Paddy Leaf Disease Detection Using Machine Learning Technique

Mustafizur Rahman Konik[1], Azaz Ahmad Swapnil[2]
Dept. of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh

***Abstract*—One of the most important key resources to prevent global warming on the planet is plants. But the plants are suffering from various diseases. In recent time, research has been begun for acknowledgement of plant disease. Paddy disease detection is the key intention of this paper. Brown Spot Disease (BSD), Leaf Blast Disease (LBD), and Leaf Blight Disease (LBD) are a few of the paddy diseases that prevent the paddy from growing and protecting every portion of the plant including diseases that can affect paddy at various stage of growth. This research examined 3 different disease kinds as well as one group of healthy paddy leaves. Bacteria, fungi, and other organisms are among those that can cause paddy disease. The Technique was created to eliminate noise automatically by decreasing the time needed to measure the impact of paddy leaf disease on humans so using machine learning techniques k-means for image segmentation and an automated detection method to get the best results for finding paddy leaf disease with the approach of machine learning using classifications with the best accuracy. To measure classification of this paper K-Fold cross validation techniques has been used. Applying 4 classes of paddy leaf's into Random Forest, Decision Tree, Logistic Regression and SVM like Support Vector Classifier (SVC), among them Random Forest gave the highest 94.16% accuracy with the using of K-fold cross validation techniques in predicting the three classes of paddy leaf disease with one group of healthy paddy leaves.**

***Index Terms*—Paddy leaf disease; image acquisitions; image preprocessing; segmentation; classification;cross validation.**

I. Introduction

Plant disease is mostly harmful things in every sector. Plant can help of all climate disaster but many diseases can be affected in plants. Many of different disease can down the growth of every plants. Over the past few years, more plant diseases have spread and because of increasing the plant disease it can affect also in agriculture sector. To choose ideal Fruit and Vegetable crops, farmers have access to a broad variety of options. Various types of disease problems have occurred in both disarray and illness. Farmers may choose from a wide selection of crops and choose the best pesticides for their plants. Plant disarray appears when a plant's regular functions are disrupted or diverted due to issues with the soil, environmental pressures, or other physical factors, it is said to have a plant disorder. The illness cannot spread from an infected plant to an unaffected plant in the case of plant diseases. Viruses, bacteria, and fungi are just a few examples of the many plant diseases that have been appearing that can harm various plant sections both above and below the earth. One fascinating area of research in the agricultural industry is the detection of different illnesses using images. In this research paddy leaf disease detection have been appeared in machine learning. In order to identify leaf diseases, image processing techniques might be applied.

There are many machine learning techniques for detecting the paddy disease. For the majority of the phases in the workflow sequence to evaluate images that may be used to accurately detect paddy diseases, image processing then machine learning give a number of methodologies and the proposed model are acquisition of image, preprocessing of image, segmentation of image, feature extraction, classification. First of all, acquisition and reading of the infected images and the images format like jpg, png, jpeg can be used to learn for the machine. The system's initial stage is paddy disease leaves picture acquisition. Then the images format has been changed in the jpg extension.

In this paper, the diseases paddy blast, brown spot, and blight are investigated. Once the picture has been identified, dissimilar types of techniques take remained used for processing to solve the specific difficulty of the paddy leaf disease image for doing a variety of tasks. The following objectives can be achieved by using image analysis:

- To identify unhealthy leaves.
- Measuring the size of the disease's impact.

In the Image pre-processing system, after the disease images are read and acquired then, Using the color conversion, Red Green Blue (RGB) photos are transformed into grayscale images. Various contrast-enhancing techniques, such as contrast adjustment, are employed to boost the contrast. And same types of disease images have been removed by the image preprocessing task. Once complete of the preprocessing, in segmentation, lowering the image's complexity and allowing each image fragment to be processed or analyzed further. And K-means clustering techniques has been proposed for segmentation. Lastly, Image classification has some model's tech-

niques which is used for identify disease images. In machine learning, image classification has various types of models like K-means, Support Vector Classifier(SVC), Random forest, Decision tree and Logistic regression.Those classification have been usable for the measure of paddy leaf diseases.

II. SOME IMPORTANT PADDY LEAF DISEASE

A. BROWN SPOT

A fungus called brown spot attacks the coleoptile, leaves, leaf sheath, panicle branches, glumes, and spikelets. The most obvious harm is the multiple large spots that can kill a whole leaf that appear on the leaves. Unfilled grains, speckled, or discolored seeds form when an infection in the seed takes place.

OCCURANCE

This disease develops in regions with high humidity (86–100)% and nutrient-poor soil. The seed's fungus has a four-year lifespan. During the crop's ripening stages, infection is crucial.

SYMPTOMS

At the beginning stage, seedlings exhibit dark brown lesions that are yellowish-brown in appearance. eventually turned purple-brown. The edge of fully formed lesions is reddish brown.

Fig. 1. Brown Spot.

B. LEAF BLAST

Blast, a disease that affects rice, is brought on by the fungus Magnaporthe oryzae. Blast affects the components of the panicle's leaf collar, leaf node, collar node, and neck.

OCCURANCE

Blast happens where there is little soil moisture, a lot of rain, extended rain showers, and cool daytime temperatures. Large day-night temperature variations in upland rice, along with generally lower temperatures, which lead to dew deposition on leaves, support the growth of the illness.

SYMPTOMS

At first, they appear as a dark green patch with a gray green border. With a gray center and a brownish rim, the lession afterwards turned elliptical. Then it grows larger and destroys the entire leaf.

Fig. 2. Leaf Blast.

C. LEAF BLIGHT

One of the most damaging diseases to affect cultivated rice is rice bacterial blight, commonly known as bacterial blight of rice. A severe outbreak can result in crop losses of up to 75%, and millions of hectares of rice are afflicted every year.

OCCURANCE

The Gram-negative bacterium Xanthomonas oryzae pv is responsible for the illness known as rice bacterial leaf blight (BB).

SYMPTOMS

Withering of seedlings or yellowing and wilting of leaves (also called kresek). Infected seedling leaves curl up and turn a grayish green color. The illness spreads, causing the leaves to wilt and turn yellow to straw-colored, which causes the entire seedling to dry out and perish.

Fig. 3. Leaf Blight.

III. RELATED WORKS

Many jobs have been done already in this sector. Commonly used aimed at analysis of rice plant diseases in agriculture sector. In general, a lot of work has been done in the agriculture sector for the finding of rice plant diseases. Numerous investigations happening different rice illnesses have been conducted in the past. Research is now being done on a number of rice illnesses and their treatments. Using K-means clustering machine learning techniques by Archana KS, Sahayadhas A [1] mainly using two types of class and disease like brown spot and bacterial leaf blight. These 2 classes mainly focused on segmentation of k-means clustering of those main disease leaf of the paddy in order to finish their work. Different steps of segmentation using k-means clustering for locating bacterial leaf blight and brown spot in rice plant in their work.

In 2001 Md. Ashiqul Islam et. al [2] from Daffodil international university published a report on Deep Learning using CNNs aimed at paddy leaf disease detection and Classification with the best accuracy. In addition to one healthy leaf class of the paddy, this article has examined four different forms of illnesses, totally 5 classes have been used. Firstly, image acquisition, image data preprocessing, deep learning CNN

models with classification and the prediction with the best accuracy of outcome.

Using machine learning techniques with SVM algorithm by Minu Eliz Pothen and Dr.Maya L Pai [3] basically focused on 3 paddy disease classes like Bacterial leaf blight, Leaf smut and Brown spot diseases which have been segmented by Otsu's method. In 2020 this research paper published only for SVM algorithm techniques with the different characteristics are separated using the divided region as a base which is SVM (HOG and LBP) various illnesses in the leaves of other plants, including rice, may be found using the polynomial kernel function. This study gets the SVM algorithm best accuracy of the following machine learning segmentation different characteristics.

Various paddy illnesses are currently causing farmers to suffer significant losses.R.P.Narmadha and G.Arulvadivu [4] mainly concentrated on how to effectively measure and manage several paddy illnesses. There are 3 classes of paddy ill leafs for measuring paddy leaf detection and identification using machine learning techniques. Basically, machine learning techniques like SVM, ANN, and fuzzy classification are a few examples of the many classification characteristics that are used. Data that was personally taken and taken from online sources.

Paddy illness detection using AI technology has received a lot of attention so far or is now receiving it.S.Pavithra et. al [5] focused on Paddy illness detection using AI technology has received a lot of attention so far or is now receiving it. With current image processing and machine learning techniques, how to identify and categorize paddy illness. Identification, Acquisition, Image preprocessing, feature extraction, and classification are the four phases they use to do their task. SVM classifiers and KNN have been used for paddy ill detection and measurement. The main topics of this analysis are paddy blast and brown spot disease. They achieved the best accuracy from SVM 95% for classification.

IV. Methodology

In this research there are some potential things or working flow for determining the analytics of data. Data gathering, data processing, data resizing and augmentation, model selection, etc. are some of the workflow phases for this disease detection,prediction and four classes of leaf's accuracy.

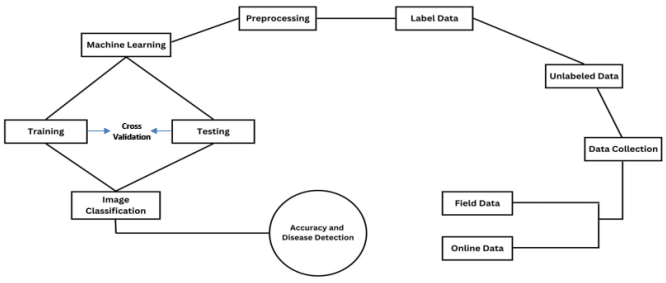


Fig. 4. Proposed model for whole research work

For implementation, we separated our effort into five key sections. These are the measures that must be taken for our project to succeed.

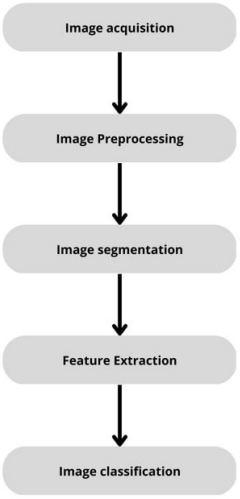


Fig. 5. Proposed model

A. Image Acquisitions

All the images of dataset which have been collected from the field raw data and some online resources data with different heights and widths capturing using Phone and Digital camera and an agricultural officer labeled the whole dataset.

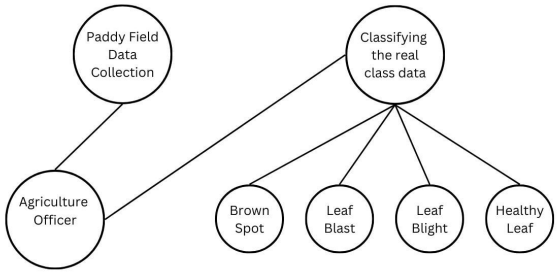


Fig. 6. Raw data collection and labeling work Flow.

There have been made a dataset of 600 images. The dataset has four classes according to the disease type and one is a healthy leaf; each class contains 150 images.The classes are Brown Spot, Leaf Blast, Leaf Blight, and Healthy Leaf.

TABLE I
Dataset Table

Image Data	**Quantity**
Total Images	600
Brown Spot	150
Leaf Blast	150
Leaf Blight	150
Healthy Leaf	150

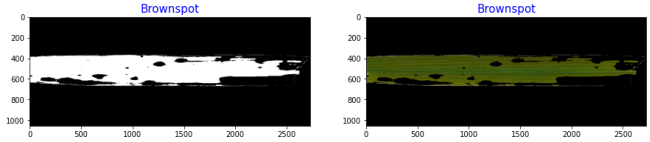
B. Image Preprocessing

Used some online dataset to prepare significant data for prediction. Some data has been scaling, rotating left or right with some degrees and flips of the same category of dataset. That the way to combining all the dataset of the different class label and complete dataset classified by an agricultural officer. The dataset's volume of pertinent data is increased as a result.Also have too changed the images into grayscale, RGB with HSV after that image have been segmented which is ready for classification. So, in order to train the model for prediction, used total segmented image dataset.

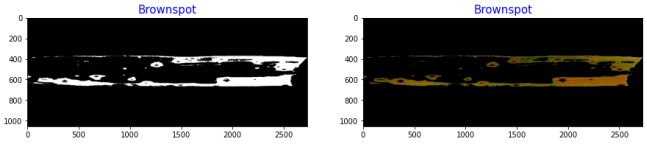
- Data augmentation.
- Fixed size images (500 × 500) pixels with code.
- jpg format conversion.
- Delete duplicate and error images.

C. Image Segmentation

In this research 4 types of classes have been used for image processing techniques. For image processing techniques, K means algorithm have been used for creating the segment of images of those classes. K means clustering is a widespread machine learning algorithm. K-means clustering can be used to create groups of observations having related properties.The sick image always be able to see the infected part of the image using ML techniques. The sounds in a picture that degrade its quality are found during image segmentation. K means algorithm is used to eliminate noise and extraneous spots k- means algorithm converted enhanced image which always produces significant quality images for the detection of paddy leaf disease. Here Fig. 7(a) shows the green good portion of the disease leaf and Fig. 7(b) shows the segmented part of infected part of the disease leaf of brown spot.

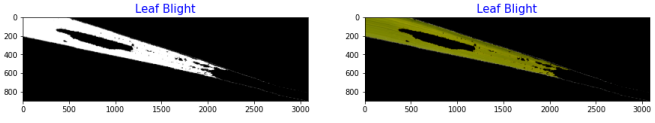


(a) Good Portion of leaf.

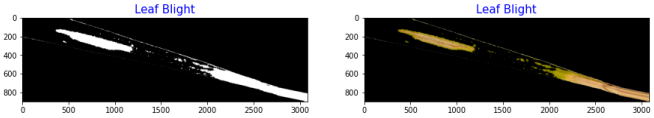


(b) Infected portion of leaf.

Fig. 7. Segmentation of Brown Spot

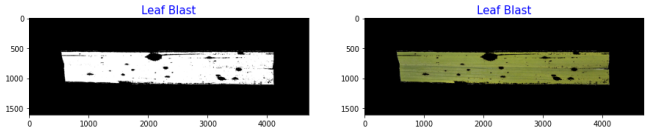


(a) Good portion of leaf.

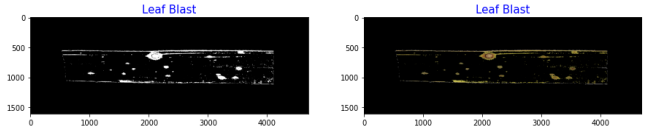


(b) Infected portion of leaf.

Fig. 8. Segmentation of Leaf Blight.



(a) Good portion of leaf.



(b) Infected portion of leaf.

Fig. 9. Segmentation of Leaf Blast.

D. Feature Extraction

Feature extraction define as the image feature like color, shape and texture have been build up for train and testing significantly. Global features extraction also has been part of this feature extraction. In this research global feature have been used for all classes of images. It's increases training and testing speed for prepare the data by removing characteristics from the input data, feature extraction progresses the accuracy of learnt models. The images converted to RGB (Red, Green, Blue) and HSV formatted for the images captured in the pixel values. It also plays significant roles for measuring the better accuracy. The image's width and length are important factors in describing the form. Measuring the number of pixels in an item is an easy way to determine the image's width and height. 3 types of global feature extraction have been used :

- Color.
- Shape.
- Texture.

E. Image Classification

This work used 4 machine learning algorithms for the accuracy of paddy disease. For the accuracy measurement we applied K-fold cross validation techniques for the means of all the model's algorithm after that a prediction result have been given by K-fold cross validation techniques.

Algorithms used Decision tree, Random Forest, Linear regression, Support Vector Machine (SVM) classifier (SVC) models and applied it on K-fold cross validation techniques. Each models used the same dataset, which included both previously accessible data and own dataset that collected from the internet and raw data, after labeling the final dataset. After completing the dataset procedure, utilized Python and its prebuilt libraries to evaluate the algorithms accuracy.

V. Experimental Result and Analysis

A. Result

Working with several ML techniques and K-fold cross validation for the average & accuracy score allowed us to accomplish our target. We employed a total of 4 algorithms for this research Support vector classifier, Decision tree, Random Forest, and linear regression.Before beginning our task, we had to search for a variety of things. We did begin working on the algorithm after selecting it. Then,learned how accurate each algorithm was. As previously noted, we labeled our dataset by an agriculture officer.

After using K-fold cross validation techniques for the following 4 models' classifier, our most accuracy was in Random Forest where we got 94.16% of accuracy score. It is most significant for this dataset after getting the highest accuracy with 4 classes of leaf's accuracy.

TABLE II
Accuracy Table

Classifier	Accuracy Score (AUC)
Decision Tree	85.83%
Random Forest	94.16%
Logistic Regression	85%
Leaf Blight	82.50%

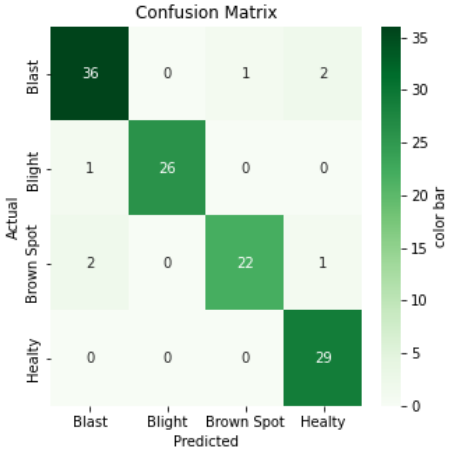


Fig. 10. Confusion matrix of Random Forest.

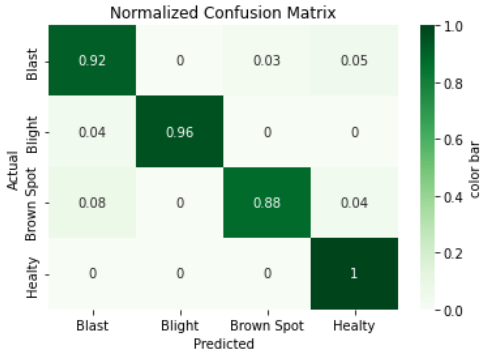


Fig. 11. Normalization Confusion matrix of Random Forest.

Here, showing only Random Forest classification all report and normalization matrix for the achieving highest accuracy.

Models Performance measurements result of three Disease leaf classes as well as healthy leaf with the value of K=10 in cross validation techniques and average cross validation accuracy by following algorithms below Fig. 12.

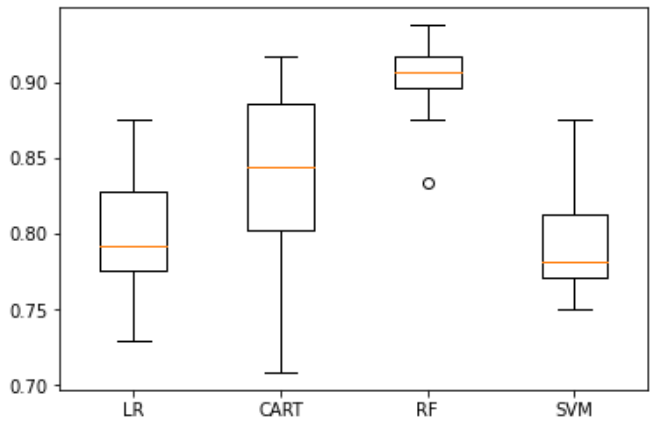


Fig. 12. Comparison of average cross validation accuracy

Performance measurements results of Brown Spot,Leaf Blast,Leaf Blight and Healthy Leaf below Table. III,IV,V,VI with the value of K=10 in cross validation techniques.

TABLE III
Performance Measurements of Leaf Blast.

Algorithms	Accu(%)	F1-score	Precision	Recall	Support	Avg. Cross Valid.
Random Forest (RF)	92%	0.92	0.92	0.92	39	0.9020
Decision Tree (CART)	69%	0.77	0.87	0.69	39	0.8375
Logistic Regression(LR)	67%	0.74	0.84	0.67	39	0.8000
SVM(SVC)	59%	0.70	0.85	0.59	39	0.7958

TABLE IV
Performance Measurements of Leaf Blight.

Algorithms	Accu(%)	F1-score	Precision	Recall	Support	Avg. Cross Valid.
Random Forest (RF)	96%	0.98	1.0	0.96	27	0.9020
Decision Tree (CART)	96%	0.98	0.96	0.96	27	0.8375
Logistic Regression(LR)	96%	0.98	1.0	0.96	27	0.8000
SVM(SVC)	96%	0.98	1.0	0.96	27	0.7958

TABLE V
PERFORMANCE MEASUREMENTS OF BROWN SPOT.

Algorithms	Accu(%)	F1-score	Precision	Recall	Support	Avg. Cross Valid.
Random Forest (RF)	88%	0.92	0.96	0.88	25	0.9020
Decision Tree (CART)	96%	0.87	0.80	0.96	25	0.8375
Logistic Regression(LR)	84%	0.89	0.95	0.84	25	0.8000
SVM(SVC)	84%	0.91	1.0	0.84	25	0.7958

TABLE VI
PERFORMANCE MEASUREMENTS OF HEALTHY LEAF.

Algorithms	Accu(%)	F1-score	Precision	Recall	Support	Avg. Cross Valid.
Random Forest (RF)	100%	0.95	0.91	1.0	29	0.9020
Decision Tree (CART)	90%	0.85	0.81	0.90	29	0.8375
Logistic Regression(LR)	100%	0.83	0.71	1.0	29	0.8000
SVM(SVC)	100%	0.77	1.0	0.63	29	0.7958

Here below Table. VII shows the 4 classes of paddy disease leafs accuracy according to classification models.

TABLE VII
AVERAGE DISEASE LEAF ACCURACY FOLLOWED BY MODELS.

Disease Leaf	Accuracy
Brown Spot	88%
Leaf Blast	71.75%
Leaf Blight	96%
Healthy	97.5%

VI. COMPARATIVE ANALYSIS

In above Table. VIII discuss about different methods with techniques for identify leaf disease of rice with using of classification. In past, research have been occurred for detection various typed of disease. In this research trained four model's accuracy gave better result for perspectives of ML in previous research and Random forest & Support Vector Machine classifier give the best accuracy of the following comparative analysis using 4 paddy classes like: Brown Spot, Leaf Blast, Leaf Blight, and Healthy Leaf. There have been various datasets used in past section of paddy leaf disease using Machine Learning techniques.Machine Learning done some good impacts of disease detection. In this study 3 terms have been done like: disease detection,prediction and four classes of leaf accuracy.

TABLE VIII
COMPARISON OF SOME PREVIOUS RESEARCH ON PADDY LEAF DISEASES.

References	Diseases	Tools and class.	Accuracy
Xiaochun Mai et al. (2016)	Brown spot, bacterial blight, leaf scald, leaf blast	Random forest	dataset2: 80%
Prabira Kumar Sethy et al. [2]	Brown spot, bacterial blight, leaf scald, leaf blast	Fuzzy logic, RF, SVM, K-means	86.35%
Mohd adzhar abdul kahar et al [6]	LBD, BSD, BLB	Neuro-Fuzzy expert system	74.21%
Suman T1, Dhruvakumar T2 [7]	blast diseases, narrow brown spot, BLB,brown spot	SVM with classifier	70%
Suresha M et al. (2017)	Various disease	(k-NN) classifier	76.59%

VII. ADVANTAGES

One of the biggest potential effects of machine learning methods is the ability to detect paddy leaf disease and improve the industry with better results. Early detection of a leafs condition will assist farmers in bringing an instant fix to the field to increase crop production. In the end, this might result in increased agricultural output and a more effective way to combat disease.There is little chance that applying machine learning techniques to identify paddy leaf illness from disease leaf data would have a significant negative effect on the environment. However, it's feasible that the creation and use of these approaches might have an indirect effect on the environment by increasing resource usage and emissions.

VIII. CONCLUSION

According to this work, the study findings and procedures that employed are excellent.Additionally,considering how to integrate algorithms in the future to provide more effective solutions to the issues raised by this research.Decision tree, Random forest, Linear regression, and Support Vector Machine classifier (SVC) those classifiers were used to classify the raw and online dataset that was completely labeled by an agricultural officer and this is the unique things of this research.After using K-fold cross validation techniques,the Random Forest classifier achieved an accuracy score of 94.16%.After achieving the greatest accuracy with four classes of leaf accuracy, it is very notable for this dataset. Brown Spot, Leaf Blast, Leaf Blight, and Healthy Leaf were the four courses utilized for this research.The identification of diseases has benefited from machine learning. In this research, three terms: disease detection using K-means algorithm,prediction using ML classifiers and four classes of leaf accuracy that have been completed examining the accuracy, precision, recall, and f1-score.Applying through this research,this will be able to learn more about area of study.This anticipate that it will further the development of paddy-lead disease detection and innovative approaches to technology that enable to contribute to agriculture.It is hope to propose a new system for paddy leaf disease detection based on this research.

IX. Future Work

There are many possibilities for further study in this field, including in our own work. It found different ways to make our work better. As mentioned, There also found some errors, and these errors are ways to make our study better. That will try to fix this error by implementing our work and increasing prediction results with better opportunities to catch good accuracy.We have plans to fix those errors we get.We also have other goals in mind.We would like to make our prediction better. That is why we can use an algorithm to work with an application that people can use to predict paddy leaf disease, and we will add features like disease detection types and other tools to help the user get the most out of this.We think we can use this kind of work to make the agriculture sector more technology-based and increase the production rate of crops using the Machine learning techniques.We might help the people grow more crops.

References

[1] K.S. Archana, and A. Sahayadhas, "Automatic rice leaf disease segmentation using image processing Techniques," Int. J. Eng. Technol, 7(3.27) ,2018, pp.182-185.

[2] M.A. Islam, M.N.R Shuvo, M. Shamsojjaman, S. Hasan, M.S. Hossain, and T. Khatun, "An automated convolutional neural network-based approach for paddy leaf disease detection," International Journal of Advanced Computer Science and Applications, 12(1), 2021.

[3] M.E. Pothen, and M.L Pai. March, "Detection of rice leaf diseases using image processing," In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, (pp. 424-430).

[4] R.P. Narmadha, and G. Arulvadivu, "Detection and measurement of paddy leaf disease symptoms using image processing," In 2017 International Conference on Computer Communication and Informatics (ICCCI), 2017 January (pp. 1-4).

[5] S. Pavithra, A. Priyadharshini, V. Praveena, and T. Monika, "Paddy leaf disease detection using SVM Classifier," International Journal of communication and computer Technologies, 3(1), pp.16-20, 2015.

[6] M.A. Kahar, S. Mutalib, and S. Abdul-Rahman, "Early detection and classification of paddy diseases with neural networks and fuzzy logic,", in 17th International Conference on Mathematical and Computational Methods in Science and Engineering, pp. 248-257, 2015.

[7] T. Suman, and T. Dhruvakumar, "Classification of paddy leaf diseases using shape and color features," vol. 7, no. 1, pp. 239-250, 2015.

[8] R. Deshmukh, and M. Deshmukh, "Detection of paddy leaf diseases," International Journal of Computer Applications, 975, p.8887, 2015.

[9] P.K. Sethy, N.K. Barpanda, A.K. Rath, and S.K. Behera, "Image processing techniques for diagnosing rice plant disease: a survey," Procedia Computer Science, 167, pp.516-530, 2020.

[10] T. Daniya, and S. Vigneshwari, "A review on machine learning techniques for rice plant disease detection in agricultural research," system, 28(13), pp.49-62, 2019.

[11] T. Bera, A. Das, J. Sil, and A.K. Das, "A survey on rice plant disease identification using image processing and data mining techniques," In Emerging Technologies in Data Mining and Information Security (pp. 365-376), 2019.

[12] A.D. Nidhis, C.N.V. Pardhu, K.C. Reddy,and K. Deepa," Cluster based paddy leaf disease detection, classification and diagnosis in crop health monitoring unit," In Computer Aided Intervention and Diagnostics in Clinical and Medical Images (pp. 281-291),2019.

[13] S. Ramesh, and D. Vydeki, "Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm," Information processing in agriculture, 7(2), pp.249-260, 2020.

[14] S. Nalini, N. Krishnaraj, T. Jayasankar, K. Vinothkumar, and A. Sagai, "Paddy leaf disease detection using an optimized deep neural network," Computers, Materials & Continua, 68(1), pp.1117-1128, 2021.

[15] S.V. Vasantha, B. Kiranmai, and S.R. Krishna, "Techniques for rice leaf disease detection using machine learning algorithms," Int. J. Eng. Res. Technol., 9(8), 2021, pp.162-166.

[16] N. Shelar, S. Shinde, S. Sawant, S. Dhumal, and K. Fakir, "Plant Disease Detection Using Cnn," In ITM Web Of Conferences (Vol. 44, p. 03049),2022.

[17] N. Mangala, P.B. Raj, S.G. Hegde, and R. Pooja, "Paddy leaf disease detection using image processing and machine learning," Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng, 7(2), pp.97-99,2019.

[18] N.S. Patil, "Identification of Paddy Leaf Diseases using Evolutionary and Machine Learning Methods," Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(2), pp.1672-1686,2021.