International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Bacterial and Virus affected Citrus Leaf Disease Classification using Smartphone and SVM

Utpal Barman, Ridip Dev Choudhury

Abstract*: Automatic detection of citrus leaves disease is very much essential for the better productivity of citrus. Citrus leaves are affected by bacteria, fungus and virus respectively. Farmer detects the diseases of the plant using laboratory, naked eyes or using expert's view. The rural farmers often face difficulties to detect these diseases due to the non availability of the laboratories in their area. Here in this paper, a computer automation system is proposed to detect the diseases of citrus leaves on an early stage. Citrus leaves images are captured using Smartphone. Captured images are used to extract the different features of the citrus leaves samples using Gray Level Co-occurrence Matrix. Finally, citrus greening and citrus CTV images are classified from citrus healthy images using Gaussian kernel based support vector machine. Accuracy of the kernel is evaluated for the different values of Gamma parameter of kernel. The Gaussian kernel gives maximum accuracy (95.5%) with Gamma value 1.

***Index Terms*: Citrus, disease, SVM, Kernel, Image Processing.**

I. INTRODUCTION

Plant leaves are mainly responsible for photosynthesis. Citrus leaves are also important for citrus plant and these are the main sources of energy. So, it's important to take care of citrus leaves. Citrus leaves are often affected by different diseases. These are basically due to the virus, fungus and bacterial infections. Citrus canker, citrus anthracnoses are the fungal affected diseases whereas citrus greening and citrus scab are bacterial affected diseases. Citrus leprosies and citrus CTV are virus affected diseases. Early detection of citrus disease is possible in laboratory. Sometimes, plant pathologists help farmers to detect the diseases of the plant. Both the methods are not approachable for rural farmers for all round the year due to the non availability of laboratories and experts. In recent years, many computerized methods are developed to classify the diseases of the plant leaves using machine learning technique[1]. Different machine learning techniques such as K-nearest neighbor, support vector machine, decision tree are used to classify the plant leaves diseases[1].

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Utpal Barman*, Research Scholar in the Department of Information Technology, Gauhati University.
Dr. Ridip Dev Choudhury, Assistant Professor in the Department of Computer Science, IDOL, Gauhati University.

In this paper, support vector machine classifier is used to classify the diseases of citrus leaves on self collected citrus images.Computer automation with machine learning is widely used in agriculture industry and especially for plant disease detection. In paper [2], authors used K nearest neighbor, probabilistic neural network, support vector machine (SVM) , neuro K nearest neighbor classifier to detect the fungal diseases of fruits, commercial fruits, vegetables, cereal crops. They achieved accurate result in case of SVM and neuro-KNN. SVM is a very powerful machine learning classification approach which is widely used in past study. SVM is used to classify the diseases of pomegranate [3], lemon, banana, bean, rose [4], apple [5], rice [6], banana, beans, guava, jackfruit, lemon, mango, potato, tomato [7], cotton [8], cotton leaf spot [9], wheat [10], citrus [11], citrus canker[12]. SVM is proposed by Vapnik [13] for two class classification but later on it is used for multi class classification. Apart from SVM, artificial neural network classifier [6], Bayes' classifier [6], Minimum distance classifier [4,7], Back propagation neural classifier [8], Fuzzy classifier[8], Deep neural network classifier [14] are also used to classify the diseases of plant leaves. Among all the classifier, SVM and deep neural network classifiers are very popular and accurate classifier. Most of the farmers detect the diseases of the plant using the laboratory, naked eyes and expert views. Expert views and laboratory methods are cost consuming process and these are not accessible for all the farmers. But, computer automation helps farmers to solve such kind of complex problem of agriculture[15,16].

In the previous study of computer automation, authors used either standardized dataset [17] or self captured images for disease classification. Most of the works of self captured datasets are with the help of digital camera[5,7, 9]. Sometimes authors collected the images from Agriculture University [2]. In this paper, citrus leaves images are captured using an Android Smartphone. Capturing Images with Smartphone is the main challenging task as the quality of the Smartphone images is not good as compared to the digital camera. The proposed system used a low budget Android Smartphone to capture the citrus leaves images in natural light condition.

In the previous study of citrus, image processing stages are used to recognize citrus diseases. In paper [4], authors used lemon leaf as one of the species for study. They used 25 lemon leaves samples to detect the sun burn disease of lemon. They applied clipping, smoothing and contrast enhancement in their preprocessing steps and finally applied the genetic algorithm and support vector machine classifier to classify the lemon disease.

Citrus canker is one of the major fungal diseases of citrus. Paper [12] focused on citrus canker. They used histogram equalization method to increase the quality of the citrus canker images. Gray Level Co-Occurrence Matrix, K-Mean and Support Vector Machine are used by the authors as a method of feature extraction, clustering and classification. Paper[18] introduced hidden markov model to classify the citrus anthracnose, canker and greening respectively. Paper [19], introduced a cluster based disease detection system for lemon. Authors of paper[11] used SVM with RBF and Polynomial kernel to classify the citrus anthracnose, canker and greening diseases respectively.

It is found that image processing and machine learning algorithms are widely used in citrus leaves diseases classification. In this paper, a new method is proposed to detect the diseases of the citrus leaves on self collected images. Images of citrus leaves are collected using a Smartphone and then preprocessing, feature extraction and classification algorithm is applied to classify the diseases of the citrus. In the proposed system, GLCM technique is applied to extract the texture features of the leaves images and SVM with Radial Base function kernel is used to classify the citrus leaves.

Citrus leaves are affected by bacteria, fungus and virus. Citrus Greening, Citrus bacterial spot, Citrus Scab, Citrus Canker, Citrus Leprosies, Citrus CTV are the bacterial, fungus and viral diseases of citrus respectively. These diseases attack in the leaves, fruit and steam parts of the citrus. Fig.1 shows the some of the citrus diseases [**20**].

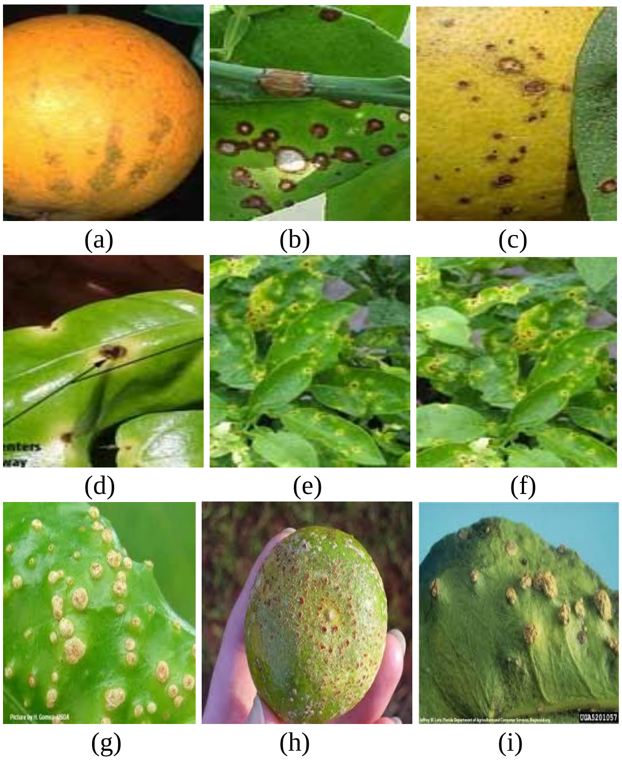

Fig. 1 (a-c) Citrus Bacterial Spot [20]; (d-f) Citrus Anthracnose [20] (g-i) Citrus Scab [20]

II. Materials and Methods

A. Sample Collection and Image Acquisition

In this paper, Assam citrus is selected for study. Disease affected citrus leaves are collected and identified as citrus greening and citrus CTV. The experiment is conducted in Horticulture Research Station (HRS), Assam Agriculture University Kahikuchi, Assam. A plant pathologist helps to identify the diseases of the citrus. Along with the affected diseases of citrus, healthy images of citrus are also used to classify the diseases of citrus from the healthy images. Fig. 2 shows some of the self collected images of citrus greening, citrus CTV and citrus healthy. The colors of healthy images are green but the colors of other two classes of leaves are not green. It changes from green to light yellow. Sometimes, the parts behind the lamina are also light yellow. The upper surfaces of the green leaves are smoothie in nature but the surface of citrus CTV and citrus greening is not much smoothie as compared to citrus healthy. So, color and texture are the important features to be considered to classify the citrus healthy images from citrus greening and citrus CTV.

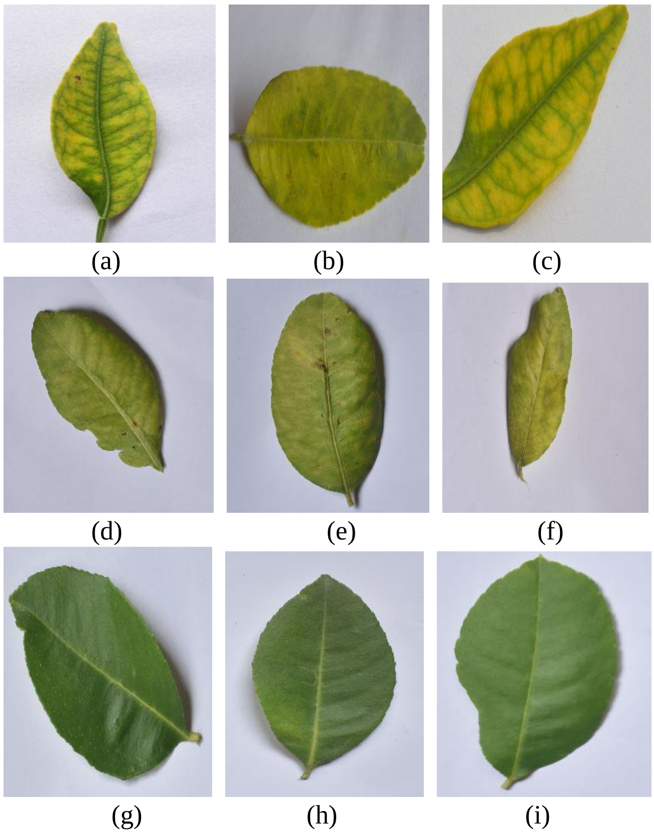

Fig. 2 (a-c) Citrus Greening; (d-f) Citrus CTV (g-i) Citrus Healthy

For the experiment, different trees of citrus are selected and 300 samples of citrus leaves are collected for diagnosis. Each 100 samples belong to citrus greening, citrus CTV and citrus healthy respectively and it is presented in Table I. Here, most of the leaves are collected from the various citrus.

Table I. Total numbers of samples off Citrus

Citrus Class	Total No. of Samples
Citrus Healthy	100
Citrus Greening	100
Citrus CTV	100

Image acquisition is a procedure to obtain different images from the samples using different image sensing devices such as camera [4], Smartphone [16], etc. The image sensing capacity of Smartphone is less as compared to digital camera. So, the quality of Smartphone image is less than digital camera image. By considering this challenge, a low budget Redmi Android Smartphone is used to capture the images. The Redmi Smartphone consists of 13MP camera with full HD display. The images are captured 1.5 feet ahead from the citrus samples to avoid blur. The F-Stop, exposure time, ISO speed of the camera are in default mode. The average dimensions of the images are 3120 X 4160. All the citrus images saved in JPEG format with 8 bit RGB mode.

B. Image preprocessing of Citrus Leaves

All the captured images are large in dimensions. In this step, dimensions of citrus leaves images are reduced to an acceptable dimension so that the processing time of the proposed model is less. The reduced dimensions of the images are 256 x 256. At the same time, the contrast of the model is increased by considering 1% upper and lower pixel value of the citrus. For the entire process, Streachlm and Imadjust functions of Matlab 2015a are used. Fig.3 shows the original and contrast enhanced images of citrus.

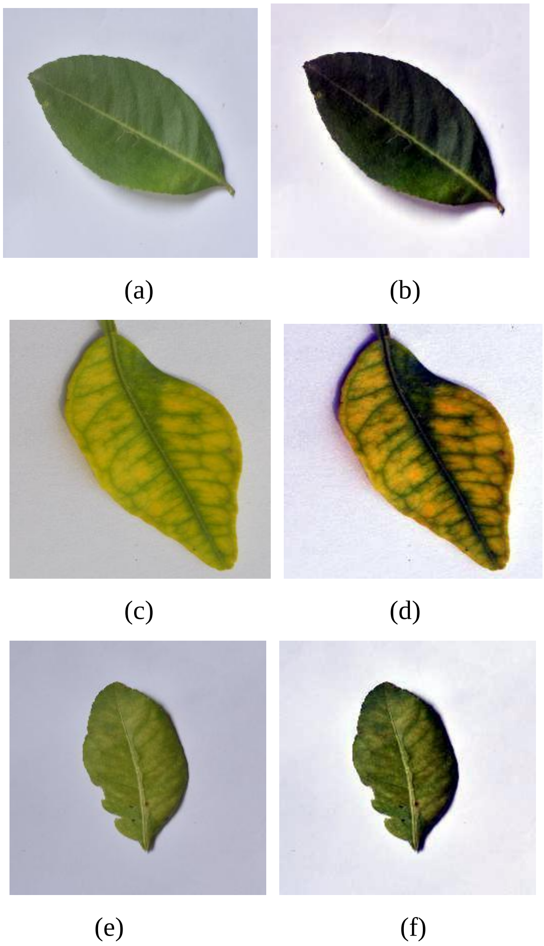

Fig. 3 (a-b) Normal Healthy image Vs contrast enhanced Healthy Image (c-d) Normal Greening image Vs contrast enhanced Greening Image (e-f) Normal CTV image Vs contrast enhanced CTV Image

C. Important Features of Citrus Leaves

The features of disease affected citrus images are defined by the color and textures features of the images. The experiment is done using Matlab 2015a. It is found that all the healthy citrus leaves images are green but the colors of the disease affected citrus leaves images are changing day to day. Again homogeneity, smoothness, kurtosis of the affected citrus image is rough as compared to the healthy image. All these properties come under color and texture feature of the image. In previous study, it is found that GLCM is a significant feature extraction technique which describes the color and texture properties of the image[7]. Both the features are defined in Color and Gray domain of the images. Total 13 numbers of features of citrus leaves images are calculated in two domains. In the gray domain, Contrast, Correlation, Energy and Homogeneity are determined. Later on color Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, and IDM are calculated. These features are calculated using the equation (1-13).

Table II. GLCM parameter and equation

Parameter	Equation	No.
Contrast	$\sum_{i,j=0}^{N-1}(i,j)^2 m(i,j)$	(1)
Correlation	$\sum_{i,j=0}^{N-1} m_{ij}\dfrac{((i-\mu)(j-\mu))}{\sigma^2}$	(2)
Energy	$\sum_{i,j=0}^{N-1} m(i,j)^2$	(3)
Homogeneity	$\sum_{i,j=0}^{N-1}\dfrac{m(i,j)}{1+(i-j)^2}$	(4)
Mean	$\sum_{i,j=0}^{N-1} i.m_{i,j}$	(5)
Variance	$\sum_{i,j=0}^{N-1} m_{i,j}(i-\sigma)^2$	(6)
Standard Deviation (SD)	$\sum_{i,j=0}^{N-1}(i-Mean)^2 m(i,j))^{1/2}$	(7)
Entropy	$\sum_{i,j=0}^{N-1} m(i,j)\ln m(i,j)$	(8)
Root Mean Square (RMS)	$(\sum_{i,j=0}^{N-1}(Mean)^2\ )^{1/2}$	(9)
Smoothness	$1-\dfrac{1}{1+(SD)^2}$	(10)
IDM	$\sum_{i,j=0}^{N-1}\dfrac{1}{1+(i-j)^2}m(i,j)$	(11)

Skewness	$\sum_{i,j=0}^{N-1} (i - Mean)^3 m(i,j)$	(12)
Kurtosis	$\sum_{i,j=0}^{N-1} (i - Mean)^4 m(i,j)$	(13)

In these equations, m(i,j) is the intensity value of the pixel and N is the Gray levels.

In this paper, entire 13 features of citrus leaves images are calculated using GLCM. The size of the feature vector of citrus is 13X300 where each 100 features belong to the each category of citrus leaves. The features of the images are presented as scatter plot in Fig. 4.

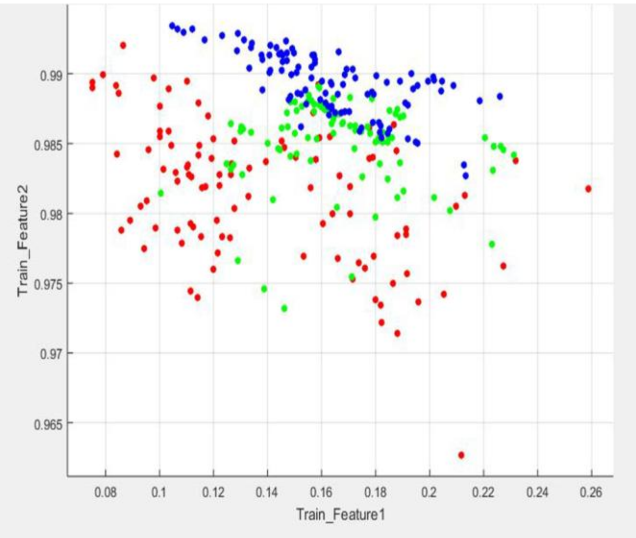

Fig. 4 Scatter plot of feature vector

All the 13 features of images are presented as Train_Feature1, Train_Feature2, Train_Feature3, and so on. In Fig. 4, Train_Feature1 represents contrast and Train_Feature2 represents energy feature of the image. Three colors are used to represent each category of the citrus image. Red color is for citrus CTV and green color is for citrus greening. Citrus healthy is represented by blue color.

III. Results and Discussion

Extracted features of citrus are saved as .mat file for the classification and a class label is assigned for each class. The class labels for Citrus CTV, Citrus greening and Citrus healthy are 0, 1, and 2 respectively. Support vector machine classifier is used to train the model and accuracy is determined in the test images.

A. SVM classifier for Citrus Leaves

Support vector machine is statistical based binary classification which is introduced by Vapnik[13]. Initially, it is used for two class classification but multiclass classification can be also done using SVM. SVM classify the data into two classes with the help of hyper plane. Hyper plane separates the linearly separable data. But, non-separable data are converted to a high dimensional space for classification. It can be done using kernel function. Kernel used the features vector information as an input and converted these into required forms with the help of some mathematical functions. In this paper, a Gaussian Radial Basis Function (RBF) kernel is used to transform the features vector into a desired form. The RBF kernel for two sample M and M$^{/}$ is represented by the equation 14. The right performance of the kernel depends on the value of gamma. Gamma is the kernel parameter which determines the actual behavior of the SVM for the classification. If the value of the gamma is large, then the value of M$^{/}$ will go away from M and the variance of the model is low but the bias will be high. When the value of the gamma is less, then the variance is more but the bias is less. So, the system performance depends on the right choice of gamma.

$$RBF(M, M^{/}) = \exp(-gamma \| M - M^{/} \|^2) \quad (14)$$

In this paper, SVM is applied with the different values of gamma parameter of the RBF. These gamma values are 0.1, 1, 10, and 100. The gamma value 0.1 is very less and the gamma value 100 is high. The model is implemented by considering the regularization parameter (C) of the SVM as 1.

A holdout cross validation is applied to find the accuracy of the model. In holdout cross validation, the entire feature values are divided into two sets such as training set and testing set. For each class of data, 70 samples are in training set and 30 samples are in testing set.

Different parameters are used to evaluate the accuracy of the model.

a) *Confusion Matrix:* The output of the model is represented in tabular form which is known as confusion matrix. It contains the value true classification and misclassification.

b) *ROC Curve:* In this parameter, true positive rates and true negative rates of the classification are plotted in terms of graph.

c) *Scatter Plot:* Scatter plot defines the true classification and misclassification in terms of symbols.

d) Overall Accuracy: It is the overall true classification of the model.

The confusion matrix is presented in Table III along with the percentage value of true positive rate and false negative rate of each class. In the table III, class label 0, 1, and 2 denotes citrus CTV, greening and healthy class respectively.

Table III. Confusion table of SVM

Gamma Value	Class Label	0	1	2
	0	0(0%)	30(100%)	0(0%)
0.01	1	0(0%)	30(100%)	0(0%)
	2	0(0%)	30(100%)	0(0%)
	0	27(90%)	0(0%)	3(1%)
1	1	0(0%)	30(100%)	0(0%)
	2	1(3.3%)	0(0%)	29(96.7%)
	0	25(83.3%)	0(0%)	5(16.7%)
10	1	4(13.3%)	24(80%)	2(6.7%)
	2	1(3.3%)	0(0%)	29(96.7%)
100	0	30(100%)	0(0%)	0(0%)

1	13(43.3%)	12(40%)	5(16.7%)
2	11(36.7%)	0(0%)	19(63.3%)

The ROC curve and Scatter plot of the classification is presented in Fig. 5(a-h) and the overall accuracy of the model are presented in Table IV. All the evaluating parameters of the model are presented for the different values of gamma.

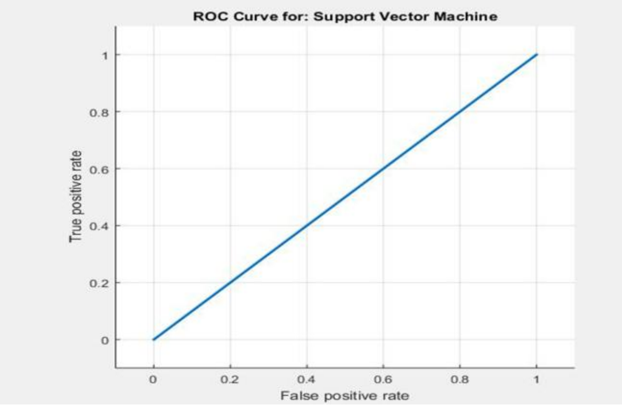

Fig. 5 (a) ROC curve at gamma=0.01

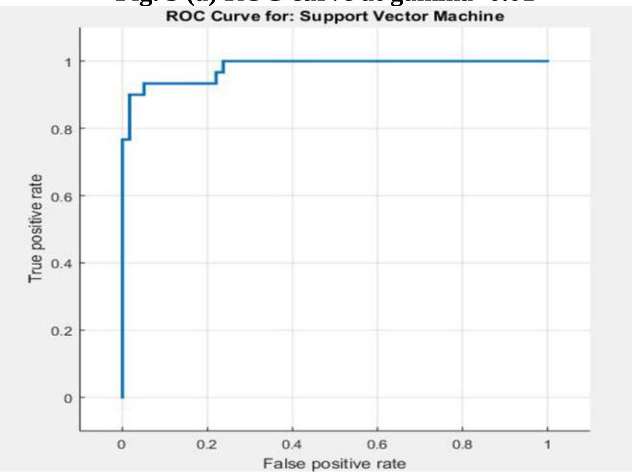

Fig. 5(b) ROC curve at gamma=1.0

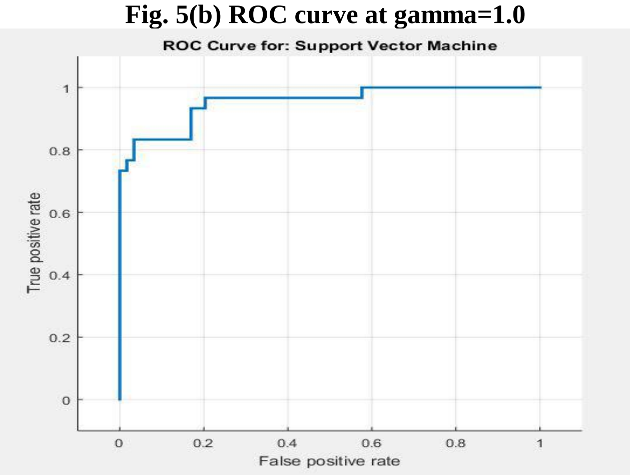

Fig. 5(c) ROC curve at gamma=10

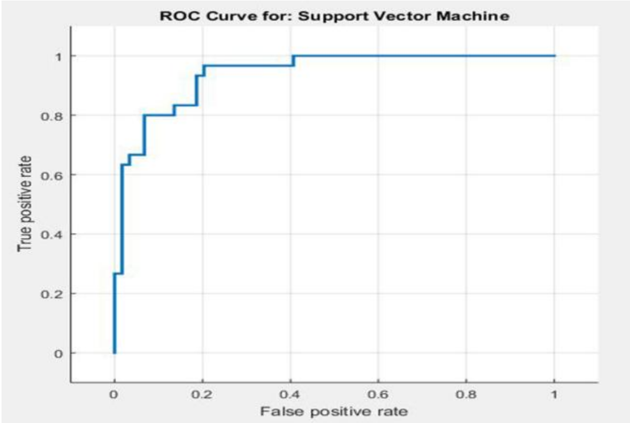

Fig. 5(d) ROC curve at gamma=100

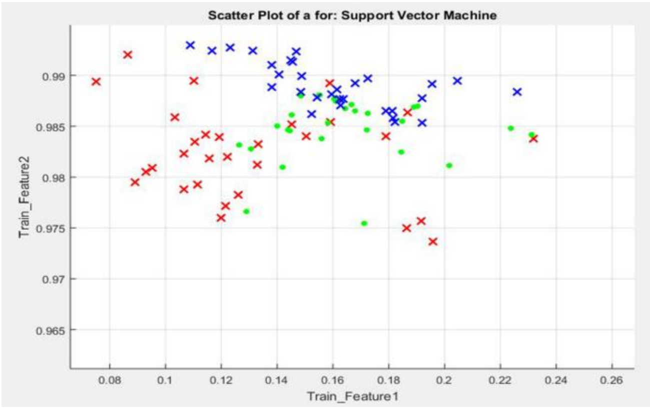

Fig. 5 (e) Scatter plot at gamma=0.01

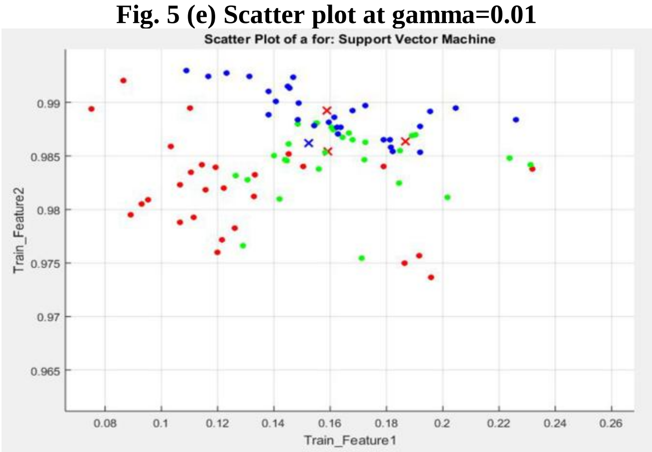

Fig. 5(f) Scatter plot at gamma=1.0

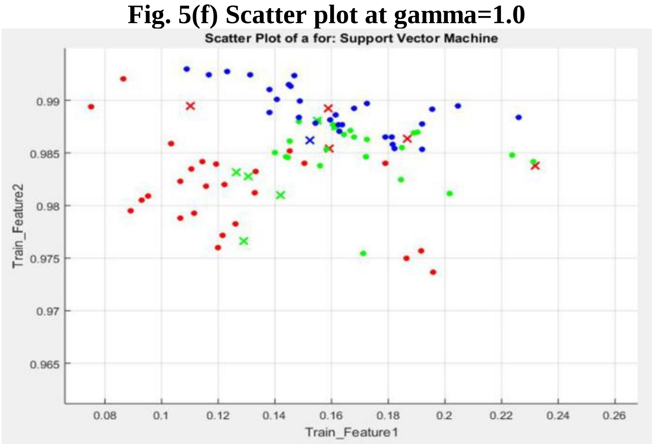

Fig. 5(g) Scatter plot at gamma=10

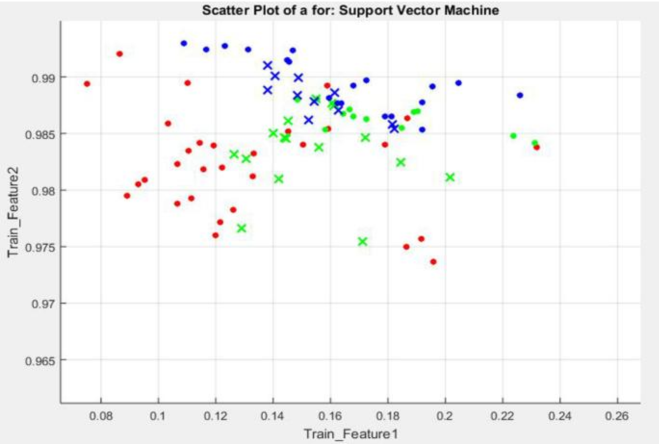

Fig. 5(h) Scatter plot at gamma=100.

From table IV, it is found that the overall accuracy of the model is maximum at gamma =1.0.

Table IV. Overall Accuracy and Error of the SVM

Gamma	C	Accuracy	Error
0.01	1	32.6%	67.4%
1	1	95.5%	4.5%
10	1	86.5%	13.5%
100	1	68.5%	31.5%
1000	1	59.6%	40.4%

B. Comparative Result Analysis

SVM with RBF kernel is used to classify the model as the features of the citrus images are non linear. To avoid the over-fitting of the model, hold-out cross validation is applied to divide the features of the citrus images and accuracy is determined in the testing set of the model. SVM always gave good result for disease classification such as 95.71% [4], 96%[11], 85% [3], and 96.87%[5].

In this paper, the maximum accuracy is found for gamma=1.0 and it is 95.5%. From table III, it is found that all the images of citrus CTV and citrus healthy are misclassified as citrus greening at gamma=0.01. So, the true positive rate of Citrus CTV is 0.0% and the false negative rate of citrus CTV is 100%. Again the true positive rate of citrus greening and citrus healthy is 100% and 0% respectively but the false negative rate of citrus greening and citrus healthy is 0% and 100% respectively at gamma=0.01.

When the gamma value is increased to 1, then the true positive rate of each class is more as compared to other class. For gamma=1, a total of 27 images of class Label 0 (Citrus CTV) is completely classified as citrus CTV but 3 images of citrus CTV is misclassified as citrus healthy. All the 30 images of citrus greening are completely classified as citrus greening so the accuracy of citrus greening is 100%. There is no misclassification for citrus greening but 1 image of citrus healthy is misclassified as citrus CTV so the accuracy of citrus greening is 96.7% (Table III).

For gamma=10, the individual accuracy of citrus CTV is 83.3% as 5 images are completely classified as citrus CTV but 5 images of citrus CTV is misclassified as Citrus healthy. So the false negative rate of citrus CTV is 16.7%. The true positive rate of citrus greening is 80% but the false negative rate of citrus greening is 20% (Table III).

For gamma=100, the true positive rate is more for citrus CTV as compared to greening and healthy images. But the misclassification is increased for the class label 1 and Class label 2 (Table III).

From Table IV, it is found that overall accuracy of the model is 95.5% at gamma=1. It means that at gamma=1, the model is neither over-fitted nor under-fitted. The model performed well at gamma =1. But, the model is not behaved well at gamma=0.01, and 100. At gamma=0.01, the testing accuracy of the model is 32.6 %, so the variance is high and the model is over-fitted. Again, at gamma=100, the testing accuracy of the model is 68.5 %. When the gamma value is increased to 1000, the testing accuracy of the mode is 59.6%. At this situation, the error is more and the bias is high for the model. It seems to be under-fitted due to the increase of gamma value.

The area under ROC curve explains the accuracy of the SVM model with RBF kernel. From Fig.5 (b), it is found that the area of the curve is more at gamma 1 as compared to other gamma values.

The Fig. 5 (e-h) show the scatter plot and it is found that the completely classified feature is denoted by the dot symbol where as the misclassification is denoted by cross symbol. All the features of citrus greening and healthy are crossed as the images are misclassified (Fig. 5 e). But for gamma=1, only 4 features of citrus dataset are crossed out of which three symbols belong to red (Citrus CTV) and 1 symbol is blue (Healthy). It is found that numbers of misclassification symbols are less at gamma=1 as compared to other gamma values. So, the accuracy of the model is accepted for gamma=1.

IV. CONCLUSION

In this paper, a simple and low cost method is presented to classify the healthy leaves of citrus from the diseases affected leaves. Smartphone image based computer automation helps farmers to recognize healthy images from diseases affected images. The proposed method gives accurate result for the self collected images and it can be used for diseases detection on an early stage. The accuracy of the model can be further increased by applying the deep neural network based classifier for large number of data samples of citrus leaves.

REFERENCES

1. T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, and L. Plümer, “Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance,” *Computers and electronics in agriculture*, vol. 74, no. 1, pp. 91–99, 2010.
2. J. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “Image processing based detection of fungal diseases in plants,” *Procedia Computer Science*, vol. 46, pp. 1802–1808, 2015.
3. M. Bhange and H. A. Hingoliwala, “Smart Farming: Pomegranate Disease Detection Using Image Processing,” *Procedia Computer Science*, vol. 58, pp. 280–288, 2015.
4. V. Singh and A. K. Misra, “Detection of plant leaf diseases using image segmentation and soft computing techniques,” *Information processing in Agriculture*, vol. 4, no. 1, pp. 41–49, 2017.
5. J. Tian, Q. Hu, X. Ma, and M. Han, “An improved KPCA/GA-SVM classification model for plant leaf disease recognition,” *Journal of Computational Information Systems*, vol. 8, no. 18, pp. 7737–7745, 2012.

6. S. Phadikar, J. Sil, and A. K. Das, “Classification of rice leaf diseases based on morphological changes,” *International Journal of Information and Electronics Engineering*, vol. 2, no. 3, pp. 460–463, 2012.
7. S. Arivazhagan, R. N. Shebiah, S. Ananthi, and S. V. Varthini, “Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features,” *Agricultural Engineering International: CIGR Journal*, vol. 15, no. 1, pp. 211–217, 2013.
8. P. Revathi and M. Hemalatha, “Cotton leaf spot diseases detection utilizing feature selection with skew divergence method,” *International Journal of scientific engineering and technology*, vol. 3, no. 1, pp. 22–30, 2014.
9. S. P. Patil and R. S. Zambre, “Classification of cotton leaf spot disease using support vector machine,” *International Journal of Engineering Research*, vol. 3, no. 4, pp. 1511–1514, 2014.
10. D. Pujari, R. Yakkundimath, and A. S. Byadgi, “SVM and ANN based classification of plant diseases using feature reduction technique,” *IJIMAI*, vol. 3, no. 7, pp. 6–14, 2016.
11. K. R. Gavhale, U. Gawande, and K. O. Hajari, “Unhealthy region of citrus leaf detection using image processing techniques,” in *International Conference for Convergence for Technology-2014*, Pune, India, 2014, pp. 1–6.
12. S. Sunny and M. I. Gandhi, “An efficient citrus canker detection method based on contrast limited adaptive histogram equalization enhancement,” *Int. J. Appl. Eng. Res.*, vol. 13, no. 1, pp. 809–815, 2018.
13. V. Vapnik and V. Vapnik, “Statistical learning theory Wiley,” *New York*, pp. 156–160, 1998.
14. S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep neural networks based recognition of plant diseases by leaf image classification,” *Computational intelligence and neuroscience*, vol. 2016, 2016.
15. U. Barman, R. D. Choudhury, A. Saud, S. Dey, and B. Kumar, “Estimation of Chlorophyll Using Image Processing,” *Int J Recent Sci Res*, vol. 9, no. 3, pp. 24850–24853, 2018.
16. U. Barman, “Prediction of Soil pH using Smartphone based Digital Image Processing and Prediction Algorithm,” *JMCMS*, vol. 14, no. 2, Apr. 2019.
17. S. P. Mohanty, D. P. Hughes, and M. Salathé, “Using deep learning for image-based plant disease detection,” *Frontiers in plant science*, vol. 7, p. 1419, 2016.
18. R. P. Shaikh and S. A. Dhole, “Citrus leaf unhealthy region detection by using image processing technique,” in *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)*, 2017, vol. 1, pp. 420–423.
19. T. N. Tete and S. Kamlu, “Detection of plant disease using threshold, k-mean cluster and ann algorithm,” in *2017 2nd International Conference for Convergence in Technology (I2CT)*, 2017, pp. 523–526.
20. [20] http://idtools.org/id/citrus/diseases/

AUTHORS PROFILE

Utpal Barman is acting as Research Scholar in the Department of Information Technology, Gauhati University. His area of Interest is Machine Learning, Application image processing in Agriculture. He has published many papers in Image Processing and machine learning domain.

Dr. Ridip Dev Choudhury is working as an Assistant Professor in the Department of Computer Science, IDOL, Gauhati University. He has published many papers in the domain of image processing and expert system. He is currently supervising 4 students as PhD scholar.