuracy (ACA) using SVM.

Accuracy : The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation Accuracy (AC)

$$accuracy = Tp+TN / TP+TN+FP+FN$$

Recall ratio: The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation Recall ratio

$$recall = TP / TP+FP$$

Precision: It is the proportion of the predicted positive cases that were correct, as calculated using the equation Precision (P)

$$precision = TP / TP+TN$$

Table 1. Performance analysis comparison

Feature	Performance		
	Recall	***Precision***	***Accuracy***
Texture	0.92	0.84	97.22
Color	0.89	0.92	96.32
Texture and Color	0.96	0.92	98.46

EXPERIMENTAL RESULTS

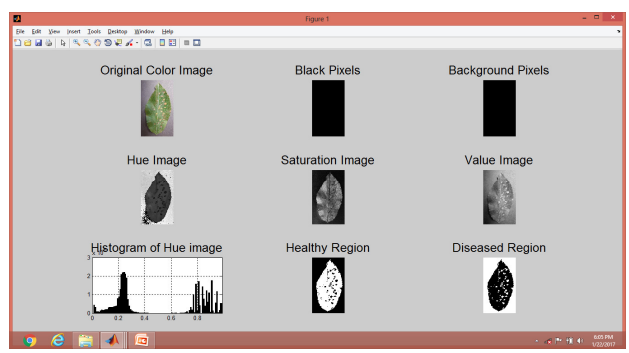

Fig 1:Color Transformations and Masking Green Pixels

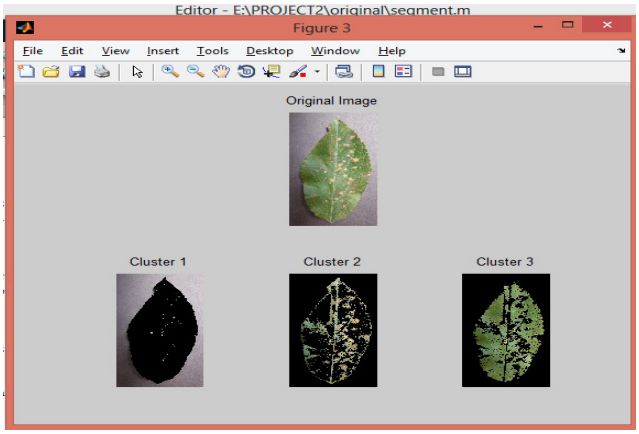

Fig.2: Segmentation using K-means Clustering

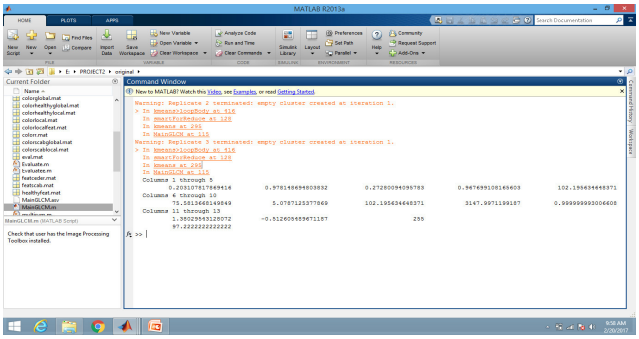

Fig 3: Feature Extraction Measures

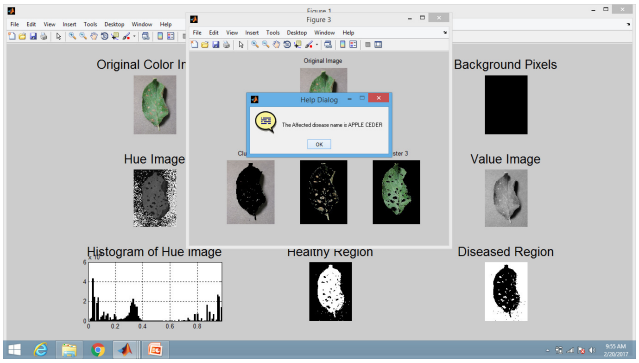

Fig 4: Classification Results

CONCLUSION

Main approach of our paper is to recognize diseases on the leaf. At first preprocessing is done which include two steps gray conversion Second stage is k-means based Image Segmentation which eventually does image analysis. Third stage is feature extraction that include color feature, shape features. And after that classification of diseases is performed victimization our projected formula. The goal of this analysis work is to develop Advance automatic data processing system which will determine the illness affected a part of a leaf spot by victimization the image analysis technique. Prediction of the diseases and discuss recommendation is finished. The producers will amend the Yield and scale back the loss. Through this projected system the farmers' burden has been reduced and saves their life. Perform better than others. Accuracy of detection can be increased when using SVM classifier with more number of features included to it.

References

[1] Jagadeesh D.Pujari, Rajesh and Abdulmunaf, Syedhusain Byadgi," SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique", International Journal of Interactive Multimedia and Artificial Intelligence,2016.

[2] Mr. Pramod S. landge, Sushil A. Patil, Dhanashree S. Khot, Omkar D. Otari, Utkarsha G. Malavkar," Automatic Detection and Classification of Plant Disease through Image Processing", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 7, July 2013.

[3] Supriya S. Patki, Dr. G. S. Sable," Cotton Leaf Disease Detection", IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) Volume 6, Issue 3, Ver. I (May-Jun. 2016).

[4] 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).

[5] Nikita Rishi, Jagbir Singh Gill." An Overview on Detection and Classification of Plant Diseases in Image Processing", International Journal of Scientific Engineering and Research (IJSER).

[6] Prof. Sanjay B. Dhaygude, Mr.Nitin P.Kumbhar,"Agricultural plant Leaf Disease Detection Using Image Processing", International Journal ofAdvanced Research in Electrical, Electronics and Instrumentation EngineeringVol. 2, Issue 1, January 2013

[7] Sagar Patil1, Anjali Chandavale," A Survey on Methods of Plant Disease Detection", International Journal of Science and Research (IJSR).