Pest Identification in Leaf Images using SVM Classifier

R. Uma Rani
Associate Professor of Computer Science
Sri Sarada College for Women (Autonomous)
Salem-16.

P. Amsini
PG student, Department of Computer Science
Sri Sarada College for Women (Autonomous)
Salem-16

***Abstract*- This paper is mainly developed to detect and calculate the accuracy of pest infected area in leaf images. In modern agricultural field, pest detection is a major role of plant cultivation. The production rate of crops is reduced in agricultural field by the presence of whitefly pests, aphids and thrips which cause leaf discoloration. The image segmentation technique is used to detect the presence of pests in leaf images. The performance of the clustering based image segmentation algorithm depends on its simplification of images. The K-means cluster algorithm has been proposed to identify the accurate location of whiteflies, aphids and thrips in various leaf images. The infected area is calculated by SVM classifier. The algorithm was developed and implemented using MATLAB 7.14 build 2012a.**

Keywords- Greenhouse crops, early pest detection, SVM (support vector machine), image processing, feature extraction.

I. INTRODUCTION

Agriculture is considered the backbone of Indian economy.However, the cultivation of crops for optimum yield and quality produce is highly essential. A lot of research has been done on greenhouse agro systems and more generally on protected crops to control pests and diseases by biological means instead of pesticides. Research in agriculture is aimed towards increase of productivity and food quality at reduced expenditure and with increased profit, which has received importance in recent time. In fact, in production conditions, greenhouse staff periodically observes plants and search for pests. This manual method is very time consuming. With the recent advancement in image processing techniques, it is possible to develop an autonomous system for disease classification of crops. There are different crops which are cultivated under greenhouse e.g. Rose, Cucumber, Tomato, Gerbera, Capsicum etc. Whiteflies, Thrips, Aphids are the most common pests which attach on greenhouse crops. White flies, thrips and Aphids are very small in size. Normally the size of adult 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 entire life cycle of whiteflies is 21-36 days. Thrips are tiny, slender pest about 1/25-inch long in length. They range in colour from light brown to black. Thrips grows on flower plants and fruit plants. Aphids are very small in size. Aphids are soft-bodied, sluggish pests. They form cluster in colonies on the leaves of the host plants. Their life span is 20 to 30 days. The only way to stop the effect of these pests is pesticides. The production of pesticides started in India in 1922 with the establishment of a plant for production of BHC near Calcutta. By using of pesticides, the production 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 derived or synthetically produced substance. Pesticides are chemical. It can be toxic to humans and lower animals, so use natural organic pesticides such as neem, salt spray, mineral oil, onion and garlic spray. Early detection of pest or the initial presence of pests is a key-point for crop management. Improved crop protection Pest infected images are acquired using cameras. Then the image processing methods are used to detection of pest in various infected leaves images.

II. LITERATURE SURVEY

Earlier papers are describing to detect mainly pests like aphids, whiteflies, thrips, etc using various approaches suggesting the various implementation ways as illustrated and discussed below. [1] Proposed early pest detection in defect tomatoes plants and identifies the borer in it by using morphology segmentation technique. [2] 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 algorithms 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] Proposed the SVM classifiers are used to classify the brain MRI images. The process consists of two components which are training phase and a testing phase. Percentages of accuracy on each parameter in SVM give the best result. [4] It classified and discussed main image segmentation algorithms and concluded that the methods are classified on the basis of the features as, homogeneity of images, spatial characteristics of the image continuity, texture, image content. Viruses are extremely tiny particles consisting of protein and genetic material with no associated protein. The term disease is usually used only for the destruction of live plants. The developed processing scheme consists of four main steps, first a color transformation structure for the input RGB image is created, and this RGB is converted to HSI because RGB is for color generation and his for color descriptor. Then green pixels are masked and removed using specific threshold value, then the image is segmented and the useful segments are extracted, finally the texture statistics is computed from SGDM matrices. Finally the presence of diseases on the plant leaf is evaluated. [5] Image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image .Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture.[6] The segmentation techniques are helps to detect the region of the image. The result of image segmentation is depends on many factors, i.e. pixel color, texture, intensity, similarity of images, image content and problem domain. Therefore, it is not possible to consider a single method for all types of images nor all methods can perform well for a particular type of image. [7] Automatic insect recognition and classification has been given extra attention especially in term of crop pest and disease control. Among the method used in the previous research includes color histogram, edge detection and feature extraction. To detects the insect identification accurately. [8]

III. THE PROPOSED WORK

The flowchart of proposed work is shown in figure 1. It consists of various stages such as image acquisition, preprocessing, segmentation and accuracy of infected area. It is calculated by SVM classifier. The proposed approach is implemented in MATLAB.

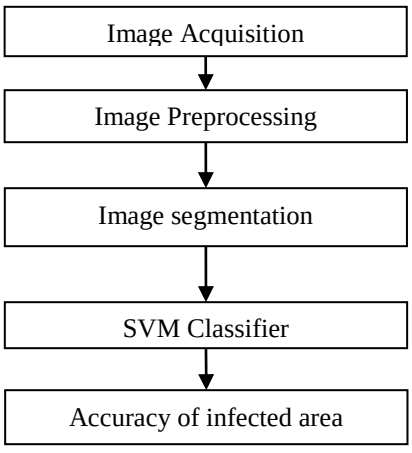

Figure 1: The basic procedure of the proposed approach

A. Image Acquisition

The data set is prepared and used in this research. The data set contains pest infected leaf images. This consists of two types of leaves. These leaves will be divided into whiteflies and aphids with whiteflies infected

leaves. The infected leaves are taken as input for processes such as preprocessing, segmentation and the SVM classifier. Select the original image in dataset shown in figure 2.

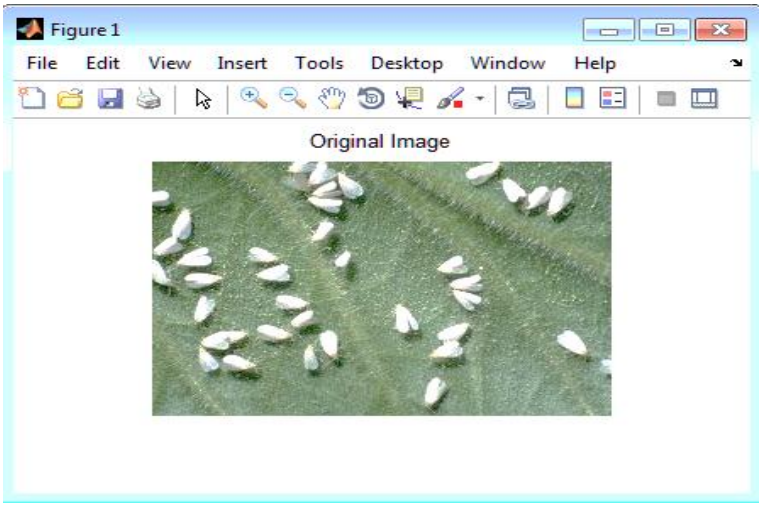

Figure2: A sample leaf image with whiteflies

B. Image Pre-Processing

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

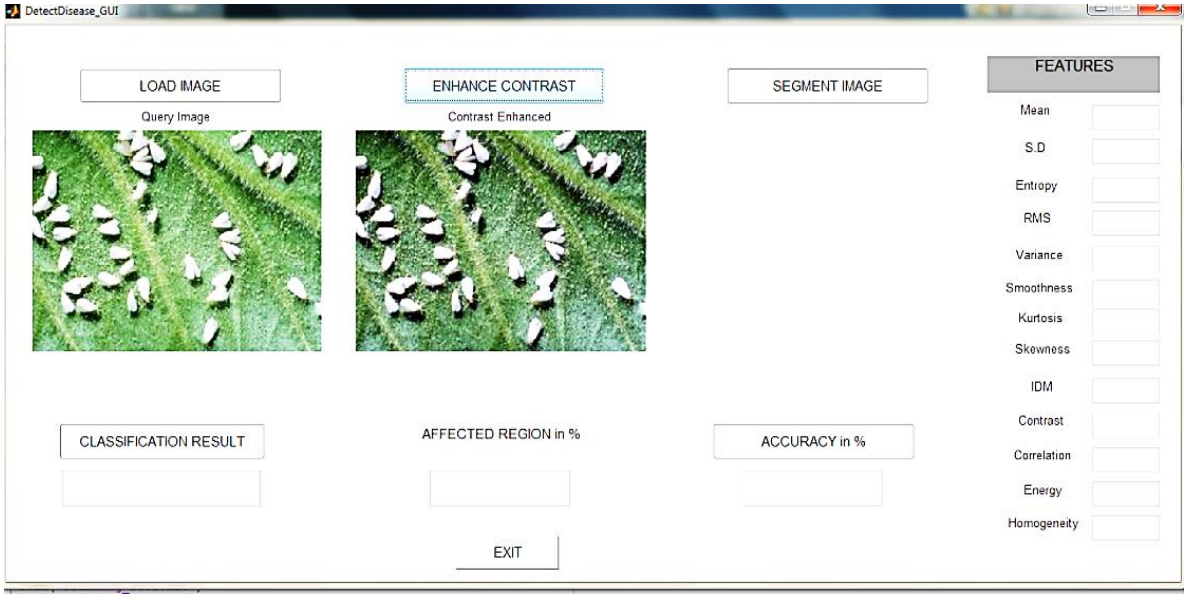

Figure 3: The leaf image after preprocessing

C. 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 4.

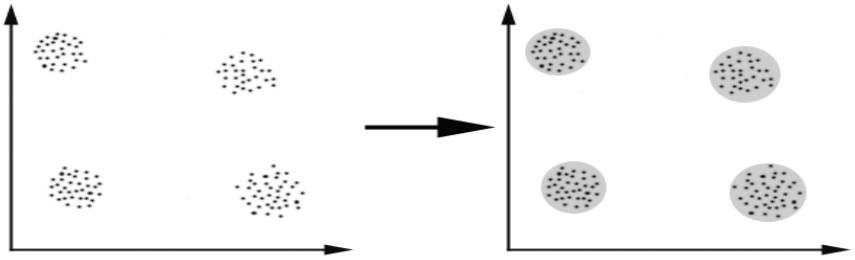

Figure 4: Clustering Process Diagram

In this case we easily identify the 4 clusters into which the data can be divided; the similarity criterion is distance: two or more objects belong to the same cluster if they are "close" according to a given distance (in this case geometrical distance). This is called distance-based clustering. Another kind of clustering is conceptual clustering: two or more objects belong to the same cluster if this one defines a concept common to all that objects. Clustering refers to the process of grouping samples so that the samples are similar within each group. The groups are called clusters. Clustering is a data mining technique used in statistical data analysis, data mining, pattern recognition, image analysis etc. Different clustering methods include hierarchical clustering which builds a hierarchy of clusters from individual elements. Because of its simplicitly and efficiency, clustering approaches were one of the first techniques used for the segmentation of (textured) natural images. In partition clustering, the goal is to create one set of clusters that partitions the data in to similar groups. Other methods of clustering are distance based according to which if two or more objects belonging to the same cluster are close according to a given distance, then it called distance is based clustering. In our work we have used K-means clustering approach for performing image segmentation using Mat lab software. A good clustering method will produce high quality clusters with high intra-class similarity and low inter-class similarity. The quality of clustering result depends on both the similarity measure used by the method and its implementation. The quality of a clustering method is also measured by its ability to discover some or all of the hidden patterns. Image Segmentation is the basis of image analysis and understanding and a crucial part and an oldest and hardest problem of image processing. Clustering means classifying and distinguishing things that are provided with similar properties. Clustering techniques classifies the pixels with same characteristics into one cluster, thus forming different clusters according to coherence between pixels in a cluster. It is a method of unsupervised learning and a common technique for statistical data analysis used in many fields such as pattern recognition, image analysis and bioinformatics. There are a lot of applications of the K-mean clustering, range from unsupervised learning of neural network, Pattern recognitions, Classification analysis, Artificial intelligent, image processing, machine vision, etc.

D. SVM Classifier

A Support vector machine is a powerful tool for binary classification, capable of generating very fast classifier function following a training period. There are several approaches to adopting SVMs to classification problems with three or more classes. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. SVM are inherently two class classifiers. The traditional way to do multiclass classification with SVMs is use one of the methods. The classifier evaluation consist the output value higher than the threshold area recorded as "true" and any SVM output value lower than the threshold are recorded ad "false". The SVM classifier consist the binary classification of images.

Table 1. 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 pest infected area accuracy is calculated after the k–means classification of segmentation regions using machine learning approaches such as multiclass SVM in SVM Classifier.

E. Accuracy of Infected Area

The pest such as whiteflies, aphids and thrips are very small in size and affect the leaves. Normally the size of adult 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 entire life cycle of whiteflies is 21-36 days. Thrips pest is 1/25-inch long in length. They range in colour from light brown to black. Thrips grows on flower and fruits plan leaves. The thrips and aphids are counts in leaf image by manually are not easy. The pest infected region accuracy calculated by the SVM classifier.

F. Parameters for Performance Evaluation

Effectiveperformance of K-Means Clustering Methods is evaluated 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) (1)

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) (2)

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

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

E= entropy (I), Entropy is defined as –sum (p.*log (p)) (3)

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) (4)

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) (5)

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

y = skewness(X) (6)

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

Contrast $= \sum_{i,j=0}^{N-1} (i,j)^2 \, \mathrm{c(i,j)}$ (7)

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.

$$\text{Correlation}= \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{\{i\times j\}\times P(i,j)-\{\mu_x-\mu_y\}}{\{\sigma_x-\sigma_y\}} \quad (8)$$

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

$$\text{Energy} = \sum_{k=0}^{n}\binom{n}{k}x^k a^{n-k} \quad (9)$$

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.

$$\sum_{i,j=0}^{N-1} C(i,j)/(1+(i-j)^2) \quad (10)$$

IV. Results and Discussions

The input image is collected manually using digital camera. The captured image is preprocessed. The preprocessed image and the image after classifying the pest infected region by using k-means color based segmentation technique are given. Finally the accuracy of infected region is calculated by SVM classifier shown in figure 5.

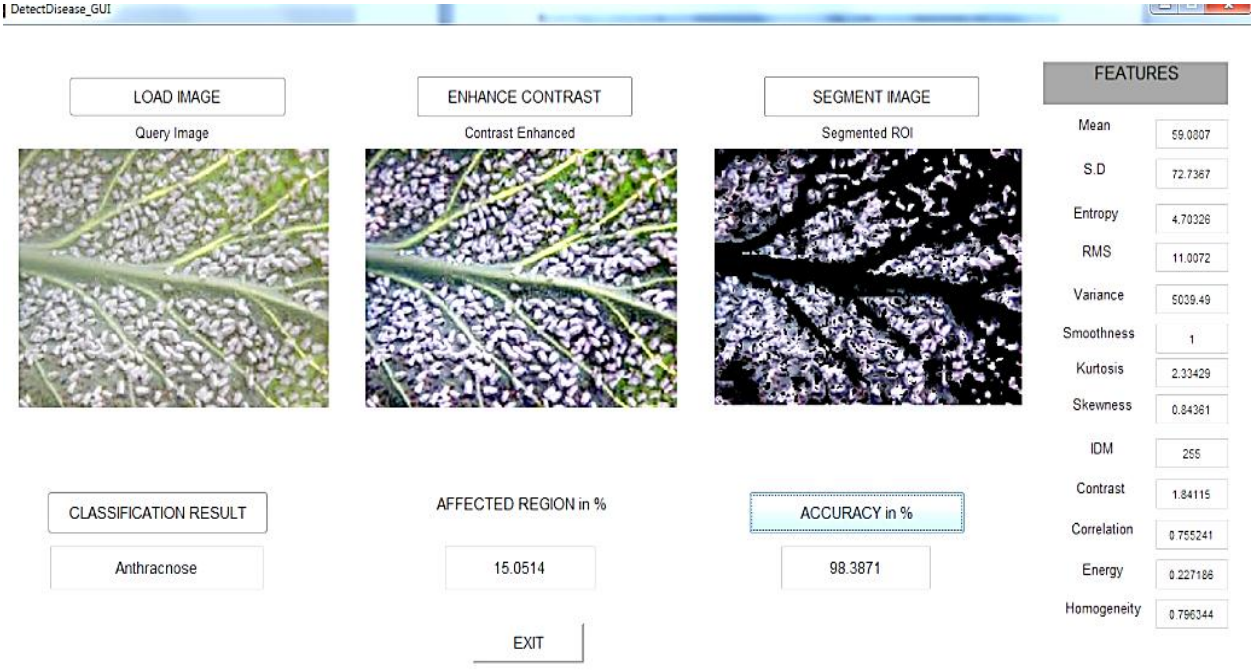

Figure 5: Accuracy of pest infected area in leaf image by using svm classifier- image I

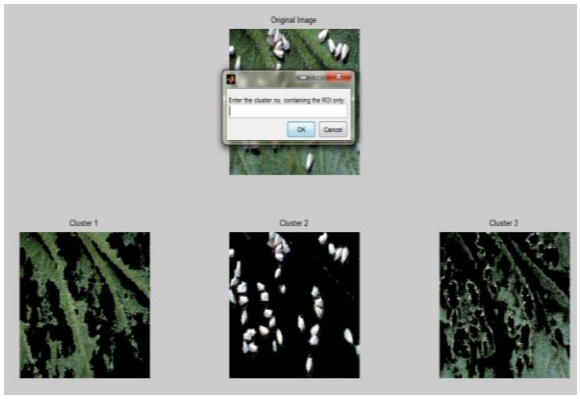

Figure 6: After Classification selected the clear Pest Image– Image II

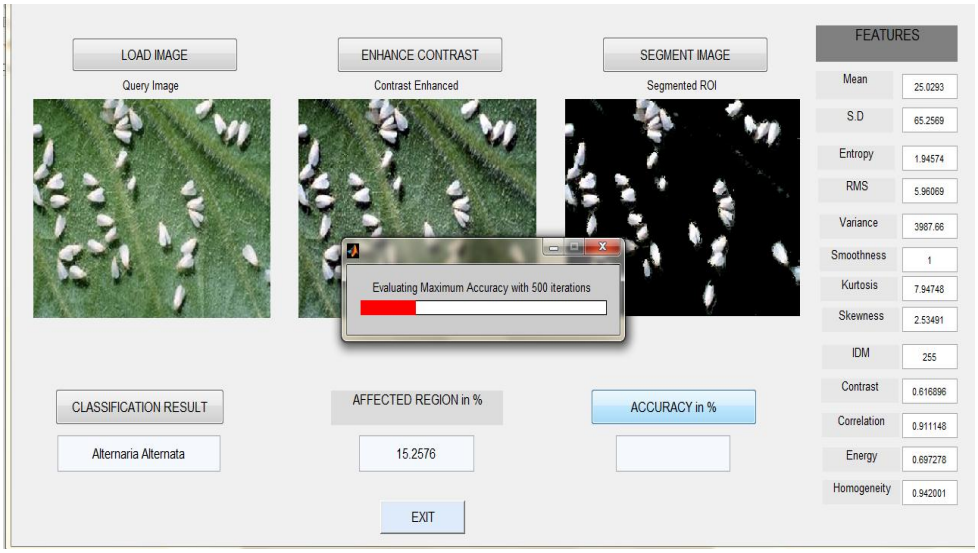

Figure 7: Value of parameter and accuracy calculated for the image II

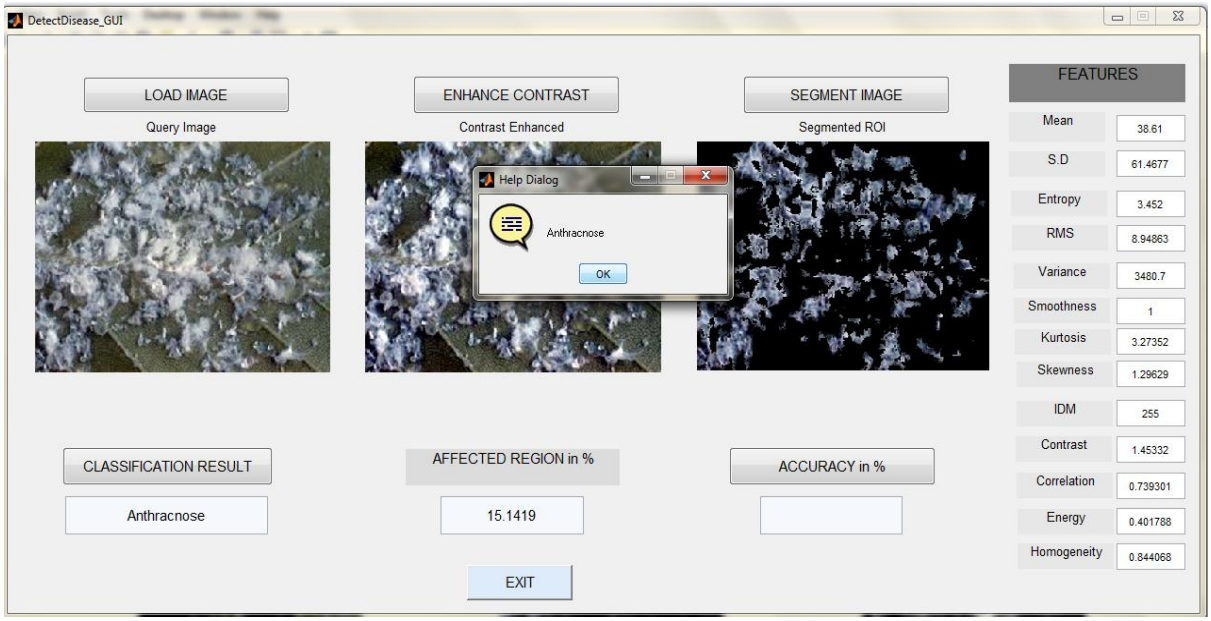

Figure 8: Accuracy of pest infected area in leaf-Image III

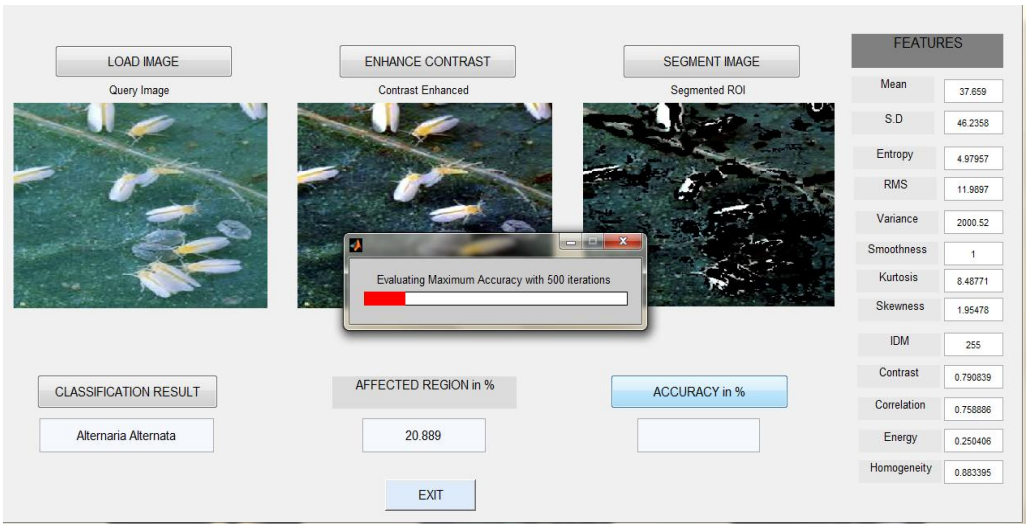

Figure 9: Accuracy of pest infected area in leaf-Image IV

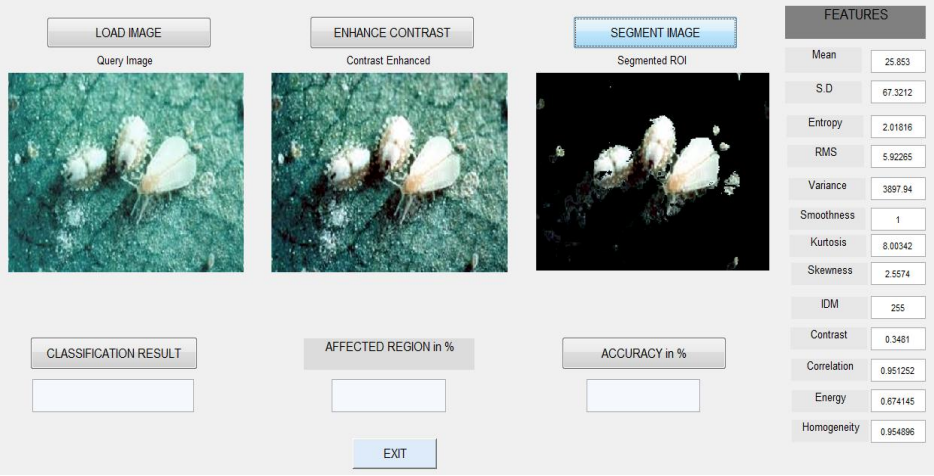

Figure 10: Apply segmentation of leaf pest infected leaf-images V

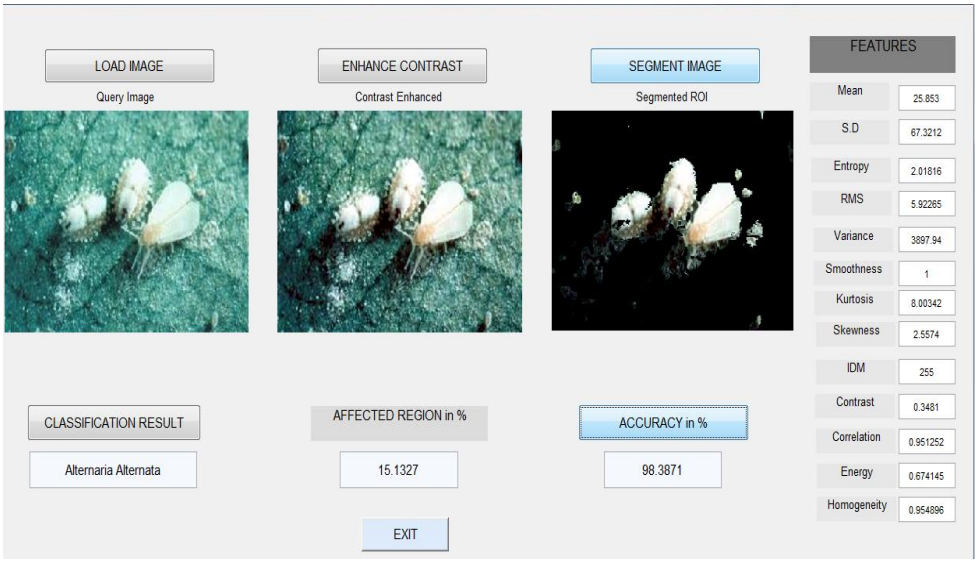

Figure 11: Value of parameter and accuracy calculated for the leaf image-Image V

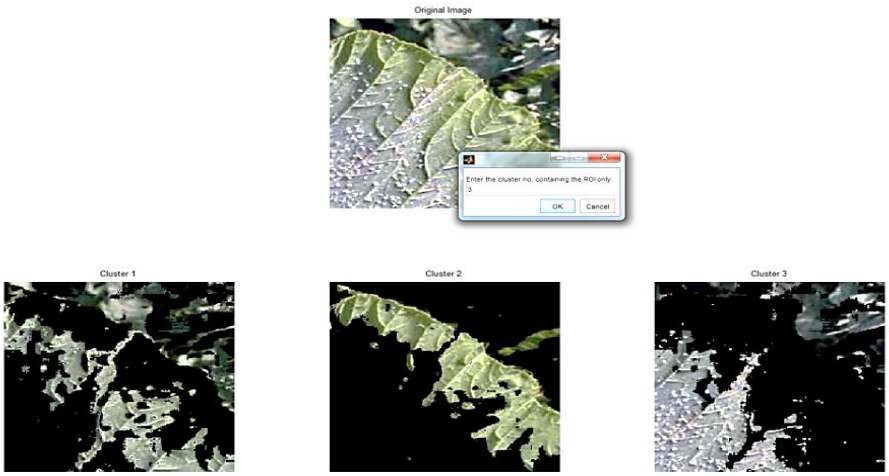

Figure 12: Selected Clear Pest Image form classification result of image – Image VI

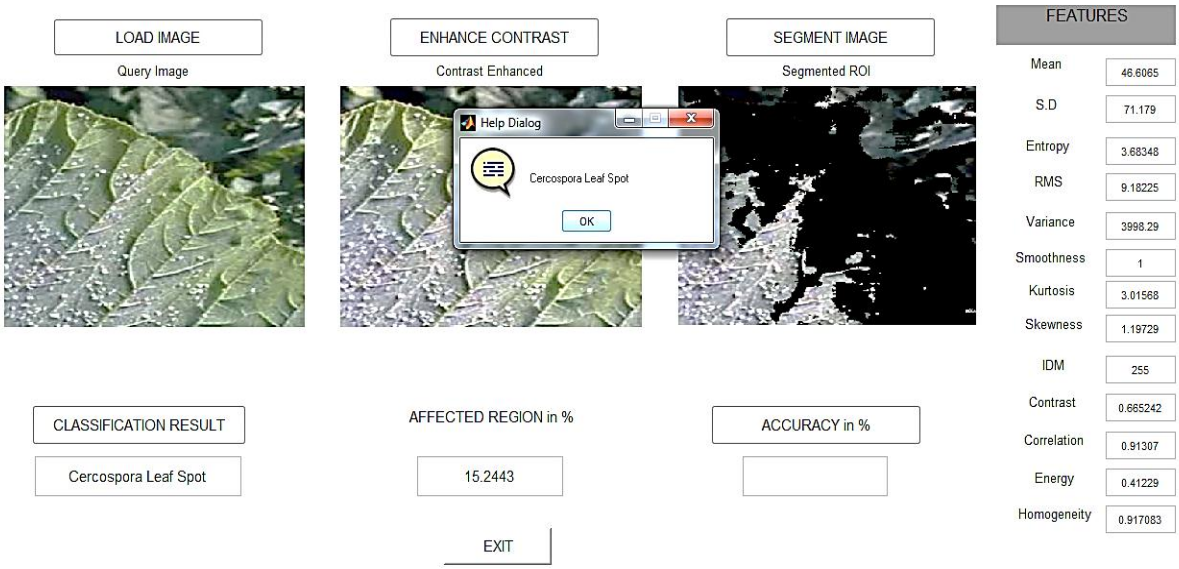

Figure 13: Classification result for the pest infected leaf image – Image VI

Table 2. Parameter values for Image-I

Parameter	**Values**
Mean	24.857
Standard Deviation	65.1178
Entropy	1.92977
RMS	5.92785
Variance	3972.57
Kurtosis	8.01108
Skeewness	2.54772
Contrast	0.619332
Correlation	0.910402
Energy	0.699636
Homogeneity	0.942531

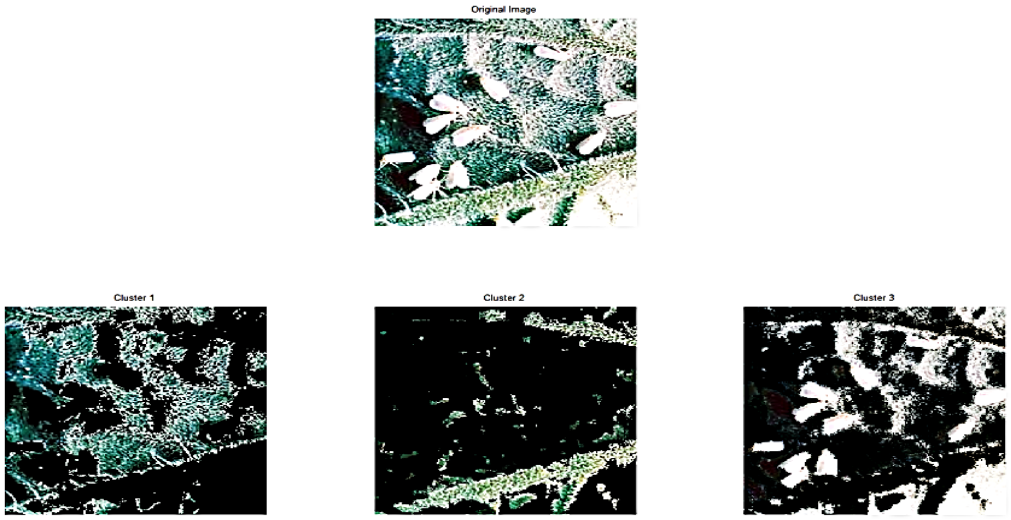

Figure 14: Apply the Classification for pest infected leaf image– Image VII

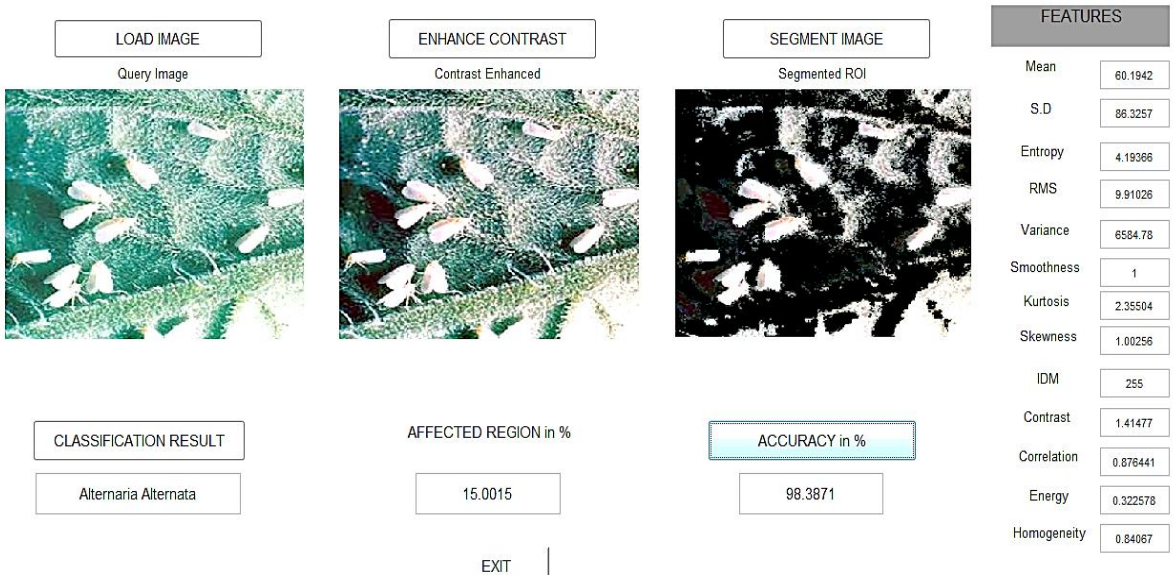

Figure 15: Accuracy of pest infected area in leaf-Image VII

Table 3. Parameter values for Image-II

Parameter	**Values**
Mean	25.098
Standard Deviation	66.2569
Entropy	19.4575
RMS	5.060069
Variance	3987.66
Kurtosis	255
Skeewness	0.616696
Contrast	0.911148
Correlation	0.6789
Energy	0.697278
Homogeneity	0.942001

Table 4. Parameter values for Image-III

Parameter	Values
Mean	38.61
Standard Deviation	62.4677
Entropy	3.452
RMS	8.9463
Variance	3480.7
Kurtosis	3.2352
Skeewness	1.29629
Contrast	0.296032
Correlation	0.904795
Energy	0.771095
Homogeneity	0.883

V. FUTURE WORK

As for future work, better clustering techniques to produce the accurate results may be aimed. K-means is not efficient because the mean of the cluster changes in the every iteration. The future work includes automatic determination of number of clusters required to segment the color images more accurately.

VI. CONCLUSION

Image processing technique plays an important role in the detection of the pests. The pest such as whiteflies, aphids and thrips are very small in size and affected the leaves. Normally the size of adult 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. Thrips pest is 1/25-inch long in length. They range in colour from light brown to black. The whiteflies, thrips and aphids measures are undertaken such as manual observation of pest infected leaves. This method does not give accurate measures. Hence automatic detection is very much important for early detection of pests. This work's main objective is to detect the pest infected region accuracy in various leaf images. The color based segmentation using k-means 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 andRichard E. Woods, "Digital Image Processing", 2nd ed., Beijing: Publishing House of Electronics Industry, 2007.

[2] S. Arivazhagan, R. Newlin Shebiah, 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] Shalihatun Azlin Aziz, "Image classification of brain MRI using support vector machine"2100 IEEE International Conference on Imaging Systems and Techniques, IST 2011-Processings 05/2011;DOI:10.1109/IST.2011.5962185

[4] Rajeshwar Dass, Priyanka and Swapna Devi, "Image Segmentation Techniques", IJECT Vol. 3, No1, 2012.

[5] Sanjay B.Dhaygude and Nitin P.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.

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

[7] 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.

[8] Jongman Cho, Junghyeon Choi, Mu Qiao, Chang-woo Ji, Hwang-young Kim, Ki-baikUhm, and Tae-soo Chon, "Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis", International Journal of Mathematics and Computers in Simulation.

BIOGRAPHY

Dr. R. Uma Rani completed her M.C.A from Regional Engineering College, Trichy, M.Phil from Mother Teresa University, Kodaikanal and Ph.D., from Periyar University in 2006. Her area of interest includes network security, data mining, image processing and mobile communications. She has published 120 research articles in international journals and authored 3 books. She has acted as resource person in national and international conferences and chaired technical sessions too.

P. Amsini completed her M.C.A in 2015 from Kongu Engineering College, at Perundurai, M.phil computer science from Sri Sarada College for Women, Salem in 2016. Her area of interest includes image processing.