Application of Image Processing for Plant Disease Identification Based on SVM Technique

M. H. Nayeem[1], MehrabIbn Newaz[1], A.K Chakraborty[2], Md Asaduz Zaman Mamun[1]*

[1]Department of Electrical and Electronic Engineering,
[2]Department of Mathematics
Shahjajal University of Science and Technology Sylhet, Bangladesh-3114

Abstract – *Generally, plant diseases are detected by plant pathologists with the eye observation of different leaf syndromes. In this paper, we propose an automatic system which will be able to detect and classify the plant diseases automatically based on the preset datasets. This project deals with the image processing techniques and SVM algorithm to identify the disease symptoms of different plant leaves based on color, texture, shape, smoothness, variance, skewness and other image properties from an affected leaf image matrix. The diseases focused in this study include Alternaria Alternata, Cercospora leaf spot, Anthracnose, Bacterial canker, Bacterial leaf streak and Bacterial Blight. The goal of this study is to accurately detect and categorize the main symptoms of plant disease by extracting the features from the disease affected portion of a leaf image. The techniques used in our present study are k-means clustering for detection and SVM algorithm for classification of diseases. The experimental results demonstrate the validity of our proposed method as a robust technique for the detection of plant leaf diseases. In this work, a unique dataset of 175 images each containing a total of eleven features of the affected portion of a leaf image was created and it is a major contribution to the study of automatic detection of plant diseases.*

Index Terms - *Image Processing, Segmentation, plant leaf disease, K-means Method, SVM classifier*

I. INTRODUCTION

Agriculture sector plays a vital role to the survival of mankind by providing the basic needs. Moreover, the economic prosperity of a country is mostly dependent on agriculture. To meet the increasing demands of the growing population in the world, a sustainable growth of agricultural products is of prime concern. But unfortunately, a significant portion of agricultural products are lost as a result of several plant diseases. For example, a disease named Rice Brown Spot (BS) causes severe loss in the crops amount and Barnwal et al., in their paper [1] reported the variation of yield losses ranging from 4 to 52 % due to rice BS. So, it becomes very important to analyse plant diseases very accurately within specific time. Some diseases are visible to human eyes and can be easily detected and procured. Some are so sophisticated that those need powerful microscopes or specific electromagnetic spectrum to detect. For detection and identification of plant diseases, the direct eye observation of experts is the main approach adopted in practice. However, this requires continuous monitoring of experts which might be prohibitively expensive in large farms. Furthermore, farmers may have to go long distances to contact experts which is too expensive and time consuming to complete. Digital technology can make it very easy task to process all kind of disease images very precisely. It also gives the facility to remote sense the diseases without having an expert on the field. As a result, automatic detection of plant diseases has become a necessity and nowadays it is, indeed, a contemporary research topic. It may become beneficent in monitoring large fields of crops, and thus automatically detecting the symptoms of diseases as soon as they appear on plant leaves may lead to exclusion or reduction of the losses due to plant diseases.

Almost in every case, a plant disease is identified by the symptoms seen in its leaves. A very few cases require pathological tests of its leaves or other organs. To detect the disease of a leaf it is required to have an in-depth knowledge on disease symptoms in the leaf. For automatic detection of plant diseases, the only element available is the leaf images and thus image processing techniques are required. There are several image processing tools available to work with and, in this study, MATLAB’s image processing tool has been used. The shapes, sizes and colors of leaves of different species of trees are different and so the disease symptoms for a same disease in different plants may vary which makes the task of automatic detection more

challenging. To overcome this problem, only disease affected portions were segmented from a leaf image and several distinctive features were extracted for further analysis in this study.
This work consists of four major sections. They are:

i) Collection of a fair number of images and preprocessing of these images;
ii) Segmentation or clustering to extract the diseased portion of the image;
iii) Extraction of several features from the selected cluster image and forming a training data; and
iv) Classification based on a classifier.

At the first stage, several images were collected from different image databases and some images were collected from the farm of the Plant Pathology department of Bangladesh Agricultural University, Mymensingh. After collecting a good number of leaf image samples, we selected leaf images of six particular diseases. Healthy leaf samples of all these diseases were also collected. The pre-processing part consists of resizing, RGB to grayscale transformation, grayscale histogram formation, denoising and filtering. After that, segmentation of the diseased part of the image was done through the clustering. In this project K-means clustering method had been used and three clusters were formed from which the selection of a proper cluster is required. Definitely the selected cluster should be that one among the three which carries the disease affected portion most accurately. Then different features regarding that clusters' color, texture and shape were extracted. Training data of diseased leaf images are formed from the extracted features of the cluster. After creating a good number of training data, the classifying system has been formed and leaf images are then tested. In this project the classifying system is a multiclass SVM classifier.
The diagram in fig.1shows the overall working system of this study.

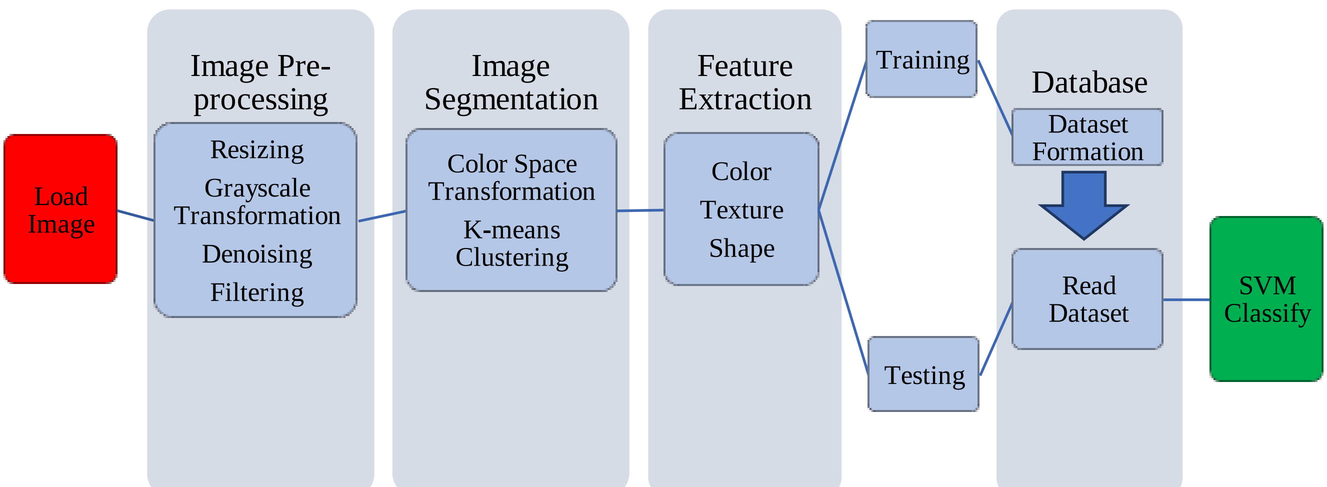
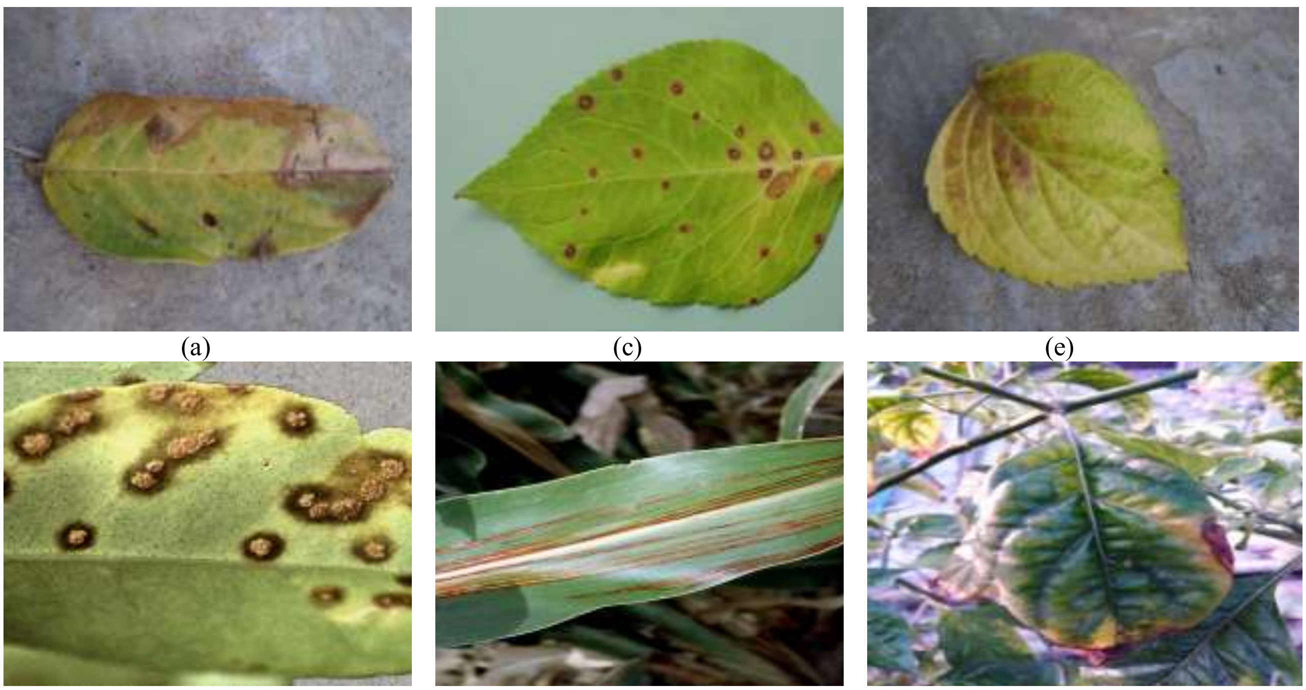


Fig.1 Block Diagram of the study

There are large number of plant diseases but ten to fifteen of those diseases are most common. Among those, six major diseases were selected for detection and classification purpose in this project. They are *Alternaria Alternata*[2], *Bacterial Canker*[4], *Cercospora Leaf Spot* [5] [6], *Bacterial leaf streak*[7] , *Anthracnose* [8] and *Bacterial Blight* [9].The images of fig.2 show the collected samples of each of these diseases.

II. BACKGROUND STUDY AND LITERATURE SURVEY

A good number of studies have been conducted on this topic of automatic detection of plant diseases during last 10 to 15 years by many scholars. The authors described several methods for detecting plant diseases. There are different approaches which are advantageous in different sectors. First, the field of this study starts with the basic work by identifying whether a leaf is disease affected or not. Then studies were carried on to identify some diseases accurately. Wenjiang Huang *et al.*,considered three different pests (Powdery mildew, yellow rust and aphids) in winter wheat for their study. The most and the least relevant wavelengths for different diseases were extracted using RELIEF-F algorithm. The classification accuracies of these new indices for healthy and infected leaves with powdery mildew, yellow rust and aphids were 86.5%, 85.2%, 91.6% and 93.5% respectively [11]. Zulkifli Bin Husin*et al*, in their paper [12], they captured the chilli plant leaf image and processed to determine the health status of the chilli plant. Their technique is ensuring that the chemicals should apply to the diseased chilli plant only. They used the MATLAB for the feature extraction and image recognition. Moreover, pre-processing was done using the Fourier filtering, edge detection and morphological operations. Computer vision extends the image processing paradigm for object classification. Here digital camera was used for the image capturing and LABVIEW software tool to build the GUI. The segmentation of leaf image is important while extracting the feature from that image. Mrunalini R. Badnakhe, Prashant R. Deshmukh compare the Otsu threshold and the k-means clustering algorithm was used for infected leaf analysis in [13]. They have concluded that the extracted values of the features are less for k-means clustering. The clarity of k-means clustering is more accurate than other method. H. Muhammad Asraf and others has proposed, Support Vector Machine (SVM) as classifier with three different kernels namely linear kernel, polynomial kernel with soft margin and polynomial kernel with hard margin [14]. Initial results show that the recognition of oil palm leaves is possible to be performed by SVM classifier. Based on the best performance result, polynomial kernel with soft margin iscapable of classifying nutrient diseases accurately in the oil palm leaves with accuracy of 95% of correct classification.



(a) (c) (e)

(b) (d) (f)

Fig.2 Collected leaf image samples showing (a)Alternaria Alternara, (b)Bacterial Canker, (c)Cercospora, (d)Bacterial leaf streak, (e)Bacterial Blight and (f)Anthracnose affected leaf symptoms

Satish Madhogaria*et al*. implemented an automatic pixel-based classification method for detecting unhealthy regions in leaf images [15]. Studies show that Machine learning methods can successfully be applied as an efficacious disease detection mechanism. Examples of such machine learning methods that have been applied in agricultural researches; Artificial Neural Networks (ANNs), Decision Trees, K-means, k nearest neighbors, and Support Vector Machines (SVMs). For example, Wang et al. in [16] predicted Phytophthora infestans infection on tomatoes by using ANNs. Also, Camargo and Smith in [17] used SVMs to identify visual symptoms of cotton diseases using SVMs. There are several classifiers available in machine learning. In our study we chose SVM classifier for classification. “Support Vector Machine” (SVM) [23, 24, 25] is a machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well. Support Vectors are simply the coordinates of individual observation which best segregates the two classes (hyper-plane/ line).

III. THE PROPOSED SYSTEM

In the introduction section it had been discussed that this study has four distinct sections. The diagram shown in Fig.3 is the overall working diagram of this study where all the parts of this work are shown.

A. Image acquisition and pre-processing

The digital images were captured directly from the tree leaves using a digital camera. Some images were collected from the album of plant pathology department of Bangladesh Agricultural University, Mymensingh. Some were collected from online image databases. For pre-processing, a color transformed structure for the RGB leaf image is created, and then, a device-independent color space transformation structure is applied. Images are normally found in RGB format in different sizes. The first task is to resize it into a fixed matrix and in this project the matrix was set to 256x256. After that, the contrast of that image is enhanced by specific tuning using the MATLAB's image processing toolbox.

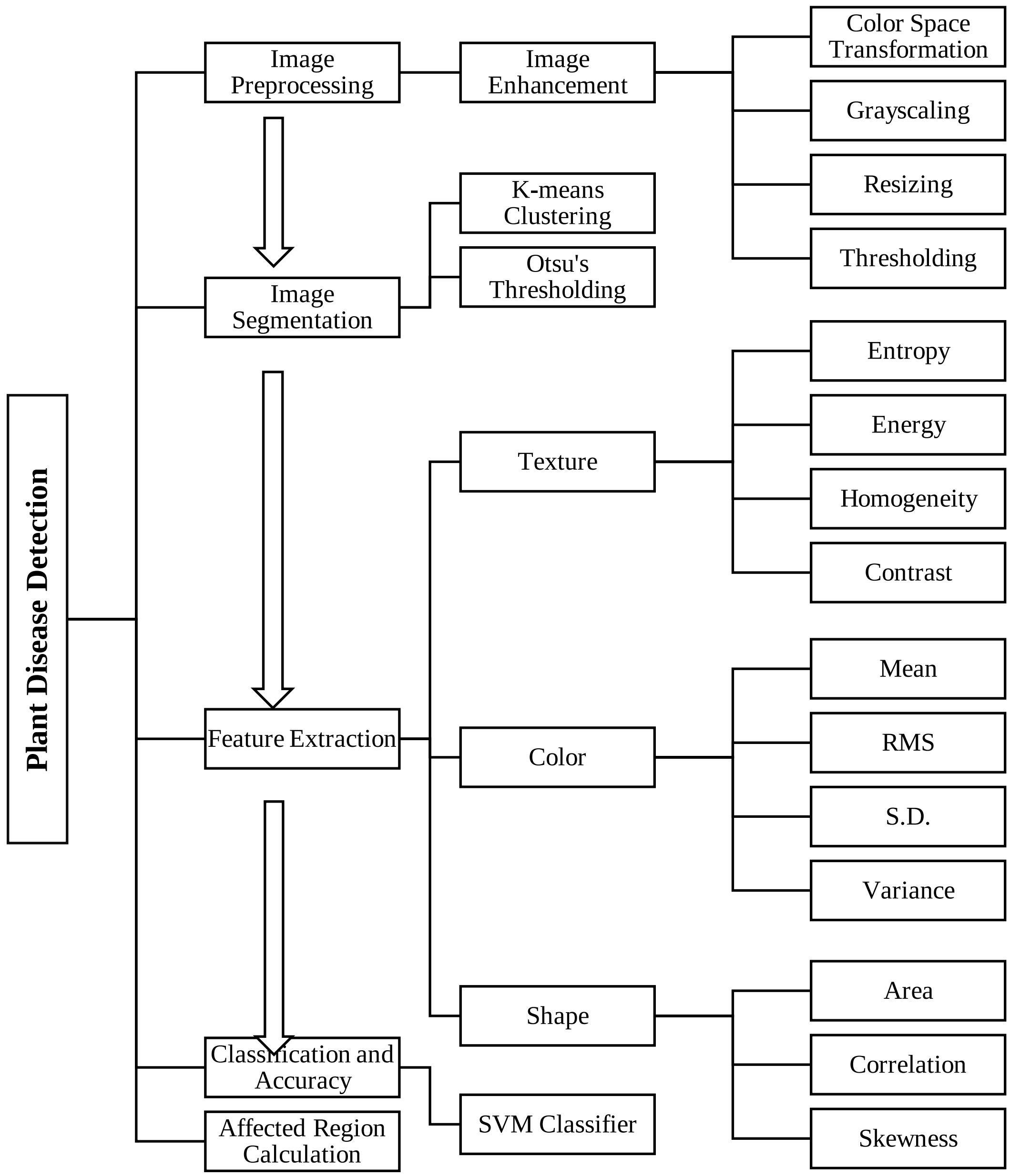


Fig.3 Block Diagram of the complete work

Pre-processing is common for operations with images at the lowest level of abstraction; both input and output are intensity images. The aim of pre-processing is an improvement of the image data that suppresses unwilling distortions or enhances some image features important for further processing, although geometric transformations of images are classified among pre-processing methods. Several pre-processing steps [18] were applied in this study and the images in Fig.4 demonstrate some of those steps. Here in Fig.4 (a) shows two samples in question and Fig.4 (b) shows the result images of these two images after the pre-processing steps.

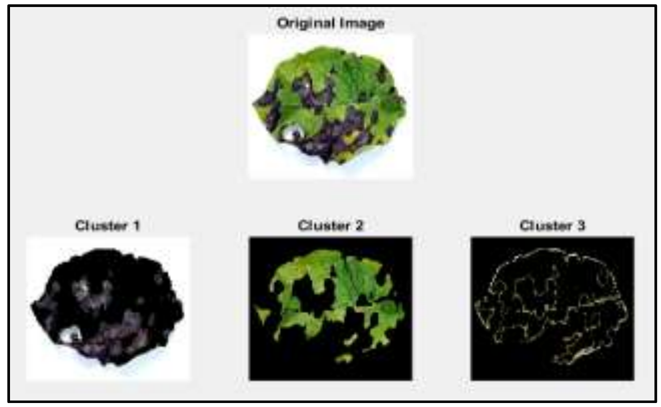
(a) (b)

Fig.4 (a) Raw image file, (b) Image after pre-processing (resizing and contrast enhancing)

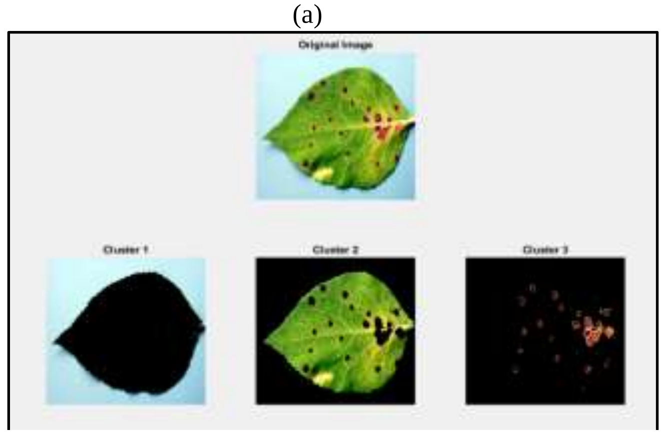
B. Image Segmentation

K-means clustering algorithm [19,20] was applied to crop the disease affected portion and cut off the green fresh portion of the query leaf image. For this clustering, the image is converted into lab color space. For example, an RGB 256x256 image is basically a 256x256x3 matrix which shows that a definite point of the matrix has three different values indicating red, green and blue pixel values. Similarly, the Lxaxbcolor space consists of a luminosity layer 'L', chromaticity layer 'a' and 'b'. All the color information is in the a and b layers. Since the image has three colors three clusters are created from the color space. The distances were measured using Squared Euclidean Distance Metric. K-means clustering is applied for partitioning the leaf image into three clusters in which one or more clusters contain the disease affected portion. A cluster is needed to be selected and the selected cluster should be among that one which contains the affected region of interest most accurately. In our project multiple number of clusters have been tested. For proper segmentation we also needed to mask the green pixels. This task consists of two steps: The mostly green colored pixels were identified, and then the global image threshold using Otsu's method [21,22] were applied in order to specify the varying threshold value which chooses the threshold to minimize the interclass variance of the threshold black and white pixels. Then, the green pixels are masked as follows: if the green component of pixel intensities is less than the computed threshold value, the red, green and blue components of this pixel are cleared. The Fig.5 shows two examples of segmentation process:

The distance measurement for k-means was done using the Squared Euclidean distance methodwhich indicates each centroid is the mean of the points in that cluster.



(a)



(b)

Fig.5 Applying k-means clustering. An original image transforming into three clusters

In the Fig.5 (a) a pre-processed image sample of the Alternaria Alternata disease with the application of k-means clustering is shown and, In the Fig.5 (b) a pre-processed image sample of the Cercospora leaf disease with the application of k-means clustering is shown. In this work, 3-means clustering was applied i.e., an image was segmented into three clusters from which one cluster contains the best segmented portion of the plant disease. Here one cluster should be chosen to extract the features or properties of the disease affected portion of the loaded leaf image.

C. Feature Extraction and Training Datasets Formation

After the cluster containing the disease affected portion of the leaf image has been selected, the features of the disease affected part are extracted. The image processing tool of MATLAB provides a good variety of features of an image. For this project, four features from the image Gray Level Co-occurrence Matrices (GLCMs) were taken. The features contrast, correlation, energy and homogeneity are taken from the GLCM. Mean standard deviation, entropy, rms value, skewness, variance, smoothness, kurtosis and IDM-these properties are taken from the segmented cluster of the RGB image matrix. For example, Correlation of an image matrix is found from:

$$d(x,c) = 1 - \frac{(x-\vec{x})(c-\vec{c})'}{\sqrt{(x-\vec{x})(x-\vec{x})'(c-c')(c-\vec{c})'}}$$

Where,

$$\vec{x} = \frac{1}{p}(\sum_{j=1}^{p} Xj)\overrightarrow{1p}$$

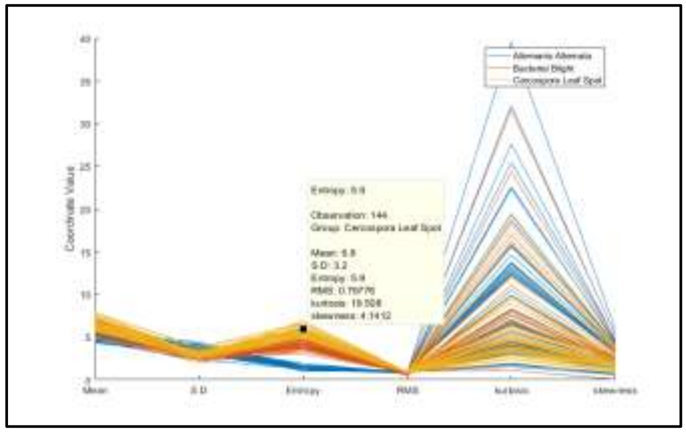
$$\vec{c} = \frac{1}{p}(\sum_{j=1}^{p} Cj)\overrightarrow{1p}$$

$\overrightarrow{1p}$is a row vector of p ones.In the formulae, x is an observation (that is, a row of X) and c is a centroid (a row vector).Similarly, other features were extracted from the image matrix. Taking these different values as different variables a training matrix was formed. At the primary stage, the project was aimed to identify six leaf diseases, healthy leaves and other unidentified diseases. First trained datasets were formed by extracting features from twenty-five to thirty pictures for each class. The area of the disease affected part was also calculated using the zero images and comparing the zero images with the selected cluster image.Five samples of trained datasets are given in TABLE Ithese are the datasets of the disease Alternaria Alternata plant disease.

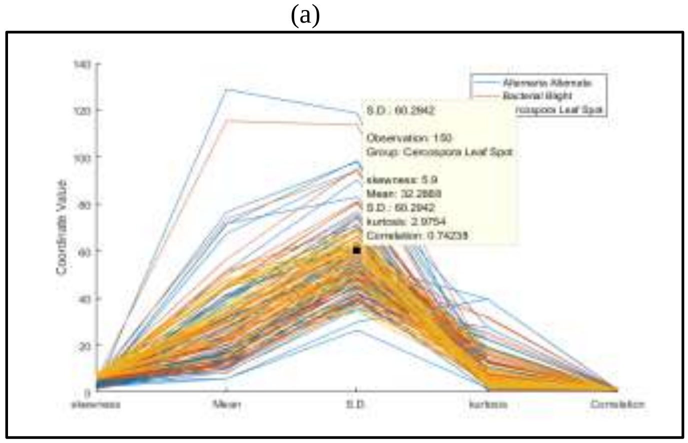
TABLE I
Five Training Datasets

Contrast	Correlation	Energy	H.geneity	Mean	S.D.	Entropy	RMS	Variance	Kurtosis	Skewness	IDM
0.08	0.98	0.76	0.97	14.84	47.81	1.71	5.57	2150.70	15.60	3.63	255
0.47	0.87	0.80	0.96	14.15	48.14	1.37	4.31	1632.22	15.77	3.67	255
0.37	0.91	0.76	0.96	16.44	51.42	1.67	5.34	2305.04	13.79	3.40	255
0.54	0.75	0.54	0.92	17.97	37.66	2.58	7.40	1306.81	10.50	2.59	255
0.51	0.71	0.89	0.97	17.12	35.52	2.84	10.45	1162.23	27.60	4.68	255

At the time of testing, the classifier uses the testing matrix to compare with the training matrix.The two graphs shown in Fig.4 (a) and (b) consecutivelyarethe parallel co-ordinate plotting of our trained datasets.



(a)



(b)

Fig.4 Parallel co-ordinate plot of the trained datasets showing variations in different leaf samples

From the plots in Fig.4 we can see the distinctions among the properties from different disease affected leaf samples. The features of a class of a disease sum up a different value which is distinctive from the other class of disease.

D. Classification

This is the most important part of the project. As our dataset had a moderate number of elements for classification we chose multi-level SVM classifier [26, 27, and 28]. A multilevel SVM classification algorighm code was formed. For example, the segmented image that was discussed in this paper previously has the properties as shown in TABLE II:

TABLE II
A Testing Data

Contrast	Correlation	Energy	H.geneity	Mean	S.D.
0.08	0.85	0.78	0.91	13.76	51.81

Entropy	RMS	Variance	Kurtosis	Skewness	IDM
1.71	5.57	2265.70	12.60	3.69	255

Which is mostly similar with the training matrix set of the disease Alternaria Alternata? So, the designed multiclass SVM model compares these data with the given trained datasets. In this project the classifier was formed to have 8 classes which could identify six diseases, healthy leaves and unidentified leaves.

Primarily the SVM algorithm is a binary classifier algorithm which separates two classes of data by a hyperplane. But some techniques are available to use SVM for multilevel classification. In our work we used a multilevel technique for which trained data needed to be formed into another dataset separated by 8 class labels. The multilevel technique is such that the features of a testing sample enter the classification segment by setting a predefined class to its variables. Then it compares each of the class one by one as a binary classifier. So this technique can be called a one vs one to multilevel classification algorithm.

IV. CONCLUSION

In this paper, the applications of K-means clustering and multi class Support Vector Machines (SVMs) have been formulated for clustering and classification of diseases that have effects on plant leaves. Identifying and classifying the disease is mainly the purpose of the proposed approach. Thus, the proposed algorithm was tested on six diseases which influence on the plants; they are: Alternaria Alternata, Cercospora Leaf Spot, Bacterial Blight, Anthracnose, Bacterial Canker and Bacterial Leaf Streak. The experimental results indicate that the proposed approach is a viable approach, which can significantly support an accurate detection of leaf diseases with a little computational effort. An extension of this work will focus on developing hybrid algorithms such as generic algorithms and Neural Networks in order to increase the recognition rate of the final classification process underscoring the advantages of hybrid algorithms; also, we will dedicate our future works on automatically estimating the severity of the detected disease.

REFERENCES

[1] M. K. Barnwal, A. Kotasthane,N. Magculia, P. K. Mukherjee, S. Savary, A. K. Sharma, H. B. Singh, U. S. SinghA, H. Sparks, M. Variar, N. Zaidi "A review on crop losses, epidemiology and disease management of rice brown spot to identify research priorities and knowledge gaps" European Journal of Plant Pathology, July 2013, Volume 136, Issue 3, pp 443–457

[2] JiaChengguo; Zhang Liping; Liu Lihong; Wang Jiansheng; Li Chuanyou; Wang Qiaomei (2013-01-01). "Multiple phytohormone signalling pathways modulate susceptibility of tomato plants to Alternariaalternata f. sp. lycopersici". Journal of Experimental Botany. 64 (2): 637–650. doi:10.1093/jxb/ers360. ISSN 0022-0957. PMC 3542053

[3] D. (1 April 1999). "DNA relatedness among the pathovars of Pseudomonas syringae and description of Pseudomonas tremae sp. nov.and Pseudomonas cannabina sp. nov. (ex Sutic and Dowson 1959)". International Journal of Systematic Bacteriology. 49 (2): 469–478. doi:10.1099/00207713-49-2-469. PMID 10319466

[4] Gardan, L.; Shafik, H.; Belouin, S.; Broch, R.; Grimont, F.; Grimont, P. A. D. (1 April 1999). "DNA relatedness among the pathovars of Pseudomonas syringae and description of Pseudomonas tremae sp. nov.and Pseudomonas cannabina sp. nov. (ex Sutic and Dowson 1959)". International Journal of Systematic Bacteriology. 49 (2): 469–478. doi:10.1099/00207713-49-2-469. PMID 10319466

[5] Brooklyn Botanic Garden. (2000): Natural disease control: A common-sense approach to plant first aid. Handbook # 164. Brooklyn Botanic Garden, Inc. 1000 Washington Avenue, Brooklyn, NY

[6] Horst, Kenneth; Rev, By R. Cynthia Westcott (2001), Westcott's plant disease handbook, Boston, Mass.: Kluwer Academic Publishers, pp. 245, 678, ISBN 0-7923-8663-9

[7] Gardiner, Donald M.; Upadhyaya, Narayana M.; Stiller, Jiri; Ellis, Jeff G.; Dodds, Peter N.; Kazan, Kemal; Manners, John M. (2014). "Genomic Analysis of XanthomonasTranslucens Pathogenic on Wheat and Barley Reveals Cross-Kingdom Gene Transfer Events and Diverse Protein Delivery Systems." Ed. TurgayUnver". PLoS ONE. 9 (1): E84995. doi: 10.1371/journal.pone.0084995. PMC 3887016. PMID 24416331

[8] Brooklyn Botanic Garden. (2000): Natural disease control: A common-sense approach to plant first aid. Handbook # 164. Brooklyn Botanic Garden, Inc. 1000 Washington Avenue, Brooklyn, NY

[9] Mathre, D.E. (1997). Compendium of barley diseases. American Phytopathological Society. pp. 120 pp

[10] Wenjiang Huang, Qingsong Guan, JuhuaLuo, Jingcheng Zhang, Jinling Zhao, Dong Liang, Linsheng Huang, and Dongyan Zhang, "New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases", IEEE journal of selected topics in applied earth observation and remote sensing,Vol. 7, No. 6, June 2014

[11] Zulkifli Bin Husin, Abdul Hallis Bin Abdul Aziz, Ali Yeon Bin MdShakaffRohaniBinti S Mohamed Farook, "Feasibility

Study on Plant Chili Disease Detection Using Image Processing Techniques", 2012 Third International Conference on Intelligent Systems Modelling and Simulation.

[12] Mrunalini R. Badnakhe, Prashant R. Deshmukh, "Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Translated J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digest 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[13] H. Muhammad Asraf, M. T. Nooritawati, M.S.B. Shah Rizam," A Comparative Study in Kernel-Based Support Vector Machine of Oil", Procedia Engineering 41 (2012) 1353 – 1359

[14] SatishMadhogaria, MarekSchikora, Wolfgang Koch, Daniel Cremers," Pixel-Based Classification Method for Detecting Unhealthy Regions in Leaf Images", Informatik 2011 - Informatikschafft Communities 41. Jahrestagung der GesellschaftfürInformatik , 4.-7.10.2011, Berlin

[15] Wang, X., M. Zhang, J. Zhu and S. Geng, Spectral prediction of Phytophthorainfestans infection on tomatoes using artificial neural network (ANN), International Journal of Remote Sensing 29 (6) (2008), pp. 1693–1706.

[16] Camargo, A. and Smith, J. S., (2009). An imageprocessing based algorithm to automatically identify plant disease visual symptoms, Biosystems Engineering, Volume 102, Issue 1, January 2009, Pages 9-21, ISSN 1537-5110, DOI: 10.1016/j.biosystemseng.2008.09.030.

[17] Digital Image Processing Techniques by P. Prabhu, Alagappa University.

[18] Hartigan, J. A.; Wong, M. A. (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C (Applied Statistics) 28 (1): 100–108

[19] Ali, S. A., Sulaiman, N., Mustapha, A. and Mustapha, N., (2009). K-means clustering to improve the accuracy of decision tree response classification. Inform. Technol. J., 8: 1256-1262. DOI: 10.3923/itj.2009.1256.1262

[20] Otsu, N. (1979). "A threshold selection method from gray-level histograms". IEEE Trans. Sys., Man., Cyber. 9: 62– 66. DOI:10.1109/TSMC.1979.4310076

[21] Otsu, N., "A Threshold Selection Method from Gray- Level Histograms," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, 1979, pp. 62-66

[22] Meyer, David; Leisch, Friedrich; Hornik, Kurt (2003). "The support vector machine under test". Neurocomputing.55 (1–2): 169–186. doi:10.1016/S09252312(03)00431-4

[23] Hsu, Chih-Wei; Chang, Chih-Chung & Lin, Chih-Jen (2003). A Practical Guide to Support Vector Classification (Technical report). Department of Computer Science and Information Engineering, National Taiwan University.

[24] Platt, John C. (1998). Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. NIPS.

[25] Duan, Kai-Bo; Keerthi, S. Sathiya (2005). "Which Is the Best Multiclass SVM Method? An Empirical Study". Multiple Classifier Systems. LNCS. 3541. pp. 278– 285. CiteSeerX 10.1.1.110.6789. doi:10.1007/11494683_28. ISBN 978-3-54026306-7.

[26] Hsu, Chih-Wei & Lin, Chih-Jen (2002). "A Comparison of Methods for Multiclass Support Vector Machines". IEEE Transactions on Neural Networks.

[27] Crammer, Koby& Singer, Yoram (2001). "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines". Journal of Machine Learning Research. 2: 265–2 Lee, Yoonkyung; Lin, Yi &Wahba, Grace (2001). "Multicategory Support Vector Machines". Computing Science and Statistics. 33.