# Plant Disease Diagnosis and Classification by Computer Vision using Statistical Texture Feature Extraction Technique and K Nearest Neighbor Classification

### Kapilya Gangadharan, G. Rosline Nesa Kumari, D. Dhanasekaran, K.Malathi

*Abstract*: *Pest attack and infectious diseases has become more common in the field of agriculture in the recent times. It has become a challenging task to identify the infection or the insect that destructs the plant growth and production. Diagnosing the disease or the insect attack on the plants in the early stage will safe guard the plant growth and the production rate. Timely intervention of technology that deals with disease detection and control method can protect the plants from usage of harmful pesticides. The higher dosage of pesticides impacts the health of human as well as other creatures like birds and animals which directly or indirectly consumes the plant or get in touch with the plants in different circumstances. A Computer vision technique which combines the Digital Image processing and Machine Learning methodology has been proposed to provide pest management solution. The disease detection is based on the statistical texture feature analysis and it is classified using K nearest neighbor classifier. Statistical PCA is combined with SIFT method to extract the key points, which eliminates the non-operational key points and SFTA is used to extract the texture. The system has achieved better result in identifying and differentiating the infection and insect attack on multiple plant taxonomy. The implementation has been performed using MATLAB.*

*Keywords: Computer Vision, Pest management, Machine learning, Diagnosis.*

## I. INTRODUCTION

The study of plant pathology is the scientific study of diseases in plants which contributes in understanding the plant health. Wherein this study throws light on numerous factors that can probably affect the plant's health such as pathogen identification, disease cycles, economic impact, and epidemiology. Diseases in plant disrupts the ecosystem as it involves the animal food chain including human beings and also affects agriculture in the country. The main purpose of this study is to contribute to agriculture which is estimated that in 2016, 15% to 25% of crop production is lost due to vigorous change in the plant health condition [5].

As we know, agriculture contributes to the growth of the entire country ranks top two in farm production in the world at recent times. And it is noticed that 52% of the jobs are available in the agricultural sector in India and 18.1% is contributed to GDP, the destruction in the agricultural sector would mean indirectly impact the shares in national income, foreign exchange, employment of individuals and mainly food. Plant health has to be studied and rectified so as to increase the productivity and growth of the country and as well as the food supply chain. Hence, close monitoring of crops are very essential for the management of diseases and insect attack that can impact the yield, production, cost management significantly and also impacts the post-harvest life of the crop. To come over the impacts and destruction of plants due to infections or insect attack, image processing with machine learning algorithms plays a major role very efficiently. Plant diseases are observed and detected based on the evidence foremost we call it as symptoms that can be visualized. These symptoms are identified on leaves or stems of the plant. Therefore, identification of plants, leaves, percentage of pest, symptoms can give successful diagnosis by means of capturing it in image. To avoid destruction of plants while capturing the images it is highly recommended to use drone images or satellite images with high quality cameras. Hence, the scope of this research is to capture and detect the diseased plant using digital image processing techniques and supervised learning techniques. The diseases being considered in this research are Cercospora leaf spots, Boll weevil, fall armyworm and European corn borers these are diseases with clear epidermis indication. This paper describes about the image processing steps carried out to detect and identify the infection or insect attack based on the symptoms in the early stage and classify the different category of the plant variety, infection or insect attack by training the dataset of sample images using supervised machine learning method.

* Correspondence Author

**Kapilya Gangadharan,** Research scholar in Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. She is currently Utility Analyst at Fidelity National Information Services, Durham, NC, USA

**G. Rosilne Nesa Kumari,** Professor in the Department of Computer Science and Engineering, in Shadan women's college of Engineering and Technology, Chennai, India

**D. Dhanasekaran,** Professor in the Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

**K.Malathi,** Associate Professor in the Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India

**Plant Disease Diagnosis and Classification by Computer Vision using Statistical Texture Feature Extraction Technique and K Nearest Neighbor Classification**

## II. LITERATURE SURVEY

This section presents the similar diagnosis efforts and draws valuable input. The usage of leaf textures as a parameter has been done in the research which is shown in the work lists the methods for plant health monitoring of agricultural crops [3]. The paper suggests by looking for multiple parameters while diagnosing. The use of k-means clustering can output n number of clusters based on value of 'k', however, to obtain the clusters with least mean deviation the value of k must be optimally set according to the given data set. There will be substantial change in the mean variance for a data set with varying value of k till the optimal value of k, if graphically represented this point will be seen where the graph starts to stabilize its slope and is called the elbow of the curve. Another method is derived by conducting model-based clustering and found 3 clusters were most effective in basic pigment separation [4].

The use of texture analysis has proved to be superior in terms of detection accuracy, adapted a method called fractal image texture analysis based on OCT images to identify Macular Edema [1]. Their algorithm derived 97.5% accuracy. However, it always helps to have means to cross verify the result, color-based conformation helps the neural network to differentiate between pests or objects with similar character, size and shape as the target pathogen using it color. suggested that the neural network can be trained to identify specific characteristic shapes and texture [2]. To identify the specific features the have used the neural network implementing root mean square deviation to detect rusting in structures.

Machine Learning could be broadly classified into Supervised and Unsupervised learning [12]. Image samples can be used as the training sets of supervised learning and different set of images were used to compute the efficiency of the system by passing the corresponding response values. In unsupervised learning evidence should be inferred first for the absent label response. Under semi supervised learning takes the advantages of both unsupervised learning and supervised methods which utilizes the labeled and unlabeled training dataset for further processing. An automatic system that detects diseases has been formulated which compares the accuracy of KNN and SVM on the training and testing dataset. It utilized SIFT algorithm to extract the features and KNN to classify the sample images, on comparison it proved that KNN performed better than SVM with the accuracy of 93.3% it had classified three stages of Bacterial diseases [13].

## III. PROPOSED METHOD

Our Proposed system combines Statistical PCA based Feature extraction and KNN Classification. Texture feature is extracted using Statistical PCA (Principal Component Analysis)- SIFT method and SFTA to identify the key points and texture. The proposed system has been classified into training phase and testing phase. The training phase consists of image gathering, color conversion, noise removal, segmentation and feature extraction. The extracted features of the segmented images are utilized for training the KNN Classifier. Once after the images are trained, testing has to be performed for different set of images stored in the database under the training process all the images processing methods should be carried out and extract the required features using Statistical PCA -SIFT and SFTA method from which select the best feature and compare the features of the images and it has to be classified based on the four different type of classes in the database. Fig.1. shows the basic flow of the system.
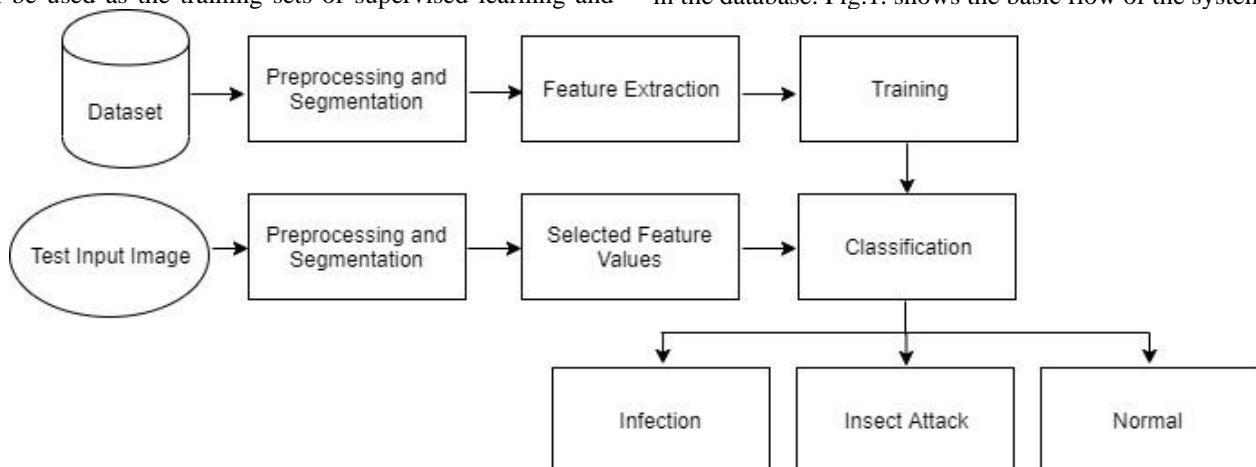


**Fig.1: Basic System Architecture**

### A. Dataset

We have collected images from various sources [9],[10] and saved it in four different folders based on the specific disease and insect attack. The dataset consists of 200 different images for each infection for testing and training. Also, we have used images captured from smart phone and few drone images. Since the dataset has images from various source which has multiple resolution. So, to standardize the resolution of all the acquired image it is converted to 640*480 standard resolution. And the images are transformed to 2-dimensional.

### B. Preprocessing

Acquired images are captured in RGB color model, which is then converted to L*A*B images. These images are preferred over RGB as these are subjected to Euclidean distance measure for k-means clustering, device-independent. A pixel by pixel 3d matrix is extracted from the lab image. Then the image is equalized using Histogram equalization method. All the pertaining noises are removed using Average filter which removes the gaussian noise very effectively.

## C. Segmentation

The 2-Dimentinal color matrix mapped images are subjected to K-means algorithm. It categorizes the object established on several groups of extracted characteristics into number of K classes. Also, it is very important to minimize the sum of square of evaluated distance among the observed target element and the related clusters.

*Algorithm*:
Step1: Evaluate the intensity distribution of histogram
Step 2: Choose the number of appropriate clusters
Step 3: Set the center of K cluster randomly
Step 4: Allot the pixels of the image to nearest cluster center
Step 5: By averaging the pixel compute the cluster center again and repeat step 3 to 5 until the coverage is attained.
When the value of K = 3, Three different clusters are generated.

Sorting of RGB with corresponding cluster id: Once k means clustering is done on the 2D matrix derived from color parameter of the image, the result is in array form. The k means index array is mapped onto a 2D matrix with the resolution of the image, which in turn helps in mapping and isolating similar pixels on the RGB image, hence forming three segments as three separate images as shown in Fig.2.
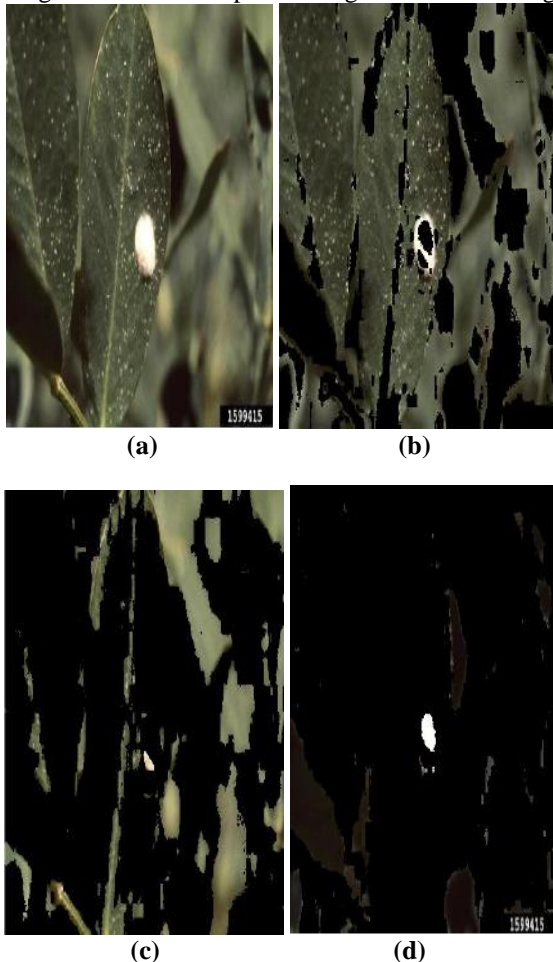


**(a)** **(b)**



**(c)** **(d)**

**Fig.2: (a) RGB Image, (b) Segment 1, (c) Segment 2, (d) Segment 3**

Segment one is the cluster that returns the leaf; therefore, it is used for all the upcoming process. Convert the returned RGB image to gray scaled image for using its intensity histogram to perform OTSU method to compute the threshold. Threshold is calculated by using the frequency distribution

value of the pixel intensity where grey scaled images are converted to binary image by assigning the value of pixels as 0 for all the pixels which is lesser than the threshold value. And assign the values of Pixel as 1 to the pixels whose value is higher than the threshold value

*Algorithm:*
Step 1: Evaluate the frequency distribution of image and initial intensity level probability for each pixel
Step 2: Compute the threshold value
Step 3: Based on the threshold split pixels into 2 clusters
Step 4: For each extracted cluster calculate the average value
Step 5: Square the contrast between the average values
Step 6: Multiply total number of pixels present in one cluster with that of the other clusters

Abnormalities of the plant in the image will show variation in the color, where greenness will be highlighted in the infected region. Hence it proves that the specific region of the plant is infected. When the Image is capture it extracts the RGB components. Appropriate threshold is evaluated using OTSU algorithm, then the greenness will be cloaked and deleted if the threshold value is greater than the intensity of the green colored pixel.

Gradience with morphological dilation and erosion: The gradience creation or boundary creation Fig .3 is done by subtracting the original grayscale from a dilated or eroded form of the image, i.e, an image whose areas of maximum contrast have been expanded or compressed based on a morphological model of the neighborhood defined in the form of a zero matrix. On subtraction only the boundary remains, forming a basic gradient if the neighborhood is a [3x3] and a horizontal or vertical gradient for row and column matrix respectively.
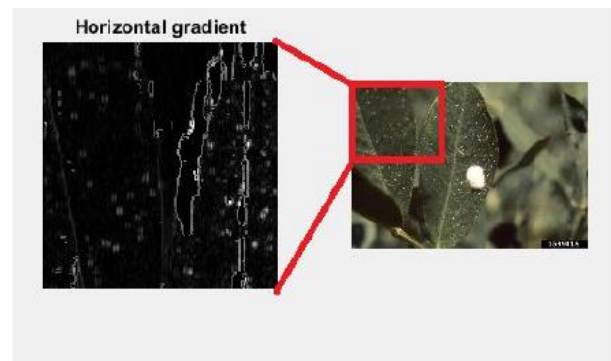


**Fig.3: Extracted horizontal gradient**

## D. Feature Extraction

The extraction of the affected regions are well known with the help of the graph that is produced in image segmentation. Scale-invariant feature transform (SIFT) is an algorithm which is used here in order to detect the features and identify them using computer vision technology. The Key points of the image are extracted by the SIFT algorithm which is patched using Statistical PCA method [6]. The extraction process is carried out through the calculated gradients of the image using mathematical morphology. Where PCA compresses the dimensons obtained by the SIFT descriptor. We use statistical PCA method because it removes the non-operational features which is not required for the selection process.

2338

That compresses the higher dimensional values to the lower dimension that will limit the occurrence of misconception. Leaf can be detected using size and texture therefore for further extraction texture of the leaf is considered. This is achieved by using SFTA Segmentation-Based Fractal Texture Analysis. In this method, there are two ways: intra-class analysis which is a single texture feature and inter-class analysis for more layers like color, shape and texture. Here, digital image toolbox is used to achieve it.

### E. Classification

K-Nearest Neighbor (KNN) classifies the diseased image based on the closest trained feature vector. A comparison of each training samples should be done before performing a new classification element. It is very simple when compared to the other methods. When some pattern of training is appended to the existing training set then retraining is not required. K-nearest neighbors are conceived during the occurrence of same class within the neighbors [11]. K values are estimated based upon the size of the data which is utilized for calculation when the value of K=1, the target will be allotted to the category of the nearest neighbor. If the value of K is larger, then it reduces the noise effect on classification but as the value of K increases it deceases the performance.

*Algorithm:*
Step 1: Load the image
Step 2: K value should be initialized
Step 3: To obtain the estimated similar group, iterate all the training data
   a. Distance between the test data and each row of training data must be evaluated using Euclidean distance metrics
   b. In ascending order sort the evaluate distances
   c. Choose the top K rows from the array of sorted item
   d. Select the class which is most frequent in the row
   e. Return the predicted class.

These images are stored in a structured database with their threshold values generated from various stages during image processing. These images are graded based on image intensity, color and texture thresholds and serve as a baseline for further training and testing of the neural web. Our work is focused on plant disease Cercospora leaf spots and insect attacks like Boll weevils, European corn borer and Fall armyworm the training set for these diseases are stored in four directories and with every confirmed detection this data set is expanded.

*Training:*
Step 1: Load the training Image
Step 2: Read the image
Step 3: Preprocess by converting the color and standardizing the resolution to 640*480 and remove the noise.
Step 4: Segment the image and estimate the gradient
Step 5: Extract the features using SIFT with Statistical PCA and SFTA method
Step 6: Construct the features using feature vector by the extracted features in a matrix.
Step 7: Repeat Step 2 to 7 until all the images in the dataset are trained
Step 8: Perform KNN method and tune the dataset.

*Testing:*
Step 1: Read the image
Step 2: perform all image processing steps

Step 3: Identify the nearest neighbor using Euclidean distance
Step 4: If all the neighboring values are the same then exit, else construct the matrix by comparing the distance between the K neighbors
Step 5: Classify the image based on the plant taxonomy and the infection or the insect attack.

## IV. RESULT AND DISCUSSION

This is the tail end of all the process, as the K nearest neighbor outputs a value from its threshold unit after considering the weighted inputs from the image processing segments, the threshold is compared with the stored indexes of the four diseases, it displays the findings along with isolated images of the pathogen or pest.
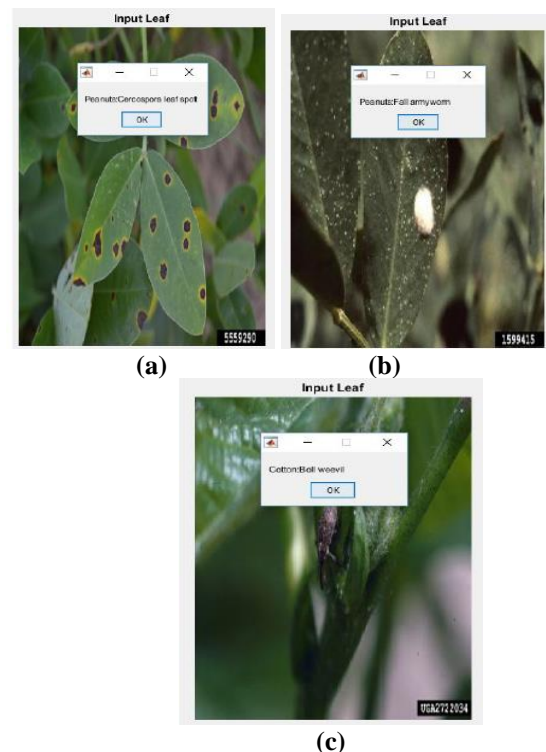


**Fig 4 Classification Result (a) Cercospora leaf spot-Peanut leaf (b) Fall army worm peanut leaf and (c) Boll Weevil Cotton plant**

### A. Performance Evaluation

The performance has been evaluated based on the valuation metrics. It can be classified as

True Positive (TP): Abnormal case appropriately evaluated as Abnormal
False Negative (FN): Abnormal case imperfectly classified as normal
False Positive (FP): Normal case imperfectly identified as abnormal
True Negative (TN): Normal case appropriately calculated as Normal

Sensitivity: TP/TP+FN                (1)
Specificity: TP/TP+FP                (2)
Accuracy : (TN+FP)/(TN+TP+FN+FP)     (3)
F-Score :

(2*Sensitivity*Specificity)/sensitivity+specificity        (4)
From the above metrics the performance of the individual disease identification and the accuracy of the entire system has been evaluated. As shown in the table 1.   The performance evaluation graph is generated for all the disease and insect attack which is shown in the Fig.5. And it is observed that the system performs the overall accuracy of 97.13%

**Table- I: Statistical Measure for Accuracy Calculation**

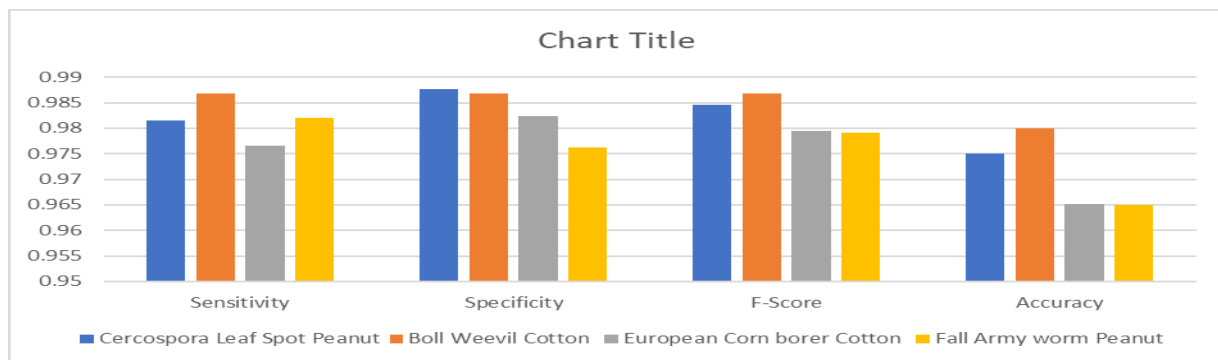| Disease/Insect Attack | Plant Taxonomy | Sensitivity | Specificity | F-Score | Accuracy |
|---|---|---|---|---|---|
| Cercospora Leaf Spot | Peanut | 0.981595092 | 0.987654321 | 0.984615 | 0.975 |
| Boll Weevil | Cotton | 0.986842105 | 0.986842105 | 0.986842 | 0.98 |
| European Corn borer | Cotton | 0.976608187 | 0.982352941 | 0.979472 | 0.965174 |
| Fall Army worm | Peanut | 0.982035928 | 0.976190476 | 0.979104 | 0.965 |
| Average | | 0.981770328 | 0.983259961 | 0.982509 | 0.971294 |



**Fig.5: Graphical representation of statistical metrics**

The proposed algorithm is compared with three other traditional methods with slight variation in the dataset. The overall accuracy is taken into consideration for the evaluation purpose because the compared algorithms vary in the set of

**Table II: Comparison of proposed method with existing tradition methods**

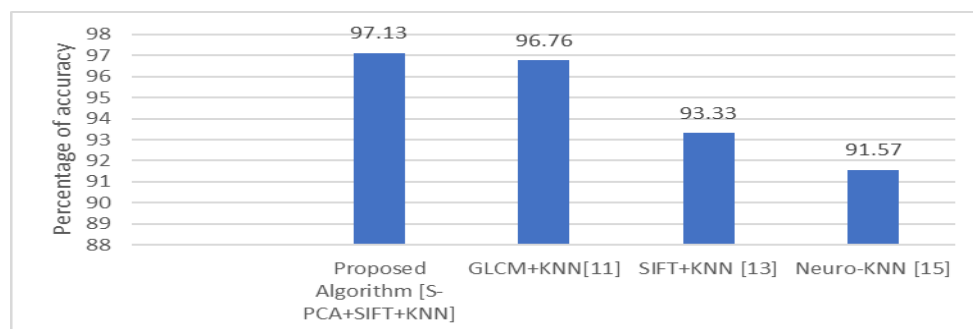| Method | Disease/Insect Attack | Accuracy |
|---|---|---|
| Proposed Algorithm [S-PCA+SIFT+KNN] | Cercospora, Boll Weevil, fall Army Worm, European Corn Borer | 97.13 |
| GLCM+KNN [11] | Cercospora, Bacterial Blight, Boll Weevil, fall Army Worm, European Corn Borer | 96.76 |
| SIFT+KNN [13] | Bacterial Blight, Leaf blast and Brown spot | 93.33 |
| Neuro-KNN [15] | anthracnose, blight, rust, and mildew | 91.57 |



**Fig. 6: Comparison of proposed method with tradition methods**

analyzed diseases and plant varieties. From the analysis it is observed that the overall percentage of the proposed system is higher than the existing methods. Percentage of accuracy of proposed system is 97.13 Table 2 shows the comparison and Fig.6 is the graphical representation of the table.

## V. CONCLUSION AND FUTURE WORK

By diagnosing four diseases this paper demonstrates the scope for quality and reliable. The KNN can be trained to identify a wide variety of diseases as shown in Fig.4 and it will keep getting better at accurate diagnosis. This can be further improved by including geolocation tags and time stamp with the picture of the disease to train the neural network to use seasonal trends of a place to differentiate between similar diseases. This application will serve as a tool to directly help plant life in ecosystems that are fighting imbalance in natural equilibrium and will indirectly help by bringing change in agricultural practices. From the analysis the proposed system achieved overall accuracy of 97.13%. The method has been processed with a clean slate in terms of machine learning and over the time it is observed that how distinctions were being made among the different types of disease and insect attack in plants being fed during the training stage. Cercospora leaf spots and Fall armyworm were being identified by the freckle like patterns on the leaf, but they were being said apart by size of the spots and more importantly the texture difference between the ventral and dorsal side of the leaves that are respectively affected by the diseases. On the other hand, the European corn borer and the Boll weevil we're being recognized by their distinct shape and host plant. A cotton bud combined with the outline of an insect hinted Boll weevil and a stem with sizable holes indicates the European corn borer.

## REFERENCES

1. S.C. Athira, Reena M. Roy,R.P. Aneesh," Computerized Detection of Macular Edema Using OCT Images Based on Fractal Texture Analysis",2018 International CET Conference on Control, Communication, and Computing (IC4 2018), 978-1-5386-4967-1@IEEE
2. Luh-Maan Chang,Po-Han Chen and Heng-Kuang Shen, "Recognition for color rust images based on artificial neural network", Automation in Construction Volume 90 , June 2018, Pages 178-187.
3. Sachin D. Khirade and A. B. Patil from Pimpri Chinchwad College of Engg,"Plant disease detection using image processing" 2015 IEEE computer society.
4. Chris Fraley And Adrian E.Raftery, "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis", Department of Statistics, University of Washington, USA, The Computer Journal, Vol. 41, No. 8, 1998
5. Kekane Maruti Arjun, "Indian Agriculture- Status, Importance and Role in Indian Economy", India.International Journal of Agriculture and Food Science echnology. ISSN 2249-3050, Volume 4, Number 4 (2013).
6. Patrik Kamencay, Robert Hudec, Miroslav Benco, and Martina Zachariasova, "Feature Extraction for Object Recognition using PCA-KNN with Application to Medical Image Analysis", TSP 2013, 978-1-4799-0404-4/13©2013 IEEE
7. Kapilya Gangadharan, G. Rosline Nesa Kumari, D. Dhanasekaran, "An Efficient Plant Disease Detection System Using Hybrid Watershed Segmentation with Extended K-Means Clustering Algorithm", International Journal of Advanced Science and Technology, Vol. 28, No. 11, (2019), pp.308-320
8. Trishen Munisami, Mahesh Ramsurn, Somveer Kishnan, Sameerchand Pudaruth, "Plant leaf recognition using shape features and color histogram with k-nearest neigbour classifier", Second International Symposium on Computer Vision and the Internet (VisionNet'15), Procedia Computer Science 58 (2015) 740-747
9. https://plantvillage.psu.edu/diseases
10. https://www.forestryimages.org/index.cfm
11. Eftekhar Hossain, Md. Farhad Hossain and Mohammad Anisur Rahaman,"A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019
12. Machine Learning. https://en.wikipedia.org/wiki/Machine_ learning. Accessed December 2019
13. K.Malathi, R.Nedunchelian, "Efficient Method To Detect And Classify Diabetic Retinopathy Using Retinal Fundus Images", International Journal of Pure and Applied Mathematics, Volume 116 No. 21 2017, 89-97 ISSN: 1311-8080 (printed version);ISSN: 1314-3395 (on-line version)
14. K Malathi1, R Nedunchelian, "A recursive support vector machine (RSVM) algorithm to detect and classify diabetic retinopathy in fundus retina images", Biomedical Research 2017, ISSN 0970-938X
15. Pujari JD, Yakkundimath R, Byadgi AS, "Image processing-based detection of fungal diseases in plants", Proc Comput Sci 46:1802–1808
16. Reena M. Roy,R.P. Aneesh," Computerized Detection of Macular Edema Using OCT Images Based on Fractal Texture Analysis",2018 International CET Conference on Control, Communication, and Computing (IC4 2018), 978-1-5386-4967-1@IEEE

## AUTHORS PROFILE

**Kapilya Gangadharan** is a Research scholar in Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. She is currently working as Utility Analyst at Fidelity National Information services, Durham, NC, USA with total of 6 years' teaching experience as Assistant Professor and 2 years' experience in IT. She received B.Tech degree in Information Technology from CSI College of Emgineering, Ketti, Nilgiris and M.E in Computer Science and Engineering from Periyar Maniammai University, Thanjavur. Her current research includes Digital Image Processing, Machine learning, Precision farming, Plant Technology and Production Engineering.

**Dr. G Rosline Nesa Kumari** received her Ph.D degree in Computer Science and Engineering at Dr MGR University Chennai, and received her M.E. Degree from Sathyabama University Chennai in 2005. She is having Sixteen Years of teaching experience. At present she is working as a Professor, Department of Computer Science and Engineering in Shadan women's college of Engineering and Technology, Chennai. She published sixty-five research publications in various International, National Conferences and Journal. She is a life member of Indian Science Congress Association (ISCA), IAENG, IET, CSI (Life Member), Red Cross. Her research interest includes Image processing, Steganography, Network security and Information Security. includes Image processing, Digital Watermarking and Security

**Dr.D.Dhanasekaran** received his PhD degree with the specialization of VLSI reconfigurable circuits. He had 29 years of teaching experience. He is working as a Principal in Saveetha school of Engineering. Saveetha Institute of Medical and Technical Sciences, Chennai. He published 50 Scopus indexed papers and two patents. His area of research includes VLSI circuits and IoT.

**Dr. K. Malathi** is an Associate Professor in the Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai with 11.6 Years of teaching experience and 1.6 years of industry experience. She graduated her B.Tech from Anna Univeristy, M.E from Sathyabama University, and Ph.D degree in computer science in 2018 from Saveetha Institute of Medical and Technical Sciences, Chennai. Her research interests include image processing, datamining, Medical Image Processing. She is life member in CSI, IEEE Computer Society, and IET. She has published around 40+ research papers in reputed journals and 4 International conferences. She has received most Enthuastic Faculty Award in women's day.