Rice Plant Disease Classification

Rumana Akter	180104106
Ashif Reza	180104107
Asifur Rahim	180104109

Project Report

Course ID: CSE 4214

Course Name: Pattern Recognition Lab

Semester: Spring 2021

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Dhaka, Bangladesh

March 2022

Rice Plant Disease Classification

Submitted by

Rumana Akter	**180104106**
Ashif Reza	**180104107**
Asifur Rahim	**180104109**

Submitted To

Faisal Muhammad Shah

Sajib Kumar Saha Joy,

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Department of Computer Science and Engineering

Ahsanullah University of Science and Technology

Dhaka, Bangladesh

March 2022

ABSTRACT

Rice is the most staple food in Bangladesh. Because of rice plant diseases, the production of rice hampers heavily. So, rice plant disease detection is a crucial topic in the context of Bangladesh. Again, research work regarding plant disease detection is a trending topic in Bangladesh. The major goal of this work is to create an automated system for identifying the five most serious diseases affecting Bangladeshi rice plants. Additionally, it will make it simpler to identify and predict Bangladeshi rice plant diseases. This research aims to provide an easy-to-use system for the detection of diseases affecting rice plants. "Dhan Somadhan" Dataset, which includes 1106 photos of a total of five different types of rice plant diseases is used. Field backgrounds and white backgrounds have been separated in the photographs.The images are splittied by 80 to 20 ratio.After that, prec-processing like resizing,normalalizing has been done over the images.A combination of gabor and sobel filter are used for feature extraction.Then several machine learning models like svm,decision tree,knn etc are applied.For comparison purposes ROC curves are plotted.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bangladesh is an agricultural country. Rice is a major food for the people of our country. Rice cultivation provides about 48% of rural development. Around 75% of overall harvested area and above 80 of overall irrigated zone is cultivated through rice plants. Consequently, Rice plays an important role for the peoples of Bangladesh. But the quality as well as quantity of rice production may be decreased because of Rice plant Diseases. As having disease in plants is quite natural. If a proper step is not taken in this regard then it causes serious effects on Rice plants. Therefore, it is the major task to identify Rice plant Disease in the early stages. Early detection of Rice plant diseases is very crucial for the crop protection system of our country. Most of the farmers of our country use their own experience to detect diseases, farmers use pesticides in excessive quantities which cannot help in the prevention of disease, but can have a malignant effect on plants. Thus their misclassification creates bad impact on rice cultivation. So, they need advice from rice disease specialists. In remote or rural areas, rice disease specialists are not able to give quick remedies or advice to the farmers in the right time and they also require expensive equipment and a large amount of time for manually identifying and classifying rice diseases. Moreover, traditional visual observation methods are mostly inaccurate. Besides that laboratory testing requires time and can be expensive. The research that has been done on Bangladeshi Rice plant disease is not sufficient. If the farmers can somehow afford a smartphone it will be easier for them touse image processing methods to detect rice plant diseases.

Chapter 2

Literature Reviews

[1] Applied 4 supervised classification algorithms to detect 3 diseases trained dataset using KNN, logistic regression, j48, naive bayes. Three rice plant diseases are detected here.About 400+ pictures are used here for training Affected parts which were separated using K-means clustering and SVM For extracting the features of an image. HOG was used. The Achieved accuracy was 96.77% .

[2] Images are classified using ANN. K-Means Clustering is used for Image Segmentation. Only one leaf disease is detected here. Testing phase accuracy is found to be 90% for the infected and 86% for the healthy images. Total 300 leaf samples are taken as dataset.

[3] Detected Rice Blast disease using KNN and ANN classification techniques.K-means Clustering is used for image segmentation. A small dataset of 451 images is used.Detected only one type of disease(Rice Blast).ANN classifier provides 99% accuracy for normal images and 100% for blast infected images.

[4] Deep convolutional Neural Network (CNN) as feature extractor. Support Vector Machine (SVM) is used as a classifier here. AlexNet deep CNN pre-trained on large ImageNet dataset was used for feature extraction.Small dataset of 619 images was used for this work. Accuracy of 91.37% was achieved here. But this accuracy is based on only 3 types of rice diseases.

[5] A hybrid network integrating Deep CNN with SVM for classification. Features are extracted using D-CNN and SVM classifier and it is trained with the features. The proposed model achieved 97.5% accuracy. Small datasets have been used here for training and testing.

Chapter 3

Data Collection & Processing

Dhan-Shomadhan: A dataset of 5 different harmful diseases of rice leaf called Brown Spot, Leaf Scaled, Rice Blast, Rice Turngo, and Sheath Blight.Dhan-Shomadhans dataset contains 1106 pictures in two different background variations named field background picture and white background picture.Dhan-Shomadhans dataset[6] can be used for rice leaf diseases classification, disease detection using Computer Vision and Pattern Recognition for different rice leaf diseases. Grayscale conversion was done to images.The images were resized 128x128.The dataset was split into training and testing datasets containing 614 train images and 155 test images.

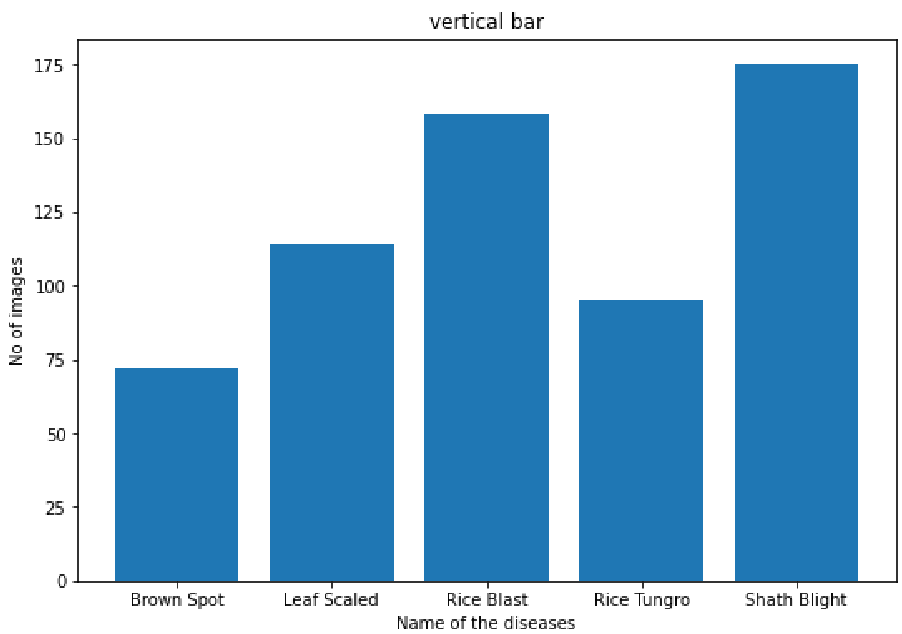

Figure 3.1: Distribution of Data Classes

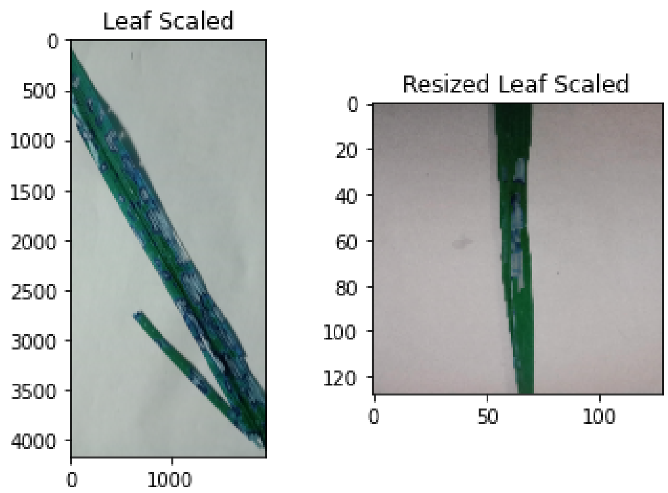

Figure 3.2: Resized Image for Leaf Scaled

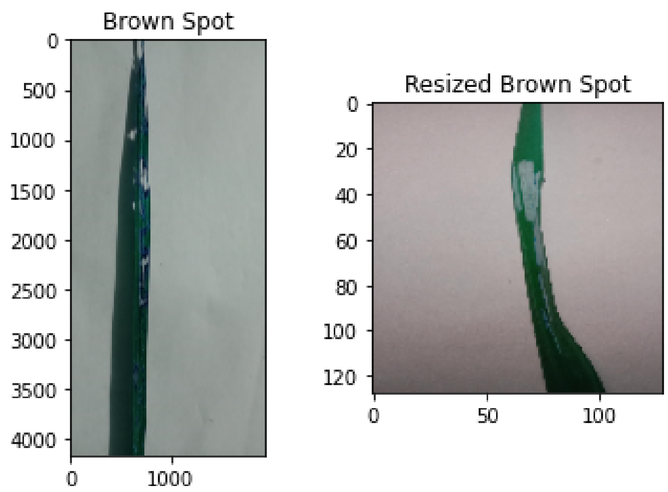

Figure 3.3: Resized Image Brown Spot

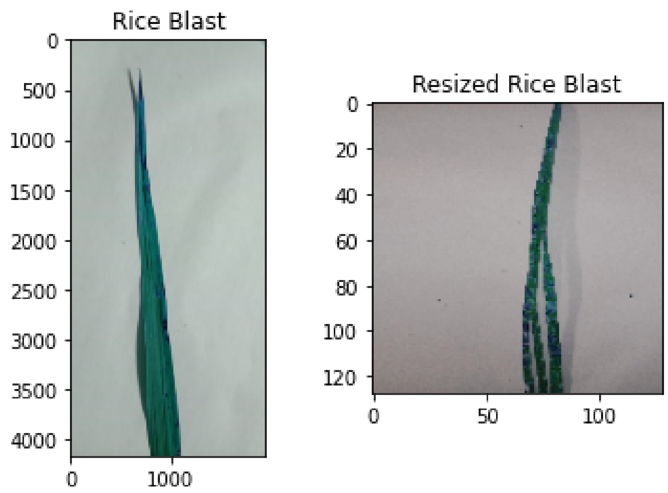

Figure 3.4: Resized Image for Rice Blast

Chapter 4

Methodology

4.0.1 Data Pre-Processing

The Dhan somadhan Dataset was significantly changed for better feature extraction. The Dataset was split into two parts. Among the two, 80% was allocated training and 20% for testing. The images were originally 4000 pixels.It would be difficult to train an image of such size since the processing would take too long. So the images were resized to about 128 pixels so that only the features needed to successfully classify the disease would be present.

4.0.2 Feature Extraction

Two types of feature extractors were used for this purpose. They are the Gabor and Sobel Filters.

4.0.2.1 Gabor Filter

The Gabor filter is an application of the Gabor transform, which combines a Gaussian window and a short-term Fourier transformation for analysis in the spatial domain. The texture information of the image is incorporated into the distortion information of content adaptive image steganography. There creates textural anomaly in an image when embedded. The result of the 2D Gabor filtering on the Gabor signal can be used to identify this textural anomaly. Due to its spatial selectivity and direction, the two-dimensional Gabor filter represents the texture information.Because of its spatial representation, a filter represents texture information.

$$g_{\lambda,\theta,\varphi,\sigma,\gamma}(x,y) = exp\left(-\frac{(x^2+\gamma^2 y^2)}{2\sigma^2}\right)cos(2\pi\frac{x}{\lambda}+\varphi)$$

where, the value of x and y denote the following

$$x = acos\theta + bsin\theta$$

$$y = -acos\theta + bsin\theta$$

λ – Wavelength of Gabor function cosine factor.
θ – Orientation of Gabor function normal to the parallel stripes.
φ – Phase offset of the of Gabor function cosine factor.
σ – Standard deviation sigma of Gaussian factor.
γ – Ellipticity of the Gaussian factor.

Based on different orientation parameters, different types of filters may be generated

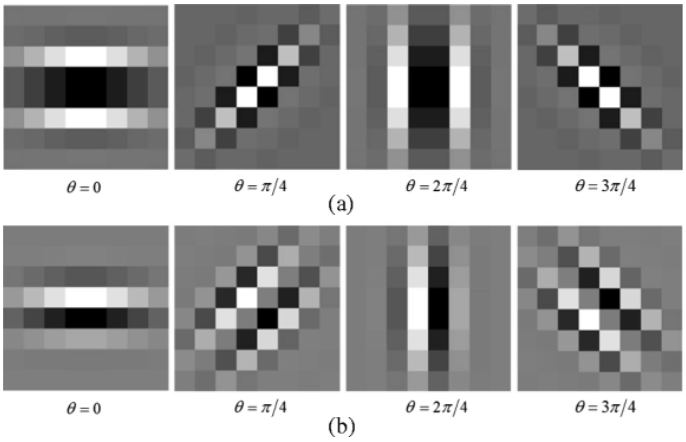

Figure 4.1: Different Orientations for Gabor Filter

4.0.2.2 Sobel Filter

Two 3 x 3 kernels are used in the Sobel filter. There are two for changes in the horizontal and vertical directions, respectively.To compute the approximate derivatives, the two kernels are convolved with the original image. The calculations are as follows if Gx and Gy are two images that, respectively, include the horizontal and vertical derivative approximations:

$$G_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} * A$$

$$G_y = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * A$$

Where A is the original source image. The x coordinate is defined as increasing in the right-direction and the y coordinate is defined as increasing in the down-direction. To compute Gx and Gy we move the appropriate kernel (window) over the input image, computing the value for one pixel and then shifting one pixel to the right. Once the end of the row is reached, we move down to the beginning of the next row.

At each pixel in the image, the gradient approximations given by Gx and Gy are combined to give the gradient magnitude, using:

$$G = \sqrt{G_x^2 + G_y^2}$$

The gradient's direction is calculated using:

$$\theta = arctan\left(\frac{G_y}{G_x}\right)$$

A θ value of 0 would indicate a vertical edge that is darker on the left side.

4.0.3 Machine Learning Models

ML models such as knn, svm, adaboost, random forest and decision tree have been used for the classification.

4.0.3.1 KNN

The KNN algorithm, which is used for both regression and classification (most frequently), belongs to the field of supervised learning. It is a flexible approach that may also be used to resample datasets and impute missing values. The K Nearest Neighbor method, as its name suggests, uses K Nearest Neighbors to forecast the class or continuous value for a new data point.

1. Unlike model-based algorithms, we use whole training instances in this case to predict output for unobserved data instead of learning weights from training data.

2. Model learning is delayed until after a prediction request is made on a new instance and does not use training data beforehand.

3. Non-Parametric: The mapping function in KNN does not have a preset form.

4.0.3.2 SVM

A supervised machine learning approach called "Support Vector Machine" (SVM) can be applied to classification or regression problems. However, classification issues are where it is most frequently utilized. When using the SVM algorithm, each data point is represented as a point in n-dimensional space (where n is the number of features you have), with each feature's value being the value of a certain coordinate. Next, we perform classification by identifying the hyper-plane that effectively distinguishes the two classes.

4.0.3.3 AdaBoost

AdaBoost, also known as Adaptive Boosting, is a machine learning method used in an ensemble setting. Decision trees with one level, or Decision trees with only one split, are the most popular algorithm used with AdaBoost. Another name for these trees is Decision Stumps. This algorithm creates a model while assigning each data piece an equal weight. Then, it gives points that were incorrectly categorised larger weights. The next model now gives more weight to all the points with higher weights. If a low error is not reported, it will continue to train models.

4.0.3.4 Decision Tree

A decision tree is a tool that can be used in many different contexts. Both classification and regression issues can be solved using decision trees. The term itself implies that it displays the predictions that come from a sequence of feature-based splits using a flowchart that resembles a tree structure. The decision is made by the leaves at the end, which follows the root node.

4.0.3.5 Random Forest

Supervised machine learning algorithms like random forest are frequently employed in classification and regression issues. On various samples, it constructs decision trees and uses their average for classification and majority vote for regression. The Random Forest Algorithm's ability to handle data sets with both continuous variables, as in regression, and categorical variables, as in classification, is one of its most crucial qualities. In terms of classification issues, it delivers superior outcomes.

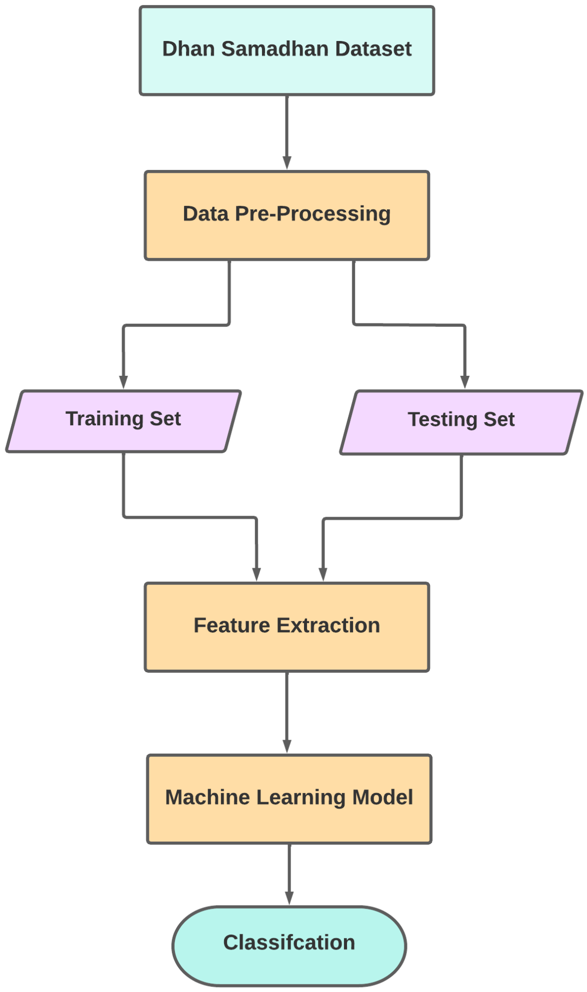

Figure 4.2: Flowchart of Methodology

Chapter 5

Experiments and Results

The Test results achieved from the different machine learning models used for classification of the images into their respective disease class after feature extraction have been displayed accordingly.

Models	Test Accuracy
KNN	.39
SVM	.43
Random Forest	.46
Adaboost	.39
Decision Tree	.34

Table 5.1: Test Accuracy for each model

5.0.1 Random Forest

It is observed in table 5.2 that the amount of True positives are greater in amount which is the cause for precision values to be high in value. Brown Spot and Shath Blight are best in terms of precision and leaf scaled is the lowest

Models	Precision	Recall	f1-score
Brown Spot	.60	.17	.26
Leaf Scaled	.14	.07	.09
Rice Blast	.46	.82	.59
Rice Tungro	.36	.33	.35
Shath Blight	.60	.57	.58

Table 5.2: Result Analysis for Random Forest

5.0.2 SVM

It is observed in table 5.3 that the amount of false positives are greater in amount which is the cause for lower precision values. Brown Spot and Shath Blight are best in terms of precision and leaf scaled is the lowest since it shows a lack of precision and recall.

Models	Precision	Recall	f1-score
Brown Spot	.43	.17	.24
Leaf Scaled	.00	.00	.00
Rice Blast	.45	.68	.54
Rice Tungro	.40	.33	.36
Shath Blight	.46	.66	.54

Table 5.3: Result Analysis for SVM

5.0.3 Decision tree

It is observed in table 5.4 that the True postives rate for rice blast is much greater whereas the Brown spot shows the least amount of precision along with recall.This shows actual postives not being correctly identified in terms of Brown spot.

Models	Precision	Recall	f1-score
Brown Spot	.25	.06	.09
Leaf Scaled	.25	.14	.18
Rice Blast	.52	.28	.26
Rice Tungro	.36	.21	.26
Shath Blight	.32	.73	.42

Table 5.4: Result Analysis for Decision Tree

5.0.4 KNN

In table 5.5 ,Rice Blast and Shath Blight show comparatively better result in terms of the precision whereas Rice Tungro shows the best recall.Brown Spot and leaf scaled however show low values for both metrics.

5.0.5 AdaBoost

in table 5.6, Shath Blight shows more correct predictions for postive identifications based on its relatively higher Precision. Rice Blast Shows more actual positives based on its high recall score.

Models	Precision	Recall	f1-score
Brown Spot	.25	.17	.20
Leaf Scaled	.28	.17	.21
Rice Blast	.47	.40	.43
Rice Tungro	.31	.54	.39
Shath Blight	.47	.52	.49

Table 5.5: Result Analysis for KNN

Models	Precision	Recall	f1-score
Brown Spot	.18	.11	.14
Leaf Scaled	.17	.14	.15
Rice Blast	.44	.62	.52
Rice Tungro	.33	.21	.26
Shath Blight	.50	.55	.52

Table 5.6: Result Analysis for AdaBoost

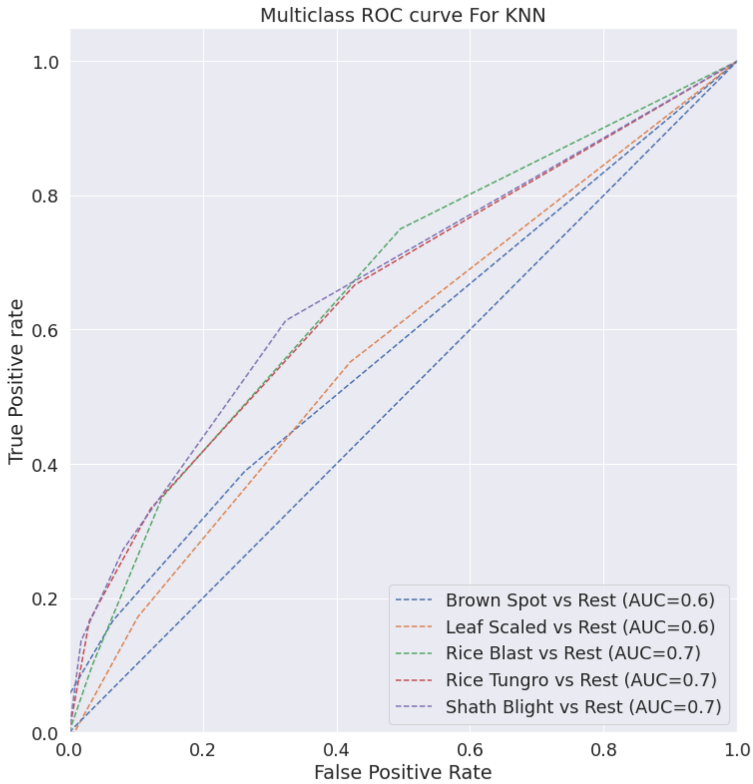

Figure 5.1: ROC curve for KNN

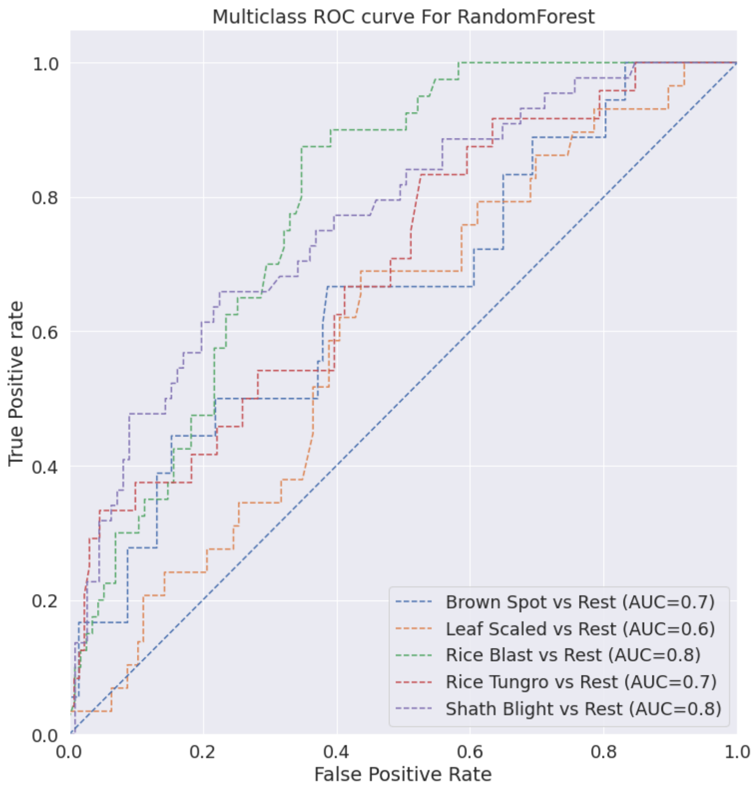

Figure 5.2: ROC curve for Random Forest

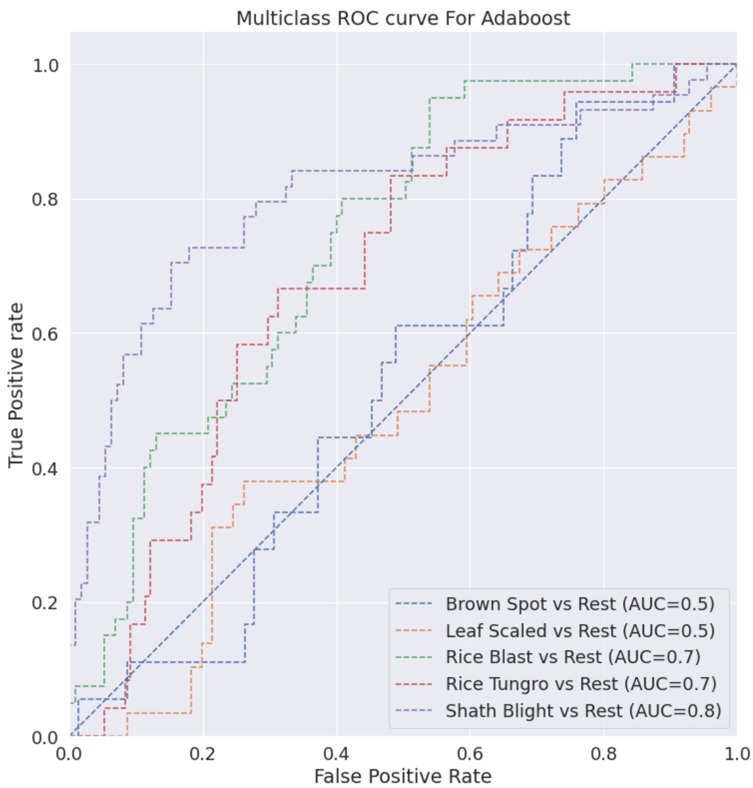

Figure 5.3: ROC curve for AdaBoost

Chapter 6

Future Work and Conclusion

Only photographs with white backgrounds were used, however field background images will also be taken into consideration in the future. Custom ensemble procedures will be applied for increased precision. For purposes of comparison, a number of machine learning algorithms will be used. Oversampling or undersampling techniques will be used to better balance the dataset. Custom Resnet/VGG-16 model and dropout layers will be added over the enriched dataset to prevent overfitting. To improve the results of the training and testing, datasets from other nations will be combined with the local dataset.

References

[1] Kawcher Ahmed,Tasmia Rahman"Rice Leaf Disease Detection Using Machine Learning Techniques.",2019

[2] S Ramesh,D.Vydeki"Rice Blast Disease Detection and Classification Using Machine Learning Algorithm.",2021

[3] S Ramesh,D.Vydeki"Application of machine learning in detection of blast disease in South Indian rice crops.",2019

[4] "Rice Plant Disease Classification Using Transfer Learning of Deep Convolutional 24 Neural Network.",2019

[5] "Rice Disease Identification and Classification by Integrating and Classification by Integrating Support vector Machine with Deep Convolutional Neural Network",2019.

[6] Md Fahad Hossain, Sheikh Abujar,"Dhan-Shomadhan: A Dataset of Rice Leaf Disease Classification for Bangladeshi Local Rice",2021

Generated using Undegraduate Thesis LaTeX Template, Version 1.4. Department of Computer Science and Engineering, Ahsanullah University of Science and Technology, Dhaka, Bangladesh.

This project report was generated on Wednesday 7th September, 2022 at 4:10am.