Detection of Crop Diseases using different Machine Learning Approaches

Mst. Rehena Khatun, Md. Palash Tai, Mrittika Mahbub, Md. Easir Arafat

Lecturer, Assistant Professor, Lecturer, Lecturer
Department of Computer Science & Engineering
Pundra University of Science &Technology, Gokul, Bogura-5800, Bangladesh

Abstract Crop diseases should be identified to meet the food demand of the world's rapidly growing population. Early detection of the crop's diseases will increase its production. The goal of this research is to reduce crop diseases. we collected various samples such as rice, wheat, maize, potato, and tomato plants. Firstly, leaf images were resized into 128 X 128-pixel values. After that, the affected region was taken out using color-based segmentation, and the AKAZE model was used for feature extraction. Then three different classifiers—KNN, SVM, and random forest—were applied to classify and detect diseases. The accuracy was measured using precession, recall, and F1 score. It was observed that the accuracy of KNN is 82%, SVM is 94 % and random forest classifier is 95%.

Index Terms - Crop diseases, Segmentation, AKAZE, KNN, SVM and random forest.

I. INTRODUCTION

Bangladesh is an agricultural country and its economy depends on it. However, the fact is that crop yields are not increasing at the same rate as the population is growing. Plant diseases, climate change, and other problems make it difficult for farmers. As the demand for food increases due to population growth, advancement in agriculture is very urgent. In this study, we attempted to reduce crop diseases by applying various machine-learning techniques to increase productivity.

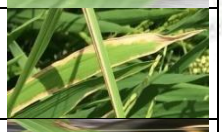
In this research, the dataset of rice plants was collected from Mendeley data [13] and the rest of the dataset such as wheat, maize, potato, and tomato were collected from Kaggle [14]. Crops are typically infected by bacteria, viruses, and fungi in which the fungus cause damage to crops at every stage of growth, symptoms appear on the rachis, glumes, leaves, and nodes; viruses discolor the leaves of plants from yellow to orange; bacteria cause wilting, yellowing and some stunting of the plants [15].

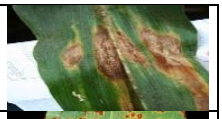
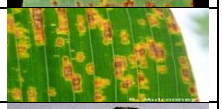
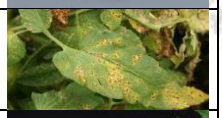
In this work, both healthy and affected leaf was used that is four types of rice plant diseases such as brown spot, bacterial blight, blast, and tungro; two types of wheat plant diseases such as septoria and stripe rust; three types of maize leaf diseases such as blight, common rust, and gray leaf spot; two types of potato leaf diseases such as potato early blight, and potato late blight; four types of tomato leaf diseases such as tomato early blight, tomato late blight, septoria leaf spot, and bacterial spot which is depicted in table 1:

With the help of the machine learning approach this paper proposes a method for identifying crop diseases. This work accepts the images of the crop diseases as input. The input data is preprocessed and segmentation is applied to extract the affected region. Then AKAZE model is used for feature extraction. Finally, different classifier such as KNN, SVM, and random forest is used to classify and detect crop diseases.

The existing crop disease techniques are described in section 2; section 3 discusses the proposed methodology; section 4 presents the results of the proposed model; section 5 concludes the crop disease detection model.

Table 1: Different crop diseases

Name of the crops	Crops plant diseases	Number of train data	Number of test data	Images
Rice	Brown spot	500	150	
	Bacterial blight	500	150	
	Blast	500	150	
	Tungro	500	150	
Wheat	Septoria	500	150	
	Stripe rust	500	150	

Maize	Blight	500	150	
	Common rust	500	150	
	Gray leaf spot	500	150	
Potato	Potato early blight	500	150	
	Potato late blight	500	150	
Tomato	Tomato early blight	500	150	
	Tomato late blight	500	150	
	Septoria leaf spot	500	150	
	Bacterial spot	500	150	

II. LITERATURE REVIEW

Several researches have been carried out to identify crop diseases using image processing and machine learning techniques. Many researchers have proposed a vision-based approach for plant analysis, including disease detection and classification [1-5]. However, most of the research has been emphasized on rice leaf diseases, with the aim of identifying and classifying plant illnesses [06].

P. Krithika et al., [7] proposed a method to detect leaf diseases using multiclass SVM. They applied K-Means clustering for segmentation and feature extraction was performed using GLCM. R. Meena et al., [8] used an image processing technique to automatic recognition plant leaf diseases using K-Mean clustering and SVM. Davoud Ashourloo, et. al. [9] proposed machine learning techniques for Wheat Leaf Rust disease detection They compared the performance of PLSR, v-SVR, and GPR with the PRI and NBNDVI.

P. Revathi, et al. [10] suggested two stages to determine the disease using HPCCDD Algorithm. Aditi Singh and Harjeet Kaur, [12] proposed a methodology to detect potato plant diseases using K-means for feature extraction purposes, and for classification purposes, the multi-class support vector machine was utilized. We [12] developed a method for rice plant disease detection using SVM. For that purpose, we used different feature extraction methods such as SIFT, FAST, ORB, and AKAZE.

In this paper, we have designed a method to detect different crop diseases such as rice, wheat, maize, potato, and tomato using the segmentation technique and AKAZE model for feature extraction. Then different machine learning technique such as KNN, SVM, and random forest is applied for the classification and detection of crop diseases.

III. METHODOLOGY

Our proposed model consists of several parts includes image acquisition, image preprocessing, segmentation, feature extraction, classification, and detection of diseases which is depicted in Figure 1:

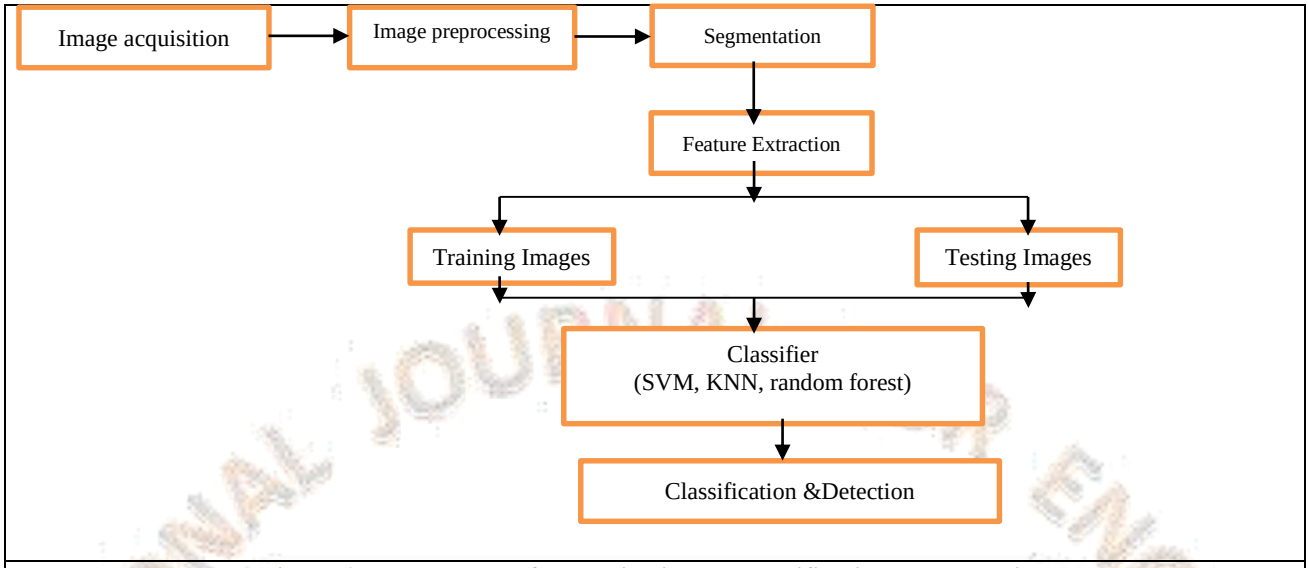


Figure 1: Methodology for crop's diseases classification and detection

IMAGE ACQUISITION

The dataset of different crop diseases was collected from Mendeley data [13] and Kaggle [14]. A total of 9750 leaf images of five different crop diseases in which there are 2600 rice plant images, 1300 wheat plant images, 1950 maize plant images, 1300 potato plant images, and 2600 tomato plant images were used.

IMAGE PREPROCESSING

Image preprocessing is essential as real-world data is noisy and uneven. Every image was resized into 128 x 128 pixels. A mean filter was used to minimize the noise. After that, image enhancement was done using histogram equalization.

IMAGE SEGMENTATION

Segmentation divides data according to similar characteristics. In this work, we used color-based segmentation to extract the affected region as most of the crop diseases have a yellow-brown appearance. To divide an image into color regions, color segmentation compares each pixel's color attribute to surrounding pixels. At first, images were converted to RGB to HSV color space. Afterward, the images of crop plants were segmented using the LAB channel.

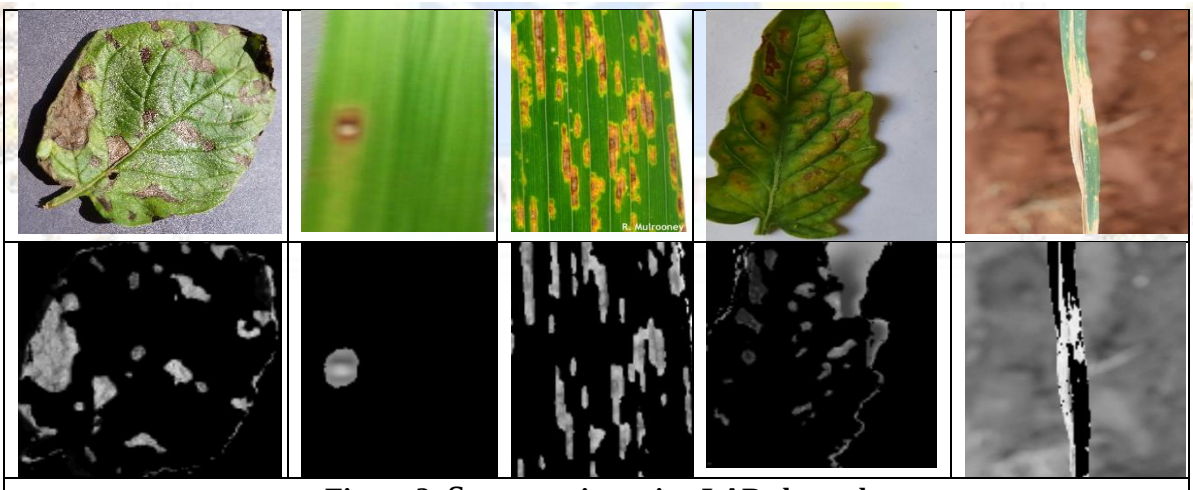


Figure 2: Segmentation using LAB channel

FEATURES EXTRACTION

The AKAZE model was applied to different crop disease images to extract features. It identifies and computes key points and descriptors of the images. At first, it creates a nonlinear scaled space pyramid of the original image, then determines key points using Hessian determinants and multidimensional derivatives in the nonlinear scale space, and computes the orientation and descriptor vectors for all key points.

CLASSIFICATION

The techniques SVM, KNN, and random forest were used to classify crop diseases. SVM is a supervised machine learning technique that is frequently used for classification. It is a useful tool for binary and multiclass classification. KNN is also a supervised machine-learning technique that is used for both regression and classification. KNN classifies new data points according to how similar they are to previously available data points. The most important step in the KNN process is determining the value of K. Random forest uses a decision tree for its implementation and combines multiple classifier outputs to improve performance. The multiclass classifier was used in this study since more than two classes need to be predicted. The dataset was divided into training and testing in which 7500 was used as training data and 1750 was used as testing.

III. RESULTS AND DISCUSSION

Precision calculates the percentage of all detected anomalies that are true anomalies. In the other words, the percentage of all true anomalies that are successfully identified is called recall. It can be defined as follows,

$$Precision=TP/(TP+FP) \text{ ----------(1)}$$
$$Recall=TP/(TP+FN) \text{ ----------(2)}$$

The F1 score is a simultaneous measure of recall and precision. It is the harmonic mean of the recall and precision scores.

$$F1\ Score= (2*Precision\ Score*Recall\ Score)/ (Precision\ Score+ Recall\ Score) \text{ ----------(3)}$$

In the following figures, various crops plant diseases are represented in X-axis and accuracy is represented in Y-axis. The SVM model performs better accuracy for brown spot, blast, common rust, gray leaf spot, tomato early blight, tomato late blight, septorial leaf spot, and bacterial spot disease than others. But in random forest model, bacterial blight, tungro, septoria, stripe rust, blight, potato early blight, and potato late blight disease images give outstanding performance. In other words, the random forest model performs well on the majority of the wheat and potato diseases.

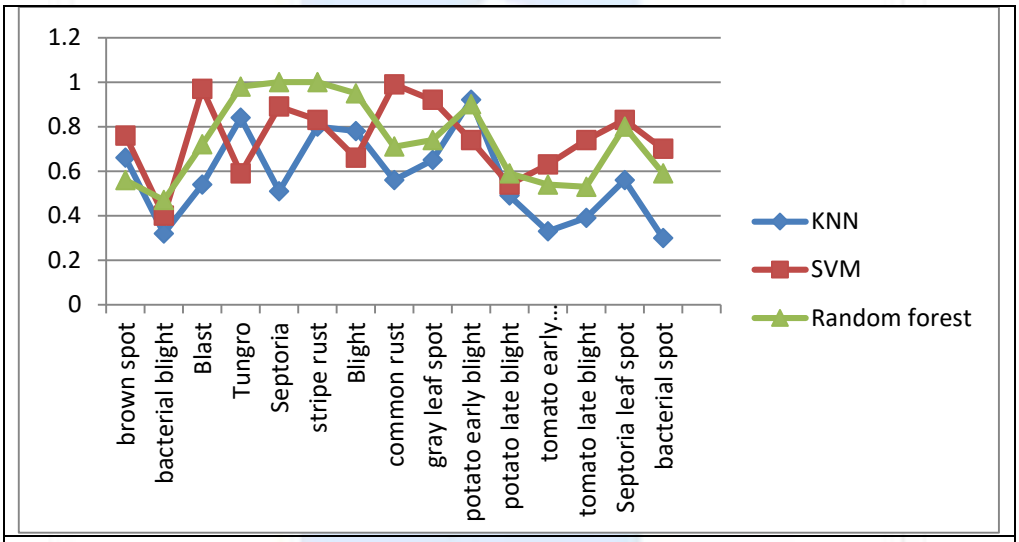


Figure 3: Precision comparison among KNN, SVM and random forest model

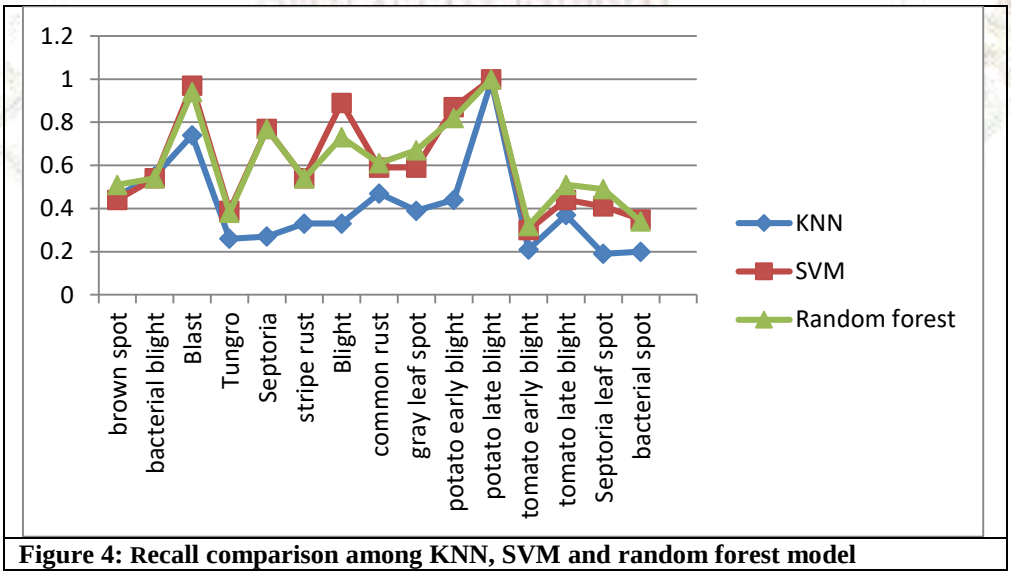


Figure 4: Recall comparison among KNN, SVM and random forest model

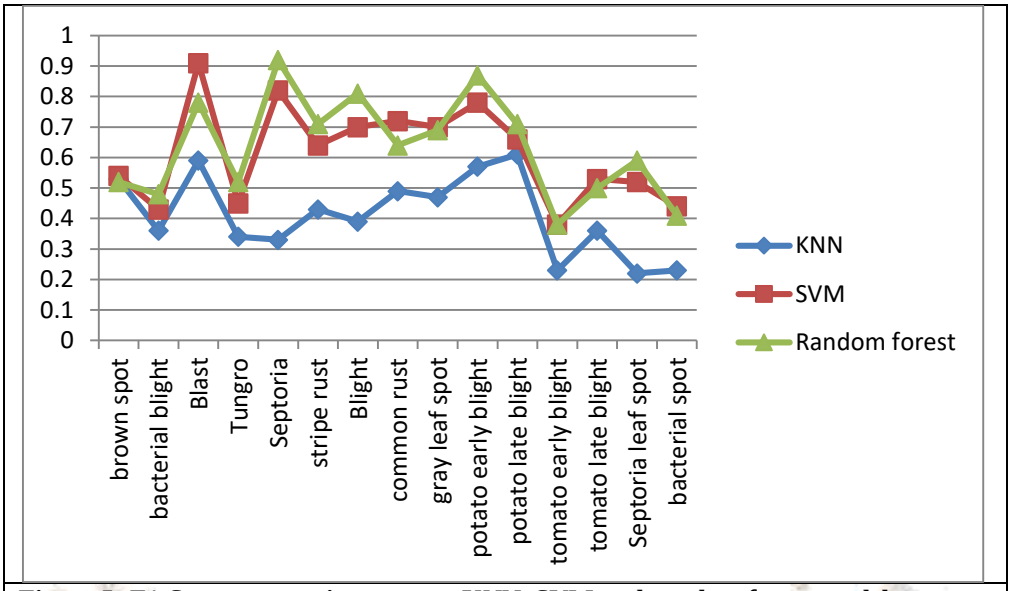


Figure 5: F1-Score comparison among KNN, SVM and random forest model

Figure -6 represents comparison among different model. Different models of KNN, SVM and Random Forest are shown in X-axis, and precision, recall and F1-score are represented in Y axis. The below comparative performance analysis clearly shows that Random Forest Classifier gives highest performance than others.

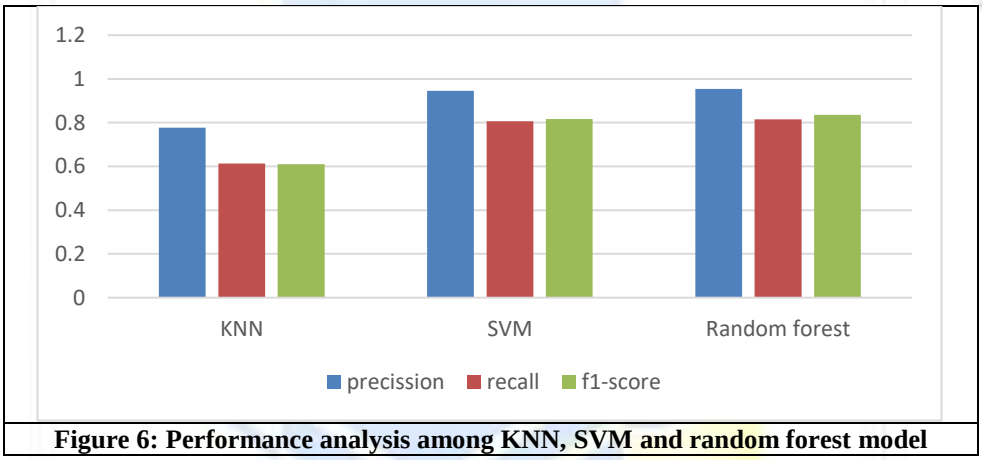


Figure 6: Performance analysis among KNN, SVM and random forest model

V. CONCLUSIONS

The dataset of crop diseases was collected from online sources. Ambiguity and anomalies were removed using pre-processing. RGB images were converted to HSV space and Segmentation was applied to identify the affected area. AKAZE model was used for feature extraction. Then, three different classifiers SVM, KNN, and random forest classified and detected crop diseases. It is observed that random forest gives the highest performance than other classifiers. In future, local data will be collected and deep learning will be applied to detect different crops diseases.

VI. REFERENCES

[1] D. Fleer and R. Möller, "Comparing holistic and feature-based visual methods for estimating the relative pose of mobile robots", *Robotics and Autonomous Systems*, vol. 89, pp. 51-74, 2017.
[2] A. S. Huang et al., "Visual odometry and mapping for autonomous flight using an RGB-D camera" in Robotics Research, Springer International Publishing, pp. 235-252, 2017.
[3] D. Nister et al., "Visual odometry" in Computer Vision and Pattern Recognition, Washington D. C.:CVPR, pp. 1-8, 2004.
[4] United States Patent Pirchheimet al., *Monocular visual SLAM with general and panorama camera movements*, June 2017.
[5] A. J. Davison et al., "MonoSLAM: Real-time single camera SLAM", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052-1067, 2007.
[6] H. Bay et al., "Speeded-up robust features (SURF)", *Computer vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, 2008.
[7]P Krithika, S Veni, Leaf disease detection on cucumber leaves using multiclass sup- port vector machine, in: IEEE International Conference on Wireless Communica- tions, Signal Processing and Networking, 2017, pp. 1276–1281.
[8] Chaitali G Dhaware, K H Wanjale, "A modern approach for plant leaf disease classi-
Fication which depends on leaf image processing", in: IEEE International Conference
on Computer Communication and Informatics, 2017, pp. 12–16.

[9] Davoud Ashourloo, Hossein Aghighi, Ali Akbar Matkan, Mohammad Reza Mobasheri, Amir Moeini Rad, "An Investigation in to Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement", IEEE Journal of SelectedTopics in Applied Earth Observations and Remote Sensing, pp. 1-7, May 2016.
[10] P. Revathi, M. Hemalatha, "Classification of Cotton Leaf Spot Disease Using Image Processing Edge Detection Technique", International Conference on Emerging Trends in Science Engineering and Technology IEEE, pp. 169-173, 2012.
[11] Aditi Singh and Harjeet Kaur, "Potato Plant Leaves Disease Detection and Classification using Machine Learning Methodologies", IOP Conf. Series: Materials Science and Engineering, pp.,2021
[12] M. R. Khatun, et al., "Conventional Machine Learning approaches for rice plant diseases classification and detection using Support Vector Machine" TIJER - INTERNATIONAL RESEARCH JOURNAL, Vol 10 Issue 12, December-2023
[13]https://data.mendeley.com/datasets/fwcj7stb8r/1
[14] https://www.kaggle.com/datasets/
[15]https://www.slideshare.net/aminulparvez/study-on-major-diseases-of-major-crops-in-bangladesh