Plant Leaf Disease Detection Using Fuzzy C-Means Clustering Algorithm

[1]A. Kumari, [2] S.Meenakshi, [3]S. Abinaya
[1] Research Scholar, [2] Head of the Department, [3]Research Scholar
[1]Department of Computer Science, JP College of Arts and Science, Agarakattu-Tenkasi, India

---

Abstract - India, the country where the main source of income is from agriculture. Farmers grow a variety of crops based on their requirement. Since the plants suffer from the disease, the production of crop decreases due to infections caused by several types of diseases on its leaf, fruit, and stem. Leaf diseases are mainly caused by bacteria, fungi, virus etc. Diseases are often difficult to control. Diagnosis of the disease should be done accurately and proper actions should be taken at the appropriate time. Image Processing is the trending technique in detection and classification of plant leaf disease. This work describes how to automatically detect leaf diseases. The given system will provide a fast, spontaneous, precise and very economical method in detecting and classifying leaf diseases. This paper is envisioned to assist in the detecting and classifying leaf diseases using Multiclass SVM classification technique. First, the affected region is discovered using segmentation by Fuzzy C-means clustering, then features (color and texture) are extracted. Lastly, classification technique is applied in detecting the type of leaf disease. The proposed system effectively detects and also classify the disease with an accuracy of 92%.

Keywords - Image Processing, Leaf diseases detection, K-means clustering, feature extraction, Multiclass SVM Classification

---

1. INTRODUCTION

India is the land of agriculture. Farmers have an option to select required crops and then find appropriate pesticides for the plant to decrease the disease and increase the production. The cultivated plants will not always be healthy. In-order to increase the production with good quality the plant need to be monitored frequently because the plant disease leads to a reduction of the product. For successful cultivation, one should monitor the health as well as the disease of the plant. Diseases in plant cause heavy loss of the product. Hence the disease needs to be identified at the early stages, recommending farmers to avoid the harm in the production of the crop to increase the yield.

Plants suffer from diseases like Alternaria alternate (fungal), Anthracnose, Bacterial Blight (bacteria), and Cercospora Leaf Spot. Plant disease will be basically identified by observing different patterns on the parts of the plant like leaf, fruit, and stem. The indications on the leaf are taken into consideration for detecting the disease. The disease can be categorized as bacterial, viral, fungal etc. The proposed work emphases on identifying and categorizing the disease like Alternaria Alternata, Anthracnose, Bacterial Blight, and Cercospora Leaf Spot which are basically found on pomegranate, rice, soya bean, carrot, rose, watermelon, mango etc., using the technique called as image processing. It automatically detects leaf diseases. This system will provide a fast, spontaneous, precise and very economical method in detecting and classifying leaf diseases.

1.1 IMAGE PROCESSING

Image processing covers a vast area of scientific and engineering knowledge. It is built on a foundation of one- and two-dimensional signal processing theory and overlaps with such disciplines as artificial intelligence (scene understanding), information theory (image coding), statistical pattern recognition (image classification), communication theory (image coding and transmission), and microelectronics (image sensors, image processing hardware). Broadly, image processing may be subdivided into the following categories: enhancement, restoration, coding, and understanding. The goal in the first three categories is to improve the pictorial information either in quality (for purposes of human interpretation) or in transmission efficiency. In the last category, the objective is to obtain a symbolic description of the scene, leading to autonomous machine reasoning and perception.

Image Processing and Analysis can be defined as the "act of examining images for the purpose of identifying objects and judging their significance". A major attraction of digital imaging is the ability to manipulate image and video information with the computer. Digital image processing is now a very important component in many industrial and commercial applications and a core component of computer vision applications. Image processing techniques also provide the basic functional support for document image analysis and many other medical applications. The field of digital image processing is continually evolving. Transform theory plays a key role in image processing. Image and signal compression is one of the most important applications of wavelets.

1.2 OVERVIEW OF LEAF DISEASE

Digital image process is the use of computer algorithms to perform image process on digital pictures. It permits a far wider vary of algorithms to be applied to the computer file and might avoid issues like the build-up of noise and signal distortion throughout process. Digital image process has terribly important role in agriculture field. it's widely accustomed observe the crop disease with high accuracy. Detection and recognition of diseases in plants mistreatment digital image method is extremely effective in providing symptoms of characteristic diseases at its early stages. Plant pathologists will analyze the digital pictures mistreatment digital image process for diagnosing of crop diseases. Computer Systems area unit developed for agricultural applications, like detection of leaf diseases, fruits diseases etc. altogether these techniques, digital pictures are collected employing a camera and image process techniques are applied on these pictures to extract valuable data that are essential for analysis. The diseases are mostly on leaves and on stem of plant. The diseases are viral, bacterial, fungal, diseases due to insects, rust, nematodes etc. on plant. It is important task for farmers to find out these diseases as early as possible. Following example shows that how diseases on cotton plant reduces the productivity.

1.3 PROBLEM DESCRIPTION

Digital camera or similar devices are use to take images of leafs of different types, and then those are used to identify the affected area in leafs. Then different types of image-processing techniques are applied on them, to process those images, to get different and useful features needed for the purpose of analyzing later.

Algorithm written below illustrated the step by step approach for the proposed image recognition and segmentation processes:

- Image acquisition is the very first step that requires capturing an image with the help of a digital camera.
- Pre-processing of input image to improve the quality of image and to remove the undesired distortion from the image. Clipping of the leaf image is performed to get the interested image region and then image smoothing is done using the smoothing filter. To increase the contrast Image enhancement is also done.
- Mostly green colored pixels, in this step, are masked. In this, we computed a threshold value that is used for these pixels. Then in the following way mostly green pixels are masked: if pixel intensity of the green component is less than the pre-computed threshold value, then zero value is assigned to the red, green and blue components of the this pixel.
- In the infected clusters, inside the boundaries, remove the masked cells.
- Obtain the useful segments to classify the leaf diseases. Segment the components using genetic algorithm

For doing clustering appropriately, the search capability of GAs can be used, to set of unlabeled points in N-dimension into K-clusters. On image data, we have applied the same idea in our proposed scheme. We have taken a color image of size $m \times n$ and every pixel has Red, Green and Blue components. Every chromosome shows a solution, which is a sequence of K cluster centers. Population is initialized in various rounds randomly and from existing chromosome best chromosome survives in each round for the next round processing.

In the first step of fitness computation the dataset of pixel is clustered according to nearest respective cluster centers such that each pixel x_i of color image is put into the respective cluster with cluster center z_j for $j = 1, 2, \ldots, K$ by the following equations

$$\text{If} \|xi-zj\|<\|xi-zl\|,$$

$i=1,2,\ldots,m\times n, l=1,2,\ldots,K, \text{and} p\neq j.$

In the further step new cluster centers are obtained by calculating the mean of each pixel of the assigned clusters. The new center of cluster Z_i is given by for the cluster C_i as:

$$Zi(r,g,b)=1ni\sum xj\in ci(xj(r,g,b))i=1,2,\ldots.k \quad \text{- - - - - - - - - -} \ (1)$$

Now the fitness function is computed by calculating Euclidean distance between the pixels and their respective cluster by using following equations

$$M=\sum Mi \quad \text{- - - - - - - - - -} \ (2)$$

$$Mi=\sum xj\in ci(xj(r,g,b)-zi(r,g,b)) \quad \text{- - - - - - - - - -} \ (3)$$

- Computing the features using color co-occurrence methodology

For feature extraction the method used is color co-occurrence method. It is the methodology in which both the texture and color of an image are considered, to come to the unique features, which shows that image.

Over the traditional gray-scale representation, in the visible light spectrum, the use of color image features provides an additional feature for image characteristic. There are three major mathematical processes in the color co-occurrence method. First, conversion of the RGB images of leaves is done into HIS color space representation. After completion of this process, to generate a color co-occurrence matrix, each pixel map is used, which results into three color co-occurrence matrices, one for each of H, S, I.

1.4 TECHNIQUES STEPS

The steps involved are

1) Training – In training all the collected images are trained to the model and all six features are extracted and stored in the database.

2) Classification – After training, the SVM will classify the given new input as which type of disease is affected.

Data Collection

The sample images of the diseased leaves are collected and are used in training the system. To train and to test the system, diseased leaf images and fewer healthy images are taken. The images will be stored in some standard format. In this study, the available images from the internet are also taken. The leaf images that are infected by Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora Leaf Spot and Healthy leaf are also included.

Pre-processing

Image pre-processing is significant for genuine data that are frequently noisy and uneven. During this phase, the transformation is applied to convert the image into another image to improve the quality that better suits for analyzing. This step represents a crucial phase in image processing applications because the effectiveness of subsequent tasks (e.g., features extraction, segmentation) depends highly on images quality. Also, it significantly improves the effectiveness of data mining techniques.

Image Segmentation

During image segmentation, the given image is separated into a homogeneous region based on certain features. Larger data sets are put together into clusters of smaller and similar data sets using clustering method. In this proposed work, K-means clustering algorithm is used in segmenting the given image into three sets as a cluster that contains the diseased part of the leaf. Since we have to consider all of the colors for segmentation, intensities are kept aside for a while and only color information is taken into consideration. The RGB image is transformed into LAB form (L-luminous, a*b-chromous). Of the three dimensional LAB, only last two are considered and stored as AB. As the image is converted from RGB to LAB, only the "a" component i.e. the color component is extracted.

Feature Extraction

From the input images, the features are to be extracted. To do so instead of choosing the whole set of pixels we can choose only which are necessary and sufficient to describe the whole of the segment. The segmented image is first selected by manual interference. The affected area of the image can be found from calculating the area connecting the components. First, the connected components with 6 neighbourhood pixels are found. Later the basic region properties of the input binary image are found. The interest here is only with the area. The affected area is found out. The percent area covered in this segment says about the quality of the result. The histogram of an entity or image provides information about the frequency of occurrence of certain value in the whole of the data/image. It is an important tool for frequency analysis. The co-occurrence takes this analysis to next level wherein the intensity occurrences of two pixels together are noted in the matrix, making the co-occurrence a tremendous tool for analysis. From gray-co-matrix, the features such as Contrast, Correlation, Energy, Homogeneity' are extracted. The following table lists the formulas of the features.

Using the statistical MATLAB commands the other properties are found out. Those are Mean Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM.

2. MATERIALS AND METHODS

2.1 EARLIER RESEARCH WORK

The existing method for plant disease detection is simply naked eye observation by experts through which identification and detection of plant diseases is done. For doing so, a large team of experts as well as continuous monitoring of plant is required, which costs very high when we do with large farms. At the same time, in some countries, farmers do not have proper facilities or even idea that they can contact to experts. Due to which consulting experts even cost high as well as time consuming too. In such conditions, the suggested technique proves to be beneficial in monitoring large fields of crops. Automatic detection of the diseases by just seeing the symptoms on the plant leaves makes it easier as well as cheaper. This also supports machine vision to provide image based automatic process control, inspection, and robot guidance.

Plant disease identification by visual way is more laborious task and at the same time, less accurate and can be done only in limited areas. Whereas if automatic detection technique is used it will take less efforts, less time and become more accurate. In plants, some general diseases seen are brown and yellow spots, early and late scorch, and others are fungal, viral and bacterial diseases. Image processing is used for measuring affected area of disease and to determine the difference in the color of the affected area.

2.2 PROPOSED RESEARCH WORK

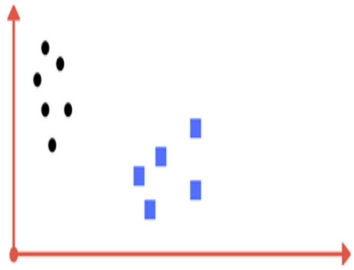
Image segmentation is the process of separating or grouping an image into different parts. There are currently many different ways of performing image segmentation, ranging from the simple thresholding method to advanced color image segmentation methods. These parts normally correspond to something that humans can easily separate and view as individual objects. Computers have no means of intelligently recognizing objects, and so many different methods have been developed in order to segment images. The segmentation process is based on various features found in the image. This might be color information, boundaries or segment of an image.

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, and 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.

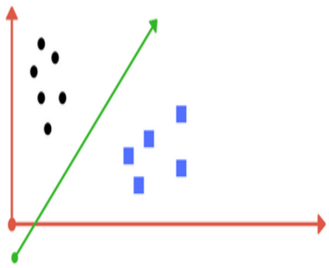
3. ALGORITHM DESCRIPTION

3.1 SUPPORT VECTOR MACHINE (SVM)

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples. In two dimensional spaces this hyper plane is a line dividing a plane in two parts where in each class lay in either side.
Suppose you are given plot of two label classes on graph as shown in Fig 3.1. Can you decide a separating line for the classes?



SVM classifier



Sample cut for SVM classifier

You might have come up with something similar to following image. It fairly separates the two classes. Any point that is left of line falls into black circle class and on right falls into blue square class. Separation of classes. That's what SVM does. It finds out a line/ hyper-plane (in multidimensional space that separate outs classes). Shortly, we shall discuss why I wrote multidimensional space.
In machine learning, Support Vector Machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

3.2 FUZZY C-MEANS CLUSTERING ALGORITHM

Fuzzy C-Means is a variation on Q-means which allows each datum partial membership in each cluster, similar to a mixture model

The steps are given below

Step 1: Choose primary centroids Ci (prototypes)

Step 2: Compute the degree of membership of all feature vectors in all the clusters:

$$u_{ij} = \frac{\left[\frac{1}{d^2(X_j, C_k)}\right]^{\frac{1}{(q-1)}}}{\sum_{k=1}^{K}\left[\frac{1}{d^2(X_j, C_k)}\right]^{\frac{1}{(q-1)}}}$$

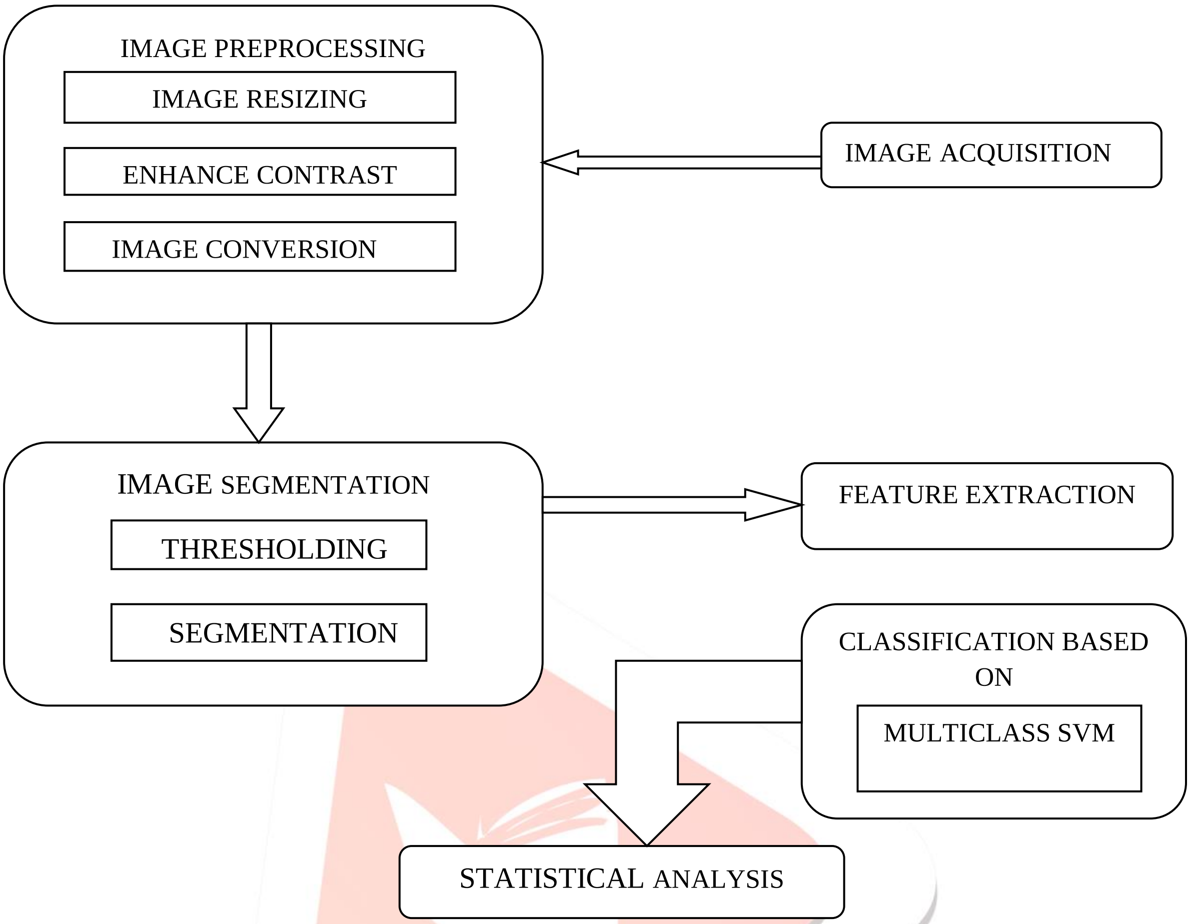
Step 3: Compute new centroids $C\hat{\ }_i$

$$C\hat{\ }i = \frac{\sum_{j=1}^{M}(u_{ij})^{q} X_j}{\sum_{j=1}^{M}(u_{ij})^{q}}$$

and update the memberships, u_{ij} to $u\hat{\ }_{ij}$ according to step 2.

Step 4: If $\max_{ij} \|u_{ij}-u\hat{\ }_{ij}\|$<tol stop, otherwise go to step 3.

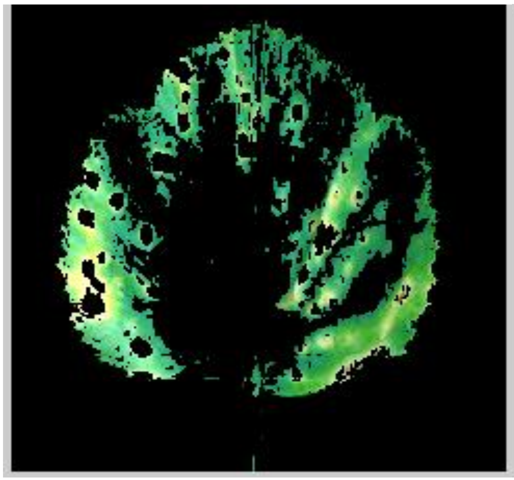
ARCHITECTURE DIAGRAM



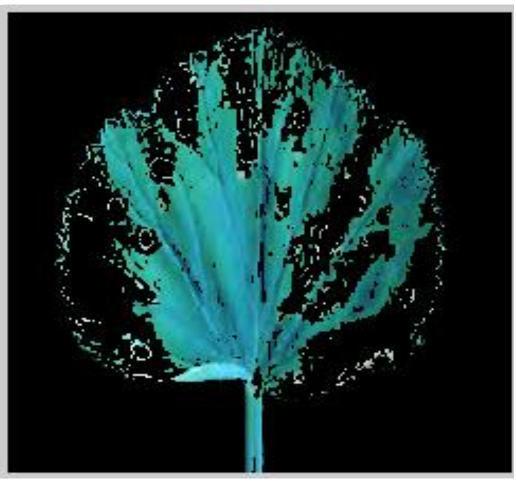
4. Experimental Result


Original Image


Segmented Image



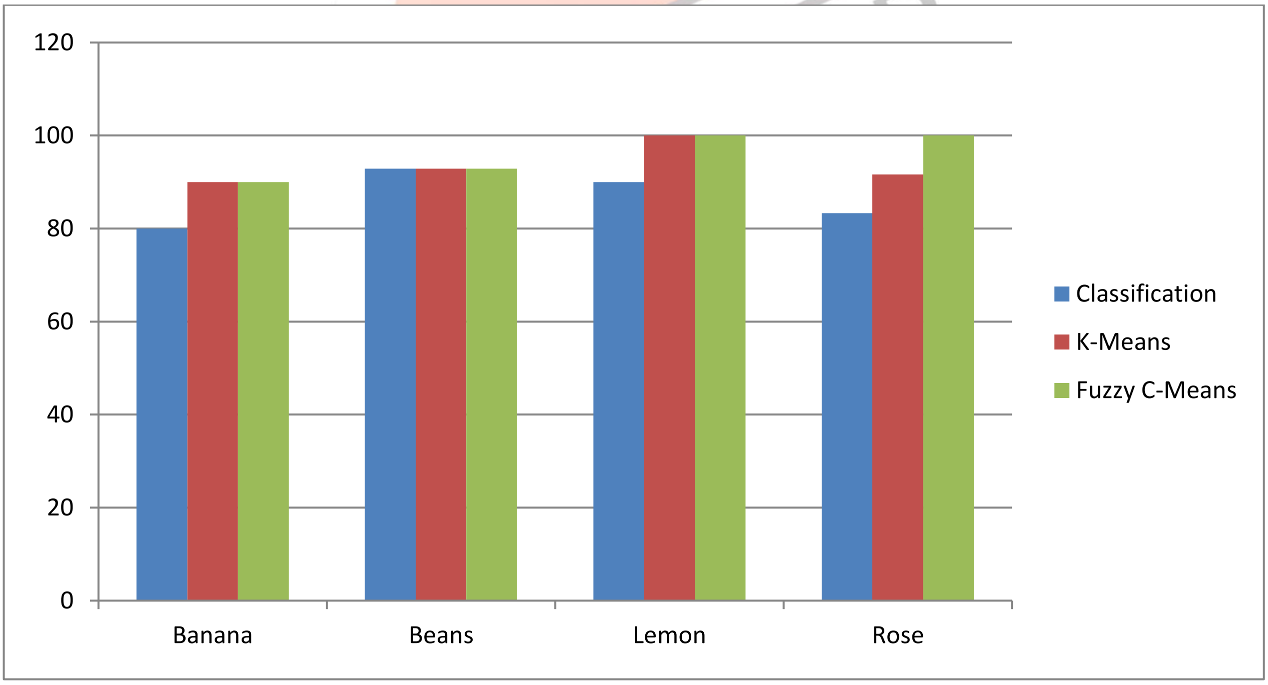
K-Means Clustering



Fuzzy C- Means Clustering

CLASSIFICATION RESULTS

Leaf disease	Bacterial leaf spot	Frog eye leaf spot	Sun burn disease	Fungal disease	Early scorch	Accuracy
Bacterial leaf spot	23	2	0	0	0	92
Frog eye leaf spot	1	24	0	0	0	96
Sun burn disease	0	0	25	0	0	100
Fungal disease	0	0	0	25	0	100
Early scorch	0	0	0	0	25	100
Average						97.6



5. CONCLUSION

This paper presents the survey on different diseases classification techniques used for plant leaf disease detection and an algorithm for image segmentation technique that can be used for automatic detection as well as classification of plant leaf diseases later. Banana, beans, jackfruit, lemon, mango, potato, tomato, and sapota are some of those ten species on which proposed algorithm is tested. Therefore, related diseases for these plants were taken for identification. With very less computational efforts the optimum results were obtained, which also shows the efficiency of proposed algorithm in recognition and classification of the leaf diseases. Another advantage of using this method is that the plant diseases can be identified at early stage or the initial stage. To improve recognition rate in classification process Artificial Neural Network, Bayes classifier and hybrid algorithms can also be used.

The world is moving more towards technology dependent era. Every day we keep hearing owes of farmers that even after using costly fertilizers the leaves were eaten away by various diseases. One of the most sensitive and costly treatments in India in terms of leaf concerned is that of pomegranate. The expertise in this field is rarely available. Since the opinion of an expert can vary from that of a novice, for the benefit of all it is advisory to make the most use of the technology available to infer or conclude for treatments.

The accuracy results in an available range from mid-90 to top 90%. This can be bettered by increasing the database. However, the results obtained from real life images are very encouraging.

6. REFERENCES

[1]"Classification of diseases plant leaves using Neural Network Algorithms" K. Muthukannan1, P. Latha2, R. Pon Selvi1 and P. Nisha1 1Department of ECE, Einstein College of Engineering, Anna University, Tirunelveli, India 2 Department of CSE, Government College of Engineering, Anna University, Tirunelveli, India

[2] S. S. Sannakki, V. S. Rajpurohit, V. B. Nargund, and P. Kulkarni, "Diagnosis and Classification of Grape Leaf Diseases using Neural Networks", IEEE 4th ICCCNT, 2013.

[3] P. Chaudhary, A. K. Chaudhari, Dr. A. N. Cheeran and S.Godara, "Color Transform Based Approach for Disease Spot" International Journal of Computer Science and Telecommunications Volume 3, Issue 6, pp.65-70, June 2012

[4] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, "Fast and Accurate Detection and Classification of Plant Diseases", IJCA, Vol-17, No.-1, pp. 31-38, March 2011.

[5] Suman T. and Dhruvakumar T., "Classification of paddy leaf diseases using shape and color features", IJEEE, Volume 07, Issue 01, PP.239250, Jan- June 2015.