Fruit & Leaf Disease Detection Using Image Processing, K-Means Clustering and SVM Classifier

Khushbu kumari, MTech Scholar, SAM College of Engineering and Technology

Bhopal (MP)

Dr. Neelesh Jain, Professor, SAM College of Engineering and Technology

Bhopal (MP)

Bity Merin Anish, Assistant Professor, SAM College of Engineering and Technology

Bhopal (MP)

ABSTRACT: The purpose of this study is to examine how different image processing techniques may be used to identify leaf disease. To identify and classify plant leaf diseases, different algorithms may be used to digital image processing, which is a quick, reliable, and accurate approach. Classifiers and support vector machines for illness classification are among the strategies presented in this study effort that have been employed by several authors to identify disease. Our research primarily focuses on the evaluation of several methods for detecting leaf disease and also gives an overview of various image processing methods. Fruit disease is also discussed in this study as a potentially disastrous issue that has the potential to harm both the economy and the agricultural industry. Because of technological advances, advanced image processing algorithms have recently been created to help identify contaminated fruit that was previously detected by hand. There are two stages: the first is for training, and the second is a testing phase. Data on infected and uninfected fruit is collected during the training phase, and during the testing phase, it is determined whether or not the fruit has been infected, and if so, by which disease. The various methods currently in use to identify infected fruit are examined in this piece of research. Farmers benefit from the use of these methods since they assist to identify fruit disease in its earliest stages.

KEY WORDS: Fruit & Leaf diseases, Segmentation, Image processing, features extraction, K-means Clustering, SVM.

1. INTRODUCTION

Agricultural production is the backbone of Indian economy. The scientific theory of plant illnesses caused by microorganisms and other environmental circumstances is known as plant pathology [1]. Direct and indirect infection diagnosis technologies have been employed in agriculture in recent years, however they are quite pricey it can only be used by qualified people. These techniques require a lot of work and also needs an extreme amount of time and cost. Even though there are automatic detection systems available, it suffers in term of accuracy which is very much essential in Computer Aided Detection (CAD) systems. This motivates for a low-cost automatic plant disease detection system which identifies the plant disease in a very accurate way. Also, a robotized system has been designed to bring out research work in an efficient way. In the agriculture area, leaf disease identification has been studied intensively by several researchers for more than two decades [2-3]. However, a large number of works have been done previously, but many issues are still open and deserve further research. The common issue like identifying the disease in a particular plant leaf have been sorted out in the proposed model. Proposed model satisfies the need of being more accurate and efficient. Considering these issues, this thesis provides a step into solving disease identification in all plant leaves. This has been carried out via a schematic technique in a very cost-effective manner [4].

1.1 Agriculture Image Processing

Agriculture is the backbone of Indian economy. Agricultural Image Processing is one of the core applications of Image processing and the most growing research area. Computer vision has much been a useful technique for analyzing data in a variety of industries, including agriculture [5]. Image processing in agriculture is done by capturing images through cameras, aircraft or satellites. These images are then processed and analyzed using computers via image processing techniques. It is made easy with new technological advancements in image capture and data processing to solve various problems in the fields of agriculture.

A digital system will improve the picture. Image is a two-dimensional collection of brightness values that is often represented by texture, shape and color. A computer can digitally process a picture. To digitally process a picture, it must first be reduced to a set of numbers that the computer can alter [6-7]. Pixels are numbers that indicate the brightness of a picture at a certain point. Digitized photos are typically 512 by 512 pixels (250,000 pixels), while bigger images are also popular. After the picture has been digitized, the computer may execute three fundamental operations on it. There is a correlation between a pixel's output value and the input image's pixel value. A pixel's value is decided by its neighbours in the input image while performing local operations. All of the pixels in the original picture are reflected in the final image pixel value. Any one or a combination of these methods may be used to enhance, repair, or compress the image [8].

These are just a few examples of how image processing may be utilised in agriculture: Identify the illness by analysing the damaged area's colour, shape, and size, and then remove the sick leaf and stem.

2. Agriculture Image Processing in Applications

There are a number of applications of image processing in the agriculture industry of which few common applications (shown in fig 2.1) are given below [9-10]:

- Crop Management (e.g., pest management detection).
- Identifying nutrient deficits as well as plant composition (e.g., Nutrient deficiencies and of plant contents).
- Inspection, sorting, and grading of fruits (e.g., categorization of agricultural goods based on fruits and vegetables).
- Crop and terrain estimate, as well as object tracking (for illustration, using a Geographic Information System) (GIS).
- Plant disease detection (e.g., Identifying plant disease based on color, shape, texture, etc).

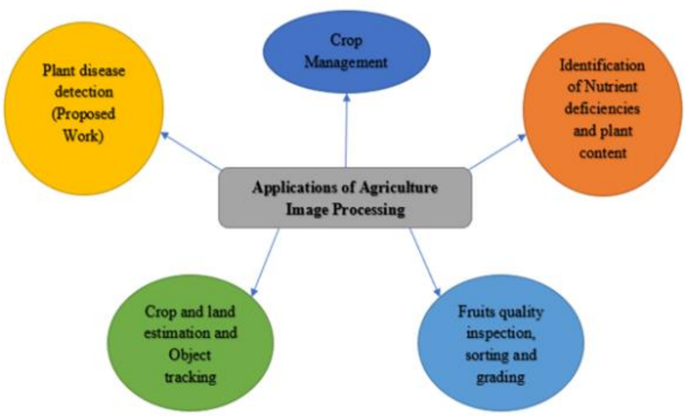

Fig 2.1 Applications of Agriculture Image Processing

2.1 Types of Plant Diseases

Plants can be categorized by their similar features. Disease identification in plants can be done on the basis of plant type. Most plant diseases are caused due to three main reasons which are bacterial, fungal and viral as shown in fig 2.2.

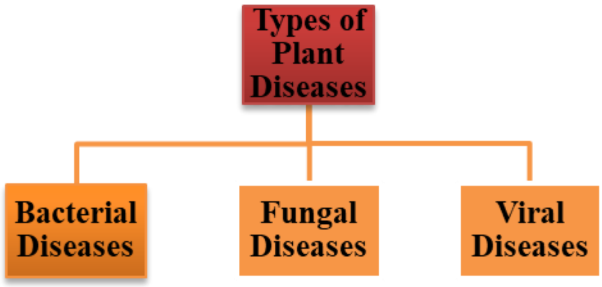

Fig 2.2 Types of Plant Diseases

3. METHODOLOGY

3.1 Introduction to Image Processing

Picture processing is a technique for applying operations on the data in order to improve it or extract relevant information from it. It's a sort of signal analysis in which the input is an image as well as the output is either that picture or its related characteristics. Image processing is one of the most quickly evolving technology today [11]. Image processing is to improve its quality and to modify it through restriction for further use in different applications through its enhancement, segmentation, feature extraction, classification, etc. Image enhancement is a process of adjusting the brightness, changing the tone of the color, removing noise and sharpening the image. Image segmentation is the process of dividing an image into multiple parts. This is typically used to identify objects in digital images. There are many different ways to perform image segmentation, including, threshold, color-based, transform, texture-based [12]. Feature extraction is a sort of dimensionality reduction in which important sections of a picture are efficiently represented as a compact feature vector. When picture files are huge as well as a reduced feature representation is needed to accomplish tasks like image matching and retrieval rapidly, this technique is effective. Image classification refers to the labelling of images into one of a number of predefined categories. In the classification categorized into two parts such as, supervised and unsupervised. Digital image processing focuses on some key aspects, Improvement of pictorial information for human interpretation, Processing of image data for storage, transmission, and representation for autonomous machine perception [13].

3.2 Image Processing Techniques

Steps may have sub-steps, even if they are essential. There are some basic steps [14-15].

- **Image Acquisition -** It might be as easy as being provided a picture that has already been converted to digital format. Pre-processing, like as scaling, is usually done at the picture capture stage.
- **Image Enhancement -** Enhancing a photograph digitally is a basic and aesthetically pleasing technique. Improving a photograph by altering its brightness and contrast might bring out previously hidden details or simply draw attention to certain aspects of the image, among other things.
- **Image Restoration -** Image restoration is the process of enhancing the look of a photograph. The majority of picture restoration approaches are based on algebraic or probabilistic image degradation models.
- **Color Image Processing -** The use of digital photos via the Internet has increased significantly as a result of color image processing. Color modelling and processing in the digital environment are examples of this.
- **Wavelets and Multiresolution Processing -** Using wavelets, you may create images with various resolutions. Images are split up into smaller chunks for the purposes of compression and pyramidal representation.
- **Morphological Processing -** A morphological processing technique extracts picture components important for representing and describing form.
- **Segmentation -** To segment a picture, it is divided into its component components. Self-segmentation in digital image processing is notoriously difficult. Individual objects may be identified using a robust segmentation method, which makes the procedure suited for imaging difficulties.
- **Representation and Description -** Normally, the output of a segmentation phase is raw pixel data, however in certain cases the region's boundaries or all of the points inside it may be represented and described. Solutions for transforming raw data into formats that may be handled by the system include representation. The process of extracting characteristics yields quantitative data, which is referred to as description. This is the foundation for distinguishing one object type from another.
- **Object recognition -** Labeling an object according on its characteristics, such as

the description of an automobile, is known as recognition.

❖ **Knowledge Base** - Guides the operation of each processing module and controls the interaction between modules. This knowledge interprets and deduces high-level information feature (semantic features) from low-level information feature (visual features).

3.3 Outline of the Proposed Research Work

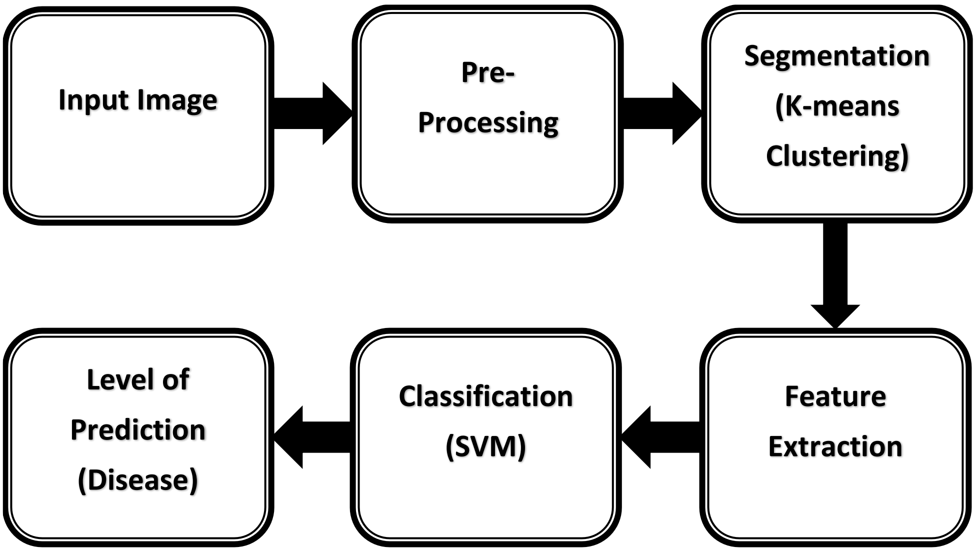

Figure 3.1 Outline of the Research Work

The proposed workflow method is shown in Figure 3.1. The processing methods include removing the background, petiole and unwanted noise using preprocessing. To segment the disease-affected region, the clustering Approach is first performed to the supplied fruit or leaf picture. Then, to represent each disease area, the numerous characteristics are retrieved. Finally, utilizing extracted characteristics, the previously trained classifier model is used to categorize the illness.

4. EXPERIMENTAL ANALYSIS AND RESULTS

4.1 Introduction

Different types of fruit and leaf diseases data sets are implemented in MATLAB 2019a Version to show the Region of Infection (ROI) as a percentage to detect whether a fruit or leaf has some damage or these are pre-symptoms of some defined diseases are as follows:

1. Alternaria Alternata
2. Anthracnose
3. Bacterial Blight
4. Cercospora Leaf Spot

4.2 Dataset Description

Dataset used in this work is collected from the Tamil Nadu Agricultural University at Tirunelveli and also uses the images from the public dataset “plant image analysis”. The dataset has been collected from weblink https://tnau.ac.in . The dataset fruit or leaf images that are collected under four categories namely fruit crops, vegetable crops, cereal crops, and commercial crops. These images are further categorized using the diseases found in the crops. The dataset considers only the diseases that are common in most of the crops. It contains fungus, the viral and bacterial disease affected images. The dataset contains a total of 2912 images are collected from web source. The data sample of the diseases in fruits and plant leaf is shown in Table 4.1. Images of few samples is shown in Figure 1.6. 80% (2329) Trained data samples and 20% (583) Test data samples are splitted using K=5 (K-Fold Cross Validation).

Table 4.1 Distribution of Sample Images into Disease Classes

S. No.	Disease Name	Sample Found
1	Alternaria Alternata	708
2	Anthracnose	896
3	Bacterial Blight	677
4	Cercospora Leaf Spot	631

(a)

(b)

(c)

(d)

(e)

(f)

(g)

(h)

Figure 4.1 (a-h) Experimental Images of Leaf Disease Detection (a)Black Rot Canker (b) Gall Midge (c) Gray Mold (d) Leaf Minor (e) Brown Spot (f) Leaf Blast (g) Alternaria Leaf Spot (h) Bacterial Blight

4.3 SVM Implementation

The output quantity higher than the cutoff region is marked as "true," whereas any SVM resulting value lower than just the threshold is marked as "false," as indicated in Table 4.2. The binary categorization of pictures is stored in the SVM classifier.

Table 4.2 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 accuracy of the infected area is assessed after k–means classification of segmented areas using machine learning methods such as multi-class SVM in SVM Classifier.

4.4 Pictures for the Model when Run for Various Test Images

In this section various images are shown that were taken during the running of the program. It mainly includes pictures of clusterification, histogram of selected image with ROI, classification result, ROI % and accuracy of classifying the disease.

Prediction for Test Sample 1

Fig 4.2 (a) represents the clusters that were formed for the taken input image.

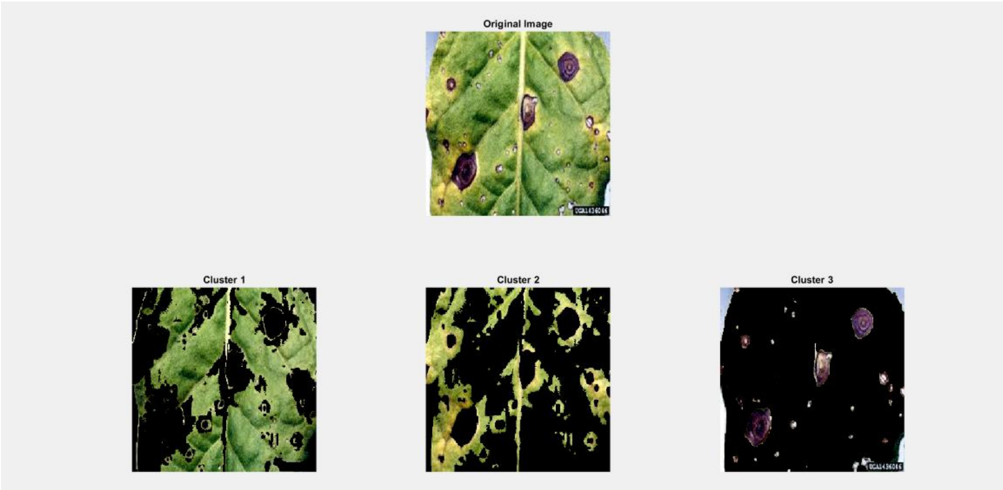

Fig 4.2 (a) Clusterification

Fig 4.2 (b) represents the histogram for the selected cluster with Region of Infection.

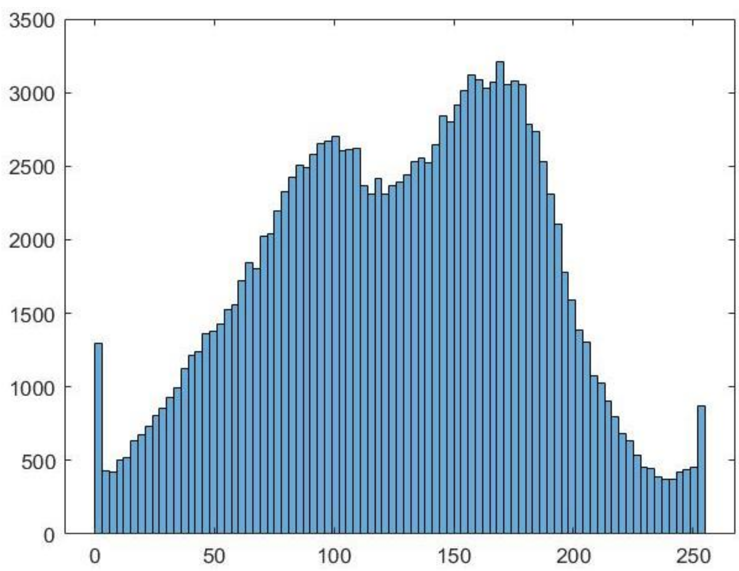

Fig 4.2 (b) Histogram

Fig 4.2 (c) represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

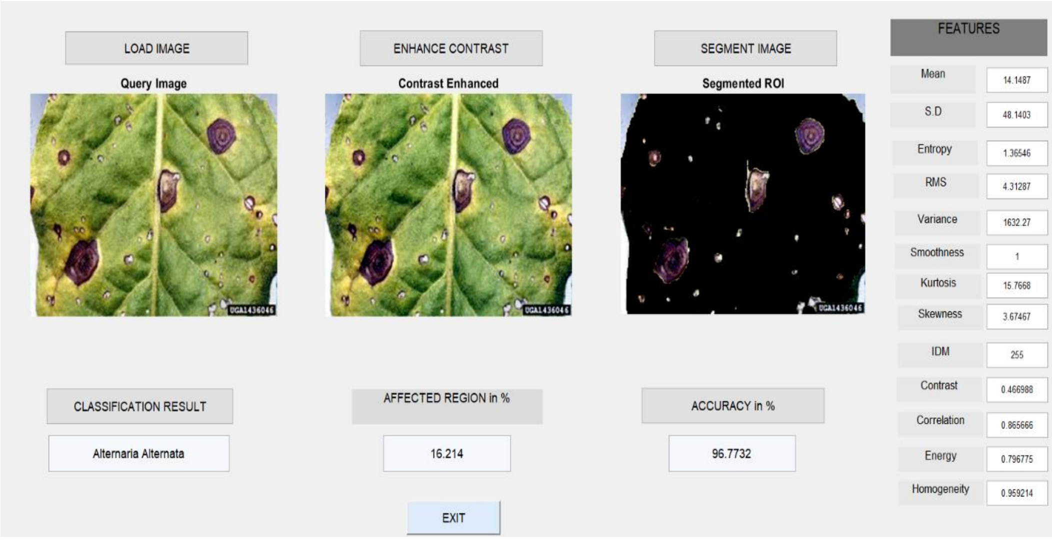

Fig 4.2 (c) Segment ROI

Disease Classification Result: Alternaria Alternata

'Affected Region Area is: 16.214 %'

Fig 4.2 (d) shows a pie chart representing the affected region in the sample.

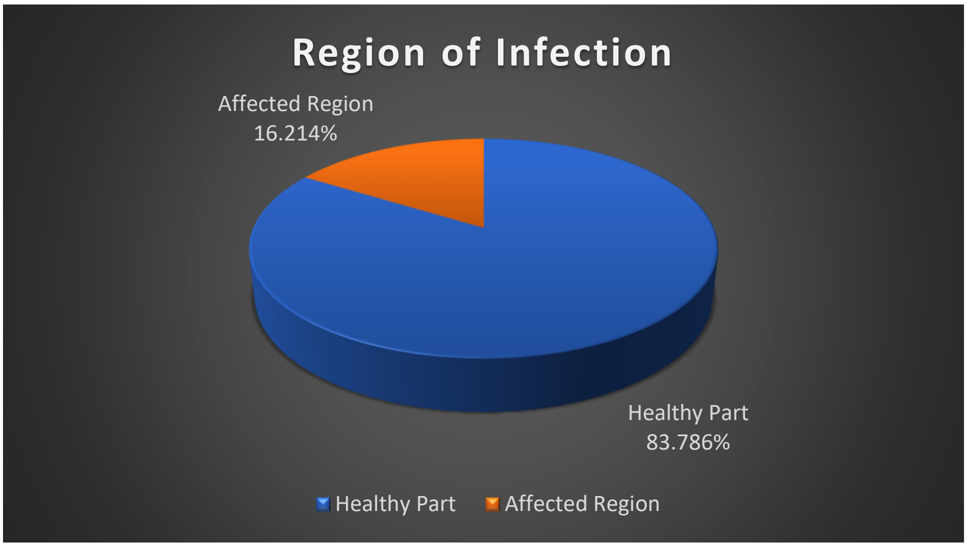

Fig 4.2 (d) Pie Chart

Prediction for Test Sample 2

Fig 4.3 (a) represents the clusters that were formed for the taken input image.

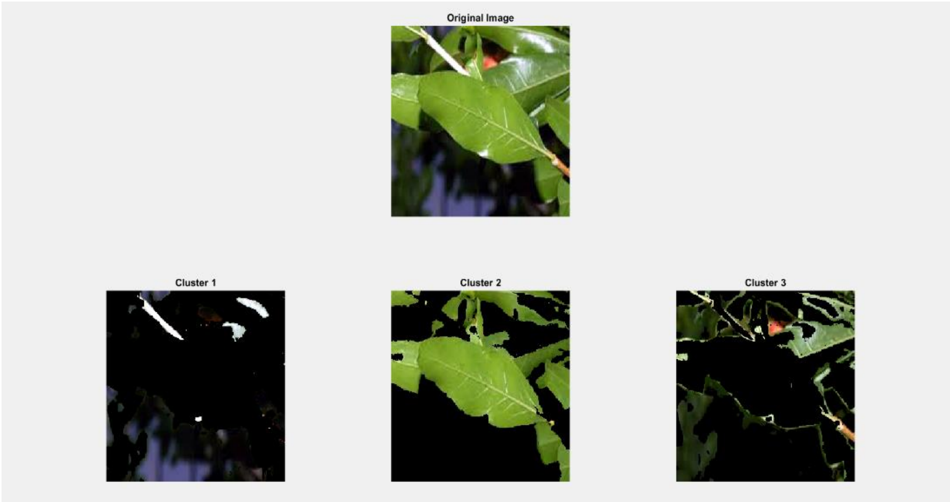

Fig 4.3 (a) Clusterification

Fig 4.3 (b) represents the histogram for the selected cluster with Region of Infection.

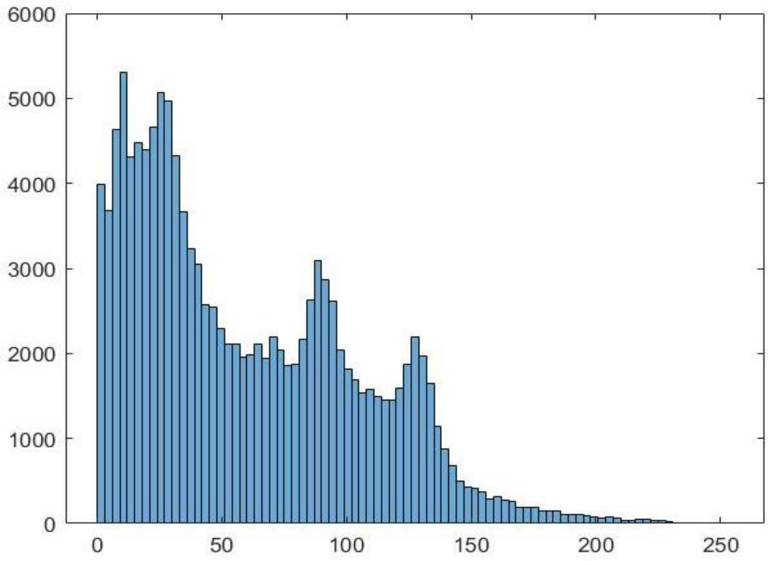

Fig 4.3 (b) Histogram

Fig 4.3 (c) represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

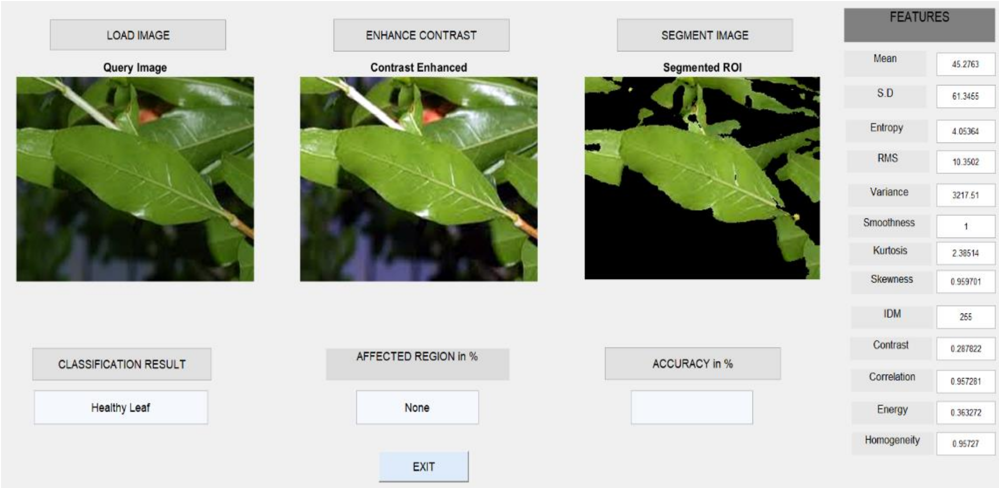

Fig 4.3 (c) Segment ROI

Disease Classification Result: Healthy Leaf

'Affected Region Area is: None'

Fig 4.3 (d) shows a pie chart representing the affected region in the sample.

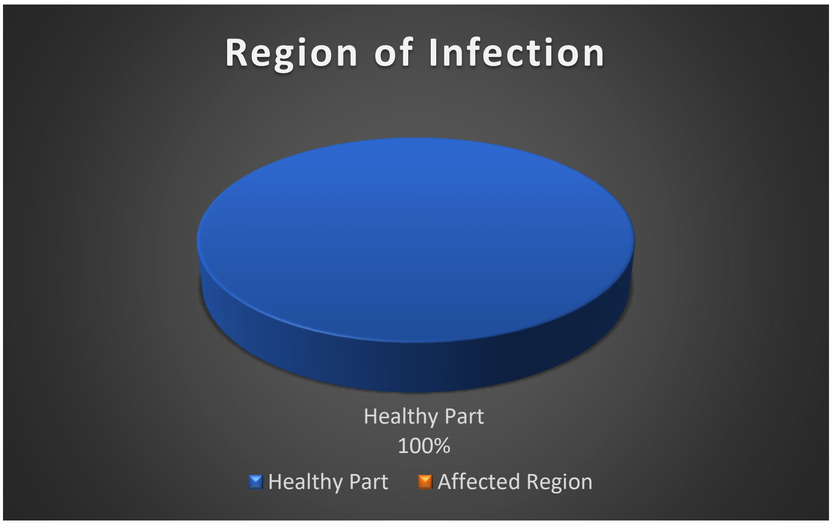

Fig 4.3 (d) Pie Chart

Prediction for Test Sample 3

Fig 4.4 (a) represents the clusters that were formed for the taken input image.

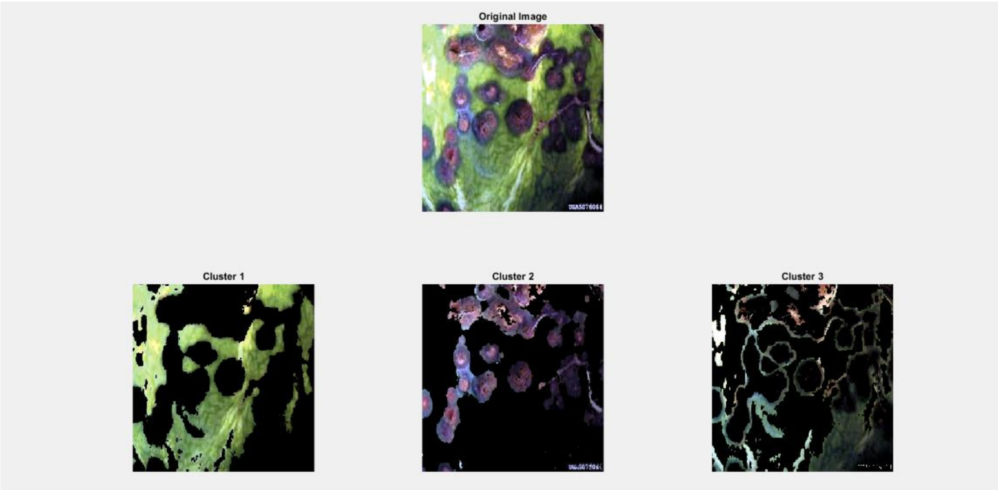

Fig 4.4 (a) Clusterification

Fig 4.4 (b) represents the histogram for the selected cluster with Region of Infection.

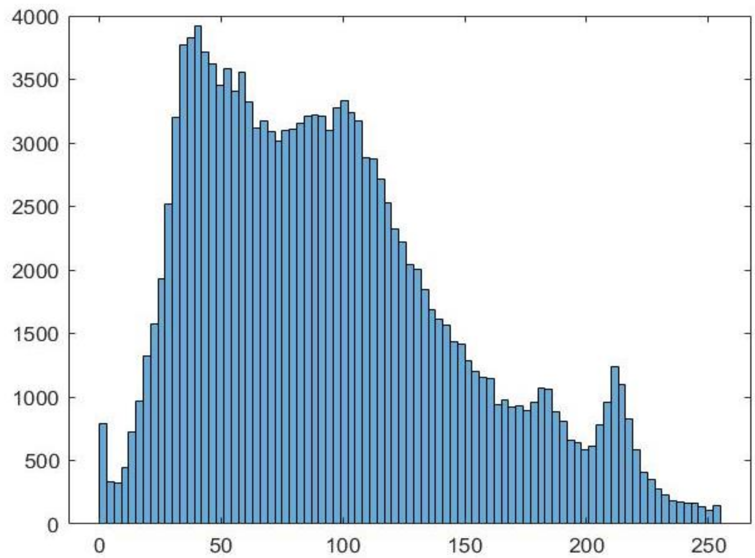

Fig 4.4 (b) Histogram

Fig 4.4 (c) represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

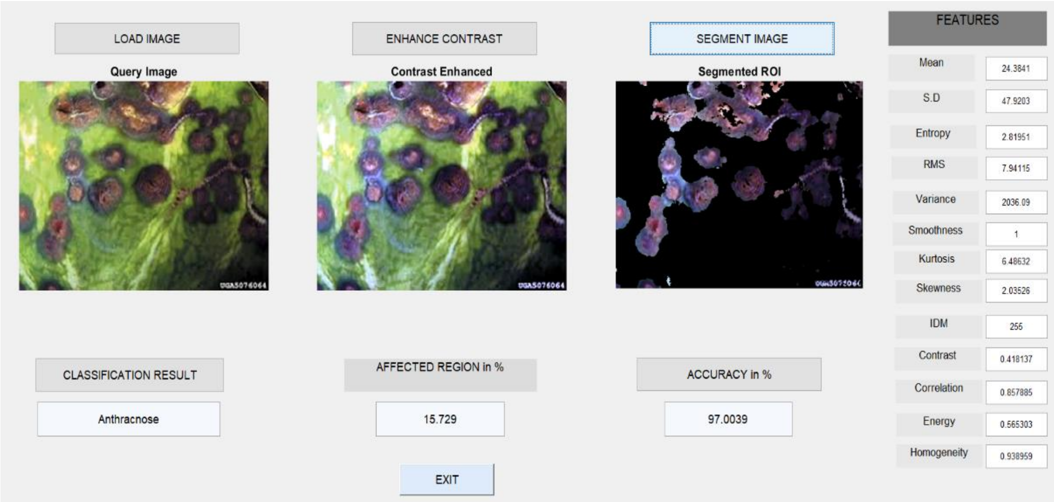

Fig 4.4 (c) Segment ROI

Disease Classification Result: Anthracnose

'Affected Region Area is: 15.729'

Fig 4.4 (d) shows a pie chart representing the affected region in the sample.

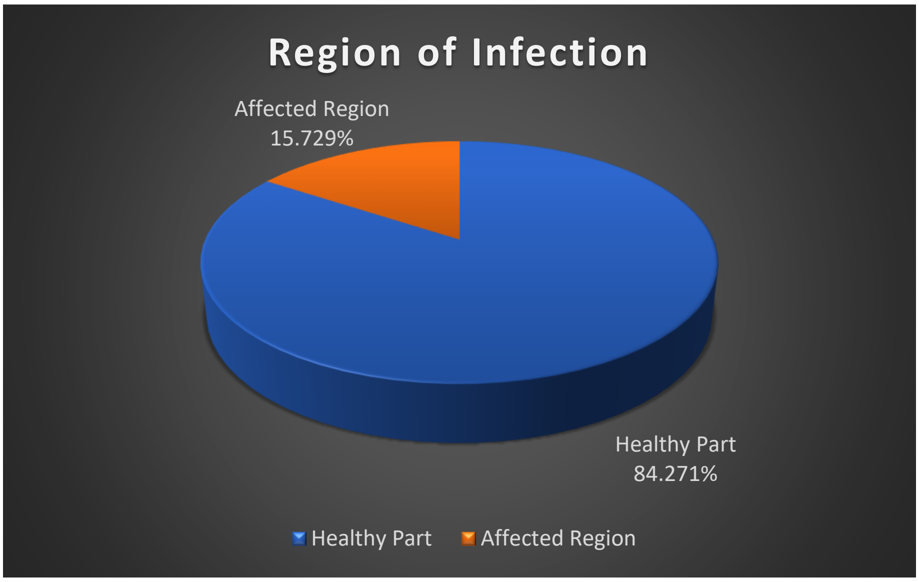

Fig 4.4 (d) Pie Chart

Prediction for Test Sample 4

Fig 4.5 (a) represents the clusters that were formed for the taken input image.

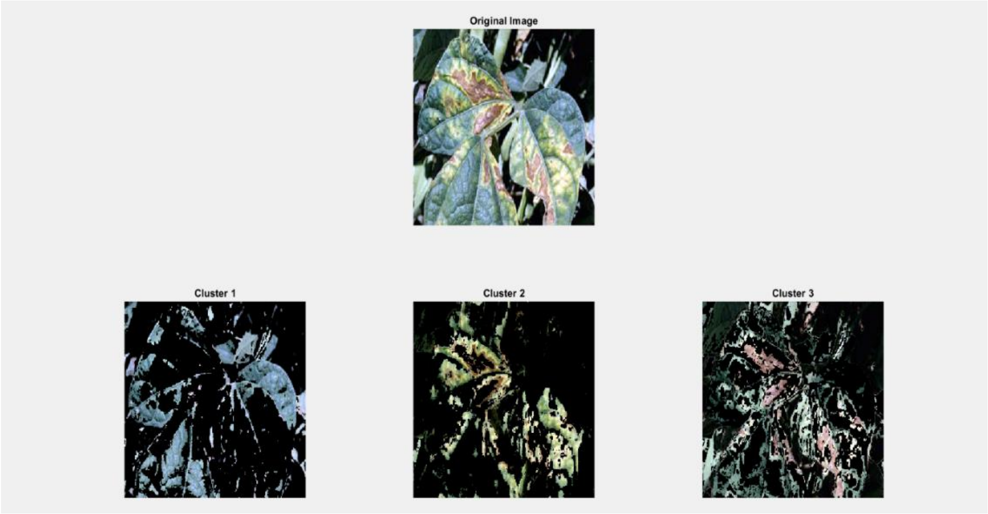

Fig 4.5 (a) Clusterification

Fig 4.5 (b) represents the histogram for the selected cluster with Region of Infection.

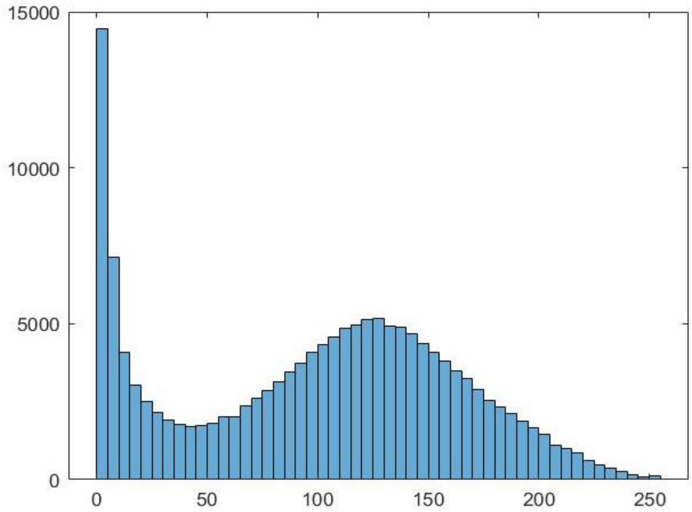

Fig 4.5 (b) Histogram

Fig 4.5 (c) represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

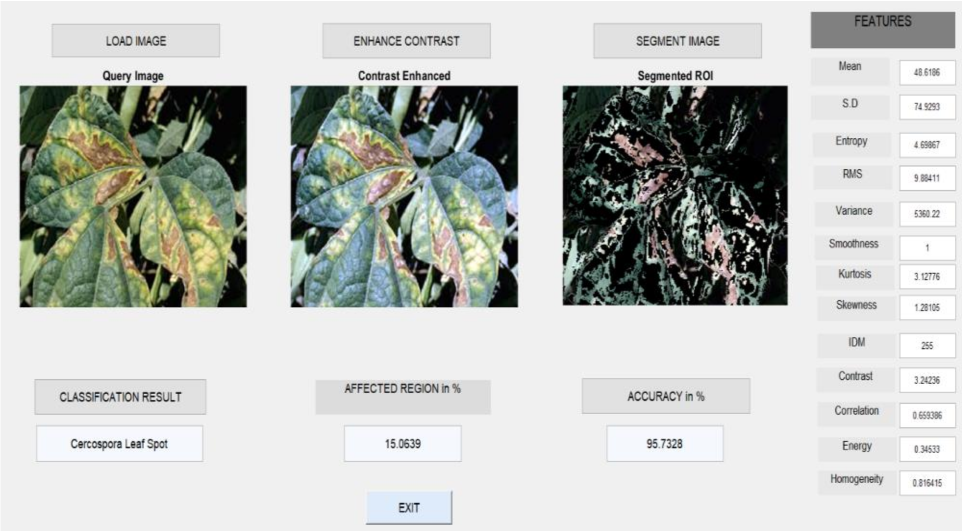

Fig 4.5 (c) Segment ROI

Disease Classification Result: Cercospora Leaf Spot

'Affected Region Area is: 15.0639'

Fig 4.5 (d) shows a pie chart representing the affected region in the sample.

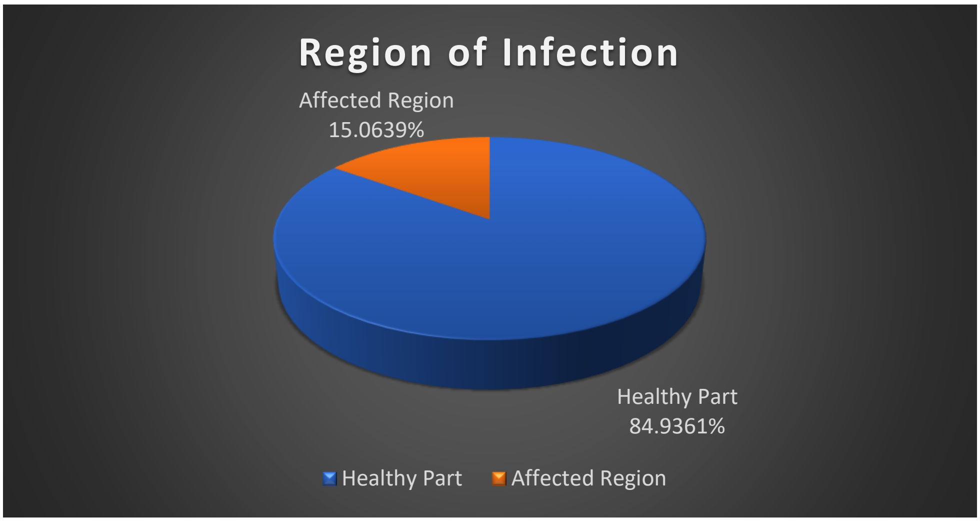

Fig 4.5 (d) Pie Chart

4.5 Pics of Various Other Results

Fig 4.6 represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

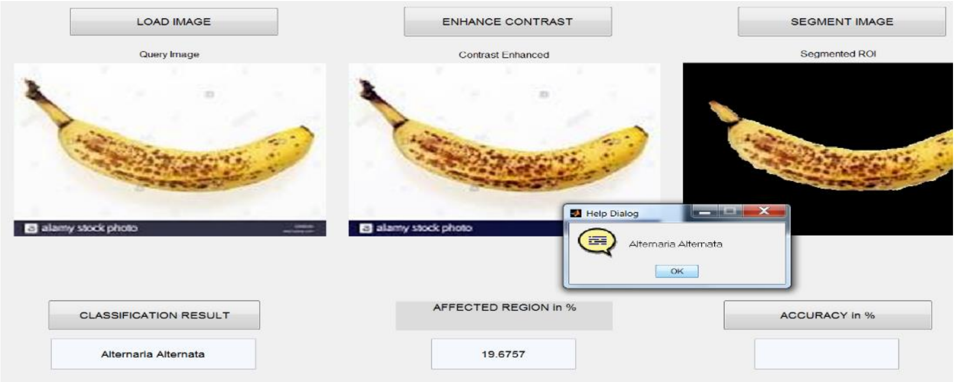

Fig 4.6 Segment ROI

Disease Classification Result: Alternaria Alternata

'Affected Region Area is: 15.0062%'

Fig 4.7 represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

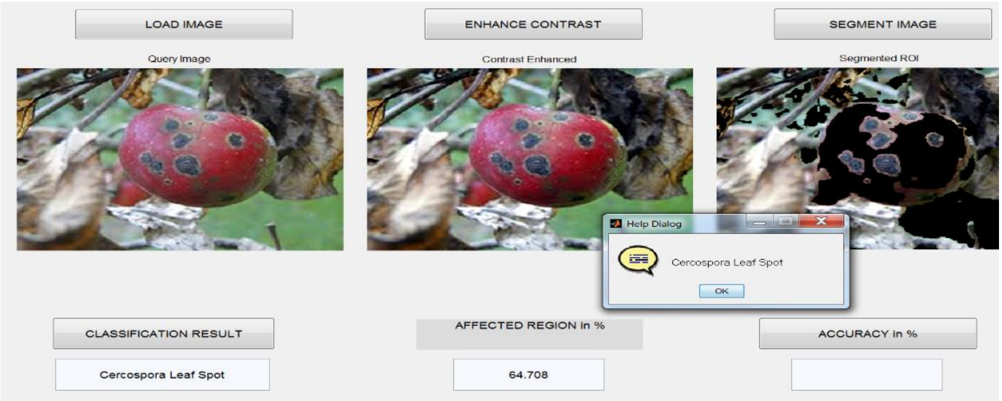

Fig 4.7 Segment ROI

Disease Classification Result: Cercospora Leaf Spot

'Affected Region Area is: 64.708 %'

Fig 4.8 represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

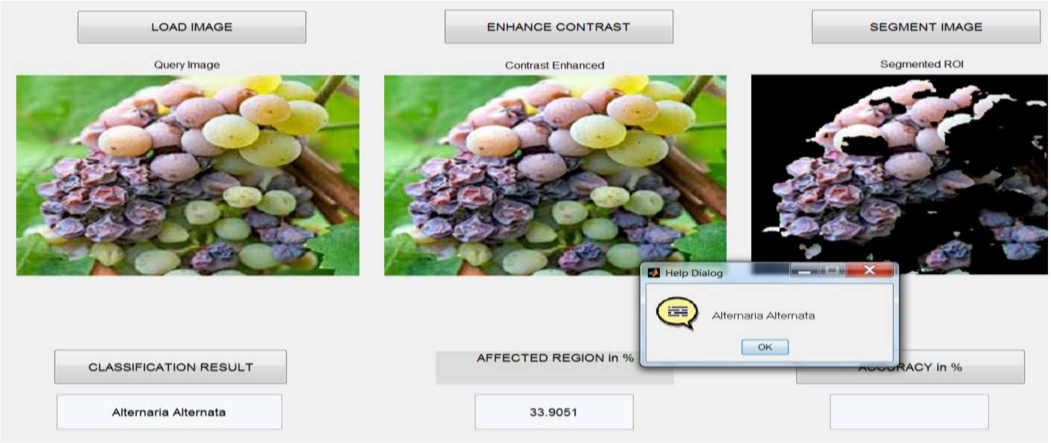

Fig 4.8 Segment ROI

Disease Classification Result: Alternaria Alternata

'Affected Region Area is: 33.9051%'

Fig 4.9 represents the region of infection in the leaf, classification result and the accuracy of classifying the disease.

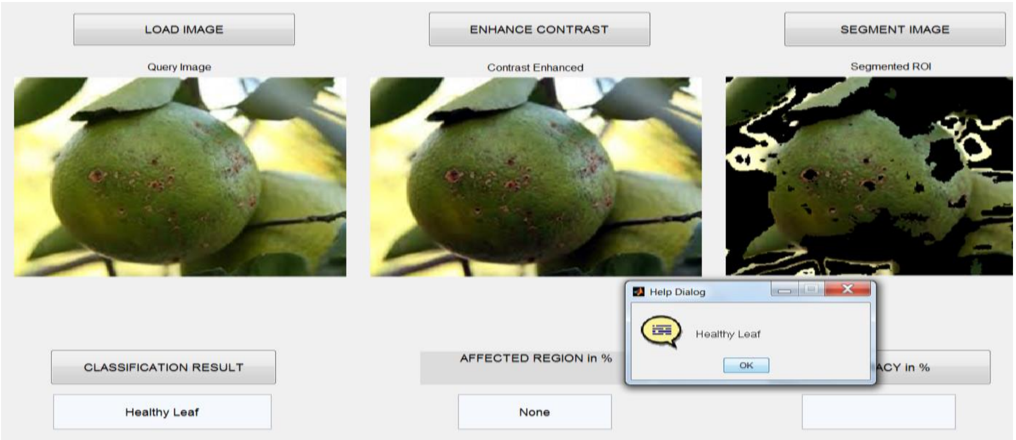

Fig 4.9 Segment ROI

Disease Classification Result: Healthy Fruit

'Affected Region Area is: None'

4.6 Experimental Results

TABLE 4.4 Result Outcome of Feature Extraction (K-means Clustering) Fruit/Leaf Disease – Disease and Region of Infection %

Fruit / Leaf	Disease	ROI %
Fruit	**Alternaria Alternata**	**15.0062%**
Fruit	**Cercospora Leaf Spot**	**64.708 %**
Fruit	**Alternaria Alternata**	**33.9051%**
Fruit	**Healthy Fruit**	**None**
Leaf	**Cercospora Leaf Spot**	**15.0639 %**
Leaf	**Healthy Leaf**	**None**
Leaf	**Anthracnose**	**15.729 %**
Leaf	**Alternaria Alternata**	**16.214 %**

The graph in fig 4.10 shows the comparison of various proposed works.

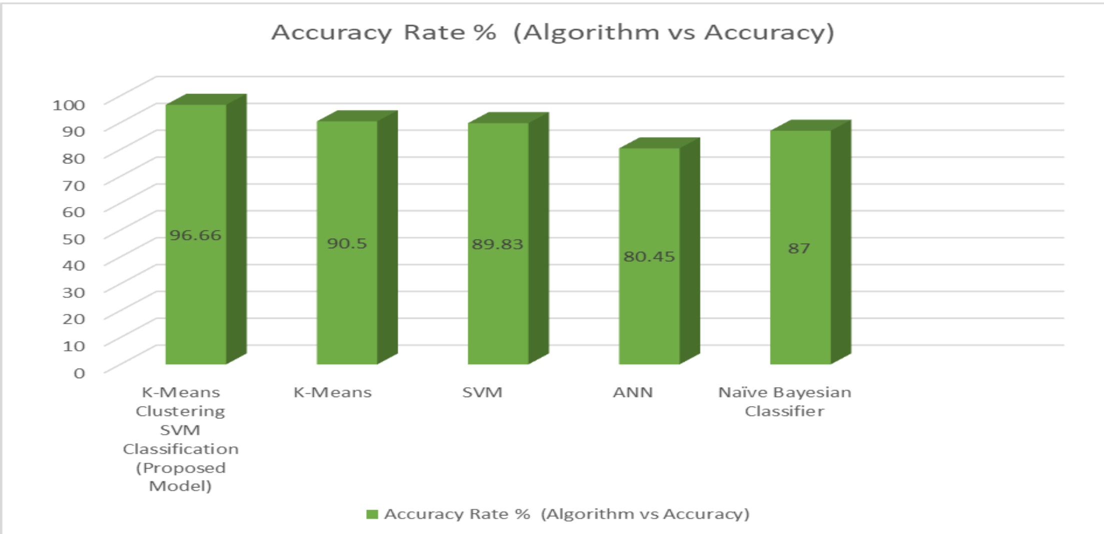

Fig 4.10 Bar Graph of Overall Accuracy of Proposed Model Compared with Other Existing Classifiers

5. CONCLUSION

It is suggested in this study to determine how much of a fruit or plant leaf is damaged, as well as to identify the illness in the defective fruit or leaf in the supplied image. This function is extremely beneficial to farmers and may be used for a variety of purposes. After tweaking the parameters of SVM, K-means clustering is employed to produce better outcomes in the classification and diagnosis of fruit illnesses. In this work, the researchers used K-means clustering with SVM to create a system that is not only good for identifying fruit or plant leaf disease, but also for detecting illnesses in vegetables and plants, making it extremely beneficial to the agricultural business. Since currently the system is trained using Fruit or Plant Leave dataset, the model is trained to detect ROI and diseases.

6. FUTURE SCOPE

• To train the system with much more data of various other Fruits & plants and diseases to further increase the scope of the system.

• By adding images of many other Fruits & plants, it will help in extracting many more features of the plants which certainly help in improving the accuracy of the system.

• The users using the system may also contribute to the system by capturing different types of fruits or plant leaf images which can be added to the dataset.

• This dataset can be further used to build better models. Also, they may be improved in terms of accuracy by implementation of better algorithms in the coming future.

• To provide some remedies for crop diseases to user by analysing the diseases.

• This will certainly help the users to avoid such diseases in the future. Also, the remedies will help the user to get rid of the diseases hence, improving their yield.

REFERENCES

[1] Ajay A. Gurjar and Viraj A. Gulhane, “Disease Detection on Cotton Leaves by Eigen feature Regularization and Extraction Technique”, (IJECSCSE 1), vol.1, Issue.1, June 2012, pp. 1-4.

[2] Alexander A. Doudkin, Alexander V. Inyutin, Albert I. Petrovsky, Maxim E. Vatkin

"Three Level Neural Network for Data Clusterzation on Images of Infected Crop Field," Journal of Research and Applications in Agriculture Engineering, vol.52, Issue .1, Jan 2007, pp. 1-14

[3] Jagadeesh D. Pujari, Rajesh Yakkundimath and Abdulmunaf S. Byadgi, "Classification of Fungal Disease Symptoms affected on Cereals using Color Texture Features", ICICES2014 - S. A. Engineering College, Chennai, TamilNadu, India J. Geralds, "Sega Ends Production of Dreamcast," vnunet.com, para. 2, Jan. 31, 2001

[4] Al-Hiary. H, Bani-Ahmad, M-Reyalat and M. Braik, "Fast and accurate detection and classification of plant diseases," IJCA, vol.17, no. 1, March 2011, pp. 31-38.

[5] Annesha Malakar, Joydeep Mukherjee, "Image Clustering using Color Moments, Histogram, Edge and K-means Clustering", International Journal of Science and Research (IJSR), India Online ISSN: 2319-7064, vol.2, issue 1. January 2013.

[6] Anand Kulkarni & Ashavin Patil R. K. "Applying image processing technique to detect plant diseases," International Journal of Modern Engineering Research (IJMER), Vol. 2, Issue 5, Sep-Oct 2012, pp. 3661- 6264.

[7] Anil K. Jain, Fundamentals of Digital Image Processing Prentice-Hall, 1989.

[8] Arati J. Vyavahare, "Canny Based DRLSE Algorithm for Segmentation", International Journal of Computer Applications, vol.102, no.7, Sept. 2014, pp.1-5.

[9] Asma Akhtar," Automated Plant Disease Analysis (APDA): Performance Comparison of Machine Learning Techniques", 11th International Conference on Frontiers of Information Technology, December 2013, pp.60- 65.

[10] Bacterial Blight of cotton" [online]. Available:http://agrihunt.com/articles/crop-diseases/angular-leaf-spot-ofcotton/.

[11] Bacterial disease in leaf information available from:https://www.ipmimages.org/browse/ subthumb.cfm?sub=17796 .

[12] Basvaraj S. Anami, J.D. Pujari, Rajesh Yakkundinath, "Identification and classification of normal and affected agriculture / horticulture produce based on combined color and texture feature extraction," International Journal of Computer Applications in Engineering, vol 1, Sept 2011, pp.356-360.

[13] Brendan A. Niemira, William W. Kirk, and Jeffrey M. Stein, "Screenng For Late Blight Susceptibiliy In Potato Tubers By Digital Analysis Of Cut Tuber Surfaces", Department of Botany and Plant Pathology, Michigan State University East Lansing, vol.83, no.5, May 1999, pp.469-473.

[14] Camargo.A and J. S. Smith, “Image pattern classification for the identification of disease-causing agents in plants,” *Comput. Electron. Agric.*, vol. 66, no. 2, May 2009, pp. 121–125.

[15] Dang. K, H. Sun, J.P. Chanet, J. Garcia-Vidal, J.M. Barcelo-Ordinas, H.L. Shi and K.M. Hou, “Wireless Multimedia Sensor Network for Plant Disease Detections”, NICST'2103 New Information Communication Science and Technology for Sustainable Development: France-China International Workshop, Clermont-Ferrand, France, Sept. 2013, pp.1-7.