A Color and Texture Based Approach for the Detection and Classification of Plant Leaf Disease Using KNN Classifier

[1]Eftekhar Hossain, [2]Md. Farhad Hossain and [3]Mohammad Anisur Rahaman
[1,2,3]*Dept. of Electronics & Telecommunication Engineering*
[1,2,3]*Chittagong University of Engineering & Technology*
Chittagong, Bangladesh
[1]eftekhar.13ete@gmail.com, [2]farhad.hridoy13@gmail.com and [3]anis00024@yahoo.com

***Abstract*—Modern organic farming is gaining popularity in the agriculture of many developing countries. There are many problems arise in farming due to various environmental factors and among these plant leaf disease is considered to be the most strong factor that causes the deficit of agricultural product quality. The goal is to mitigate this issue through computer vision and machine learning technique. This paper proposed a technique for plant leaf disease detection and classification using K-nearest neighbor (KNN) classifier. The texture features are extracted from the leaf disease images for the classification. In this work, KNN classifier will classify the diseases like alternaria alternata, anthracnose, bacterial blight, leaf spot, and canker of various plant species. The proposed approach can successfully detect and recognize the selected diseases with 96.76% accuracy.**

***Index Terms*—Plant disease, KNN, GLCM, Color segmentation, DSC, Confusion Matrix.**

I. Introduction

Agriculture is the most important driving forces for a country economy and it is the medium of livelihood of almost nearly two thirds population of a developing country. The economy depends on the quality of agricultural products and the quality depends on the weather and other environmental factors. As varieties of farming products are produced and exported to different countries, thus it is necessary to yield high quality products with a reasonable yield. Plant disease is one of the most alluded sign for the degradation of products quality [1]. In coming days disease mitigation would be considered to be an influential factor in the case of farming. As, the diseases of the plants are inevitable, detection of plant disease are indispensable in the field of Agriculture. The easiest way of finding a disease affected plant is done by checking its leaf condition. The primary diseases of plants are due to viral, fungus and bacterial diseases like alternaria, anthracnose, bacterial blight, canker, and leaf spot etc. Plant disease recognition is a very challenging topic for agriculture experts which requires the adoption of scientific methods and longtime of observance. It is not an easy task to detect a plant disease by observing only its leaves because many diseases have similar signs. Thus, there is a paucity of an automated system that can execute the operations of plant disease recognition and gives an effective solution. To achieve this target image processing techniques are widely used [2].

Color image segmentation plays an important role in color image analysis. Image segmentation is the process of extracting a region of interest from an image. There are exists a number of color image segmentation techniques and k-means clustering is a widely used color image segmentation algorithm [3]. K-means clustering is a process of arranging the similar pixels into k number of clusters. The selection of different cluster number gives different segmentation result of an image and it becomes difficult to find the appropriate cluster number.

Various segmentation and machine learning algorithms are recently proposed for the recognition and classification of plant diseases from leaf images. These techniques have created a path to remove the problems but the challenged being faced is the performance of the results obtained. This paper approaches a method for plant disease detection and classification using k-nearest neighbor classifier.

The rest of the paper is organized in the following manner: Section II gives a brief glimpse of the relevant work that was carried out all in the fields of plant leaf disease detection and classification. Section III gives a brief description of the methodology adopted for the detection and classification process. Section IV illustrates the results and discussion. Section V puts forward the conclusions followed by future enhancements and references.

II. Literature Review

There are subsists a number of research where digital image processing and machine learning algorithms were proposed for plant leaf disease detection and classification. In this section, few of them are summarized.

Chaitali G. Dhaware and Mrs. K.H. Wanjale applied a color and cluster-based segmentation technique for plant leaf disease detection. In this method the classification of sun burn, yellow mosaic, and grasshopper leaf disease was done by using support vector machine (SVM) classifier [4]. P. Krithika and S. Veni applied k-means clustering algorithm to segment the disease portion from the cucumber leaves. Multi SVM classifier was taken for classifying the leaf minor, leaf spot

978-1-5386-9111-3/19/$31.00 ©2019 IEEE

and mosaic disease on the cucumber leaves. GLCM features of the segmented diseases were used for the training and testing of the SVM [5]. R.Meena Prakash, G.P.Saraswathy, G.Ramalakshmi, K.H.Mangaleswari, and T.Kaviya segmented the region of disease from the RGB image by applying color conversion and k-means clustering. The experiment was applied upon the citrus leaves for distinguishing the normal and abnormal leaves. GLCM features and SVM classifier was introduced for the classification of citrus leaves [6]. Shriroop C. Madiwalar and Medha V. Wyawahare used color based segmentation on YcbCr color plane for mango leaf disease detection and the experiment also used SVM and minimum distance classifier (MDC) to classify the disease like anthracnose and leaf spot. In this experiment three different types of features like color, GLCM and gabor features were extracted for the performance comparison of SVM and MDC in mango leaves disease classification [7]. Pooja V, Rahul Das, and Kanchana V applied RGB to HSI conversion, k-means clustering and otsu thresholding to segment the disease region. The experiment was also used GLCM features and SVM classifier to classify the five different types of leaf disease [8]. Jobin Francis, Anto Sahaya Dhas D, and Anoop B K applied a composed segmentation technique for disease detection on pepper leaves. The technique includes green pixel masking and threshold based segmentation. This method was conducted to find the diseases like berry spot and quick wilt disease. GLCM features and neural network was used for the classification of pepper leaves disease [9]. Pranjali B. Padol and Prof. Anjali A. Yadav used k-means clustering algorithm for the segmentation. The gaussian filtering was applied over the images before segmentation. Linear SVM classifier and 54 features were taken for the leaf disease classification. The method was applied on grape leaf image for classifying the downy mildew and powdery mildew diseases [10]. S. S. Sannakki, V. S. Rajpurohit, V. B. Nargund,and P. Kulkarni detect plant leaf disease by using k-means clustering algorithm. Anisotropic diffusion filter was adopted for removing the noise from the images. The method also used neural network for the disease classification and the experiment were conducted on grape leaf images [11]. Chaudhary, Piyush and Chaudhari, Anand K and Cheeran, AN and Godara, Sharda used color transform to segment the disease region from the RGB image. Three color space comparison was done in their experiment and those are YcbCr, CIELAB and HSI. The methods were applied on dicot and monocot leaves [12]. J. Pang, Z.-Y. Bai, J.-C. Lai, and S.-K. Li applied a local threshold and seeded region growing based integrated method for the segmentation of leaf spot disease [13].

III. Proposed Methodology

The proposed method of this paper is illustrated in Fig. 1. It is divided into two phases: the training phase and the testing phase. The training and testing phases comprise of five fundamental stages which are image acquisition, color conversion, color segmentation, morphological operation, and feature extraction. The dataset consists of five different types of plant leaf disease image. In training phase, the extracted feature of the segmented plant leaf images are used for the training of the classifier. After the creation of the trained model the testing phase takes an input image and complete all the processing steps on the image up to feature extraction. The new features are given to the classifier model for performing the comparison to give the correct recognition of the disease. Different plant leaves images have bee tested, to be classified into five classes- alternaria alternata, anthracnose, bacterial blight, leaf spot and citrus canker affected.

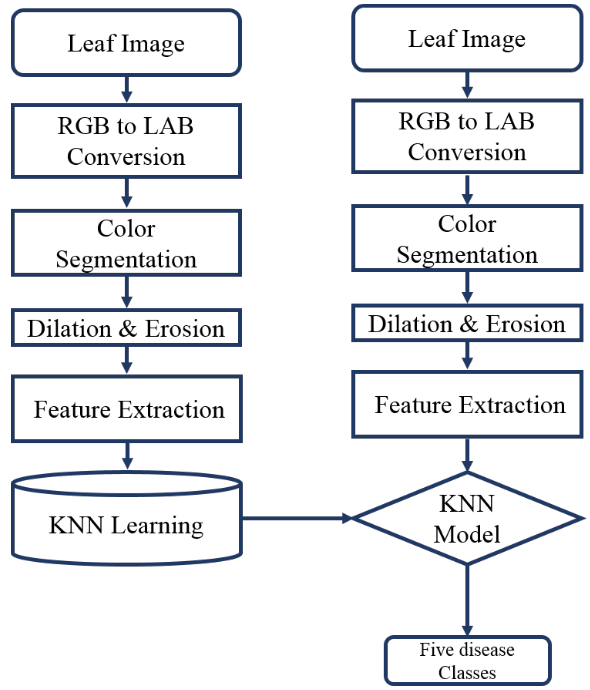


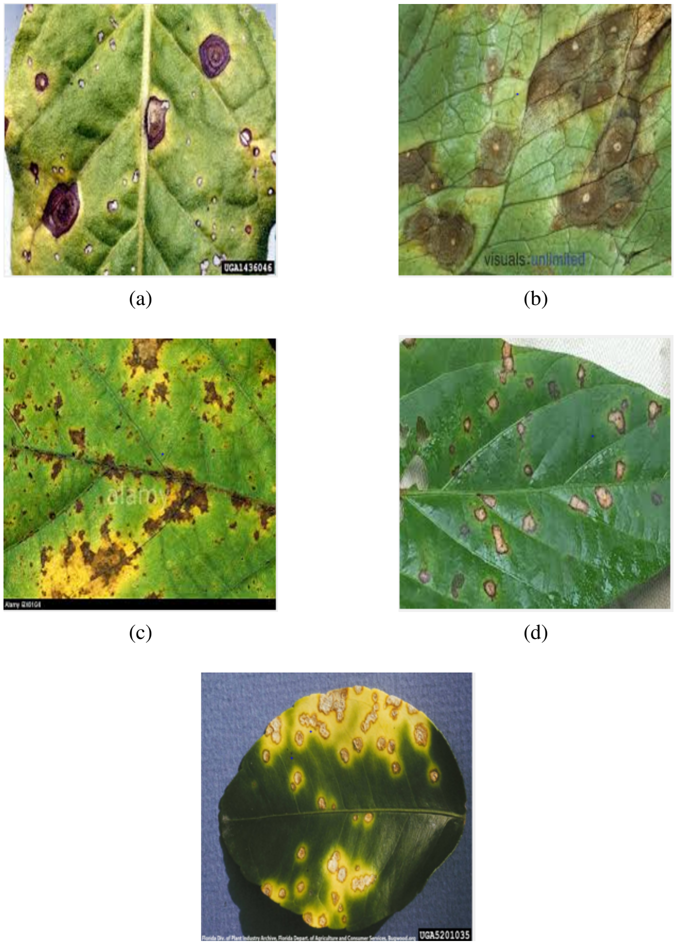
Fig. 1. Block diagram of the proposed method

A. Database Description

The database consists of 237 leaf images of plant disease. Five types of disease affected images are collected from two largest plant disease image database website. The Arkansas plant disease database and Reddit-plant leaf disease datasets are used for the acquisition of disease affected images [14] [15]. Alternaria alternata, anthracnose, bacterial blight, leaf spot, and citrus canker leaf images are taken for this experiment. Among these images, total 177 leaf images are alternaria alternata, anthracnose, and bacterial blight disease affected and each type has 59 leaf images. Other, 60 images are leaf spot and canker affected where each disease type image numbers are 30. The sample images of each type of plant leaf disease is shown in Fig. 2.

B. Color Space Conversion

Color space conversion of an image is required for providing a way to identify more intuitive color information. In this work the RGB images of leaves are transformed into Lab color space. The L*a*b* space consists of a luminosity 'L*' layer,



(a) (b) (c) (d) (e)

Fig. 2. Sample images from the database of plant disease: (a) Alternaria Alternata (b) Anthracnose (c) Bacterial Blight (d) Leaf spot (e) Canker

chromaticity layer 'a*' indicating where color falls along the red-green axis, and chromaticity layer 'b*' indicating where the color falls along the blue-yellow axis. This paper approach is to choose three small sample region for different color and to calculate each sample region's average color in 'a*b*' space. The main goal is to separate different colors in leaf image by analyzing the L*a*b* color space [16].

C. Color segmentation

In this paper, k-nearest neighbor classifier is used for the color image segmentation. The k-nearest neighbor classifier is the most widely used algorithms in machine learning. It is a learning method based on instances that does not required a learning phase. It tries to find the similarity of a pixel with its nearest neighbor pixels and besides that, a distance matric is used for the calculation of similarity. In this work, Euclidean distance matric is used. At first stage three sample regions are selected from leaf image and after conversion into LAB space, the similar pixels regions are segmented by three different colors [17]. In this work, k-nearest neighbor classifier with three neighbors are used for the segmentation. The segmented regions are represented by green, red, and blue color respectively for leaf, disease, and background portion.

D. Disease Detection using Morphology

The primary interest portion of the leaf image is the disease affected part and in the color segmented image it is represented by green color. So, for the disease detection the green channel is separated from the color segmented image and then apply simple thresholding without any fixed level to generate the binary image. Morphological opening which is followed by dilation and erosion are applied to the binary image for outlining the disease region upon the original leaf image. Dilation increases the width of maximum areas, so it can eliminate bad imprudent noises from the image. Erosion is used to minimize the object in the image and it reduces the width of smallest region. The binary image size is used as a structuring element for performing the morphological operation.

E. Feature extraction

Total six features including GLCM and color features are extracted from the segmented disease part. Let i and j are the coefficients of co-occurrence matrix, M(i,j), is the element in the co-occurrence matrix at the coordinates i and j and N is the dimension of the co-occurrence matrix.

1) Mean, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
2) Standard Deviation, $s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$
3) Energy = $\sum_{i=1}^{n}\sum_{j=1}^{n}(p(i,j))^2$
4) Contrast = $\sum_{i=1}^{n}\sum_{j=1}^{n}(i,j)^2 p(i,j)$
5) Homogeneity = $\sum_{i,j=1}^{n}\frac{p_{ij}}{1+(i-j)^2}$
6) Correlation = $\sum_{i,j=1}^{n} p_{ij}\frac{(i-\mu)(j-\mu)}{\sigma^2}$

F. Classification

Classification is the process of imposing a class on a new sample on the basis of learning attained by the classifier during training. Its task is to assign associate input pattern drawn by a vector to one of the diverse predefined patterns. In this paper, the classification plant disease into five classes are done by using the KNN classifier.

1) K-nearest neighbor Classifier: KNN is the simple and useful classifier for different classification problems. It does not require any prior knowledge of training like SVM or other machine learning algorithm. If the new training pattern is affixed to the subsisting training set then it doesn't require retraining. Before classifying a new element vector, a comparison should be finished with the training sample using distance metrics. Its k-nearest neighbors are then considered where the class that appears most among the neighbors is given to the element to be classified. A new element is classified on the basis of the neighbors are weighted by the distance measure. The appropriate working of the scheme depends on the proper selection of the appropriate parameter such as the k which represents the number of neighbors chosen to assign the class to the new element and the distance used.

2) KNN Classifier Phases & Rules:
KNN classifier comprises of two phases. One is training phase where the bone MR images labels with their class (benign or malignant) and another one is testing phase where the bone image are unlabeled and algorithm generates the list of k nearest data point (training data point) to label the unlabeled point and classifies their class [18].

KNN rules are:

- The set of stored training and testing data.
- Use euclidian distance which can given by (1) as distance parameter to measure the distance between stored records & unknown record to classify.

$$d(p_i, q_j) = \sqrt{\sum_{r=1}^{n} (p_{ir} - q_{jr})^2} \quad (1)$$

- Find k nearest neighbors & use class labels of nearest neighbors to determine the class label of unknown record by taking majority vote

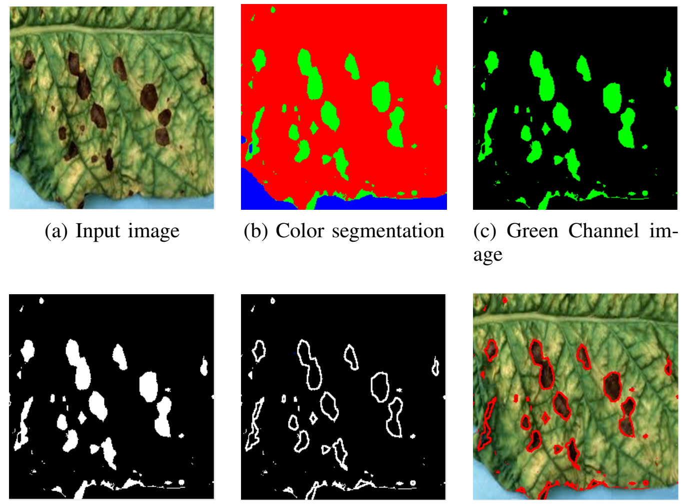
G. Performance Measures

As the performance of the classifier has to measure in terms of training and testing. So, four performance parameters accuracy, precision, recall, and f1-score are taken to evaluate the performance of the KNN classifier on the testing set. Receiver operating characteristics (ROC) curve is used to measure the trained classifier performance. The parameters are measured by determining the number of true false positives and negatives from the confusion matrix.

- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
- Precision =$\frac{TP}{TP+FP}$
- Recall = $\frac{TP}{TP+FN}$
- F1-Score = $2 * \frac{Precision*Recall}{Precision+Recall}$

IV. Results & Discussion

The result section is discussed in two parts- First one is the leaf disease segmentation output and its performance comparison in terms of various parameters and the second one is the disease classification performance analysis of the KNN classifier on leaf dataset. The leaf disease detection and classification result has been simulated on MATLAB. The output of the plant disease segmentation has been figured out in Fig. 3. In this paper, disease part detection or segmentation of leaf image has been performed by using KNN classifier. The segmentation output of the KNN classifier has been compared with the manually segmented result of the leaf image. Manual segmentation output of the leaf images has been obtained by using image segmenter application in MATLAB. The dice similarity coefficient (DSC), minimum square error (MSE), and structural similarity index measurement (SSIM) parameters have been taken out for performance calculation between the KNN segmentation output and manual segmentation output.



(a) Input image (b) Color segmentation (c) Green Channel image

(d) Binary image (e) Eroded image (f) Detected disease part

Fig. 3. Steps of plant leaf disease detection

TABLE I
Performance parameter values of the segmentation Result

Sl.no	DSC	MSE	SSIM
1	0.9894	0.0106	0.9626
2	0.9963	0.0037	0.9921
3	0.9863	0.0137	0.9756
4	0.9932	0.0068	0.9951
5	0.9692	0.0308	0.9469
6	0.9812	0.0188	0.9811
7	1.00	0.0003	0.9989
8	0.9924	0.0076	0.9849
9	0.9969	0.0031	0.9982
10	0.9751	0.0249	0.9799

The average performance of the segmentation output has been shown only for the ten leaf samples and the Table. I shows that the three parameters provide acceptable result in leaf disease detection. The mean performances of the DSC, MSE and SSIM are 98.8%, 1.203%, and 98.153% respectively.

The used dataset in the classification section consists of 237 leaf images with six features of each images. In training dataset 200 leaf images of five disease classes have been taken for training of the KNN classifier and 37 independent leaf images have been held out for the testing of the classifier performance. The testing dataset consisted nine alternata alternaria, nine anthracnose, nine bacterial blight five canker

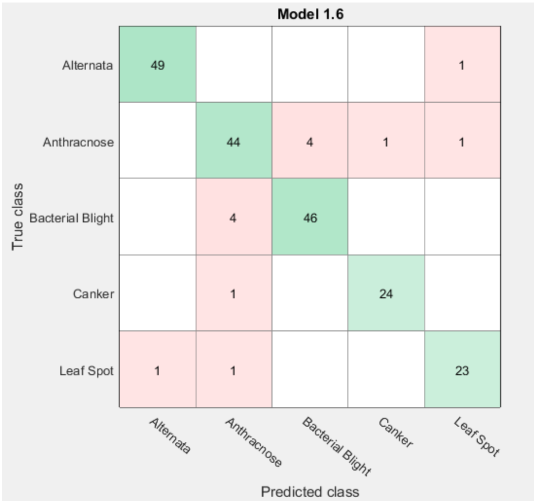


Fig. 4. Confusion matrix of the trained KNN classifier

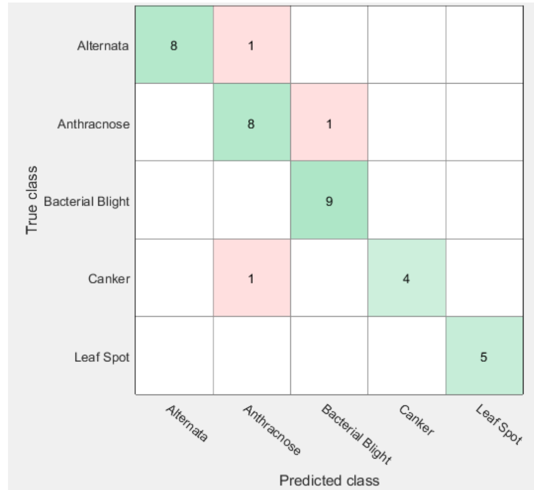
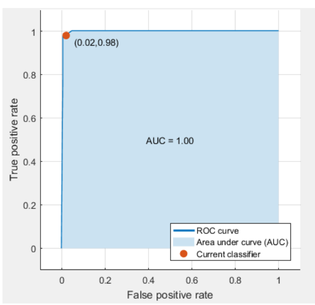
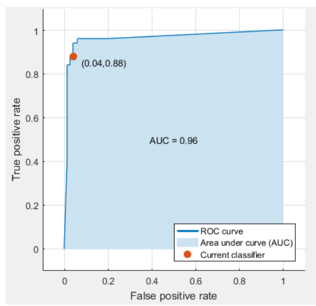


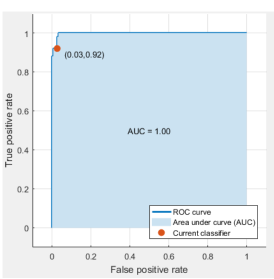
Fig. 6. Confusion matrix of the testing dataset



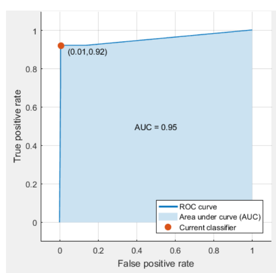
(a) Positive class: Alternaria Alternata



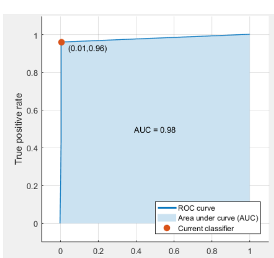
(b) Positive class: Anthracnose



(c) Positive class: Bacterial blight



(d) Positive class: Leaf spot



(e) Positive class: Canker

Fig. 5. ROC curves of trained classifier model

and five leaf spot disease affected leaf images. Before using the training dataset, the fivefold cross validation has been performed on the training dataset to ensure that the created classifier model not been over fitted. The Euclidean distance and three nearest neighbor has been selected as the classifier parameters. The performance of the trained classifier has been evaluated by determining the confusion matrix and ROC curve. The confusion matrix of the trained classifier model has been shown in Fig. 4 where the shaded green portion represents the correctly classified disease class and shaded pink portion represents the falsely classified disease class. The numeric number of correct and false classification has also been indicated in the confusion matrix. By considering each time one disease type as a positive class and others as a negative class, the ROC curves of the classifier model for the five disease classes have been plotted in Fig. 5. The value of area under the curve (AUC) for each of the classes has also been indicated in the plotted ROC curve.

The performance of the created classifier model on the testing dataset has been evaluated from the confusion matrix and the accuracy, precision, recall, and f1-score parameters have been calculated for the performance evaluation. The testing dataset consists of 37 leaf images of five disease class. The confusion matrix in Fig. 6 shows the number of

TABLE II
PERFORMANCE PARAMETER VALUES OF THE CLASSIFICATION RESULT

Sl.no	Accuracy	Precision	Recall	F-Score
Alternaria Alternata	0.9730	1.00	0.889	0.9412
Anthracnose	0.9189	0.800	0.889	0.8421
Bacterial Blight	0.9730	0.900	1.00	0.9474
Leaf Spot	0.9730	1.00	0.8000	0.889
Canker	1.00	1.00	1.00	1.00
Average	96.76%	94.00%	91.56%	92.39%

TABLE III
COMPARISON TABLE OF THE PROPOSED & EXISTING PLANT DISEASE DETECTION ALGORITHMS

Segmentation, Feature extraction & Classification Algorithms	**Research Gap**	**Performance**
Color & Cluster based + SVM	1. Dataset description is missing 2. Feature extraction technique is missing 3. No performance Comparison	Not Mentioned
K-means + GLCM + SVM	1. Undefined the cluster number 2. Undefined training & testing data 3. No performance Comparison	Not Mentioned
Color based + GLCM & Gabor + SVM & MDC	1. Applicable to the images where disease pixels are brown-black. 2. Two classifiers can not classify all the diseases correctly.	SVM Provides 100% accuracy
K-means & otsu Thresholding + Statistical feature + SVM	1. Different cluster numbers for ROI selection 2. No comparison	92.1% accuracy
K-means + GLCM + SVM	1. Cannot classify different class of diseases. 2. Dataset size is very small	90% accuracy
K-means + Color & texture + Linear SVM	Too much features for one image.	88.89% accuracy
KNN Classifier + GLCM + KNN Classifier (Proposed Method)		Provides 96.76% accuracy

correct and false classified disease. The diagonal shaded green portion represents the correctly classified disease among the total number of images of each class. The parameters have been calculated by considering the one individual disease class as true class and all other classes are false class. After calculating the individual class classification performances, the average of the five class performance gives the total performance of the classifier on the testing dataset. Table. II shows the performance parameters values. The comparison of the proposed method with other existing methods has been given in the Table. III and it illustrates that the proposed KNN classifier provides good performance on the leaf disease classification.

V. CONCLUSION

This work proposed a method which uses KNN approach to detect and classify various diseases that are present in plant leaves. Diseases such as alternaria alternata, anthracnose, bacterial blight, leaf spot, and canker of plant leaves are considered for the experiment. The segmentation of the disease portion is done by using the k-nearest neighbor classifier and GLCM texture features are used for the classification. The KNN classifier based segmentation result provides optimum accuracy in plant disease detection and the quantitive performance of the proposed algorithm is obtained by measuring the DSC, MSE and SSIM parameters. On the other hand, the classification performance of the KNN on plant leaf disease are also provides 96.76% accuracy and the proposed KNN based approach gives better results in comparing with some of the existing methods. Future work is to be carried out for classification of many more diseases in different plant and crops and to improve the classification accuracy neural network is deployed.

REFERENCES

[1] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[2] J. G. A. Barbedo, "A review on the main challenges in automatic plant disease identification based on visible range images," *Biosystems Engineering*, vol. 144, pp. 52–60, 2016.

[3] M. K. Yadav, D. D. Manohar, G. Mukherjee, and C. Chakraborty, "Segmentation of chronic wound areas by clustering techniques using selected color space," *Journal of Medical Imaging and Health Informatics*, vol. 3, no. 1, pp. 22–29, 2013.

[4] C. G. Dhaware and K. Wanjale, "A modern approach for plant leaf disease classification which depends on leaf image processing," in *Computer Communication and Informatics (ICCCI), 2017 International Conference on*. IEEE, 2017, pp. 1–4.

[5] P. Krithika and S. Veni, "Leaf disease detection on cucumber leaves using multiclass support vector machine," in *Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 International Conference on*. IEEE, 2017, pp. 1276–1281.

[6] R. M. Prakash, G. Saraswathy, G. Ramalakshmi, K. Mangaleswari, and T. Kaviya, "Detection of leaf diseases and classification using digital image processing," in *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017 International Conference on*. IEEE, 2017, pp. 1–4.

[7] S. C. Madiwalar and M. V. Wyawahare, "Plant disease identification: a comparative study," in *Data Management, Analytics and Innovation (ICDMAI), 2017 International Conference on*. IEEE, 2017, pp. 13–18.

[8] V. Pooja, R. Das, and V. Kanchana, "Identification of plant leaf diseases using image processing techniques," in *Technological Innovations in ICT for Agriculture and Rural Development (TIAR), 2017 IEEE*. IEEE, 2017, pp. 130–133.

[9] J. Francis, B. Anoop *et al.*, "Identification of leaf diseases in pepper plants using soft computing techniques," in *Emerging Devices and Smart Systems (ICEDSS), Conference on*. IEEE, 2016, pp. 168–173.

[10] P. B. Padol and A. A. Yadav, "Svm classifier based grape leaf disease detection," in *Advances in Signal Processing (CASP), Conference on*. IEEE, 2016, pp. 175–179.

[11] S. S. Sannakki, V. S. Rajpurohit, V. Nargund, and P. Kulkarni, "Diagnosis and classification of grape leaf diseases using neural networks," in *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. IEEE, 2013, pp. 1–5.

[12] P. Chaudhary, A. K. Chaudhari, A. Cheeran, and S. Godara, "Color transform based approach for disease spot detection on plant leaf," *International Journal of Computer Science and Telecommunications*, vol. 3, no. 6, pp. 65–70, 2012.

[13] J. Pang, Z.-y. Bai, J.-c. Lai, and S.-k. Li, "Automatic segmentation of crop leaf spot disease images by integrating local threshold and seeded region growing," in *Image Analysis and Signal Processing (IASP), 2011 International Conference on*. IEEE, 2011, pp. 590–594.

[14] U. of Arkansas Cooperative Extension Service, "Arkansas plant disease database," https://www.uaex.edu/yard-garden/resource-library/diseases/, Accessed October 3, 2018).

[15] R. Database, "Plant leaf disease datasets," https://www.reddit.com/r/datasets/comments/5uljlp/plant_leaf_disease_datasets/, Accessed October 9, 2018).

[16] M. Mignotte, "Segmentation by fusion of histogram-based k-means clusters in different color spaces," *IEEE Transactions on image processing*, vol. 17, no. 5, pp. 780–787, 2008.
[17] J. Li, X. Li, B. Yang, and X. Sun, "Segmentation-based image copy-move forgery detection scheme," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 3, pp. 507–518, 2015.
[18] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient knn classification with different numbers of nearest neighbors," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 5, pp. 1774–1785, 2018.