RICE LEAF DISEASES CLASSIFICATION USING CONVOLUTIONAL NEURAL NETWORKS

Submitted by
Sadia Afrin Rimi
191-35-2632
Department of Software Engineering
Daffodil International University

Supervised by
Musabbir Hasan Sammak
Lecturer,
Department of Software Engineering
Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering.

ACKNOWLEDGEMENT

At the outset, I want to thank Almighty Allah for granting me the capacity to complete the final thesis.Another thanks to my family who have always been supporting me and confident about me that I could implement anything which I have learned.

Also, special thanks to my honorable Supervisor,Musabbir Hasan Sammak, Lecturer, Faculty of Science and Information Technology, Department of Software Engineering, Daffodil International University.Sir was nice enough to accept us as his thesis students and mentor us along the road to get to where we are now. He warned us in advance about the problems we would encounter and advised us to take the correct course.

Special thanks to all my classmates who helped me in this research.

So gradually I want to express my appreciation to all of my course teachers who have built us as a good human being and a good researcher.

Abstract

In this agricultural country of ours surrounded by greenery, the population is increasing day by day. As a result of which the level of arable land is decreasing and increasing residential houses, industrial factories.Food crisis is becoming the main threat for us in the upcoming days.Because on the one hand the population is increasing and on the other hand the amount of food crop production is decreasing due to the attack of diseases.Rice is one of the most significant cultivated crops since it providing foods more than half of the world's population.Early disease detection is the main difficulty in rice crop cultivation. However, when productivity is affected, it might be challenging to diagnose sickness with the apparent. The goal of this project is to increase production by more than 30% by the early diagnosis of Rice leaf diseases.The proposed system in this paper is to detect and classify diseases like Bacterial leaf blight disease,Brown Spot,Tungro Virus by using Convolution Neural Network and deep learning approach.Due of the disease's ability to damage any leaf, regardless of size, recognizing it may be extremely difficult. In order to train the KNN model, a dataset of 1122 photos was collected. The leaf is initially divided into three kinds of categories using KNN. CNN will be utilized in the second phase to classify the diseases.Through this method,we got the justifying accuracy for finding and classifying those diseases. I will work on automatic ways to get rid of diseases in the future.

Table of content

ACKNOWLEDGEMENT ii

Abstract iii

Table of content iv

List of Table v

List of Figures v

CHAPTER 1 1

INTRODUCTION 1

1.1 BACKGROUND 1

1.2 MOTIVATION OF THE RESEARCH 2

1.3 PROBLEM STATEMENT 3

1.4 RESEARCH OBJECTIVE 3

1.5 RESEARCH SCOPE 4

1.6 THESIS ORGANIZATION 4

CHAPTER 2 5

LITERATURE REVIEW 5

2.1 INTRODUCTION 5

2.2 PREVIOUS LITERATURE 5

2.3 CONCLUSION 9

CHAPTER 3 10

RESEARCH METHODOLOGY 10

3.1 RESEARCH METHODOLOGY 10

3.2 DATA COLLECTION 10

3.3 DATA PREPROCESSING 13

3.4 Image Augmentation 14

3.5 Image Segmentation 14

3.6 CONVOLUTIONAL NEURAL NETWORK (CNN) 15

3.7 Transfer Learning 16

3.8 Evaluation Methods 16

3.9 Features Extraction 17

3.9.1 Accuracy 17

3.9.2 Precision 17

3.9.3 Recall 18

3.9.4 F1 Score 18
3.9.5 Confusion Matrix 19

CHAPTER 4 20
RESULTS AND DISCUSSION 20
4.1 INTRODUCTION 20
4.2 RESULT DISCUSSION 20
4.3 Model Discussion 20

CHAPTER 5 27
CONCLUSION AND LIMITATIONS 27
5.1 CONCLUSION 27
5.1 LIMITATIONS 27

6. REFERENCES 28

7. ACCOUNT CLEARANCE 30

List of Table

Table 1: STANDARD CLASSIFICATION **2**

List of Figures

Figure 1: All type of the leaf's sample picture 11

Figure 2: Augmentation Configurations 12

Figure 3: Layers of CNN 13

Figure 4: Confusion matrix diagram 17

Figure: 5 Output of processing system 20

Figure 6: Matrix of data 20

Figure 7: Results of labeled data 21

Figure 8: Results of before resizing data 21

Figure 9: Results of after resizing data 22

Figure 10: Results of validation_accuracy 23

Figure 11: Results of training and validation_accuracy	23
Figure 12: Results of training and validation_loss	24
Figure 13: Confusion Matrix	24

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND

In Bangladesh, the primary source of revenue is agriculture. The second-largest agricultural output production sector in the world is agriculture. Farmers can fail to notice infections or have trouble diagnosing them, which results in crop loss. There is a distinct treatment for every ailment that has to be developed. The present method of illness detection is manual.The traditional methods for diagnosing leaf diseases rely on human eyesight. In these situations, getting professional guidance requires a lot of time and money. The human vision-based approaches have several shortcomings. The eyesight of the individual or expert employed will determine the accuracy and precision of the human vision method. The best choice may be made while choosing a treatment and identifying the different sorts of disorders. The fact that machine learning-based methods accomplish jobs more consistently than human specialists is one of its benefits.it is therefore necessary to develop a new machine learning-based categorization methodology in order to address the shortcomings of existing approaches. Paddy leaf disease detection and classification is the rarest type of plant leaf disease, and there have been very few recent breakthroughs in this sector utilizing a machine learning technique.The biggest problem is that the rice plants are not being continuously inspected. One of the factors influencing the quality and output of agriculture is the illnesses of the rice plant. The phases of development for each plant disease vary.Every time the When a plant becomes ill, farmers must monitor the infection. Bangladeshi farmers are not excellent for informing people about the condition and it period of occurrence also the method for ongoing observation my struggle with infection-related illness. classification of Most important is the discovery of a disease in rice plants.We investigated both picture gathering and agricultural research fields.processing methods, convolutional neural network techniques.Additionally, our planned study in the same direction shows the usefulness of the offered image processing and discoveries.

Leaf Type	Diseases Name
Rice Leaf	Bacterial Blight
	Tungro Virus
	Brown Spot

Table 1: STANDARD CLASSIFICATION

1.2 MOTIVATION OF THE RESEARCH

Bangladesh depends heavily on agriculture. The majority of people's diet is rice, which accounts for 93 percent of all meals produced, 70 percent of typical strength consumption, and 35 percent of family spending. The Bangladesh Bureau of Statistics estimates that it has 85.77 lac hectares of arable land. In the fiscal year 2017–2018, 13992874 metric tons of Aman rice were produced. Identification of the illnesses affecting rice plants is the key to lowering the amount of production loss and the amount of agricultural produce. Utilizing an automated method to identify illness is advantageous since it shortens time-consuming tasks like monitoring agricultural fields and may identify disease signs when they are first noticed on plant leaves. Plant disease identification is essential because, if treated well and promptly, it can boost harvest yield. The necessity to reduce rice harvest loss due to disease, which will assist to reduce the food shortfall, is the driving force for our study.The need for a system that can help rice farmers identify rice diseases early is evident from the foregoing. Such a system would not only boost output and quality but also lower costs, which would be advantageous to both farmers and consumers. The improvement of productivity and quality is the primary focus of technology research in the agricultural sector. It has been said that the characteristics of the rice plant, which distinguish it from other agricultural plants and call for special attention from researchers to target rice plant illnesses and aid in the prevention of early detection.

1.3 PROBLEM STATEMENT

The cost of life in a country rises as a result of rice plant disease, and the quality and quantity of agricultural crops also decline. The yield of crops across the country may be impacted by these two factors happening at the same time. These difficulties have sparked a great deal of interest among academics to find involuntary approaches that may assist in accurately identifying illnesses connected to rice plants and assisting stakeholders in detecting diseases early to minimize economic damage.The farmers can therefore select the particular fertilizer. Furthermore, a number of models exist to identify different plant diseases using machine learning and image classification approaches. To identify and diagnose rice plant illnesses in both their early and present stages, we have to use CNN in this research. The first thing to keep in mind is that throughout Asia, rice output is slowed by rice plant diseases by between 10 and 20 percent. Studies have led researchers to conclude that fungus and bacteria are the main causes of rice plant illness.

1.4 RESEARCH OBJECTIVE

The major goal of this research is to create a system that reliably detects, categorizes, and diagnoses rice illness automatically and without human involvement. The ultimate goal is to also suggest new approaches that are more accurate at diagnosing problems than existing approaches that employ comparable or comparable-sized datasets.Our thesis is for,

- Creating a computerized system.
- Building a system that is affordable for a developing country like Bangladesh.
- Classifying the affected leaf which is infected by fungus,virus or bacteria.
- Install on any agri management system.

1.5 RESEARCH SCOPE

The main research areas are as follows:

● Developing a system that can identify diseases by looking at pictures of plant leaves.

● Such economic losses from plant diseases may result in lower crop revenues for growers and distributors as well as increased costs for consumers.It will help in maintaining the standard of living.

● We can easily identify the disease and give proper treatment according to the disease.

● Anyone can use it easily. Farmers can also use it who don't have extra knowledge about ICT or the internet.

1.6 THESIS ORGANIZATION

In particular detail about the illnesses and classification system and its application, the first chapter covers the backdrop of the study, the justification for the research, the issue description, the research questions, and the research aim. Additional elements of our research include the following:

The literature review, which enables us to see prior research on issues connected to rice leaf diseases, the methodology used, any gaps, and a comparison of my work and other researchers' work based on our shared interpretation of their findings, will be covered in the second chapter.

The third chapter will discuss the research approach we used. In the portion of my paper devoted to technique, I'll go over data collecting, data preparation, and work analysis.How the techniques function is also addressed.

The methodology's results and a variety of outcome accuracy scores will be detailed in chapter four. We can quickly comprehend this document and comprehend the accuracy of disease detection and disease categorization for the sake of this debate.

The chapter that follows is the conclusion. In this section, I'll give the conclusion and give a thorough description of my work. I've discussed here the work I'll do in the future to enhance the work.

CHAPTER 2
LITERATURE REVIEW

2.1 INTRODUCTION

An analysis of previous work, research, conference papers, books, journals, etc. is done in a literature review. With this knowledge, one may find out what research has previously been done on the problem,
Recap it and point out any gaps in the work. Upon research, they may concentrate on limitations and discover workarounds to go around them to provide higher outcomes.A literature review is a summary of the body of knowledge about your study issue. The research question and primary research are both based on this synthesis, which combines the findings of several sources to describe the entire understanding of the subject. An annotated bibliography is not a literature review, even though it will cite sources and should analyze the reliability of the sources used. Regardless of whether they support the assertions you ultimately intend to make, your literature study has to examine all the important sources on a subject.

2.2 PREVIOUS LITERATURE

This section provides an analysis of a number of research that have been done in the past on image processing. In this overview of the literature, several methods and techniques for employing image processing to identify rice leaf disease are discussed. This research also takes into account the effects of Deep Learning and Machine Learning techniques by taking into account numerous algorithmic models.The following list includes a few current studies on the identification and categorization of illnesses affecting rice plants.

Pallapothula Tejaswini and et al. [1] worked on classifying rice leaf diseases.The accuracy of their proposed 5-layer convolutional network is approximately 6 percent more than the other standard deep learning models. They observed that by adjusting the training parameters like learning rate, epochs, and optimizer methods, they can get significant accuracy with a handmade model having less number of layers. The better they can detect infections, the simpler it will be for farmers to protect their crops.The 5-layer convolution model had the best accuracy of 78.2 %.

In this study,SVM classifier is used to identify rice illnesses, and recommendations for pesticides and fertilizers are given based on the disease stage. The accuracy of this approach is 90.95 percent for brown spot, 94.1 percent for leaf blast, and 85.7 percent for bacterial leaf blight. Although this investigation was accurate, just three illnesses were identified.

This was fixed in the study by using a fungal dataset and an artificial neural network to fix the issue in the current system before developing an intelligent system for real-time smart farm monitoring.They also get the best accuracy by using CNN and deep learning model than others model.

In [10], PNN was employed to identify rice sickness. With accuracy rates of 92.31%, 96.25%, and 97.96%, respectively, this study demonstrated promising performance in identifying Brown Spot, Bacterial Leaf Blight, and Tungro. There were no form features used in this study, which might have improved accuracy by helping to identify disorders.

In this study area,the detection of rice leaf disease has been compared using four machine learning techniques, including KNN, Decision tree, Logistic Regression, and Naive Bayes.Among them Decision Tree give more accuracy than others.

Additionally, it has been noted that among other methods, the Deep feature based SVM method is displaying a higher F1-score and Accuracy. .not models that are inexpensive to implement.

In this paper,the performance and accuracy of the KNN and SVM architectures, which are employed to build the models, are also researched and analyzed.The accuracy of this model is good.

In a research by Pujari et al [22], fungi-related illnesses of economic crops such as chili, cotton, and sugarcane were found. Principal wavelet analysis (PWA) was utilized to further reduce the number of features after they had been extracted using the discrete wavelet transform (DWT).

In addition, the author of study [11] described the four main approaches for developing the processing scheme. The RGB input picture is first given a color transformation structure. RGB is utilized to produce color because HSI is employed as a color descriptor. In order to transport or convert RGB images, it is also necessary.
Second, a threshold value is used to mask and eliminate green pixels. The third step is segmenting the image using the usable segments that were retrieved in the previous stage.
The major stage is finished, and segmentation should be completed.

Hexagonal matching is used, in accordance with [9], to diagnose plant diseases. Given that plant illnesses mostly affect the leaves, the method deals with edge detection as well as color for histogram matching. Layer separation is used throughout the training process, and the provided samples are trained. In order to do this, divide the layers of the RGB picture into layers of blue, red, and green. Edge detection technology is also used to create the edges of the layered pictures. It uses spatial gray-level dependence matrices to construct the color co-occurrence texture analysis technique.

The literature study may be used to infer that in order to improve efficiency, researchers have investigated and applied a number of techniques for detecting illnesses in rice plants that are based on traditional machine learning, such as the support vector machine (SVM) and k-nearest neighbor (NN). Deep CNN is gradually becoming more effective day by day with a larger success rate and more effective results when categorizing among the classes. Despite the lengthy training period, the trained models are able to differentiate and categorize pictures fast and simply. CNN is today's ideal model for pattern recognition with large amounts of data.

In addition, many image processing algorithms are provided in the work [12] to identify plant diseases. The objective of existing techniques research is to reduce subjectivity output from plant disease diagnosis and detection conducted by naked eye observation.

The research R-CNN was utilized by Zhang et al. (2021) to look into how to identify soybean fungal infections from a fake image. Accurate detection of soybean plant disease is essential for the sustainability of soybeans and the health of the agribusiness industry. Although several researches have been done on the issue, it is more difficult to identify diseases in soybeans owing to a lack of data and technical difficulties. The goal of this study is to overcome the issue of a small database by building a synthetic soybean plant leaf picture database. Model execution resulted in an accuracy percentage of 83.34 percent.

According to [Mohammad, et al. (2019)], image processing techniques can be used to detect rice blast illness early. The disorders in lab color space were classified using enhanced k-NN and K-means in this study.

For pictures that were partitioned, the Otsu approach was employed for segmentation. Extracted and utilized for categorization are shape and color characteristics. The efficacy of the k-NN algorithm and k-means is determined by considering sensitivity, specificity, and accuracy. The shown method had a 94% accuracy rate.

To recognize and categorize the illnesses Rice blast, bacterial blight, and sheath blight, [Zhou, et al. (2019)] presented a method. The Otsu threshold approach was employed to lessen the background portion's influence with the identification of the necessary region. The optimal values of k are determined using the FCM-KM algorithm. For the detection of rice illnesses, a combination of FCMKM and Faster R-CNN is utilized. For rice blast, bacterial blight, and blight, the accuracy of disease detection was 96.71%, 97.53%, and 98.26%, respectively.

Triangle threshold approaches and fundamental threshold methods are also well detailed in [6]. These two techniques are employed to identify the contaminated region as well as to divide the leaf area. The triangle threshold and simple threshold approaches are utilized in infested areas. The leaf region is also divided into segments. Last but not least, illnesses are classified by calculating the ratio of leaf area to affected area. The computation of leaf area also employs threshold segmentation. As a result, the research concludes that the suggested approach for estimating the degree of illness severity is both time-consuming and quick.

In [23], a novel method for creating a deep convolutional neural network-based model for plant disease identification was put forward. For the experiment in this study, 4500 photos were employed. With some fine-tuning, the accuracy attained using the suggested approach increased to 96.3% from 95.8%. 2589 photos were utilized for validation after the model had been trained on 30880 images.

A brand-new rice blast identification technique has been put out by the authors in ref. [13]. In order to train and test the CNN model, a dataset of 2906 positive and 2902 negative samples was supplied, and the accuracy rate was 95.83%. They also performed comparison studies for various models, with the results favoring high-level features recovered by CNN over manually created features like local binary pattern histograms (LBPH) and wavelet transform (Haar-WT). In light of this, they came to the conclusion that their CNN model performs better than alternative methods in identifying rice blast illness. Since this study only detects rice blast illness, a significant restriction, it is not suited for use in the treatment of other diseases.

2.3 CONCLUSION

In order to offer that information in the form of a written report, a literature review's goal is to get an overview of the current research and discussions that are pertinent to a certain topic or field of study. You may advance your understanding in your subject by carrying out a literature review. There are several types of algorithms. To get a respectable result, they employed a number of approaches, including feature extraction, augmentation, annotation, and more. Pass up better precision and real-time performance. We also tried to categorize the three different rice leaf illnesses as simply as possible in our attempts.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 RESEARCH METHODOLOGY

The approach we suggest using to identify illnesses of rice leaves is described in this section. The entire procedure is broken down into several parts, starting with the creation of a brand-new training dataset, followed by the creation of a brand-new CNN model, deep feature extraction for the model training, and categorization of the illnesses affecting rice leaves. We used our own dataset to create the dataset.

3.2 DATA COLLECTION

Every research project begins with the collecting of data. These individuals gather data for a variety of reasons. Some governments do so for academic research, while others do it for business purposes. These are the impacted rice leaves in my data, which is primarily agricultural data. In order to use the data and produce useful results for the nation, I was considering gathering some of it. Due to the fact that it is the Amon rice season and my town has a large number of rice fields, I headed there for this.Then, I captured almost a thousand images of three different types of infected leaves: those with Leaf blight, Tungro virus, and Brown spot.
All photos were taken with a smartphone. The information is gathered from several angles. During the intense heat, some data was gathered. The most information was gathered throughout the day.







Figure 1: All type of the leaf’s sample picture

3.3 DATA PREPROCESSING

Before the implementation phase, this step entails working on the database. It is essential to pre-process the data for each research project before using it for modeling. To utilize the photographs in the right format and facilitate modeling more quickly, normalization of the images is done in this research effort. as a result of the differences in the original sizes of the photos.The key benefit of picture scaling is that it makes it easier to train the model quickly since it does so more quickly with smaller images. The photos are scaled to 416*416 because they were originally much larger. We have tested with the numbers in an effort to find the optimum match that would support test findings that were more accurate. As was said, several techniques have been used to improve the identification of illnesses affecting rice plants.

3.4 Image Augmentation

The act of creating fresh photos for our deep learning model to be trained is known as image augmentation. We don't need to manually gather these fresh photos because they are created using the training images that are already available.Different sizes, orientations, rotations, shearing, levels of magnification, etc. might be present in the training images. The accuracy of our learning model must not be hampered by these circumstances. By creating several versions of a picture from a single one, we have used the image augmentation approach to enhance the training set and overcome this problem. Figure 2 illustrates the settings that we have utilized.

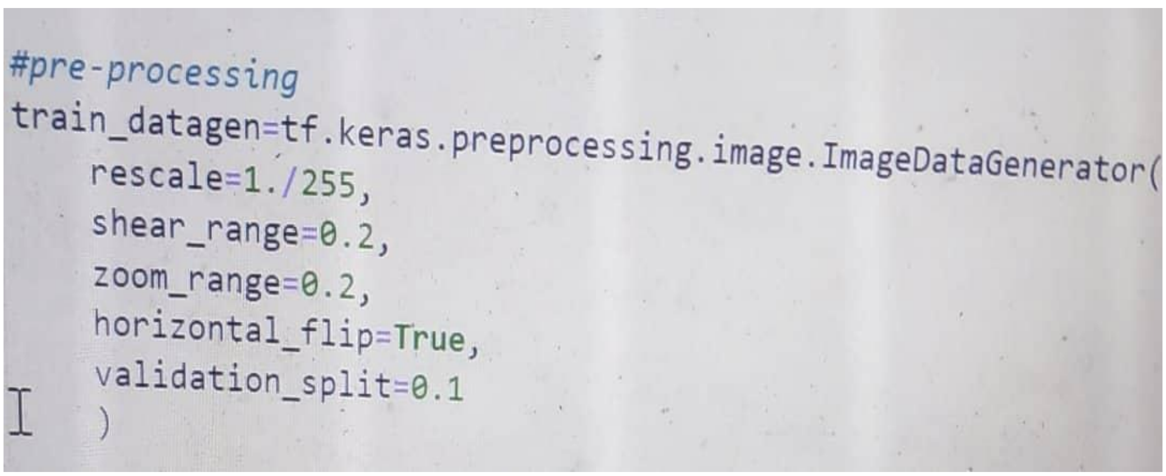

Figure 2: Augmentation Configurations

3.5 Image Segmentation

By breaking up a picture into smaller, more manageable pieces, segmentation aims to make an image's representation more understandable and research-friendly. Because Otsu's thresholding determines the threshold value automatically, it is helpful. The diseased areas' margins are reddish and have a brownish tint, as can be seen. In order to do this, the photos' thresholds were first set according to how much the red channel differed from the other two. The value of the red plane is regarded as disease-affected if it is considerably higher than the green and blue plane values.

3.6 CONVOLUTIONAL NEURAL NETWORK (CNN)

In a variety of industries, including agriculture, convolutional neural networks have advanced significantly in recent years. Backpropagation is a technique used by CNN to build conceptual spatial-temporal hierarchies of features. CNN constructs these hierarchies utilizing a number of blocks of convolutional layers, pooling layers, and fully connected layers. CNN does not need manual feature extraction, in contrast to other rudimentary methods where it is done. These traits are automatic learning processes for it.
While the kernel weights and activation weights are learnt during training, the CCN's other parameters, which must all be fixed, are not. Less overfitting is possible when the size of the data collection grows and the data are made more regular.
The following layers make up the majority of our architecture:

- Convolution layer
- Pooling layer
- Fully connected layer

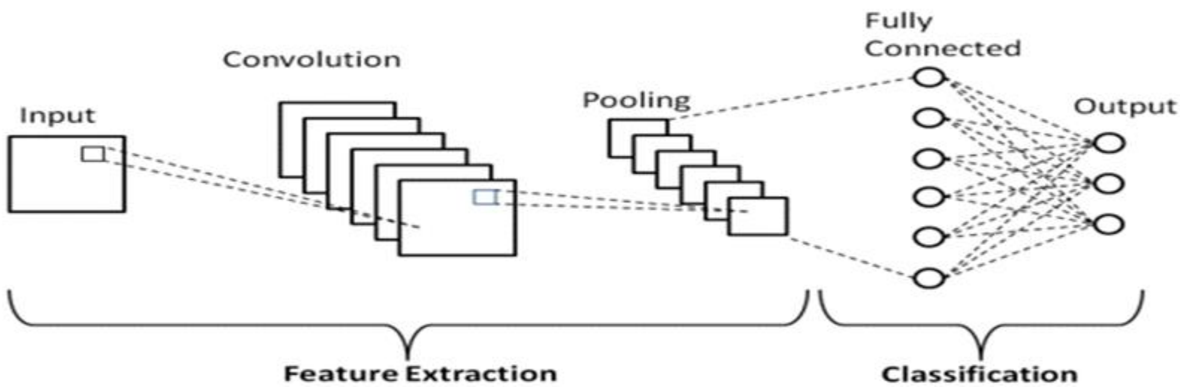

Fig 3: Layers of CNN

An illustration of how CNN operates is seen above. After data has been preprocessed and the necessary features have been extracted, the input in the form of an image is submitted to CNN, which processes it via three layers of CNN to provide an accurate representation. Following that, the result is shown.

● **Input Layer:** The dataset makes up the input layer of a CNN. A 3X3 matrix will be used to display the supplied data.

● **Convolution Layer:**A layer that extracts characteristics from a picture by using filters to learn from more manageable portions of the input data.
● **Pooling Layer:**By reducing the number of dimensions in the image, this layer can reduce the amount of processing power needed for next layers. There are two distinct pooling methods. As follows:
● **Max pooling:**As the input is being parsed, the pixel with the highest value is chosen and moved to the output. Comparing average pooling, it is the strategy that is most frequently utilized.
● **Fully Connected Layer (Dense):**One of CNN's last layers, it is capable of identifying characteristics that have a strong relationship to the output class. After flattening the pooling layer results, a one-dimensional vector is the end result.
● **Dropout Layer:**By eliminating a random group of neurons from that layer, used to mitigate the problem of model overfitting. With the FC layer, it is linked.
● **SoftMax Layer:**The last layer of the network is responsible for helping to divide the dataset's input photos into several classes based on the network's learnt attributes.
● **Output Layer:**The ultimate categorization outcome is stored in the output layer.

3.7 Transfer Learning

Applying a model that has already been learned to new problems is referred to as transfer learning. It is now highly popular in deep learning since it can train deep neural networks with only a little amount of data. This is very useful for data scientists because the bulk of real-world scenarios frequently lack the millions of labeled data points needed to train such intricate models. The technique of reusing model weights that have already been taught is known as transfer learning. Using a learnt model of one problem to solve another is the function of a neural network, a particular kind of computer system.The common properties of photographs may be recognized using transfer learning techniques. The system is still operational even though it achieved cutting-edge performance.
Transfer learning has several benefits, but its three main advantages--a shorter training period, better neural network performance, and the absence of input requirements are the most important ones.

3.8 Evaluation Methods

I have visualized the confusion matrix to evaluate the results. The assessment process necessitates both true positive and negative results in addition to false positive and negative values. This is a genuine plus as the actual amount was as anticipated. The number is expected to be positive but is instead a false positive, which is a true-negative that must be correctly rejected. False-negative is unfairly rejected.

3.9 Features Extraction

Techniques for features extraction should be used when dealing with large amounts of data because most of them would be redundant and unimportant, making calculations difficult and producing inaccurate results. Feature extraction may also be a general term for the process of putting together a number of different factors to exacerbate these problems while still portraying information through advertising. The assumption and agreement among many AI experts and machine learning practitioners is that correctly improved component extraction results from successful model building.

3.9.1 Accuracy

Depending on the model's accuracy, a computer's ability to predict outcomes will vary. It indicates that something is noteworthy when each class has equal weight. In terms of my career, I value each class equally. The accuracy is therefore vital in assessing the model's correctness. As can be seen below, accuracy may also be assessed in terms of positives and negatives for binary classification.

$$Accuracy = \frac{TrueNegatives + TruePositive}{TruePositive + FalsePositive + TrueNegative + FalseNegative}$$

3.9.2 Precision

Precision is a measure of how many positive class forecasts really fit into that category.Precision is calculated by subtracting the true positive value from the overall positive value.A definition of accuracy is provided below:

$$precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

3.9.3 Recall

The exact true positive identification can be measured using the metric recall. The recall is calculated by multiplying the real positive value by the total number of relevant documents that are currently in existence.

Here is what "recall" means:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

3.9.4 F1 Score

The F1 Score is produced by adding precision and recall. False positives and false negatives are therefore taken into account while calculating this score. Though it is less intuitively straightforward to understand, F1 is frequently more advantageous than accuracy, especially if you have an unbalanced class distribution. In order for accuracy to operate at its best, false positives and false negatives should roughly cost the same.If the costs of false positives and false negatives differ significantly from each other, it is advisable to include both Precision and Recall.An explanation of F1 score is provided below:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

3.9.5 Confusion Matrix

The performance of a classifier can be evaluated considerably more effectively by looking at the confusion matrix.
Finding out how frequently examples of class A are classified as class B examples is the key objective.
A confusion matrix must have the following four characteristics:

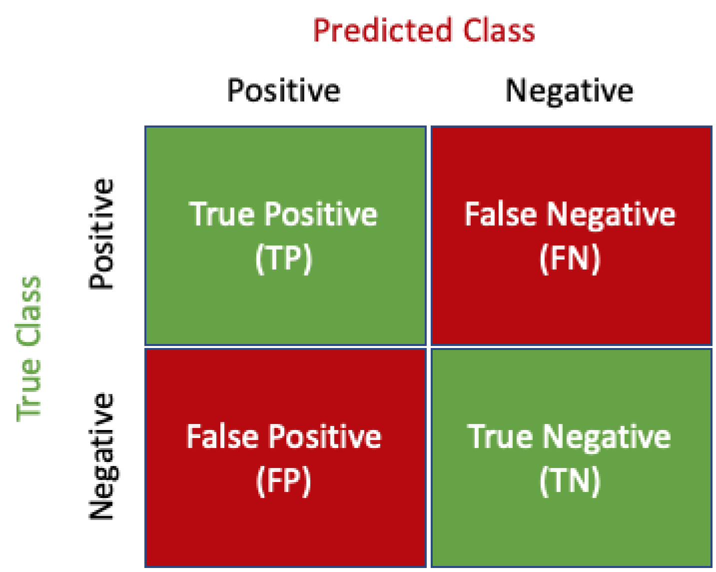

Figure 4: Confusion matrix diagram

CHAPTER 4
RESULTS AND DISCUSSION

4.1 INTRODUCTION

The methods and models for spotting rice leaf and identifying illnesses are provided. I described the models I used for the detection strategy after the data collection and preparation phase.There will be a discussion of the model's results that I employed.

4.2 RESULT DISCUSSION

In this part, experimental findings using the suggested methodology are reported together with pertinent analyses. Our personal data collection is one of the three datasets on which the suggested approach has been tested. The classification models for the aforementioned three datasets have been trained using three CNN models: VGG16, ResNet50, and Dense Net. The suggested method's classification performance for the three aforementioned datasets for the three aforementioned CNN models is reported in terms of validation accuracy and test accuracy. We used 10-fold cross-validation on the datasets and hold-out on the model to determine its full potential.The model was trained using all of the gathered and taken photographs in the final experiment, and the test accuracy was 78.84%, while the validation accuracy was 82.03%. Because separate datasets' photos were captured under various situations using various types of cameras, the accuracy suffered. Because of this, the colors of the diseased leaves varied between datasets. Therefore, when we mix them, the neural network becomes confused and cannot extract key information.

4.3 Model Discussion

A deep CNN architecture with several layers is called Visual Geometry Group (VGG).

16 and 19 convolutional layers, respectively, make up VGG-16 and VGG-19. In order to improve the network depth, these designs are built utilizing incredibly tiny convolutional filters. A 224 224 picture with three color channels is the input for both VGG16 and VGG19. The input is then sent to convolutional layers with a 3x3 maximum receptive field and max-pooling layers. The first two VGG sets with conv3-64 and conv3-128 correspondingly have a shorter training period because of the ReLU activation function in the VGG network. ReLU is a feature used in AlexNet, an extension of LeNet, to accelerate learning, use max-pooling rather than average, minimize the size of the network by overlapping pooling filters, reduce overfitting, and increase generalization. AlexNet's architecture comprises 8 layers: 5 convolutional networks and 3 FC layers. Conv3-256, Conv3-512, and Conv3-512 are used, respectively, in the last three sets with the identical activation function.

To retain spatial resolution, a max-pooling layer is added after each pair of convolutional layers with a stride of 2 (the number of pixels that are moved across the input matrix), creating a 2x2 window. Additionally, the convolutional layers' utilized channel count varies between 64 and 512. By using multilayer feature concatenation for all ensuing layers, DenseNet, an extension of Res-Net, makes the training of deep networks easier by minimizing the number of parameters in the trained model.As a result, the model's efficiency is increased and direct summing of the preceding layers is avoided. In this work, the DenseNet-201 architecture, which consists of 4 dense blocks with sets of 1x1 and 3x3 convolutional layers, is implemented. A transition block with a 1x1 convolutional layer and a 2x2 pooling layer follows each dense block, with the exception of the last block, which is followed by a classification layer with a 7x7 global average pool. A complete network with 4 outputs follows this final section.

The VGG19 network consists of five max pooling layers and 16 convolutions with ReLUs. The convolutions' filter maps start off at 64 and increase to 512 in number. A linear classifier with three fully-connected layers and a 50% dropout between the first and second FC layers follows the convolutions.

By employing inception modules and working at the same layer, the GoogleNet design also enables the network to select from a variety of convolutional filter sizes in each block, improving computing efficiency. The design has a total of 27 levels made up of 9 stacked inception modules and 22 layers of parameters. The inception module serves as the foundation layer for GoogleNet, which is subsequently built upon by other layers, where parallel filtering of the input layer from the preceding layer is implemented. The four classes are classified using SoftMax loss functions.

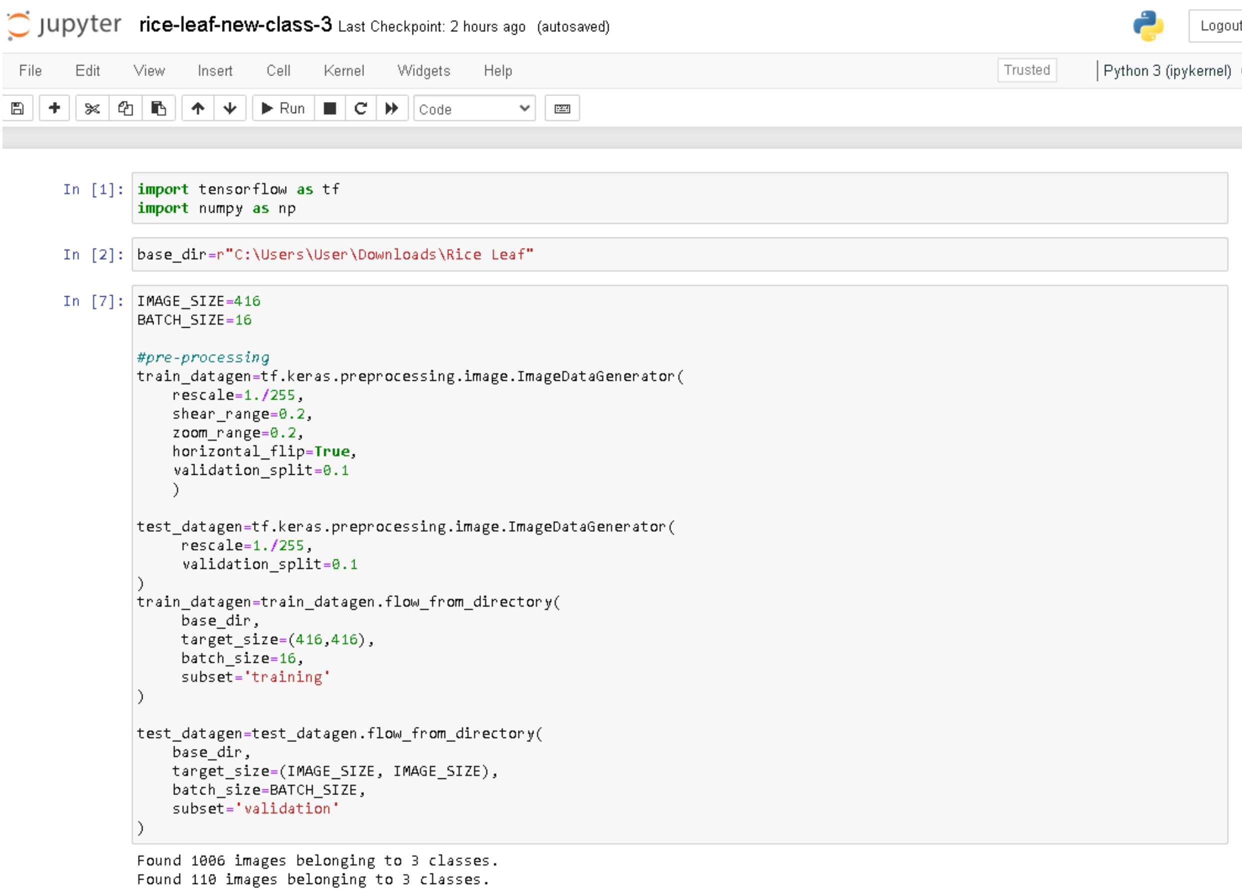

Figure: 5 Output of processing system

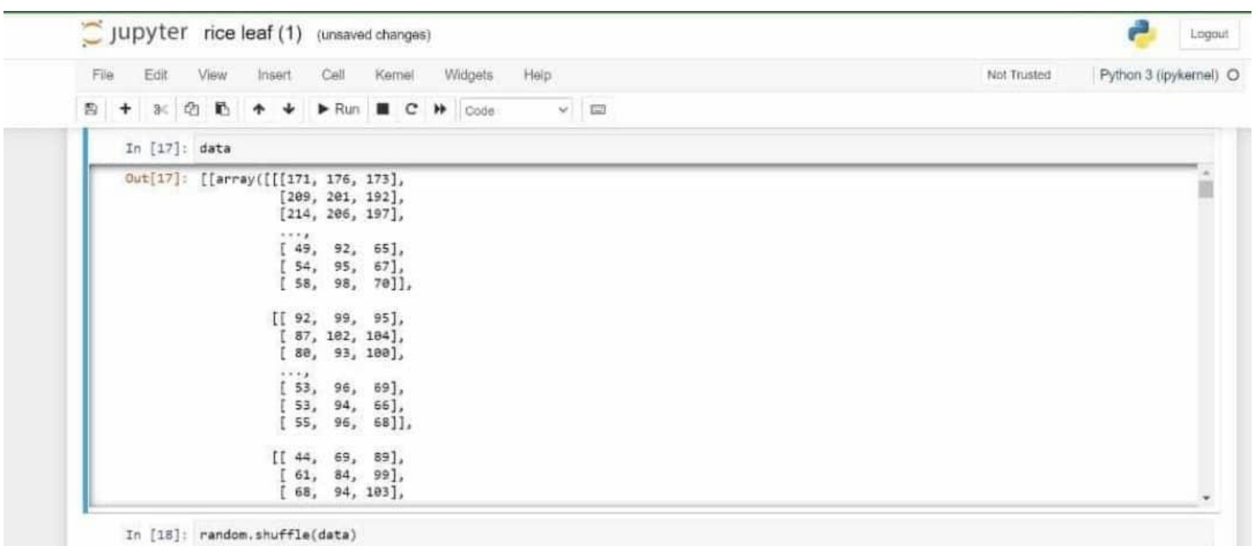

Figure: 6 matrix of data

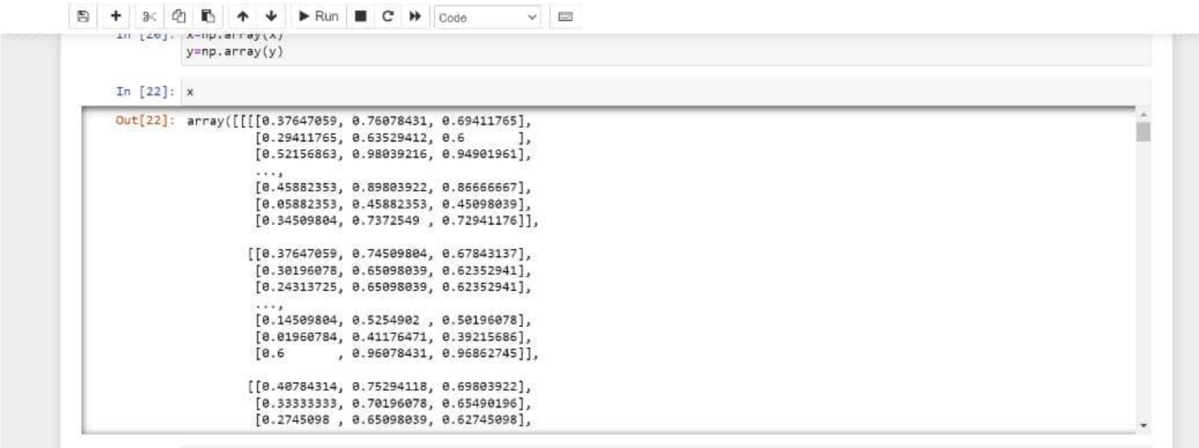

Fig 7: Results of labeled data

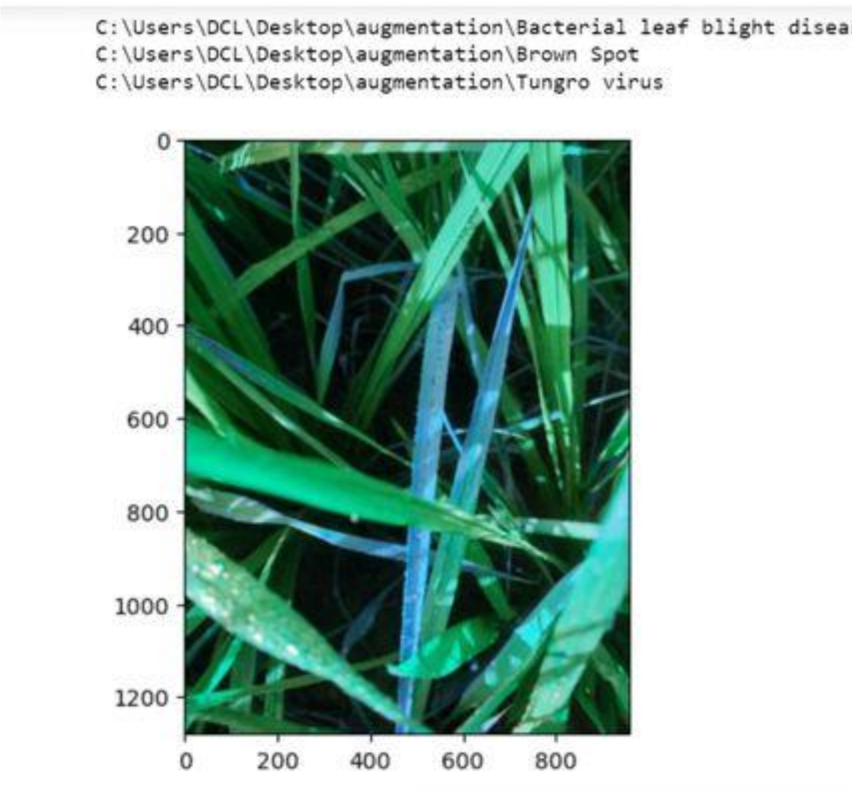

Figure 8: Results of before resizing data

```
C:\Users\DCL\Desktop\augmentation\Bacterial leaf blight disease
C:\Users\DCL\Desktop\augmentation\Brown Spot
C:\Users\DCL\Desktop\augmentation\Tungro virus
```

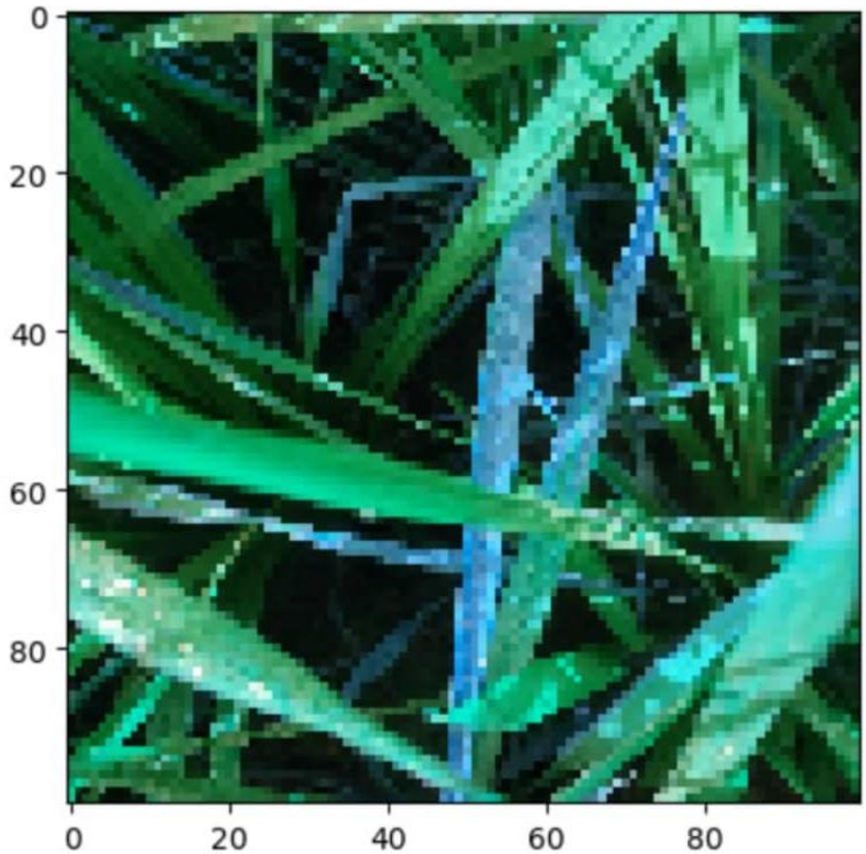

Figure 9: Results of after resizing data

```
Epoch 1/10
63/63 [==============================] - 81s 1s/step - loss: 3.6882 - accuracy: 0.4632 - val_loss: 0.9571 - val_accuracy:
0.6182
Epoch 2/10
63/63 [==============================] - 82s 1s/step - loss: 0.8680 - accuracy: 0.6362 - val_loss: 0.7210 - val_accuracy:
0.8091
Epoch 3/10
63/63 [==============================] - 81s 1s/step - loss: 0.6950 - accuracy: 0.7237 - val_loss: 0.6968 - val_accuracy:
0.8455
Epoch 4/10
63/63 [==============================] - 80s 1s/step - loss: 0.6456 - accuracy: 0.7316 - val_loss: 0.6864 - val_accuracy:
0.7909
Epoch 5/10
63/63 [==============================] - 80s 1s/step - loss: 0.6039 - accuracy: 0.7634 - val_loss: 0.8192 - val_accuracy:
0.7727
Epoch 6/10
63/63 [==============================] - 81s 1s/step - loss: 0.5650 - accuracy: 0.7704 - val_loss: 0.7662 - val_accuracy:
0.7273
Epoch 7/10
63/63 [==============================] - 80s 1s/step - loss: 0.5449 - accuracy: 0.7873 - val_loss: 0.7481 - val_accuracy:
0.7636
Epoch 8/10
63/63 [==============================] - 80s 1s/step - loss: 0.5106 - accuracy: 0.7942 - val_loss: 0.6690 - val_accuracy:
0.8091
Epoch 9/10
63/63 [==============================] - 80s 1s/step - loss: 0.4619 - accuracy: 0.8270 - val_loss: 0.7064 - val_accuracy:
0.8000
Epoch 10/10
63/63 [==============================] - 81s 1s/step - loss: 0.4573 - accuracy: 0.8340 - val_loss: 0.7578 - val_accuracy:
0.7818
```

Figure 10: Results of validation_accuracy

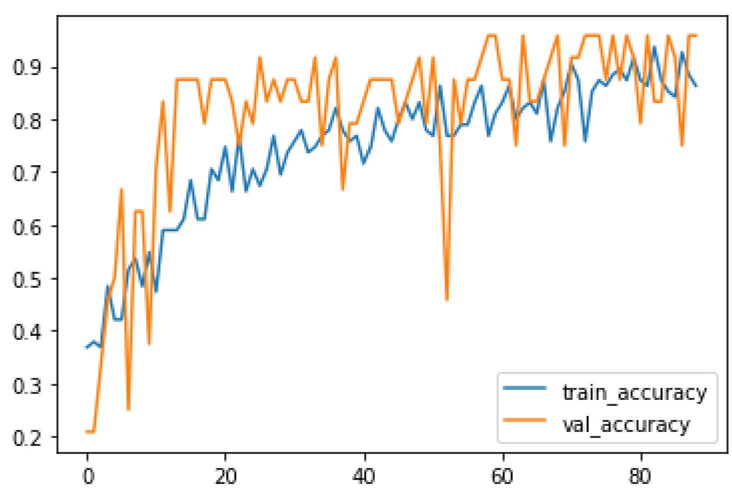

Figure 11: Results of training and validation_accuracy

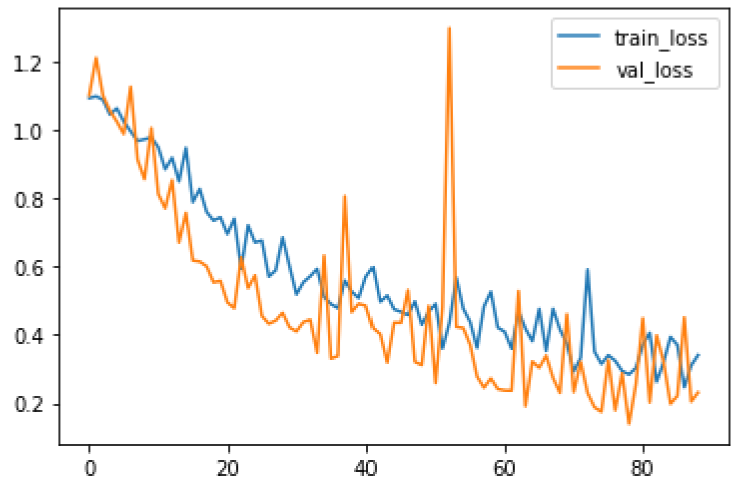

Figure 12: Results of training and validation_loss

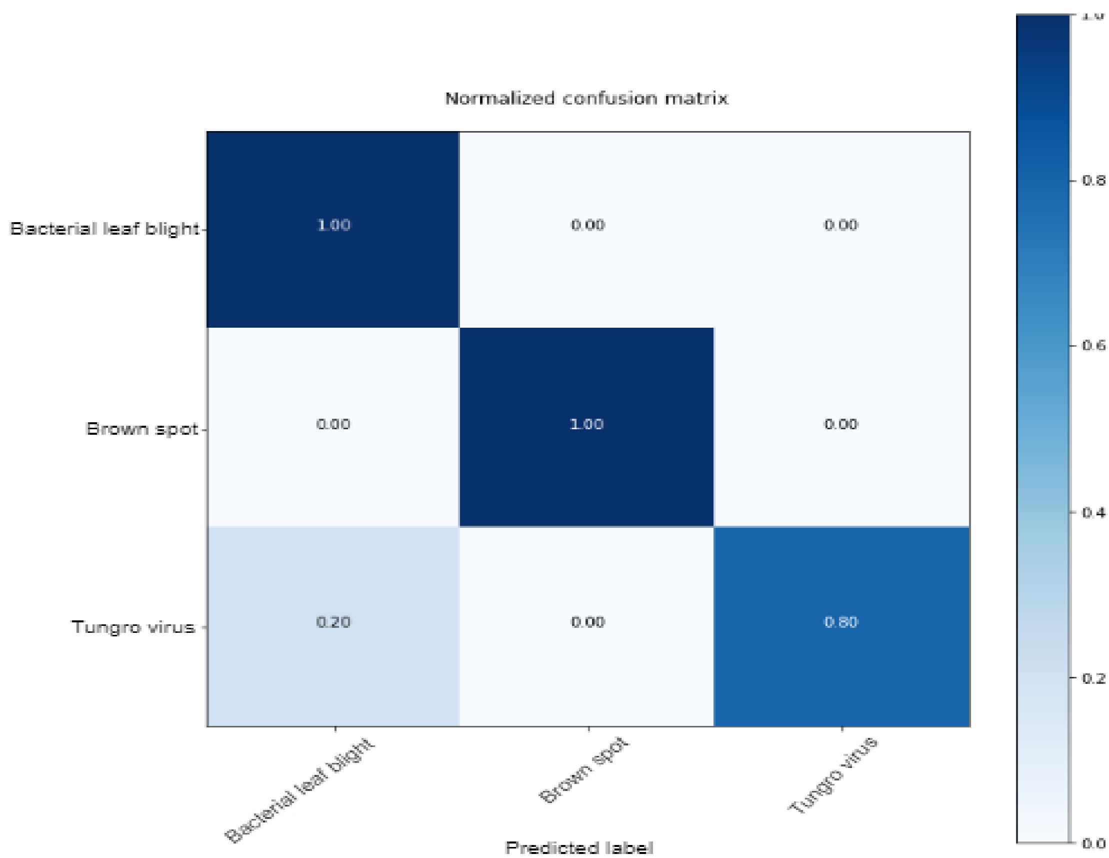

Figure 13: Confusion Matrix

CHAPTER 5
CONCLUSION AND LIMITATIONS

5.1 CONCLUSION

In this work, we did the categorization of several rice leaf illnesses utilizing a few DL approaches for four rice leaf diseases. We utilized a dataset of different rice leaves with diseases, which we then analyzed using a variety of well-known deep learning techniques, including VGG19, VGG16, Xception, Resnet, as well as a custom 5-layer convolutional network.
We found that among all of them, the 5-layer convolutional network performs the best at detecting rice leaves. We may infer from Fig. 6 that our suggested 5-layer CNN model has around 6% greater accuracy than the other common deep learning models.
Additionally, we found that by modifying training parameters like learning rate, epochs, and optimizer techniques, we could get a substantial level of accuracy with a custom model that had less layers than other conventional models. The easier it is for farmers to preserve their crops, the better we will be able to identify illnesses. In the future, we'll expand the horizon to include new illnesses and algorithms, which will vastly improve, accelerate, and simplify disease identification.

5.1 LIMITATIONS

Overfitting was a significant challenge for us in this project. When models perform exceptionally well on the training sample but badly on the testing sample, overfitting has occurred. High variance and low bias are the results. The training photos are fine adjusted with minimal background noise, which is one of the causes of this issue. The training phase used 80% of the training pictures, the testing phase 10%, and the validation phase 10%. The photos in each picture came from the same collection of carefully edited photographs. With ResNet-152, GoogleNet, VGG, and CNN, this produced results with extremely high accuracy. We have utilized distinct photos for validation and have optimized the image augmentation parameters to get around this issue. Both GoogleNet and ResNet-152 are incredibly effective in predicting the result.

6. REFERENCES

[1] W.j. Liang, H. Zhang, G.-f. Zhang, and H.-x. Cao, “Rice Blast disease recognition using a Deep Convolutional Neural Network,” 2019.

[2] Pallapothala Tejaswini et al 2022 IOP Conf. Ser.: Earth Environ. Sci. 1032 012017 “Rice Leaf Disease Classification Using Cnn"

[3] S. Ramesh, D.Vydeki - “Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm” on 28 August 2019.

[4] Latif, G.; Abdelhamid, S.E.; Mallouhy, R.E.; Alghazo, J.; Kazimi, Z.A. Deep Learning Utilization in Agriculture: Detection of Rice Plant Diseases Using an Improved CNN Model. Plants 2022,

[5] Naresh Cherukuri,G.Ravi Kumar, Ongole Gandhi, Venkata Sai Krishna Thotakura and others - “Automated Classification of rice leaf disease using Deep Learning Approach” in 2021.

[6] Syed Mohammod Minhaz Hossain-”Rice Leaf Diseases Recognition Using Convolutional Neural Networks” January 2021.

[7] Shubham Raje - “Detecting Diseases in Rice Leaf Using Deep Learning and Machine Learning Techniques” 23rd September 2021 .

[8] 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.

[9] Chouhan, S. S., Singh, U. P., Sharma, U. and Jain, S. (2021). Leaf disease segmentation and classification of jatropha curcas l. and pongamia pinnata l. biofuel plants using computer vision based approaches, Measurement 171: 108796.

[10] Jiang, F., Lu, Y., Chen, Y., Cai, D. and Li, G. (2020). Image recognition of four rice leaf diseases based on deep learning and support vector machines, Computers and Electronics in Agriculture 179: 105824.

[11] Sethy, P. K., Barpanda, N. K., Rath, A. K. and Behera, S. K. (2020). Deep feature based rice leaf disease identification using support vector machine, Computers and Electronics in Agriculture 175: 105527.

[12] Yadav, S., Sengar, N., Singh, A., Singh, A. and Dutta, M. K. (2021). Identification of disease using deep learning and evaluation of bacteriosis in peach leaf, Ecological Informatics 61: 101247.

[13] M. Hossain, M. Ali, and D. Hossain, “Occurrence of blast disease in rice in Bangladesh”, American Journal of Agricultural Science, vol. 4, pp. 74–80, Aug. 2017.

[14] M. R. KOMAL BASHIR and M. BARI, “Detection and classification of rice diseases: An automated approach using textural features”, Mehran University Research Journal of Engineering Technology, vol. 38, no. 1, 2019.

[15] Mojjada, R. K., Kumar, K. K., Yadav, A. and Prasad, B. S. V. (2020). Detection of plant leaf disease using digital image processing, Materials Today: Proceedings .

[16] Hasan, M. J., Mahbub, S., Alom, M. S., & Nasim, M. A. Rice disease identification and classification by integrating support vector machines with deep convolutional neural networks. In 2019.

[17] Zekiwos, M., & Bruck, A. (2021). Deep Learning-Based Image Processing for Cotton Leaf Disease and Pest Diagnosis. Journal of Electrical and Computer Engineering, 2021.

[18] Ramesh, S., Hebbar, R., Niveditha, M., Pooja, R., Shashank, N., & Vinod, P. V. (2018, April). Plant disease detection using machine learning.

[19] Tian, L., Xue, B., Wang, Z., Li, D., Yao, X., Cao, Q., Zhu, Y., Cao, W. and Cheng, T. (2021). Spectroscopic detection of rice leaf blast infection from asymptomatic to mild stages with integrated machine learning and feature selection, Remote Sensing of Environment 257: 112350.

[20] S. Phadikar, J. Sil, and A. K. Das, “Rice diseases classification using feature selection and rule generation techniques,” Comput. Electron. Agric., vol. 90, pp. 76–85, 2013, doi: 10.1016/j.compag.2012.11.001.

7. ACCOUNT CLEARANCE