EasyChair Preprint

№ 2256

The capitalized title is Plant Leaf Disease Identification Using Image Processing and SVM, ANN Classifier Methods

S. Sivasakthi

December 26, 2019

PLANT LEAF DISEASE IDENTIFICATION USING IMAGE PROCESSING AND SVM, ANN CLASSIFIER METHODS

Mrs. .S.Sivasakthi, MCA, M.Phil.,
Assistant Professor, Department of Computer Science,
G.Venkataswamy Naidu College (SFC), Kovilpatti
E-Mail: sakthiravi04@gmail.com

***Abstract*- Agricultural yield is something on which economy extremely depends. This is the reasons that disease detection in plants plays significant roles in agriculture field, as having disease in plants are quite usual. The identification of disease on the plant is a very essential key to avoid a heavy loss of harvest and the quantity of agricultural product. The signs can be detected on the parts of the plants such as stems, leaf, lesions and fruits. The leaf shows the symptoms by varying, showing the spots on it. This identification of the disease is done by manual statement and pathogen detection which can consume extra time and may prove expensive. The steps necessary in the process are Preprocessing, Training and Identification. Computers have been used for systematization and mechanization in different applications of agriculture/horticulture is done with the development of expert system (decision support system) using computer vision methods. The color and texture topographies have been used in order to work with the trial images of plant diseases. Algorithms for extraction of color and texture features have been established, which are in chance used to train support vector machine (SVM) and artificial neural network (ANN) classifiers. This paper is mainly developed to identify and calculate the correctness of pest infected area in leaf images. The image segmentation procedure is used to perceive the presence of pests in leaf images. The presentation of the clustering based image segmentation algorithm depends on its overview of images.**

Keywords- Image Processing, SVM (Support Vector Machine), ANN (Artificial Neural Network), Feature extraction, Segmentation

I.INTRODUCTION

Agriculture is considered the strength of Indian budget. On the other hand, the farming of crops for optimal harvest and quality produce is very important. A lot of examination has been done on conservatory agro structures and more generally on protected yields to control pests and diseases by organic means instead of pesticides. Exploration in agriculture is pointed towards rise of productivity and food quality at reduced expenses and with improved profit, which has received status in recent time. This manual method is very time overwhelming. With the recent development in image processing techniques, it is promising to develop separate system for disease classification of crops.

Whiteflies, Thrips, Aphids are the most common pests which attach on greenhouse crops. Generally the size of fully-grown whitefly is 1/12 inch in length. The female of whitefly is sap-sucking pest may lay 150 eggs at the rate of 25 per day. The complete life span of whiteflies is 21-36 days. Thrips are small, slim pest about 1/25-inch long in length. They variety in from light brown to black. Thrips grows on flower plants and fruit plants. Aphids are very small sizes.

Aphids are soft-bodied, inactive pests. They form cluster in gatherings on the leaves of the host plants. Their life extent is 20 to 30 days. The only way to stop the outcome of these pests is pesticides. The production of pesticides started in India in 1922 with the creation of a plant for production of BHC near Calcutta. By using of pesticides, the making of grains should be increased. A pesticide is any substances used to destroy, suppress or alter the life cycle of any pest.

A pesticide can be a naturally consequent or synthetically formed substance. Pesticides are chemical. It can be poisonous to humans and lower animals, so use natural organic such as neem, salt spray, mineral oil, onion and garlic spray. First detection of pest or the initial presence of pests is a key-point for yield management. Improved crop protection Pest infected images are attained using cameras. Then the image processing approaches are used to detection of pest in different infected leaves images.

II.THE IMAGE PROCESSING METHODOLOGY

Digital signal processing is the procedure to achieve fast and correct result about the plant leaf diseases. It will decrease many agricultural facets and improve productivity by identifying the suitable diseases. Used for diseases finding image of an infected leaf should examine through the set of procedures. After then some classifier methods should be used to classify the diseases according to the specific data set. The existing technique for plant

disease revealing is simply naked eye reflection by specialists through which identification and detection of plant diseases is done.

The step-by-step procedure of the system:

1. RGB image acquisition
2. Convert the input image from RGB to HSV format.
3. Masking the green-pixels
4. Removal of masked green pixels
5. Segment the components
6. Obtain the useful segments
7. Computing the features using color-co-occurrence methodology
8. Evaluation of texture statistics

The flowchart of suggested work is shown in figure 1. It consists of several stages such as image acquisition, preprocessing, segmentation and correctness of infected area. It is measured by SVM classifier. The methodology is executed in MATLAB.

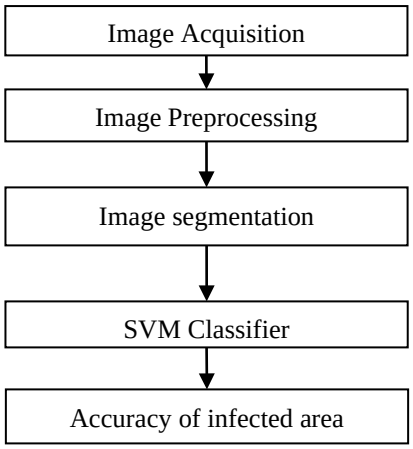

Figure 1: The basic procedure of the proposed approach

A.Image Acquisition

Image Acquisition is the procedure in which developed and transformed to the desired output format. For this application an analog image is first caught and then converted to the digital image for further handling.

The data set is equipped and used in this research. The data set holds pest infected leaf images. This consists of two types of leaves. These leaves will be separated into whiteflies and aphids with whiteflies infected leaves. The infected leaves are taken as contribution for processes such as preprocessing, segmentation and the SVM classifier. Select the original image in dataset shown in figure 2.

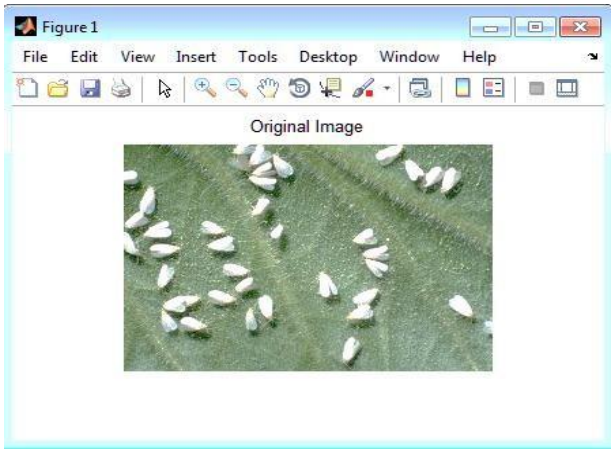

Figure 2: A sample leaf image with whiteflies

B. Image Pre-Processing

Preprocessing Segmentation covers process for image segmentation, image enhancement and color space translation firstly image digital image is improved by filter. Leaf image is filtered form the background image. Then filtered image's RGB colors are changed into color space constraint. Hue Saturation Value (HSV) is a good method for color insight. Further image is segmented to a meaning full part which is easier to examine. Any of the model based, edge based, threshold based, Region based and feature based segmentation has been prepared on the images.

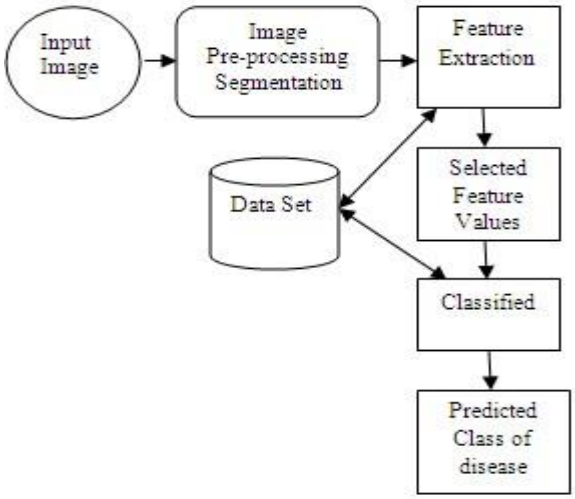

Figure 3: Classification Model

Contrast stretching is an image enhancement technique that progresses the contrast in an Image increasing the dynamic range of intensity values it contains. Before applying k-means algorithm, first stretching augmentation is applied to the image to progress the quality of the image. Apply the contrast enhancement for original image shown in figure 3.

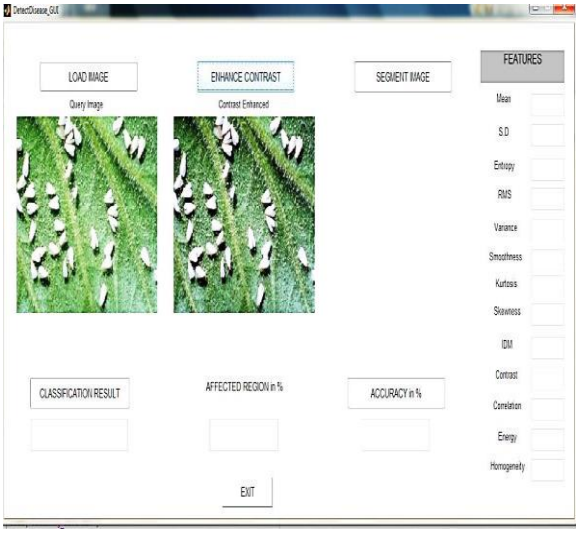

Figure 4: The leaf image after preprocessing

C. Feature Extraction

Features are extracted from image using Gabor filtering technique. These features are very significant for the morphology of the leaf spots and they supply critical information about its visual depiction. The features matches to features are the mean and variance of the gray level of the red, green and blue channel of the spots; and further features correspond to morphological and geometrical characteristics of the spots. It is assumed the shape of leaf is the same to the ellipse. The parameters are represented in Ellipse shown in Figure 4.

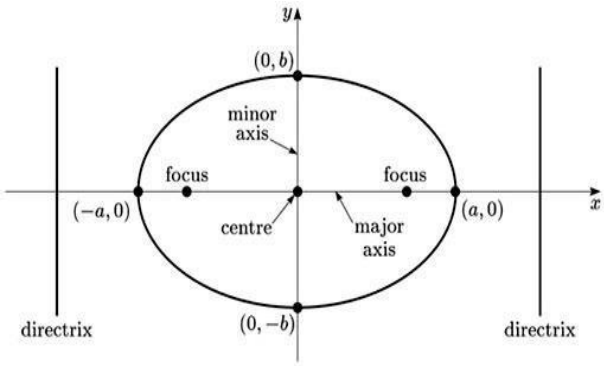

Figure 5: Ellipse and its parameters

According to the segmented data and predefined dataset some skins of the image should be extracted. This extraction could be the any of structural, statistical, fractal or signal processing. Grey Level Co-occurrence Matrices (GLCM), Spatial Gray-level Dependence Matrices (SGDM) method, Color co-occurrence Method, Gabor Filters, Wavelets Transform and Principal component analysis are some approaches used for feature extraction.

Technique	**Advantages**	**Disadvantages**
K-Nearest Neighbor (KNN)	Simpler classifier as exclusion of any training process. Applicable in case of a small not trained dataset.	More training samples- more speed of computing distances sensitive to irrelevant inputs so expensive testing every time.
Radial Basis Function (RBF)	Faster Training. Hidden layer is easier to interpret.	It is slower in execution speed
Probabilistic Neural Networks (PNN)	Tolerant to noisy inputs. Instances classified to many output adaptive to change data.	Long training time. Complex network structure. Excessive memory for training data.
Back propagation Network (BPN)	Easy to implement. Applicable to wide range of problems. Able to form arbitrarily complex nonlinear mappings	Learning can be slow. It is hard to know how many neurons as well as layers are required.
Support Vector Machine (SVM)	Simple geometric interpretation and a sparse solution. robust, when sample has some bias.	Slow training. Difficult to understand. For classification large support vector required

Table 1. Classification Techniques

D.Color Based Segmentation Using K-Means Clustering

Image Segmentation is the classification of an image into different groups. K-Means clustering algorithm is an unsupervised algorithm and it is used to segment the interested area from the background. K-means clustering is a method of vector quantization. A cluster is a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters. The simple graphical example is shown in figure 5.

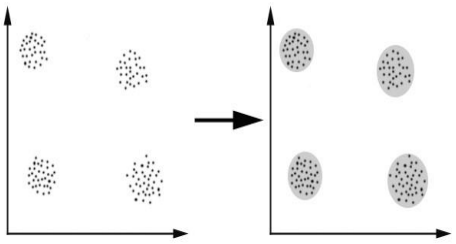

Figure 6: Clustering Process Diagram

In this case we easily classify the 4 clusters into which the data can be divided; the resemblance criterion is distance: two or more objects fit in to the same cluster if they are “close” rendering to a given distance (in this case geometrical distance). This is called distance-based clustering. Another one kind of clustering is conceptual clustering: two or more objects belong to the similar cluster if this one describes a concept common to all that objects.

Clustering mentions to the process of grouping samples so that the samples are similar within each group. The groups are named clusters. Clustering is a data mining technique used in statistical data analysis, image analysis, pattern recognition, etc.

Different clustering approaches include hierarchical clustering which shapes a hierarchy of clusters from individual elements. Because of its ease and efficacy, clustering approaches were one of the first methods used for the segmentation of (textured) natural images.

In partition clustering, the objective is to create one set of clusters that partitions the data in to comparable groups. Further methods of clustering are distance based conferring to which if two or more objects fitting to the same cluster are close according to a given distance, then it called distance is based clustering. In this paper used K-means clustering method for performing image segmentation using Matlab software.

A good clustering method will yield high class clusters with high intra-class similarity and low inter-class similarity. The value of clustering result depends on both the similarity measure used by the method and its implementation. The excellence of a clustering method is also measured by its skill to discover some or all of the hidden patterns. Image Segmentation is the base of image analysis and accepting and a crucial part and an oldest and toughest problem of image processing.

Clustering means classifying and characteristic things that are providing with similar properties. Clustering techniques categorizes the pixels with same features into one cluster, thus forming different clusters according to consistency between pixels in a cluster. It is a method of unsupervised knowledge and a common method for statistical data analysis used in many fields such as pattern recognition, image analysis and bioinformatics.

There are a lot of applications using K-mean clustering, collection from unsupervised learning of neural network, Classification analysis, Pattern recognitions, Artificial intelligent, image processing, machine vision, etc.

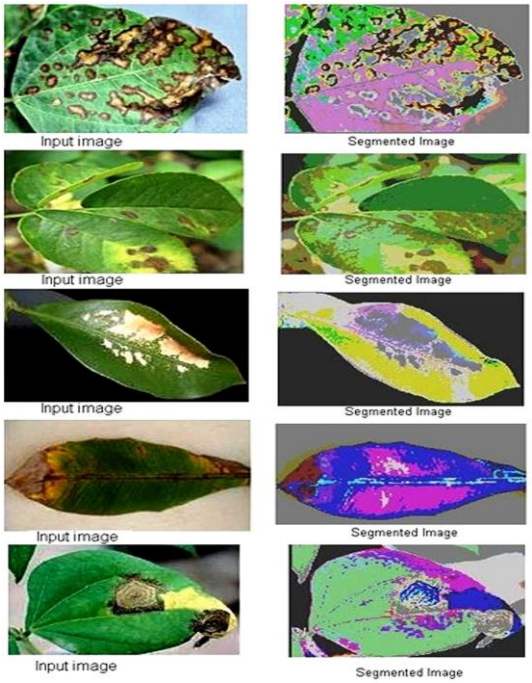

Figure 6: A sample leaf input – output images

E. SVM Classifier

A Support vector machine is a great implement for binary classification, capable of producing very fast classifier function subsequent a training period. There are several methods to implementing SVMs to classification problems with three or more classes.

Support vector machines are supervised learning simulations with associated learning algorithms that examine data used for classification and regression analysis. SVM are characteristically two class classifiers. The traditional method to prepare multiclass classification with SVMs is use one of the methods.

From the graph, it is detected that the maximum classification accuracy of 91% has happened with images of normal. The minimum classification accuracy of 78% has happened with images of nematodes disease. The average classification accurateness of 84% is realized irrespective of the image types of plant disease affecting agriculture/horticulture crops.

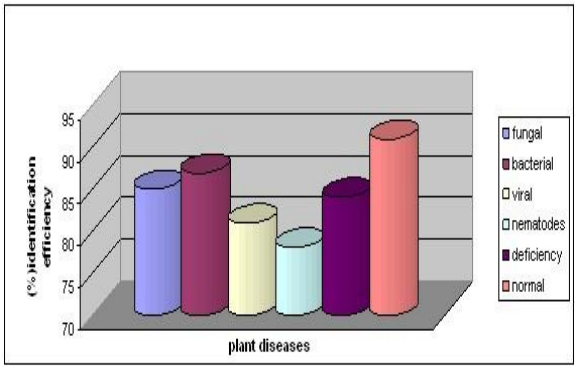

Fig.7: Classification efficiency for plant diseases using color features with SVM

The classifier assessment contain the output value higher than the threshold area noted as "true" and any SVM output value lower than the threshold are noted ad "false". The SVM classifier contain the binary classification of images.

Table 2. SVM binary classification

	Positive(+1)	**Negative(-1)**
Positive	True positive(TP)	False negative(FN)
Negative	False positive(FP)	True negative(TN)

- Recall=TP/ (TP+FN)
- Precision=TP/ (TP+FP)
- False alarm = FP/ (FP+TN)
- Accuracy = (TP+TN)/ (TD+TN+FP+FN)

The infected area accurateness is calculated after the k–means classification of segmentation areas using machine learning methods such as multiclass SVM in SVM Classifier.

F. ANN Classifier

In order to validate the accuracy of classification gained from SVM classifier, the training has considered a multilayer BPNN as an alternative model to find plant disease symptoms affecting agriculture/horticulture crops.

A typical multilayer neural network includes an input layer, output layer, and hidden (intermediate) layer of neurons. BPNN has the latent to classify dissimilar forms (patterns) of random complex input/output mappings or decision sides. Among the mainstream of multilayer feed-forward artificial neural network (ANN) algorithms, BPNN looked as the most important *algorithm for the supervised training and has been the workhorse for many classification problems in different applications of machine learning.* The number of neurons in the input layer matches to the number of input pattern vectors and the number of neurons in the output layer matches to the number of pattern types (classes). In the hidden layers sigmoid functions have been used.

From the graph, it is detected that the maximum classification accuracy of 94% has happened with images of normal. The minimum classification accuracy of 72% has happened with images of deficiency disease. The average classification accurateness of 82% is realized irrespective of the image types of plant disease affecting agriculture/ horticulture crops.

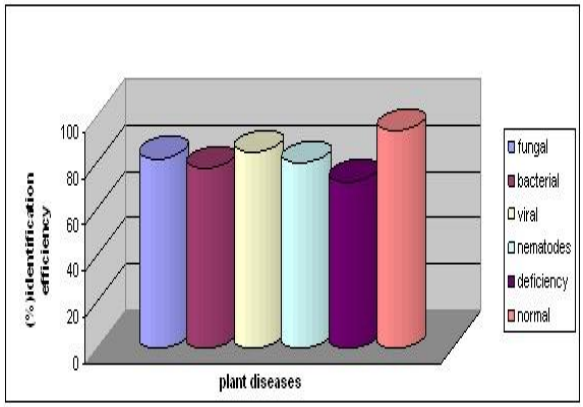

Fig. 8: Classification efficiency for plant diseases using color features with ANN

The comparison of classification correctness of both the classifiers using texture, reduced color and combined features is given in the Table 3. From the Table 3, it is manifest that the SVM based classifier has given improved classification accuracy than the ANN classifier.

TABLE 3: PERFORMANCE OF SVM AND ANN CLASSIFIERS USING COLOR, TEXTURE AND COMBINED FEATURES

Classification performance rate (%)	**Classifiers**					
	SVM			**ANN**		
	color	**texture**	**combined**	**color**	**texture**	**combined**
minimum accuracy	78	83	86	72	78	82
maximum accuracy	91	96	98	94	92	92
average identification accuracy	84	89	92	82	84	87

G. Accuracy of Infected Area

The pest such as aphids, whiteflies and thrips are very tiny in size and affect the leaves. Usually the size
of adult whitefly is 1/12 inch in length. The female of whitefly is sap-sucking pest might place 150 eggs at the rate of 25 per day. The complete life time of whiteflies is 21-36 days. The length of Thrips pest is 1/25-inch long. They variety in from light brown to black. Thrips grows on fruits and flower plant leaves. The thrips and aphids are amounts in leaf image by physically are not easy. The pest infected area accuracy calculated by the SVM classifier.

H. Constraints for Performance Evaluation

Effective presentation of K-Means Clustering Approaches is assessed based on parameters such as mean, Standard deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, IDM, Contrast, Correlation, Energy, Homogeneity which are described below.

1. Mean (mean Average or mean value of array): Its returns the mean values of the elements along different dimensions of an array

   M = mean (A)

2. Standard deviation: The result s is the square root of an unbiased estimator of the variance of the population from which X is drawn, as long as X consists of independent, identically distributed

   S = std (X)

   where, X is a vector, returns the standard deviation.

3. Entropy: It is returns E, a scalar value representing the entropy of gray scale imageI.

E= entropy (I), Entropy is defined as –sum (p.*log (p))

4. RMS (Root Mean Square): RMS is defined as the Standard Deviation of the pixels intensities. Its returns the root-mean-square (RMS) level of the input, X.

   Y = rms(X)

5. Variance: The variance is normally used to find how each pixels varies from the neighboring pixel and it is used in classify into different region.
6. Kurtosis: Kurtosis returns the sample kurtosis of X. For vectors, kurtosis(x) is the kurtosis of the elements in the vector x.

   k = kurtosis(X)

7. Skewness: It is returns the skewness of X.

   y = skewness(X)

8. Contrast: Contrast returns a measure of the intensity contrast between a pixel and its neighbor over the whole image.

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

   Range = [0 (size (SGDM, 1)-1) ^2], Contrast is 0 for a contrast image.

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

   $$\text{Correlation} = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{\{i \times j\} \times P(i,j) - \{\mu_x - \mu_y\}}{\{\sigma_x - \sigma_y\}}$$

10. Energy: Energy returns the sum of squared elements in the SGDM Range = [0 1].Energy is 1 for a constant image.

    $$\text{Energy} = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}$$

11. Homogeneity: Homogeneity returns a value that measures the closeness of the distribution of elements in the SGDM to the SGDMdiagonal. Range = [0 1] Homogeneity is 1 for a diagonal SGDM.

    $$\sum_{i,j=0}^{N-1} C(i,j)/(1 + (i-j)^2)$$

IV.RESULTS AND DISCUSSIONS

All the experiments are performed in MATLAB. For input data disease, examples of plant leaves like beans leaf with bacterial disease, rose with bacterial disease, lemon leaf with Sun burn disease, banana leaf with early scorch disease and fungal disease in beans leaf are measured.

The co-occurrence features are designed after mapping the R, G, B elements of the input image to the threshold images. The co-occurrence structures for the leaves are extracted and compared with the consistent feature values that are stored in the feature library. The classification is first complete using the Minimum Distance Criterion with K-Mean Clustering. In the second segment classification is done using SVM classifier and shows its efficiency with accuracy of 95.71%.

The input image is collected physically using digital camera. The caught image is preprocessed. The preprocessed image and the image after classifying the infected area by using k-means color based

segmentation technique are given. Finally the accuracy of diseased region is calculated by SVM classifier.

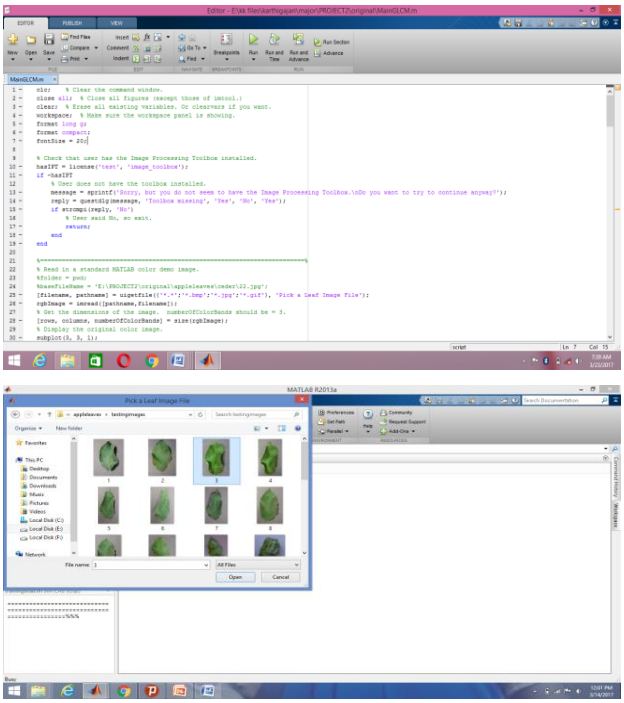

Figure 9: Color Transformation and Green Pixels Masking

The concept of RGB and HSV component extraction is presented in the Algorithm. There are 24 color features extracted from the images and they are listed in the Table 4.

Sl.No.	Features	Sl.No.	Features
1	Red mean	13	Hue mean
2	Red variance	14	Hue variance
3	Red range	15	Hue range
4	Red standard deviation	16	Hue standard deviation
5	Green mean	17	Saturation mean
6	Green variance	18	Saturation variance
7	Green range	19	Saturation range
8	Green standard deviation	20	Saturation standard deviation
9	Blue mean	21	Intensity mean
10	Blue variance	22	Intensity variance
11	Blue range	23	Intensity range
12	Blue standard deviation	24	Intensity standard deviation

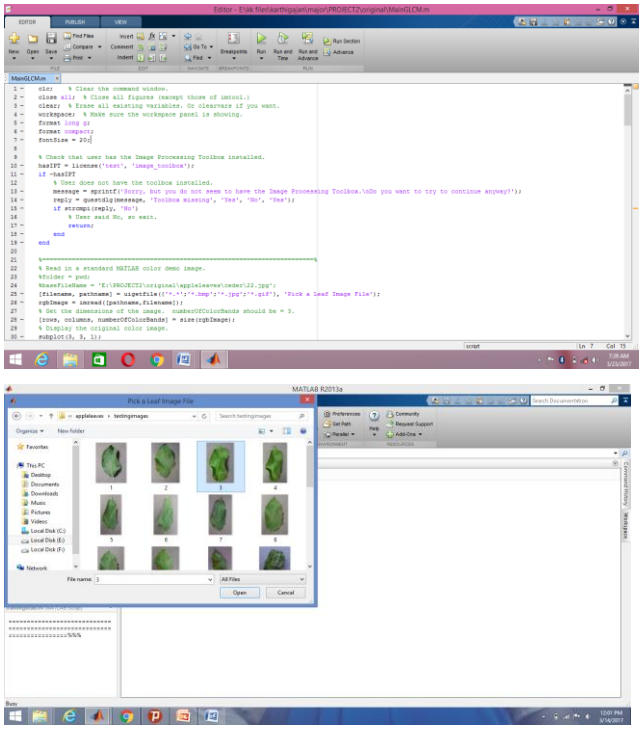

Figure 10: Color Transformation and Green Pixels Masking

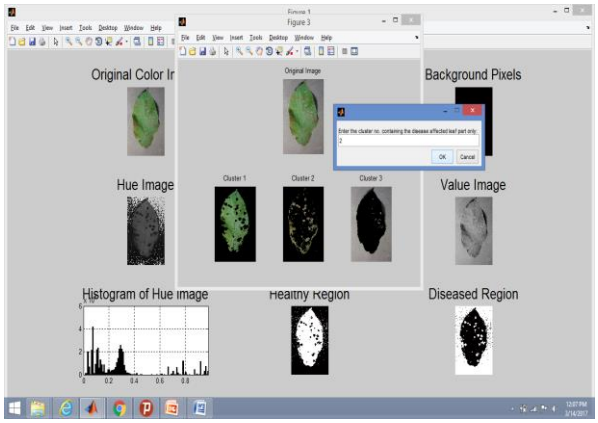

Figure 11: Segmentation

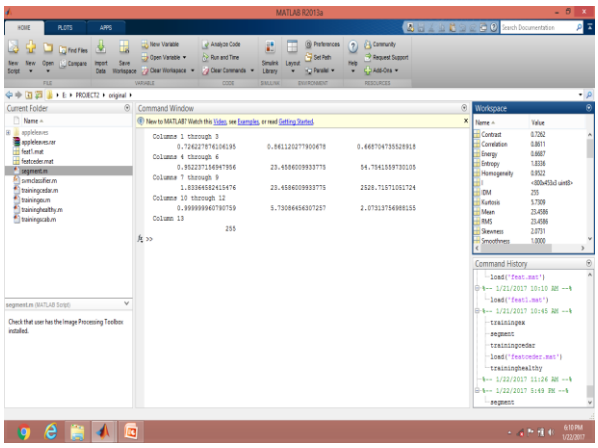

Figure 12: Feature Extraction

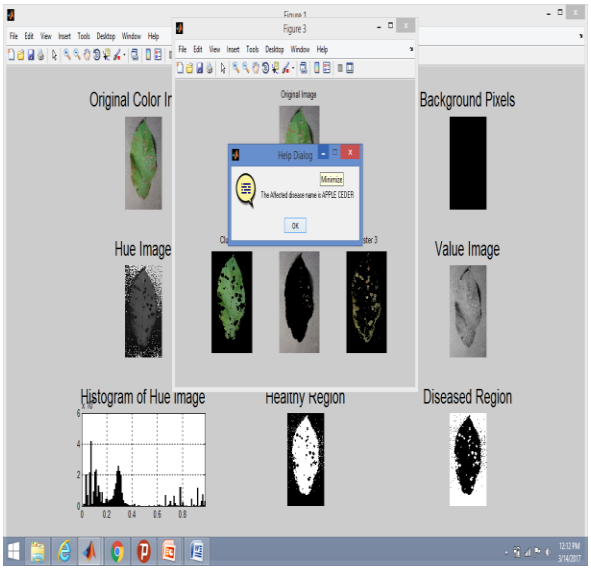

Figure 13: Classification Result

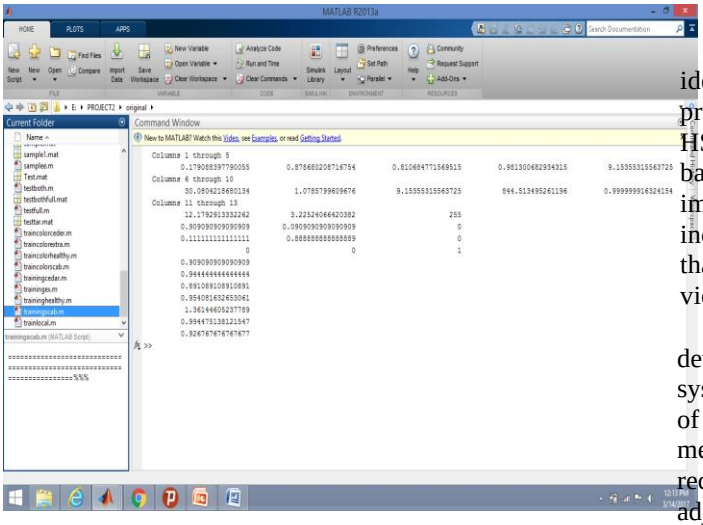

Figure 14: Performance Analysis

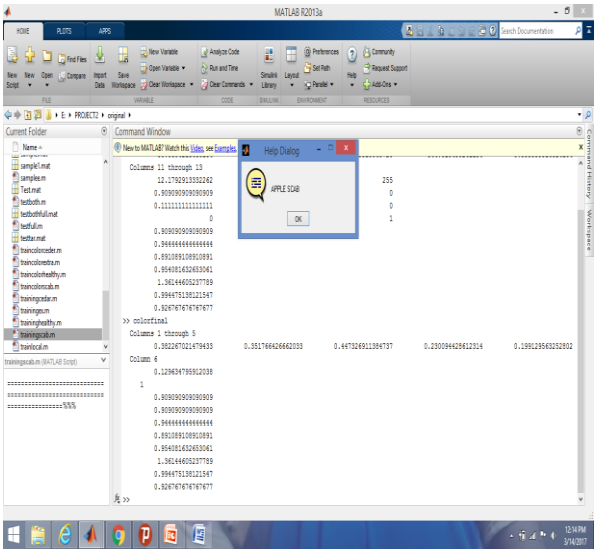

Figure 15: Color feature values

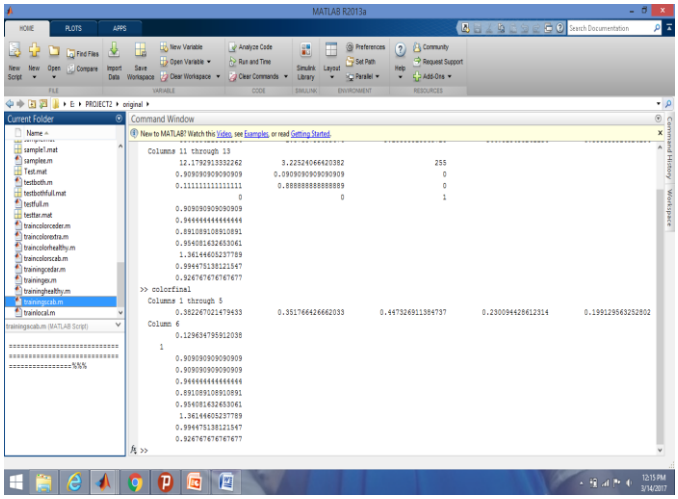

Figure 16: Performance Measurement

VI.CONCLUSION AND FUTURE WORK

Main methodology of our system is to identify diseases on the plant leaf. At first preprocessing is completed which contain two steps HSV transformation and Second phase is k-means based Image segmentation which ultimately does image analysis. Third phase is feature extraction that includes color features and texture features. And after that classification of diseases is accomplished victimization our projected formula

The objective of this analysis work is to develop Improvement automatic data processing system which will regulate the illness affected a part of a leaf spot by victimization the image analysis method. Estimate of the diseases and discuss recommendation is finished. The producers will adjust the Yield and measure back the loss. Through this predictable system the farmer's problem has been reduced and protects their life. Accuracy of recognition can be increased when using SVM classifier with advance number of features included to it. The color constructed segmentation using kmeans classifier is achieved to separate the different area with classify in the image. The SVM classifier is used to calculate the accuracy of diseased leaf region.

In the future the same system can be recycled to test all kind of plant leaves and catch out whether the plant is being affected by any disease or not .If it is affected then it furthermore displays the name of the disease.

The color based segmentation using kmeans classifier is performed to separate the different region with classify in the image. The multiclass svm classifier is used to calculate the accuracy of infected leaf region.

REFERENCES

[1] Rafael C. Gonzalez and Richard E. Woods, “Digital Image Processing”, 2nd ed., Beijing: Publishing House of Electronics Industry, 2007.

[2] S. Arivazhagan, R. NewlinShebiah, S. Ananthi and S. Vishnu Varthini, “Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features” Department of Electronics and Communication Engineering, March, 2013. Vol. 15, No.1, 2013.

[3] RajeshwarDass, Priyanka and Swapna Devi, “Image Segmentation Techniques”, IJECT Vol. 3, No1, 2012.

[4] Sanjay B.Dhaygude and NitinP.Kumbhar, “Agricultural plant Leaf Disease Detection Using Image Processing”, International Journal of Advanced research in Electronics, Electronics and Instrumentation Engineering. Vol.2, No 1, 2013.

[5] H.P. Narkhede, “Review of Image Segmentation Techniques”, International Journal of Science and Modern Engineering (IJISME) ISSN: 2319-6386, Vol-1, No-8, 2013.

[6] K.K. Singh and A. Sing, “A Study of Image Segmentation Algorithms for Different Types of Images”, International Journal of Computer Science Issues, Vol. 7, No 5, 2010.

[7] Jongman Cho, Junghyeon Choi, Mu Qiao, Chang-woo Ji, Hwang-young Kim, Ki-baikUhm, and Tae-soo Chon,

[8] “Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis”, International Journal of Mathematics and Computers in Simulation.

[9] JagadeeshD.Pujari, Rajesh and Abdulmunaf, SyedhusainByadgi,” SVM and ANN Based Classification of Plant Diseases Using Feature Reduction Technique”, International Journal of Interactive Multimedia and Artificial Intelligence, 2016.

[10] 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.

[11] 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).

[12] 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).

[13] 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).