Plant Leaf Disease Detection and Classification using Image Processing

Author(s): *[1] Yin Min Oo, [2] Nay Chi Htun,
Affiliation(s): [1] Department of Information Technology,
Pyay Technological University
[2] Department of Information Technology,
Pyay Technological University
*Corresponding Author: yinminoo.it.85@gmail.com,

ORIGINAL ARTICLE

Abstract - Myanmar is an agricultural country and then crop production is one of the major sources of earning. So, more than half of our population depends on agriculture for livelihood. Due to the factors like diseases, pest attacks and sudden change in the weather condition, the productivity of the crop decreases. Automatic detection of plant diseases is essential to automatically detect the symptoms of diseases as early as they appear on the growing stage. This paper proposed a methodology for the analysis and detection of plant leaf diseases using digital image processing techniques. The experimental results demonstrate that the proposed system can successfully detect and classify four major plant leaves diseases: Bacterial Blight and Cercospora Leaf Spot, Powdery Mildew and Rust.

Keywords: Image Processing; K-Means; Local Binary Pattern; Support Vector Machine; Plant Disease Detection

I. INTRODUCTION

The Union of Myanmar is an agricultural country, and agriculture sector is the backbone of its economy. Its economy has traditionally been based on agriculture. That is because nature has blessed it with vast areas of fertile land and abundant success of water, which are the principal ingredients of an agro-based economy. Diseases and insect pests are the major problems in agriculture. These require careful diagnosis and timely handling to protect the crops from heavy losses. In plant, diseases can be found in various parts such as fruit, stem and leaves. Plant diseases cause periodic outbreak of diseases which leads to large scale death and famine [5].

Diseases are the natural factor that can cause some serious effects on plants which ultimately reduces productivity, quality and quantity of products. The main part of plant to examine the plant diseases is leaf. The detection and classification of leaf diseases accurately is the key to prevent the agriculture loss. Different plant leaf bears different diseases. The major categories of plant leaf diseases are based on viral, fungal and bacteria [18]. The most common plant diseases are Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora Leaf Spot, Powdery Mildew, Downy Mildew and Rust.

The naked eye observation of experts is the main approach used in practice for detection and identification of plant diseases. But, this needs continuous monitoring of experts. When there is a large farm, this approach might be prohibitively expensive. Further, in some developing countries, farmers may have to go long distances to contact experts, this makes consulting experts too expensive and time consuming and moreover farmers are unaware of non-native diseases [3].

Automatic detection of plant diseases is an important research topic as it may prove benefits in monitoring large field of crops, and thus automatically detect diseases from symptoms that appear on plant leaves. Thus automatic detection of plant disease with the help of image processing technique provides more accurate and robot guidance for disease management. Comparatively, visual identification is less accurate and time consuming [14].

This automated system is designed to overcome the problems of manual techniques. The image could be captured using a regular digital camera or high resolution mobile phone camera. This image is given as an input to the system for obtaining the leaf features. The system consists of several steps like segmentation, feature extraction, identification and classification [8]. Rest of the paper is organized as follows. Section 2 describes a brief review on Literature. Section 3 presents a detailed description of basic steps to detect and classify disease on plant leaf. Section 4

describes the proposed system design and Section 5 concludes this paper.

II. RELATED WORKS

Agriculture is the mother of all nations. Research in agriculture domain is aimed towards increase the quality and quantity of the product at less expenditure with more profit. The quality of the agricultural product may be degraded due to plant diseases. These diseases are caused by pathogens viz.., fungi, bacteria and viruses. Therefore, to detect and classify the plant disease in early stage is a significant task. Farmers require constant monitoring of experts which might be prohibitively expensive and time consuming. Depending on the applications, many systems have been proposed to solve or at least to reduce the problems, by making use of image processing and some automatic classification tools [9, 11].

Suhaili Kutty et al. [16] proposed the process to classify Anthracnose and Downey Mildew, watermelon leaf diseases. For this region of interest need to be identified from infected leaf sample based on RGB color component. Then to reduce noise and for segmentation median filter is used. And for classification, neural network pattern recognition toolbox is used. Proposed method achieved 75.9% of accuracy based on its RGB mean color component.

The goal of Sanjeev Sannaki et al. [15] is to diagnose the disease using image processing and artificial intelligence techniques on images of grape plant leaf. They classify mainly two diseases, downy mildew and powdery mildew of grape leaf. Masking is used to remove background to improve accuracy. For preserving information of affected portion of leaf, Anisotropic Diffusion is used. Segmentation is carried out using k-means clustering method. After segmentation, feature extraction take place by calculating Gray Level Co-occurrence Matrix. And finally classification is done using Feed Forward Back Propagation Network classifier. They have used only Hue feature which gives more accurate result.

Akhtar et al. [2] have used the support vector machine approach for the classification and detection of rose leaf diseases as black spot and anthracnose. Authors have used the threshold method for segmentation and Ostu's algorithm was used to define the threshold values. In this approach, features of DWT, DCT and texture based eleven haralick features are extracted which are further used with SVM approach and shows efficient accuracy value.

S. Dubey and R. Jalal [4] explored the concept of detection and classification of apple fruit diseases, namely, scab, apple rot and apple blotch. For that, segmentation is done using K-means clustering technique. Then features are extracted from the segmented image. For classification Multiclass Support Vector Machine (SVM) is used.

Usama Mokhtar et al. [17] described technique of Tomato leaves diseases detection and diseases are: Powdery mildew and Early blight. Image preprocessing involved various techniques such as smoothness, remove noise, image resizing, image isolation and background removing for image enhancement. Gabor wavelet transformation is applied in feature extraction for feature vectors also in classification. Cauchy Kernel, Laplacian Kernel and Invmult Kernel are applied in SVM for output decision and training for disease identification.

Sachin Khirade and A. B. Patil [13] discussed about the main steps of image processing to detect disease in plant and classify it. It involves steps like image acquisition, image preprocessing, image segmentation, feature extraction and classification. For segmentation, methods like, otsu's method, converting RGB image into HIS model and k-means clustering are there. Among all, k-means clustering method gives accurate result. After that, feature extraction is carried out like, color, texture, morphology, edges etc. Among this, morphology feature extraction gives better result. After feature extraction, classification is done using classification methods like Artificial Neural Network and Back Propagation Neural Network.

Bhog and Pawar [7] have incorporated the concept of neural network for the classification of cotton leaf disease analysis. For segmentation, K-means clustering has been used. Different cotton leaf diseases are like Red spot, white spot, Yellow spot, Alternaria and Cercospora on the Leaf. For experimentation, MATLAB toolbox has been used. The recognition accuracy for K-Mean Clustering method using Euclidean distance is 89.56% and the execution time for K-Means Clustering method using Euclidean distance is 436.95 second.

Ms. Kiran R. Gavhale et al. [10] presented number of image processing techniques to extract diseased part of leaf. For Pre-processing, Image enhancement is done using DCT domain and color space conversion is done. After that segmentation take place using k-means clustering method. Feature extraction is done using GLCM Matrix. For classification of canker and anthracnose disease of citrus leaf, SVM with radial basis kernel and polynomial kernel is used.

In this paper, the automated plant leaf disease detection system is performed by five main steps: image acquisition, image preprocessing, segmentation, feature extraction and classification. Diseased leaf images are captured and stored for experiment. Then images are applied for preprocessing for image enhancement. Captured leaf images are segmented using K-Means clustering method to form

clusters. GLCM and LBP features are extracted after applying K-Means and SVM has been used for classification and detection of plant leaves diseases namely Bacterial Blight, Cercospora Leaf Spot, Powdery Mildew and Rust.

III. PROPOSED SYSTEM

```
           ┌─────────────┐
           │    Start     │
           └─────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │  Input Disease Affected Image │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │     Resize and Filter Image   │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │  Enhance Image Contrast using │
    │    Histogram Equalization     │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │  Apply K-Means Clustering for │
    │      Image Segmentation       │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │  Image is segmented into three│
    │            regions            │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │  Select the disease affected  │
    │            region             │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │   Extract GLCM and LBP features│
    │          from image           │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │  Apply SVM for classification │
    │      and disease detection    │
    └──────────────────────────────┘
                  │
                  ▼
    ┌──────────────────────────────┐
    │   Display disease type and    │
    │ percentage of affected portion│
    └──────────────────────────────┘
                  │
                  ▼
           ┌─────────────┐
           │     End      │
           └─────────────┘
```
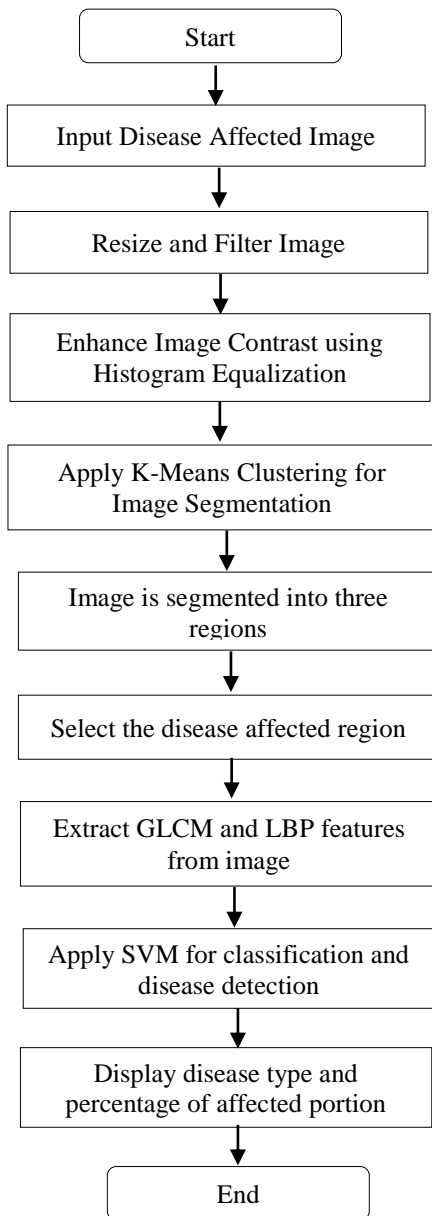
Fig. 1. Flowchart of Proposed System

In this proposed work, we have applied image processing techniques for detection of plant leaf diseases. The proposed methodology that applied in this work is shown in Fig. 1, in which we followed the image acquisition, preprocessing, segmentation, feature extraction using GLCM and LBP method and then sample images was tested by SVM classifier.

A. Image Acquisition

The first step in the proposed approach is image acquisition. In this system, the available images from the digital camera or internet are also taken. Total 560 image samples are captured of four classes namely Bacterial Blight, Cercospora Leaf Spot, Powdery Mildew and Rust leaf disease which are basically found on chilli, grape, rice, soya bean, wheat, rose, cotton, apple, mango etc. From these, 115 are of Bacterial Blight, 120 images are of Cercospora Leaf Spot, 200 are of Powdery Mildew and 125 are of Rust leaf disease. Fig. 2 below shows the sample diseased leaf images.
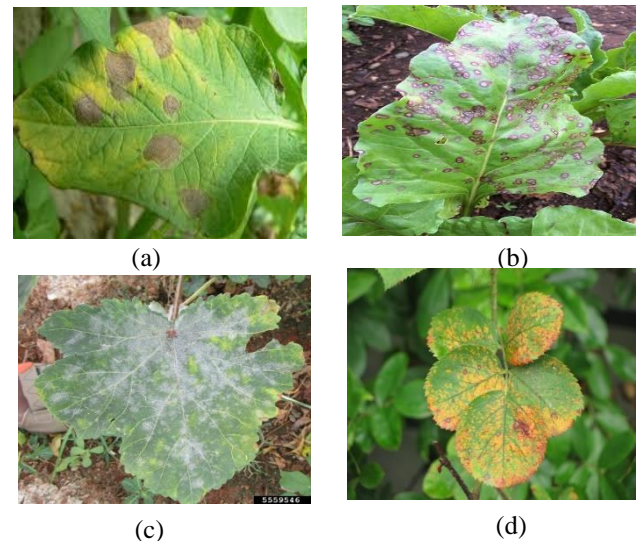

Fig. 1. Sample diseased leaf images for (a) Bacterial Blight, (b) Cercospora leaf spot, (c) Powdery Mildew and (d) Rust.

B. Image Preprocessing

Fig. 2. Preprocessing results of cercospora leaf spot disease image, (a) Original image, (b) Enhanced diseased leaf image.

Preprocessing step is to improve image data by removing background, noise and also suppressing undesired distortions. It enhances image features for processing and analysis. The images stored in RGB format are resized to standard size. These resized RGB images are then converted to HSV format [1]. The median filter is used for image smoothing, removal of noises and highlighting some information. Image enhancement is carried out for increasing the contrast. The histogram equalization which distributes the intensities of the images is applied on the image to enhance the plant disease images. Fig. 3 shows the preprocessing result of Cercospora Leaf Spot disease image as an example.

C. Image Segmentation

Image segmentation is applied to simplify the illustration of image with segments so that it can be easily analyzed. Image segmentation is performed to segment the disease affected and unaffected portions of the leaf. K-Means clustering method is used for partitioning of images into clusters in which at least one part of cluster contain image with major area of diseased part. The k-Means clustering algorithm is applied to classify the objects into K number of classes according to set of features.

The K-Means clustering algorithm tries to classify objects based on a set of features into K number of classes. The classification is done by minimizing the sum of squares of distances between the objects and the corresponding cluster or class centroid [3]. In our experiments multiple values of number of clusters have been tested. Best results were observed when the number of clusters was three. So, the image is partitioned into three clusters for good segmentation result. An example of the output of K-Means clustering for a leaf infected with Cercospora Leaf Spot disease is shown in Fig. 4. It is observed from Fig. 4 that cluster image 2 contains infected object of Cercospora Leaf Spot disease.

(b)

(c)   (d)
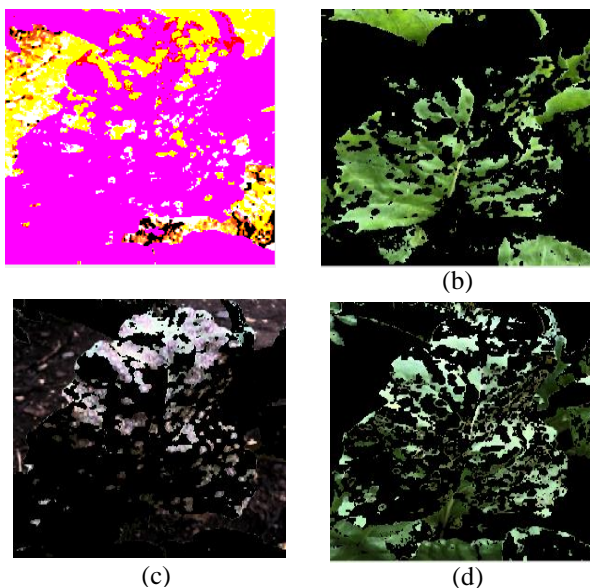
Fig. 3. Segmentation result of cercospora leaf spot image, (a)image conversion, (b) segmented cluster image 1, (c) segmented cluster image 2, (d) segmented cluster image 3.

D. Feature Extraction

After segmentation the infected region, various features are extracted to describe the infected region. HSV color and texture features are used for region description. HSV color features are important to sense image environment, recognize objects and convey information [5]. Texture is one of the most important feature which can be used to classify and recognize objects [12].

Grey Level Co-occurrence Matrices is a statistical method. It is an old and used feature extraction method for texture classification. It has been an important feature extraction method in the domain of texture classification that computes the relationship between pixel pairs in the image. The textural features can be calculated from the generated GLCMs, e.g. contrast, correlation, energy, entropy and homogeneity. Extract the disease symptoms by calculating the GLCM texture feature values of Skewness, Standard Deviation, Homogeneity, Contrast, Smoothness, Correlation, Kurtosis, Energy, Entropy, Mean, Variance, RMS, and IDM [3].

Local Binary Pattern (LBP) is also a type of texture feature used for classification in computer vision. LBP is the case of the Texture Spectrum model. A local binary pattern is called uniform if the binary pattern contains at most two 0-1 or 1-0 transitions. Using uniform patterns, the length of the feature vector for a 3x3 window reduces from 256 to 59. Concatenate normalized histograms of all cells. This gives the feature vector for the window [4]. With the help of texture features, plant diseases are classified into different types [19].

E. Classification

After extracting color and texture features, the classification is performed by using Support Vector Machine (SVM). A support vector machine constructs a hyper-plane or set of hyper-planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. SVM is supervised learning model with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis [18]. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.

The training and validation processes are among the important steps in developing an accurate process model using SVM. The dataset for training and validation processes consists of two parts: the training feature set which are used to train the SVM model and the testing features sets are used to verify the accuracy of the trained SVM model. Finally disease type with accuracy value is analyzed and percentage of disease affected region is evaluated by the ratio of disease data and leaf data.

IV. EXPERIMENTAL RESULT AND ANALYSIS

Total 560 images are used for learning of the system. Learning is a process by which the system learns the input parameter and classifies the input images into different classes. In this first infected plant leaf is loaded by using load image. Then the image is enhanced to reduce the noise from the image and after that segmentation is done using k-Means clustering. Then infected portion of leaf is extracted by masking green pixel values and then GLCM and LBP texture features of that infected area is calculated. After that classification tools is used which indicate the disease name and accuracy is calculated.

The classification of diseased plant leaves performance of various classification techniques which have been analyzed for the 560 input leaf images. From these, 115 are of Bacterial Blight, 120 images are of Cercospora Leaf Spot, 200 are of Powdery Mildew and 125 are of Rust leaf disease. The performance evaluated for the classification techniques which have been used in this paper from the confusion matrix of their respective classifier. In confusion matrix, class label 1 represents Cercospora Leaf Spot, class label 2 represents the Bacterial Blight disease, class label 3 represents Powdery Mildew and class label 4 represents Rust. The Fig. 4 shows the confusion matrix for Support Vector Machine classification results for 560 input leaf samples.
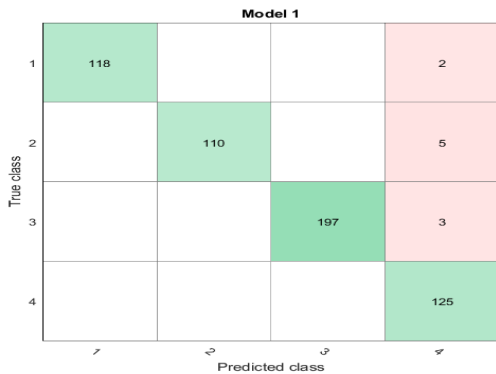
Fig. 5. Confusion Matrix for SVM Classification.

In SVM classification, 118 images are of Cercospora Leaf Spot disease are correctly classified and misclassified samples are 2.The correct classification rate for Cercospora Leaf Spot disease is 98%. For Bacterial Blight disease, 110 samples are correctly classified out of 115.The correct classification for Bacterial Blight disease is 96%. For Powdery Mildew disease, 197 samples are correctly classified out of 200.The correct classification for Powdery Mildew disease is 99%. For Rust leaf disease, all 125 samples are correctly classified. So the correct classification for Rust leaf disease is 100%. The overall system classification rate for the above leaf samples are 98.2% and the error rate of the system is 0.8%.
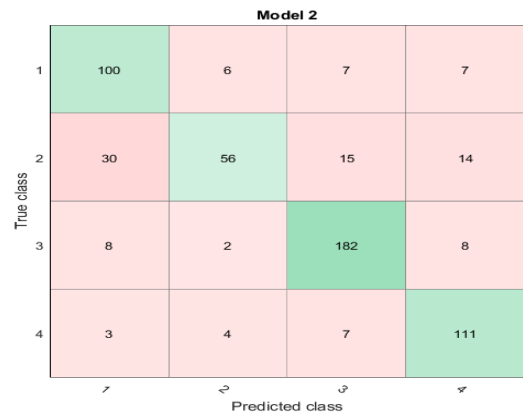

Fig. 6. Confusion Matrix for KNN Classification.

In KNN classification, 100 images are of Cercospora Leaf Spot disease are correctly classified and misclassified samples are 20.The correct classification rate for Cercospora Leaf Spot disease is 83%. For Bacterial Blight disease, 56 samples are correctly classified out of 115.The correct classification for Bacterial Blight disease is 49%. For Powdery Mildew disease, 182 samples are correctly classified out of 200.The correct classification for Powdery Mildew disease is 91%. For Rust disease, 111 samples are correctly classified out of 125.The correct classification for Rust leaf disease is 89%. The overall system classification rate for the above leaf samples are 80.2% and the error rate of the system is 19.8%.
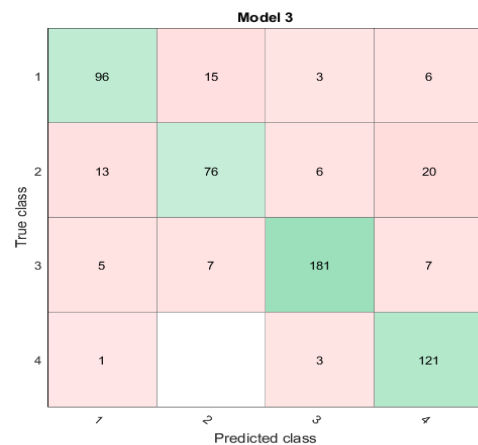

Fig. 7. Confusion Matrix for Ensemble Classification.

In Ensemble classification, 96 images are of Cercospora Leaf Spot disease are correctly classified and misclassified samples are 24.The correct classification rate for Cercospora Leaf Spot disease is 80%. For Bacterial Blight disease, 76 samples are correctly classified out of 115.The correct classification for Bacterial Blight disease is 66%. For Powdery Mildew disease, 181 samples are correctly classified out of 200.The correct classification for Powdery Mildew disease is 91%. For Rust leaf disease, 121 samples are correctly classified out of 125.The correct classification for Rust leaf disease is 97%. The overall system classification rate for the above leaf samples are 84.6% and

the error rate of the system is 15.4%.

TABLE I. ACCURACY COMPARISON OF THREE CLASSIFIERS

Disease Class	SVM Classifier	KNN Classifier	Ensemble Classifier
Cercospora Leaf Spot	98%	83%	80%
Bacterial Blight	96%	49%	66%
Powdery Mildew	99%	91%	91%
Rust	100%	89%	97%
Average	98.2%	80.2%	84.6%

TABLE II. ERROR RATE COMPARISON OF THREE CLASSIFIERS

Disease Class	SVM Classifier	KNN Classifier	Ensemble Classifier
Cercospora Leaf Spot	2%	17%	20%
Bacterial Blight	4%	51%	34%
Powdery Mildew	2%	9%	9%
Rust	0%	11%	3%
Average	0.8%	19.8%	15.4%

In case of SVM, the best accuracy is obtained under cubic kernel. From the experimental result it can be observed that best accuracy is achieved for SVM classifier. For all classifiers 10-folds cross validation is considered. The performances of these classifiers are summarized in Table 1. From the above results, the proposed system is efficient enough for detection and classification of plant leaf diseases. This comparison of results can also be show in graphical form as in Figure 8.
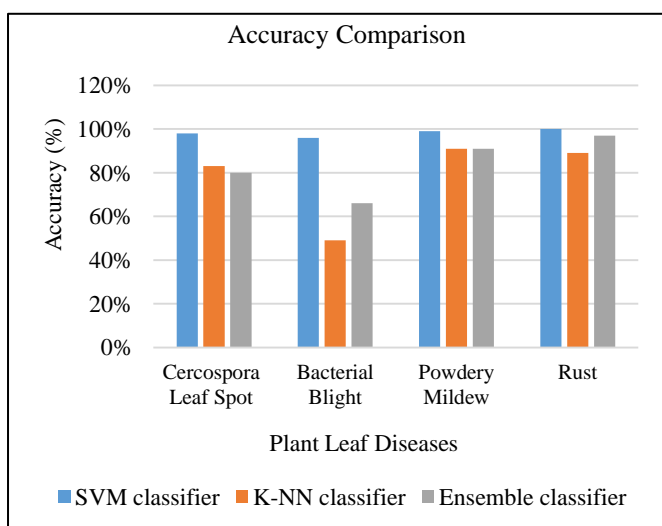
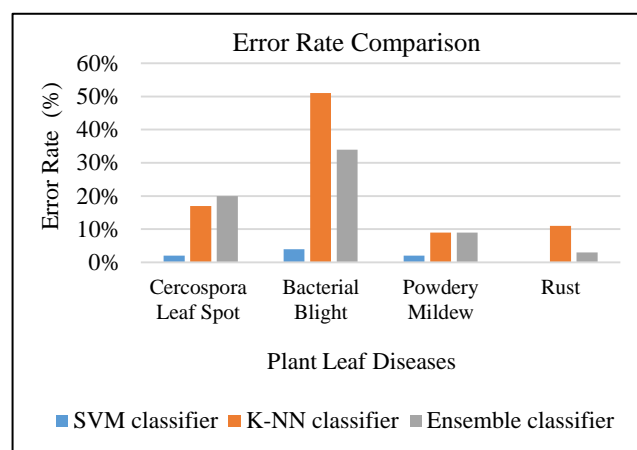Fig. 8. Accuracy comparison of three classifiers.

Fig. 9. Error rate comparison of three classifiers.

V. FINDINGS AND DISCUSSIONS

This section elaborates the evaluated results with proper discussion. Windows 10 based system with 4GB of RAM, 1000 GB of HDD, an Intel(R) Core(TM) i7 CPU, is used for conducting the experiments. MATLAB is used for the simulation of work. In MATLAB, a Graphical User Interface (GUI) base interface is generated for the experimentation. The stepwise process of image processing for plant leaf disease detection is explained in previous section.

The main goal of this work is to develop an image processing system that can identify and classify four types of plant disease namely Cercospora Leaf Spot, Bacterial Blight, Powdery Mildew and Rust. Using SVM classifier, the experimentation is performed for more than 500 images. From the evaluated results, we have analyzed that the percentage of diseased portion also affect the overall crops/agriculture land. Accuracy also varies from different images.

In disease detection, the disease affected portion of the plant leaf is first identified using GLCM and LBP features and then SVM classifier. The detection accuracy rate is 98.2%. In addition, the plant disease type is classified using GLCM and LBP features and classifiers namely k-Nearest Neighbor (k-NN) and Ensemble classifier. The disease classification accuracy rate is 80.2% using K-NN and 84.6% using Ensemble classifier. So, GLCM and LBP features and then SVM classifier obtained the high accuracy rate for classification of plant diseases.

VI. CONCLUSION

Human life is completely dependent upon nature and plants. So, there should be special methods to save plants from diseases. The decrease in crop production also affects the economy of the country. There is the need of appropriate research method that can automatically detect the plant leaf disease. The main purpose of this system is to improve the efficiency of automatic plant disease detection. Experimental results show that the proposed system can successfully detect and classify the plant disease with accuracy of 98.2%. In future work, we will extend our database for more plant disease identification and use large number of data as training purpose in classification. As we increase the training data, the accuracy of system will be high. And then we can compare the accuracy rate and speed of system.

DECLARATION

ACKNOWLEDGEMENT

REFERENCES

[1] Rafael C. Gonzalez. and Richard E. Woods. Digital Image Processing, Pearson Education, Third Edition.

[2] Akhtar, Asma, AasiaKhanum, Shoab Ahmed Khan, and ArslanShaukat. "Automated Plant Disease Analysis (APDA): Performance Comparison of Machine Learning Techniques." IEEE International Conference on Frontiers of Information Technology (FIT), pp. 60-65, 2013.

[3] Al-Hiary H., S. Bani-Ahmad., M. Reyalat., M. Braik. and Z. AlRahamneh. 2011. Fast and accurate detection and classification of plant diseases. International Journal of Computer Applications. Vol 17, No 1, pp 31-38.

[4] Anand Singh Jalal, Shiv Ram Dubey "Detection and Classification of Apple Fruit Diseases Using Complete Local Binary Patterns" IEEE Third International Conference on Computer and Communication Technology, pp. 978-0-7695-4872, 2012.

[5] Arivazhagan S., Newlin Shebia R. "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features". Agricultural Engineering Institute: CIGR journal, 2013. Volume 15, No.1.

[6] Athanikar, Girish, and Priti Badar. "Potato leaf diseases detection and classification system." International Journal of Computer Science and Mobile Computing 5.2, pp. 76-88, 2016.

[7] Bhong, Vijay S., and B. V. Pawar. "Study and Analysis of Cotton Leaf Disease Detection Using Image Processing." International Journal of Advanced Research in Science, Engineering and Technology, 2016.

[8] Ghulam Mustafa Choudhary and Vikrant Gulati "Advance in Image Processing for Detection of Plant Diseases" International Journal of Advanced Research in Computer Science and Software Engineering 5(7), [ISSN: 2277 128X], pp. 1090-1093, July- 2015.

[9] Hrishikesh Kanjalkar P. and Prof. Lokhande S., "Detection and Classification of Plant Leaf Diseases using ANN". International Journal of Scientific & Engineering Research. 2013, ISSN: 2229 5518.

[10] Ms. Kiran R. Gavhale, Prof. Ujwalla Gawande, Mr. Kamal O. Hajari, "Unhealthy Region of Citrus Leaf Detection Using Image Processing Techniques," IEEE, International Conference for Convergence of Technology, Pune, pp 1-6, 2014.

[11] Priya, P., and Dony A. Dsouza. "Study of Feature Extraction Techniques for the Detection of Diseases

of Agricultural Products." International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering 3, 2015.

[12] Robert M. Haralick., "Textural Features for Image Classification". IEEE transactions on Systems, 1973 Vol.SMC-3, No.6, pp 610-621.

[13] Sachin D. Khirade, A. B. Patil, "Plant Disease Detection Using Image Processing," IEEE, International Conference on Computing Communication Control and Automation, Pune, pp 768-771, 2015.

[14] Sandesh Raut and Amit Fulsunge, "Plant Disease Detection in Image Processing Using MATLAB" International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Volume 6, Issue6, June 2017.

[15] Sanjeev S. Sannaki, Vijay S. Rajpurohit, V B Nargund,Pallavi Kulkarni, "Diagnosis and Classification of Grape Leaf Diseases using Neural Network," IEEE , Tiruchengode, pp 1 – 5, 2013.

[16] Suhaili Beeran Kutty, Noor Ezan Abdullah, Dr. Hadzli Hashim, A'zraa Afhzan Ab Rahim, Aida Sulinda Kusim,Tuan Norjihan Tuan Yaakub, Puteri Nor Ashikin Megat Yunus, Mohd Fauzi Abd Rahman, "Classification of Watermelon Leaf Diseases Using Neural Network Analysis," IEEE, Business Engineering and Industrial Applications Colloquium (BEIAC), Langkawi, pp 459 – 464, 2013.

[17] Usama Mokhtar, Mona A. S. Alit, Aboul Ella Hassenian, Hesham Hefny "Tomato leaves diseases detection approach based on support vector machines" IEEE pp. 978-1-5090-0275-7/15, 2015.

[18] Weizheng, S., Yachun, W., Zhanliang, C., and Hongda, W. (2008). "Grading Method of Leaf Spot Disease Based on Image Processing". In Proceedings of the 2008 international Conference on Computer Science and Software Engineering (CSSE). Washington, DC, pp. 491-494, 2008.

[19] Yasikka, M., and E. Santhi, " Foliage Measurement Using Image Processing Techniques", International Journal For Trends in Engineering and Technology, Volume 5, Issue 2, May 2015.