USING MACHINE LEARNING TECHNIQUES FOR RICE LEAF DISEASE DETECTION

BY

MD MARUF HASAN ASHIK

ID: 201-15-13662

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

DR. FIZAR AHMED
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

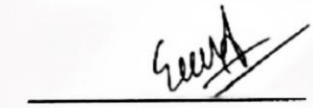
JANURAY 2024

APPROVAL

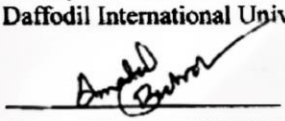
This Research titled **"USING MACHINE LEARNING TECHNIQUES FOR RICE LEAF DISEASE DETECTION"** submitted by **Md. Maruf Hasan Ashik** to the Department of Computer Science and Engineering, Faculty of Science and Information Technology, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science & Engineering and approved as to its style and contents. This Presentation has been held on 24 January 2024.

BOARD OF EXAMINERS

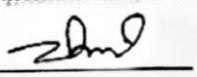
Narayan Ranjan Chakraborty (NRC)
Associate Professor & Associate Head — **Chairman**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Md. Sazzadur Ahamed(SZ)
Assistant Professor — **Internal Examiner**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Amatul Bushra Akhi (ABA)
Assistant Professor — **Internal Examiner**
Department of CSE
Faculty of Science & Information Technology
Daffodil International University



Dr. Md. Zulfiker Mahmud (ZM)
Associate Professor — **External Examiner**
Department of Computer Science and Engineering
Jagannath University

DECLARATION

We so certify that we conducted this study under the guidance of **Dr. Fizar Ahmed, an associate professor** of the Daffodil International University's Department of Computer Science and Engineering, Faculty of Science and Information Technology. I further declare that no portion of my study, nor any portion of it, has been submitted for consideration for a degree elsewhere.

Supervised by:

06.02.2024

Dr. Fizar Ahmed
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Md. Maruf Hasan Ashik

Md Maruf Hasan Ashik
ID: 201-15-13662
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

Firstly, let me begin by sincerely thanking the Almighty God for His divine favor, which has enabled us to successfully finish the final year thesis.

I would want to express my sincere gratitude and debt of gratitude to my supervisor, **Dr. Fizar Ahmed**, an associate professor in the CSE department of Daffodil International University in Dhaka's Faculty of Science and Information Technology. In the end, I finished my work on the " **Using Machine Learning Techniques For Rice Leaf Disease Detection**" thanks to his profound knowledge, strong attention, and encouraging guidance. It would not have been possible to finish this job without their unending patience, scholarly advise, consistent support, energetic oversight, constructive criticism, insightful counsel, and reading several subpar versions and editing them at every step.

I would also like to extend my sincere gratitude to the other faculty members and the staff of the CSE department at Daffodil International University for their invaluable support and guidance in helping me to complete my project. **Dr. Fizar Ahmed, Associate Professor**, and **Professor Dr. Sheak Rashed Haider Noori,** Professor and Head, Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, are both really appreciated for their valuable advice and support.

Lastly, once again, I would want to express my gratitude to all of my supporters, friends, family, and elders for their encouragement and support. Work hard and thanks to everyone who inspired and helped with this study.

Finally, I would want to respectfully thank my parents for their unwavering support and patience.

ABSTRACT

In order to prevent global warming on Earth, one of the most important resources is plants. Nonetheless, a multitude of ailments plague the plants. Just recently has study begun on the identification of plant diseases. To identify rice sickness is the major purpose of this article. Brown Spots Diseases, Leaf Blast Disease, and Leaf Blight Disease are a few illnesses that can affect rice. at different stages of growth. These infections impede the rice's ability to spread and protect its whole plant. Three different disease kinds were examined in this study along with one set of healthful rice leaves. Many different species, including fungi, bacteria, and other microorganisms, can cause rice disease. By cutting down on the amount of time required to ascertain the effect of rice leaf illness on humans, the technique was designed to consequently eliminate noise and produce the best results for leaf disease identification using ML with a greatest level of accuracy. This was achieved by applying ML techniques, including a computer-based detection approach. K-Fold validation procedures were used to measure the classification of this study. Random Forests, Decision Trees, Logistic Regression, and other support vector classifiers (SVCs) were trained with four classes of rice leaves. When K-fold cross validation tactics were applied to forecast 3 types rice leaf disease using one class of normal rice leaf, Random Forest produced the highest accuracy of 94.16%. At last, rice leaf detection through classification is accomplished using the CNN InceptionV3 model.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

Most plant diseases are harmful to businesses of all kinds. The plants can help with almost every disaster, but they may also get sick from a variety of illnesses. Numerous illnesses can impede the development of plants. The spread of several plant diseases in recent years may have an effect on agriculture as a whole. When picking the greatest Fruit & Food crops, farmers have a lot of options to choose from. All types of diseases have emerged in both disorder and illness. Growers may choose the best pesticides for the plants that they cultivate from a wide range of cultivars. The word "plant disturbance" refers to situations in which Soil problems cause routine functions of plants to be disturbed or misled. as a whole, external pressures, or other internal factors. Plant upheaval is the term for this. Regarding plant diseases, the virus is unable to spread from a sick plant to a good one. Microbes, viruses, and fungus are among the several diseases of plants that have been identified documented to be able to harm various plant parts from above to below ground. In the agricultural sector, researchers are exploring the fascinating topic of photo diagnosis. This work has led to the development of machine learning for rice leaf disease identification. Leaf diseases may be detected using image processing techniques.

Many machine learning techniques may be applied to identify the rice-related sickness. Computation image processing and ML offer several ways for most phases of a workflow sequence, and the proposed model encompasses photo capture, preprocessing, segmentation, feature extraction, and classification. To compare images that may be used to accurately identify rice illnesses, these steps are required. The first step in training the computer is to obtain and scan tainted images in JPG, PNG, and JPEG formats. The initial step of the technique is obtaining images of rice diseased leaves. The file type for the power source jpeg extension was then changed.

This essay looks at diseases including blight, brown spot, and rice blast. Following picture recognition, a variety of processing techniques are applied to tackle the particular difficulties presented by the field sick leaf picture in order to achieve a number of goals. The following objectives can be accomplished with the help of analysis:

• To recognize ill days.
• To ascertain the illness's full effects.

Once the disease images have been read and gathered, the red, blue, green, and red (RGB) photos are transformed into grayscale photographs in the image processing system utilizing a color conversion tool. Contrast manipulation is one of the brightness-enhancing techniques used to boost the contrast. Additionally, the work of preparing pictures has removed the same kind of images that show disease. The picture is segmented once preprocessing is complete in order to simplify it and make it possible for each component to go through further processing or analysis. Furthermore, it has been recommended to use K-means clustering techniques. Lastly, there are image classification methods that identify images of disease. For image recognition, ml employs a range of models. The diseases of rice leaves may be measured with the use of these classes.

1.2 Objectives

The modern era is distinguished by rapid technological development. It is possible for technology to resolve every problem. As a result, a number of scientific techniques have been used to advance the agricultural sector. Plant disease has been the biggest threat to the farming industry in recent years. There are several diseases that may affect plants. In order to identify illnesses in rice leaves, this study mainly used ML techniques. My original thought is to correctly detect different diseases in order to predict the sickness leaf. In addition to healthy leaves, there are several disease classifications, including blight, blast, and brown spot. My plan is to utilize ml of processing techniques to difference the rice disease. Consequently, I may rank these objectives as follows:

- To assist farmers and add to my economy's agriculture sector.
- A rice plant's health may be determined by examining its leaves.
- To use ML techniques to Prediction and classify too kinds of ill.
- To obtain data pertinent of illness prognosis.
- Acquiring thinghs about every aspect of the ml industry.
- Applying diverse methodologies to enhance outcomes.

- Need identify specific ailments affecting rice leaves.

1.3 Motivation

I wanted to support farmers because they provide my food source mostly. I was inspired to work in agriculture since I was raised in a rural area. I use DL to detect rice diseases and ML, DL, and AI to build machine learning systems to help farmers. Before I attempted to come across any concepts that would meet my research needs, everyone was debating what I should write on for my essay topic. So, I talked to one of my great teachers. She was happy that I decided to study agriculture, and she recommended that I look into paddy-related things because maize is one among her main source crops. This is why I chose the topic " **Using Machine Learning Techniques For Rice Leaf Disease Detection Creating Web Application** " for my essay. In addition, I observe that modern society is employing cutting-edge technology to enhance the farming industry, and a growing body of research is being conducted on this sectors to help my farming business. My curiosity in carrying out this type of research was piqued by this. Machine learning techniques have a tight relationship with human reality.

1.4 The study's justification

Numerous investigations in the domains of entity categorization and image analysis have definitely been conducted. But there is aren't many completed papers on the subject of " **Using Machine Learning Techniques For Rice Leaf Disease Detection Creating Web Application.**" As a result, my study uses a variety of computing and classification techniques. Since I have a clear plan in place to complete my objective, I have established my own interpretation and prediction for this issue to keep the work going.

Image processing is a multifaceted and intricate approach including several fields such as compression, restoration, metric extraction, and picture enhancement. It helps to lower the memory requirements for storing an electronic image. Images might have errors. Errors during the digitization process might result in defects in photographs. Image enhance methods it be useful to recover bad images. Moreover, classifiers and ml algorithms may be used to identify them.

1.5 Research Question

This study has required a considerable lot of commitment and labor to complete. This was a really tough job for me to finish. It is difficult to find a workable, precise, and acceptable solution to the problem. To address this issue and have a conversation about these concepts, the investigators would want to ask the following questions:

- Can I obtain raw image data for research purposes?
- Is the raw image data pertaining to rice illness being labeled by an Administrative Officer?
- Is it feasible to use a machine technique for processing the raw data?
- Can the agriculture industries be improved by these techniques and approaches?
- How may these initiatives and approaches benefit farmers?

1.6 Expected Outcome

This section contains many facts, assuming that they were the least anticipated outcomes. Several categorization techniques have been used to forecast the actual impact of the leaf disease based on the results of the follow-up study. The anticipated outcome of this scientific research is the development of a technique or thorough, efficient procedure that will categorize rice leaf disease in line with the model created for training and evaluating the dataset. As a consequence, the anticipated outcomes are listed as follows:

- Once the rice leaf has been examined, I will show you if it is sick.
- Improved think of rice disease prediction using ML.
- The segment pertaining to recognizing undernourished foliage.
- By comparing my results with those of earlier research, I wish to do so using old image data.
- Choosing the optimal machine learning technique based on the available data.

1.7 Layout of the Report

The first part included the presentation of thinking goals, objectives, study challenges, and anticipated study outcomes. The overall structure of the report is also covered in this section.

Chapter 2 covers what is being done in this field at the moment. The breadth that follows in the latter part of this second segment, as a result of their limitation of this field. The primary obstacles or challenges this study has faced are addressed in the concluding section. This section's remaining content addresses a variety of subjects, such as related studies, a summary of the study, and the challenges encountered while developing the project.

The conceptualization of this research endeavor is explained in the third part. To address the theoretical component of the study, this chapter provides further details on the statistical methods that were employed. This chapter also demonstrates the usage of ml approaches with procedural procedures. The next chapter describes the process for obtaining datasets and the data preparation system. Impedance matrices are also introduced in the later portion of this branch in order to assess the model and show the appropriate tag of the prediction. When using the ml and dl techniques, deployment analysis must also be included to ensure genuine accuracy. The section that follows covers a wide range of topics, including the study's topic and instruments, workflow, data collection process, handling of information, suggested structure, method of production, and the operational requirements that must be met in order for this project to move forward. There is a thorough explanation provided for each step of the ML techniques and classifiers used in this study.

The exam results, performance assessment, and explanation of outcomes are included inside Chapter 4. A few test photos are included in this chapter to aid in the project's implementation. An analysis of the outcomes utilizing ml algorithms wraps up this chapter. An online prototype for identifying the existence of a specific rice leaf virus was made using a different DL.

Chapters 5 and 6 provided an overview of the entire research effort, information on the next events, and a conclusion. This chapter provides a clear illustration of how the work report conforms with

the criteria throughout. Impact on the community, the environment, and sustainable development: The paper's conclusion highlights my efforts' shortcomings, which may have an impact on future workers who hope to enter this sector.

CHAPTER 2

Background

2.1 Introduction

The study's findings, challenges, pertinent literature, and research summary are all included in this section choice. I'll examine research articles written by other writers and how their methodologies and conceptual correctness relate to each other in the section entitled "Associated Works." In the section on similar works, I'll discuss the articles, approaches, and reliability of other academic publications that are relevant to my study. In the research summary unit, a summary of my associated work will be presented. I discuss the ways in which I overcame each obstacle encountered during the investigation and how I enhanced the precision of each layer during the demanding stage. All of this has previously been covered.

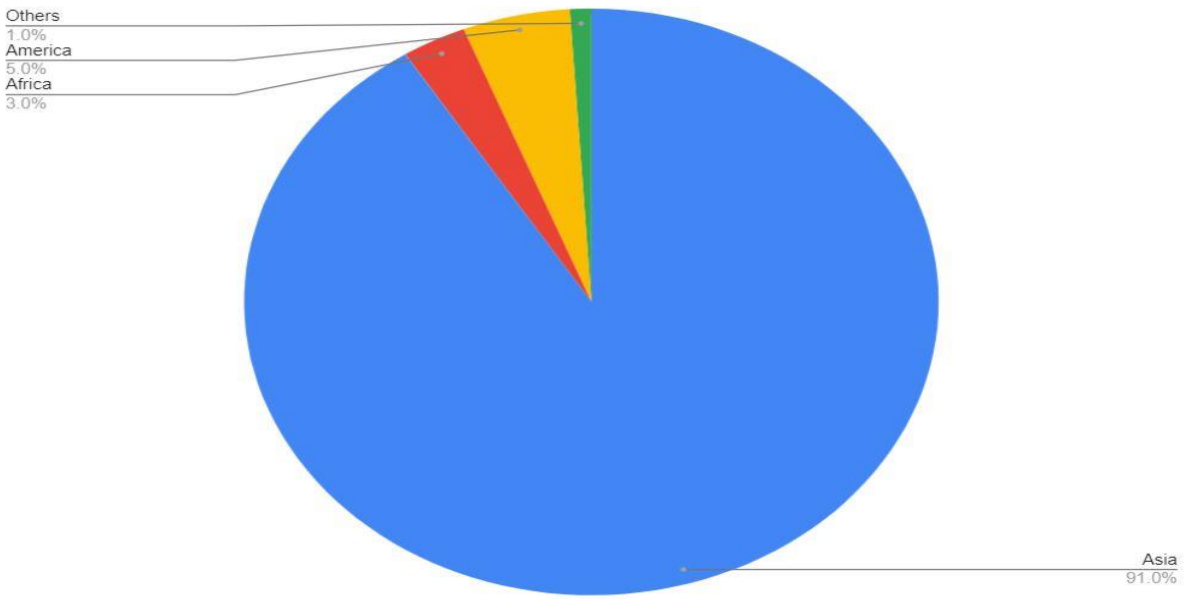


Fig 2.1: The productivity of paddy worldwide

In addition, rice is second-most widely planted crop in the world and a key source of sustenance for people who live in rural regions. The two main grain types, Japonica and Indica, are members of the Panacea group of plants. Rice was inexpensive, quite popular, and nutrient-dense across Asia. The five regions of the globe where rice is grown are Asia, America, Africa, Europe, and

Oceania. According to a report by the UN's Food and Agriculture Development Agency (FAOSTAT), Asia accounts for 91% of the world's rice production and consumption. The remaining rice is distributed around the globe. Europe gave 0.67 percent, Pacific supplied 0.15 percent, America contributed 5.19%, and Africa contributed 3.5%.

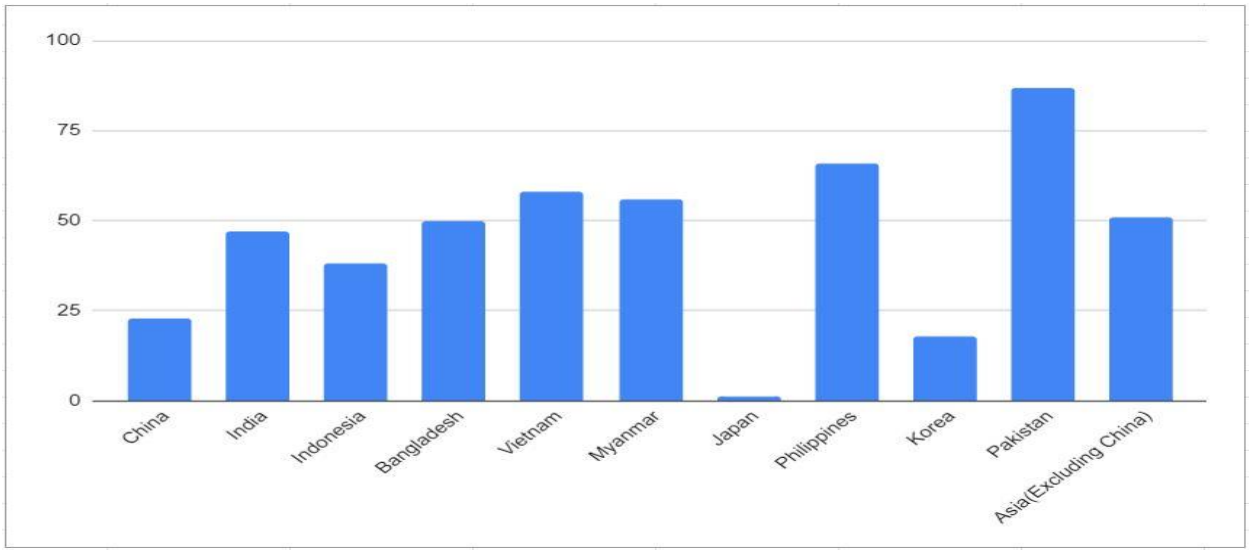


Fig 2.2: Projected rice consumption in Asian nations as a percentage of GDP, 1995–2025.
(Source: World Bank, Demographic Telepathy, 1995–2025)

According to World Bank projections, there will have been a 51% rise in the demand for wheat demand by 2025, which is now significantly higher than the pace of population growth. Figure 2.2 above shows a rice intake prediction for the major Asian nations for the years 1995 through 2015.

2.2 Literature Review

Numerous works have already created by a business. often useing in the agricultural sector to analyze illnesses affecting rice plants. Generally speaking, the agricultural sector has worked very hard to discover illnesses that affect rice plants. Numerous research investigations Studies on various rice ailments have been carried out previously. Nowadays, a variety of rice disorders and their therapies are the subject of research. The two main classifications of illnesses which Sahayadhas A [1] employed clustering K-means techniques to evaluate were brown spot and leaf blight brought on by bacteria. These two courses largely focused on dividing the k-means

groupings that constituted the major disease leaf of the paddy plant in order to complete their assignment.

Daffodil International University's Md. Ashiqul Islam et al. [2] published a study on DL using CNNs in 2001. The aim of the study was to precisely detect and categorize illnesses affecting rice leaves. Five leaf classes have been examined in this article, including one typical leaf class and four different disease kinds. Prior to applying deep CNN techniques with classification and optimal result dependability for prediction, images must first be captured and prepared.

Otsu's methodology, which classified the three disease of rice classes by using ml for learning using SVM method was the main focus of Minu Eliz Pothen & Dr. Maya L. Pai [3]. This study paper, which was published just in 2020, explores the splitting of two SVM algorithm approaches, HOG and LBP, utilizing the division region as a basis. The function of the kernel of harmonics may be used to detect a number of illnesses on the leaf surfaces of various plants, especially rice. The current study yields the highest precision in the SVM approach for the residual division characteristics.

Many illnesses that damage rice are already costing producers a great deal of money. The effective assessment and treatment of different rice disorders was the main emphasis of R.P. Narada and G. Arulvadivu [4]. There are three categories of rice ill leaves that may be employed for rice leaf recognition and identification using techniques. Basically, a variety of classification characteristics may be used, with SVM, ANN, and fuzzy classification being just a few options. information that was gathered from both direct and online sources.

Recently, and even now, there is a lot of interest in AI-based rice disease detection. The study of S. Pavithara et al. [5] on using AI technology to diagnosis agricultural disease has drawn many of thus far and is still receiving attention. How to identify and categorize rice disease with the use of contemporary image processing techniques. They use four steps to finish their work: obtaining identification evidence, preprocessing photos, feature extraction, and categorization. KNN and SVM learners has used for rice illness monitoring of identification. The final things of here study

has rice brown spot virus. Using SVM, it were able to get the maximum classification accuracy rate of 95%.

2.3 Research summary

My project endeavor's main focus is on the range of options available in the area. With my special dataset, I was able to apply extra strategies on top of the four ways I employed overall. In this instance, my data collection from the field and the World Wide Web serves as the main source of information. As previously mentioned, the data in my collection includes both recently collected and previously utilized data. I'll be able to evaluate the precision of each of the four methods I employed, as well as factors such as the impact of more data I provided from the original source. The recently added data and the previously combined dataset are identical. Given that the same types and categories of tags are present, it indicates that. Python served as my primary language of choice, and I employed ML techniques to develop my feature extraction strategies. creating a web application that precisely uses DL to identify rice leaf rot.

2.4 Challenges

Since this study is using photo data, managing it found to be too challenging, hence the main challenge is collecting the data along with to processing it. I had to make many journeys to the government's agricultural office to the fields in order to gather data because it was rather tough to get information on this topic. I used several tools and processes to clean and organize what information I had. The program required a long time to process the massive information sets with several layers and variable epoch scopes. I thus required waiting a long time for the verdict. Although there had been other data sets relating to this sector, they did not accurately reflect my knowledge after several tests and government-level experience on fields, so I had to obtain information from genuine fields. Since I wasn't performing a lot of study-related research, I had to work really hard to determine the best ways to complete the task swiftly.

CHAPTER 3

Research Methodology

3.1 Introduction

The second section provides a detailed description of the approach and methods used to classify the agricultural diseases of foliage that we have looked at. The main elements are the data collection and analysis, as well as the proposed model, which is further clarified by the relevant a system that chart, tabletop, and explain. Using my own raw dataset, it used the ml classification framework for division and prediction that produced the best accuracy to the investigation. I wrapped up the chapter by summarizing the statistical assumptions that I used. In this investigation, I used four different class types to construct my representations. Although there are numerous varieties of rice illnesses, I focused my research on three main disease classes: leaf blast, leaf blight, and brown spot. Healthy leaf is a distinct type of rice sickness. In my study, all of those involved were trained utilizing all of the photographs via a total of four kinds of training sessions.

3.2 Subject of Study and Equipment

Study areas that have been looked at and researched to help make concepts more clear are known as inquiry topics. Not only for accomplishment but also for accomplishment, data gathering, model creating, model manage, and template teaching. I discuss my tools and techniques in the subject of measuring. The Python code and Windows operating system were combined with NumPy as SkLearn, OpenCV as a, and other packages. Google's Co Lab technology is being used for testing and training purposes only. On Google's Colab, Python programmers may create applications for ML and DL methods as well as data mining.

Used libraries of all kinds:

- **Matplotlib:** Pyplot is a set of functions that is part of Matplotlib, which is a utility for creating charts. It assists you, among other things, when making forms, in establishing the limits of a plan or finding lines inside a plot.

- **NumPy:** A well-liked tool for handling vectors utilizes the NumPy Scripting library. It goes over the fundamentals of Fourier change, matrix processing, and linear algebra are included. In order to facilitate working with matrices of different lengths, the NumPy mathematical library for Python offers methods and functions. NumPy makes it possible to apply arrays in a logical and scientific manner. In short, NumPy is a numerical computation package for Python. This phrase, which is also known as "a bunch of Python," is utilized.

- **Scikit-learn:** This tool is useful and easy to use for prognostic data analysis. Anyone may use it without restriction and customize it to meet their own needs. Throughout development, Matplotlib, SciPy, and NumPy were utilized.

- **Seaborn:** This Python visualization utility is widely recognized for its compatibility with matplotlib. It is a simple-to-use tool for making insightful and eye-catching data visualizations.

- **CV2:** OpenCV-Python is a set of Python bindings designed to handle automated vision issues. It enables the analysis of images and videos to identify people, things, and even handwriting in human.

- **Joblib:** Joblib provides a more effective way to prevent making the same mistakes. calculation more than once, which might result in substantial time and cost savings.

- **H5py:** The h5py package offers a Python containment for the raw HDF5 data. Managing and storing enormous amounts of academic data in HDF5 is made simple with NumPy.

- **OS:** Developers may communication with the operating system of the computer they work on by using the many tools that the Python OS unit provides.

3.3 The process of work

There are several things or approaches that might be used to determine the data analysis in this study. The current research's workflow consists of several steps, including selecting a model, scaling and improving, processing, and data gathering.

Step 1: Data Collection: To construct my own data, I analyzed actual field data that I had obtained. Since it is challenging to categorize the dataset and gather data for the specific rice illness, there isn't a sizable, comprehensive dataset accessible in this subject.

Step 2: Preprocessing: Every sort of data was acquired in its unprocessed state from diverse sources and managed independently. Numerous data sets may have errors and noise. I physically process the selected data set before moving on to the next phase.

Step 3: Data Reduction and Enhancement: The data have remained increased and resized following the processing of each class. To get it to function, I had to add data and resize. Because I was worried about overfitting, I limited the amount of alterations I made to the largest and most moderate.

Step 4: Model choosing: Choose a model, use the data to train and test it. I've given to improve accuracy. In ML and DL, many classifiers are employed. Before deciding on one design to employ for information accuracy calculations, several designs were tried out to improve accuracy using my equipment setup.

Step 5: Assessment performance: This section covers all of the findings. After testing and training, these strategies gave us only a certain level of reliability for the following four courses. The confusion matrix and a graphic showing the recall rate, efficacy, and f1 were also computed measurement. is an internet resource for diagnosing rice leaf disease.

Step 6: Conclusion and future scope: The develop schedule and final are given in the all part.

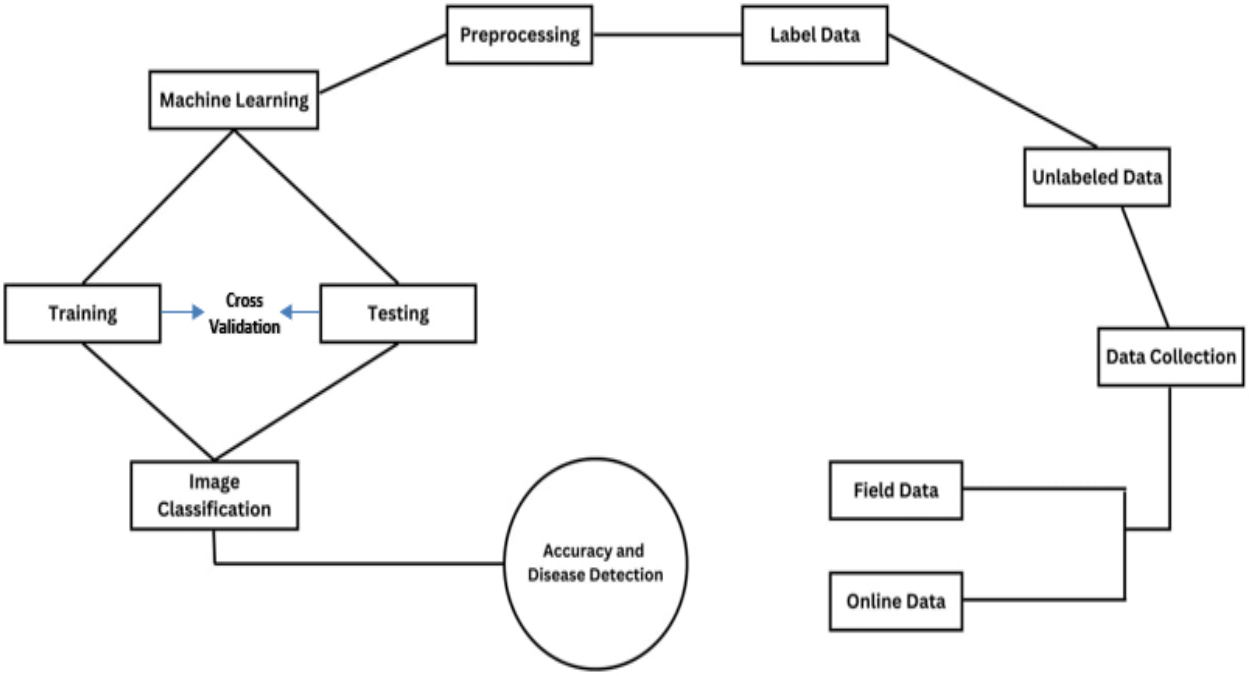


Fig 3.1: Model recommended for the full research project.

The fundamental method to the classification of leaf disease is shown in Figure 3.1. The initial raw data from the rice fields and the digital data from the internet are merged to form a dataset. The agricultural officer then annotated and classified the shot, adjusting its proportions and information. After then, this data might be obtained by the equipment. I may make use of this data to test my proposed machine learning techniques on new datasets. Because the procedure is clear, I will be able to achieve my aim of using ML models to diagnose rice sickness with the highest level of efficiency.

3.4 Procedure for Gathering Data

I've put up a series of 600 photos. Four classes are identified in the dataset based on the kind of illness; one class has 150 pictures and features a leaf that is in good health. The categories included Healthy Leaf, Leaf Blight, Leaf Blast, and Brown Spot.

I trained using 480 photographs, or 80% of the dataset, and tested using 20% of the images, or 120 shots. Each class in the test data collection has roughly 30 photographs, while each class in the initially created dataset contains about 120 photos.

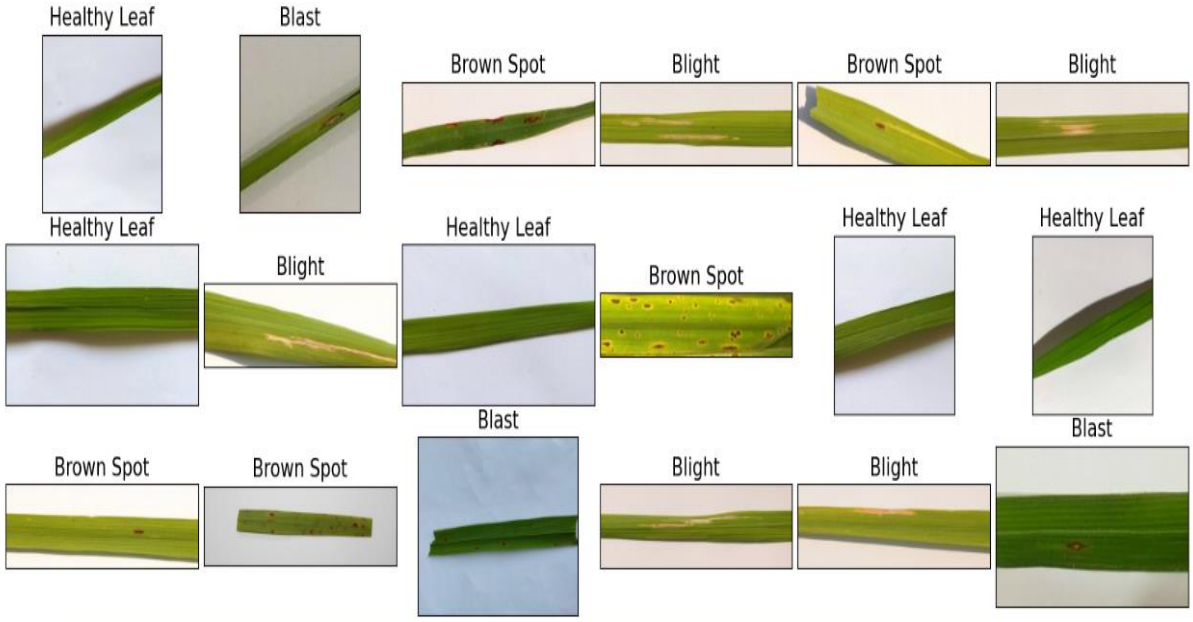


Fig 3.2: Sample of classes: Healthy Leaf, Blast, Blight, and Brown Spot.

Table 3.1: Table of dataset

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

Labeled Maker:

I frequently work with datasets in ML that has big number of labels, either in just one row or several columns. This label can be sentence or numbers. In an attempt to make the data more legible by humans, sentence is frequently used to label the data that is learned.

The act of translating tags into an alphanumeric form that an electronic device can comprehend is known as label encoding. This process includes converting the labels into a numerical format. Ultimately, machine learning algorithms might decide whether to use these labels or not. It is important to finish this phase of the dataset design for supervised training.

3.5 Analytical Statistics

3.5.1 All Data manipulation

The majority of information becomes the result of data processing. The data processing approach used on a data collection is important. Data that has been processed is a significant help when working with raw data. I'm going into rice fields for this study in order to get information on sick leaf rows. After that, the data was merged with information from online sources to produce a robust dataset with four categories. Pretreatment of the data is frequently critical to the effectiveness of dataset modification. More adept preparation of the data will yield more accurate results. A data handling system consists of two phases: data collection and data reinforcement. In other words, it is the primary obstacle for this kind of study-based job.

I. **The prepare of data and build:** The widths and heights of each photograph in my collection were collected from a few different internet sources in addition to actual field data. As every image in my model needs to have a specific pixel size, I used a reform script to resize the data to 500 × 500 pixels, which is a consistent size. Furthermore, every image in my model has been preprocessed by appending a "jpg" suffix. Along with segmenting and preparing the images for categorization, I also transformed the pictures to RGB, HSV, and shades of gray. As a result, I taught the framework using the segmented form of the entire datasets.

- Fixed-size on codes.
- Format extension to ".jpg".
- Get rid of inaccurate pictures.
- Deleted redundant images.

Farm officers classified the unmarked without classes dataset first, then divided the data into four main illness groupings. The farming officer divided them into four classes: healthy leaf, brown patch, leaf blast, and leaf blight.

```
[ ]  # get the training labels
     train_classes = os.listdir(train_path)

     # sort the training labels
     train_classes.sort()
     print(train_classes)

     # empty lists to hold feature vectors and labels
     global_features = []
     labels          = []
```

```
['Blast', 'Blight', 'Brown Spot', 'Healthy Leaf']
```

```
[ ]  len(train_classes)
```

```
4
```

Fig 3.3: Rice leaf of Four types class.

3.5.2 Feature extraction

A visual feature, including color, shape, and texture, that has undergone extensive training and testing is called a characteristic extractor. Global characteristics extraction is also a part of this characteristic extraction process. In this study, global characteristics have been used for every category of pictures. Obtaining features enhances the accuracy of learnt models and expedites testing and training by removing some characteristics from the supplied data. After being captured in pixel values, the images were converted to rgb and hsv representations. It's crucial for determining enhanced accuracy as well. The picture's width and length play a major part in defining the design. One can quickly determine an image's height and breadth by counting the number of pixels in each component.

3.5.3 Image segmentation

In this work, four distinct class types were used to methods of processing images. In The K implies procedure has been employed in image processing techniques to create the picture categorization of those groups. K is an acronym for clustering, a popular machine learning technique. The algorithm for K-means clustering can produce units of linked occurrences. When it comes to the segmentation or categorization of photo sickness recognition in ML techniques, K-means regularly makes a significant impact. The afflicted region of the image can always be seen by the sick image according to ML techniques. For the sickness image data that needs to be arranged, some normalization is required. picture segmentation locates noises in a picture that degrade its quality. This approach is useful to delete noise and unnecessary things.

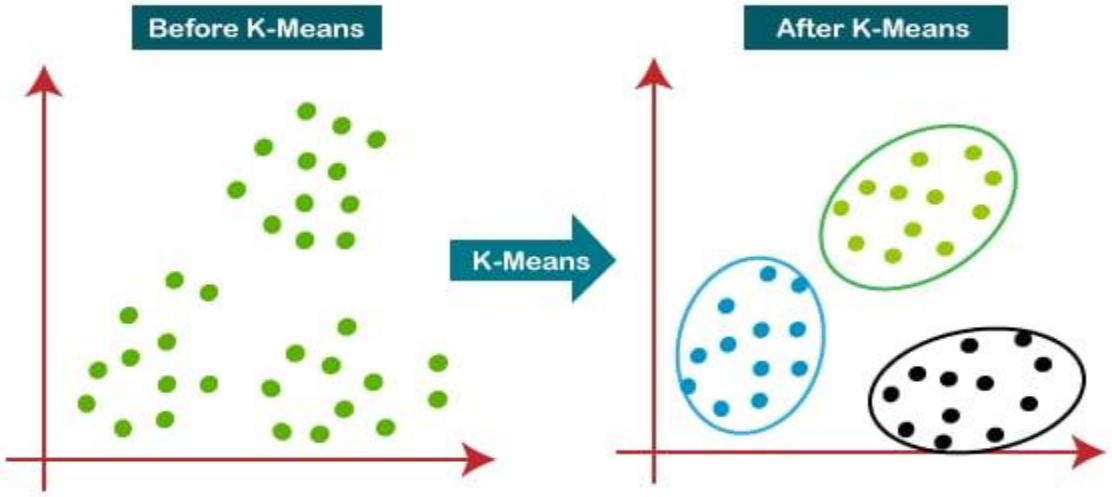


Fig 3.4: k-means cluster segmentation.

3.5.4 Train & Test

Of the most prominent ML activities is researching and creating algorithms that may forecast things on the basis of that knowledge and extract historical data. The algorithms in question use the input data to create an algorithm, that they then use to evaluate or extrapolate conclusions from what is provided in order to achieve their objectives. Before being utilized in the process of creating a model, these inputs are frequently divided into multiple data sets. In the process of constructing a model, 3 data sets are typically used: train, verification, and test. I train used 80% to initial dataset and tested it with the remaining 20%. Out of the 600 complete images in this data set, I captured 480 training and 120 testing images.

3.5.5 Applied Classifier

1. **Random Forest:** For both categories of categorization, use the tree-based approach provided by the RF Classifier. It is a procedure that creates an order according to trees in the context of ML. This method is used by AI to construct hierarchical "decision trees". Using an ensemble technique, the random forest detectors produce an enormous amount of decision trees, which are subsequently aggregated. This makes the issues with overfitting clear. ML has grown to be most talked about subjects in system today from it should be useable to any way you have a big data. Due of the many benefits that RF Classifier offers over alternative techniques, it is widely used. When the approach was 1st punlish in 1997, it has made it work with big datasets.

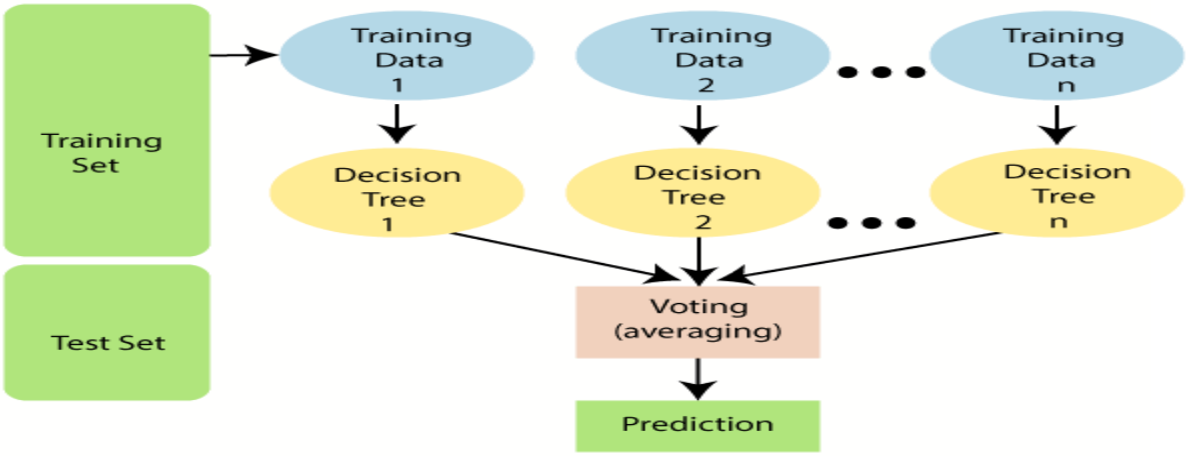


Fig 3.5: RF classifier workflow

Any classification-related issue may be managed as lengthy as each training info set has at least two sample sets and has labels for every possible class. The rf classification method is an unregulated classifier as it needs training sets that are supplied by a supervisor. once there are enough variables, the aforementioned method used to choosing, breaks down, and then uses the out assessment to provide suggestions about data.

2. **LR:** A supervised ml technique called LR is used to solve discrete class data issues in which of result is constant but dependent on the frequency of a particular kind of input. For integers between zero and one, the normal ideas of LR is demonstrated using a function. LR is one kind of classification method. By thought, a dual output of a situation which has actually just 2 potential results: it has either 1 additionally it is unlikely to 0. In a nutshell, distinct variables are any factors that might influence the research's outcome in addition to the variables that are dependent under investigation. As such, when dealing with data that is binary, logistic regression is the suitable analytical technique. You know When the outcome or deciding factor is binary or categorized you know you are working with binary data. You can also tell when the data belongs to one of two groups.

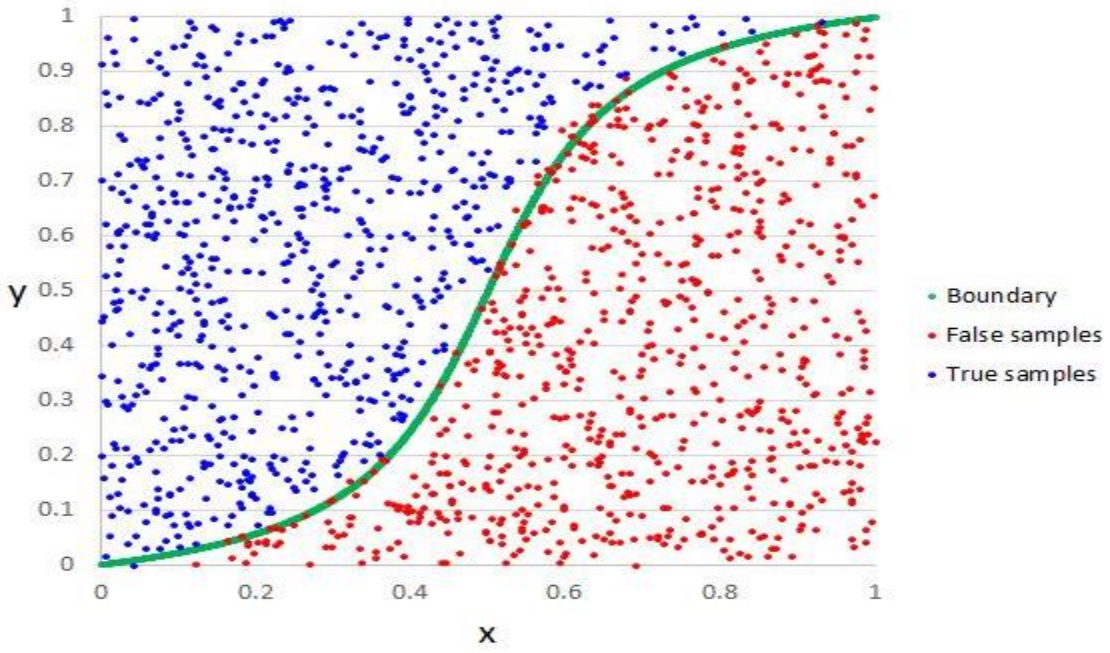


Fig 3.6: Logistic regression an example.

3. **DT:** This technique is directed ml is the DT. This approach may be used to tackle problems with regression and classification. This classifier is built on a tree structure, where each leaf node represents the result, internal nodes the dataset's attributes, and branching the classification rules. The making part with foundational part are the two parts make up a dt. The result of decision that using a choice's part, that usually has different branch, is

represented by a leaf part, that doesn't have Not another branch. To run examinations or make decisions, the features of the dataset are utilized.

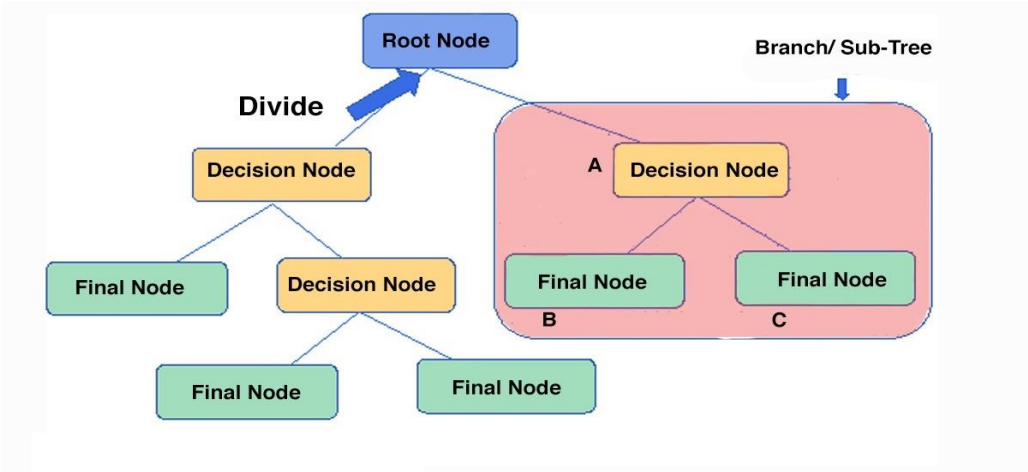


Fig 3.7: Decision trees Workflow

Because DT illustrate how individuals reason when making decisions, they are usually easy to understand. The decision tree's logic is easy to understand since it shows a structure similar to a tree.

4. **SVC:** SV Encoder is the name given to it. SVC is a stochastic approach of clustering by nature, even if it makes no assumptions about the underlying structure of the data, such as the amount of groupings and their respective sizes. As a result, you will often need to do a first processing step, such as the analysis of main components, if the amount of information is substantial. As other research have shown, it performs well with low-dimensional data. A number of modifications have been made to the original method, which offer particular methods for computing the clusters by examining only a portion of the border areas in the adjacency matrix. Many updates have been made available.

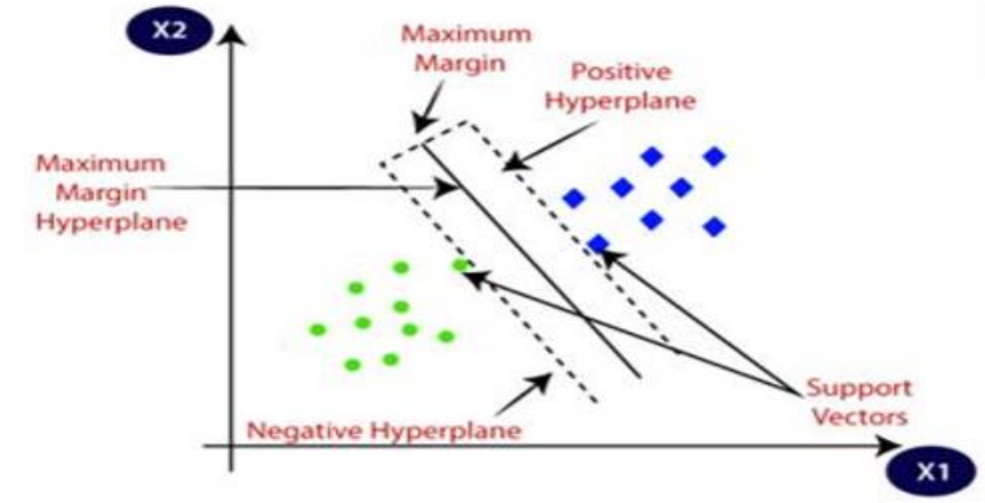


Fig 3.8: SVM Figure.

5. **InceptionV3:** The CNN framework InceptionV3 is used for picture classification and recognition. It belongs to the set of Inception DL frameworks, which the staff of the firm created in order to grow. The InceptionV3 system has become well-known for its intricate design and creative application of the "Starting section," a unique feature that boosts the system's overall accuracy and effectiveness. The InceptionV3 large neural network design consists of 48 layers. Convolutional layers, max-pooling layers, fully linked layers, and auxiliary classifiers are all combined. Networks are able to gather data at many levels of abstraction because to the Imagine module's temporal mixing and multilayer configurations (1x1, 3the dimensions of x3, and finally 5x5).

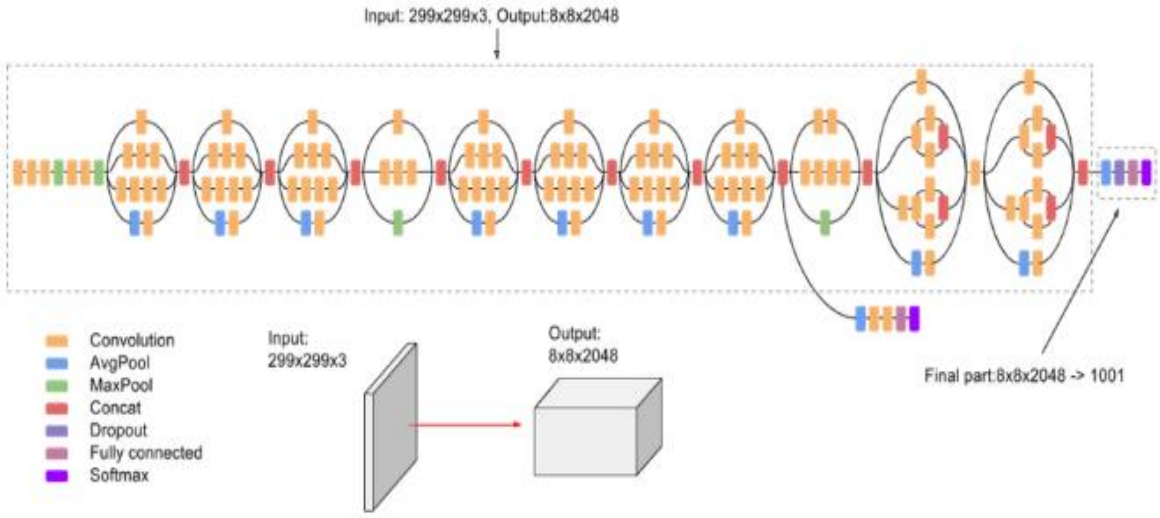


Fig 3.9: Inception V3 model full figure.

3.6 Implementation

The data set must be applied after any further tasks are finished in order to guarantee correctness. To carry out the execution, I divided my work across seven primary sections. These procedures must be taken in order to guarantee the fulfillment of my project.

- Collection data.
- Pre-processing images data.
- Feature Extraction.
- The Implementation of Algorithms
- CNN InceptionV3 model used to build web app for rice leaf disease detection.
- Evaluations of accuracy and outcomes.

I have to go to rice fields and take pictures of brown spots, leaf blast, leaf blight, and leaf health besides online in order to collect data. A top-notch dataset has been combined with these photos.

After that, I started working on the data preparation. Here, I removed any unnecessary components from the info I collected, such noise and erroneous pictures, out-of-scale images, and so on. I further use the extraction of features to make up for the lengthier data for the training and testing timeframes.

The first step in putting the concept into practice was for me to start experimenting with its code. I assessed the precision of four distinct algorithms that I employed. After the approach was finished, I evaluated its accuracy. After weighing the accuracy, I decided that one was going to be more appropriate for my needs. This has shown to be quite predictable in the case of leaf disease. A comprehensive A collection of prerequisites that are required for every attempt at picture classification has been developed as a result of an examination of all relevant theoretical and numerical methodologies and ideas. The subsequent outcomes could be required:

1. Hardware & Needs software

- Operating System: Windows 7 or later;
- Hard Drive: 1 TB or more;
- RAM: 4 GB or less

2. Tool Development Requirements

- Python Environment
- PyCharm.
- Google Colab.
- VS Code

CHAPTER 4

Experiment Results and Discussion

4.1 Introduction

This section explains the development process of the rice leaf disease categorization system. The whole process of building the model included gathering the photographs, processing the data, enhancing the data, modifying the amount of the data, recommending models, and providing instructions with model accuracy. The findings of my tests are shown and discussed in this chapter.

4.2 Experimental Result

The rice problem epidemic id detection using a variety techniques. I thus employed a variety of techniques during the procedure. I investigated and assessed a number of approaches before deciding on the most effective methodology for the research. I experimented with a number of methods to raise the standard of my work. I utilized two separate datasets: one from the internet and another gathered from an individual's raw gathering of data that an agricultural officer had categorized into four categories: Healthy, Leaf Blight, Leaf Explosion, Brown Spot, and Leaf Explosion. I made use of pre-existing Python tools, dictionaries, and content classification strategies. The illness technique used to evaluate the dataset determines the important outcome or disease reliability.

4.3 Using ML for Descriptive Analysis and CNN model for web prototype

Depending on the categorization methods, I got different results that employed. Four machine learning methods have been applied to accurately diagnose rice illness. I assessed the algorithmic means of the model's accuracy using cross-validation K-fold techniques. The cross-validation K-fold processes then resulted in a prediction result. I used Support Vector Machine (SVM) with classifier, Random Forest, Decision Tree, and Linear Regression in K-fold cross validation methods. Every version used the same information set after the final dataset was determined. This information set included both public ally accessible data and a dataset I created using the raw data.

Using Py & its pre-built libraries, I evaluated the accuracy once the dataset procedure was complete.

Table 4.1: Table os accuracy

Classifier	Accuracy Score
DT	86.83%
RF	94.16%
LR	85%
SVC	82.50%

The results of many technique are shown in that part. PyCharm & CoLab, free software packages, were utilized at every stage of the procedure. SVC, RF, DT simulations, and LR were the 4 classifiers utilized overall.

```
print(classification_report(testLabelsGlobal,y_predict))
```

```
              precision    recall  f1-score   support

           0       0.88      0.95      0.91        39
           1       1.00      0.96      0.98        27
           2       1.00      0.84      0.91        25
           3       0.94      1.00      0.97        29

    accuracy                           0.94       120
   macro avg       0.95      0.94      0.94       120
weighted avg       0.95      0.94      0.94       120
```

Fig 4.1: Random Forest's Classification Report.

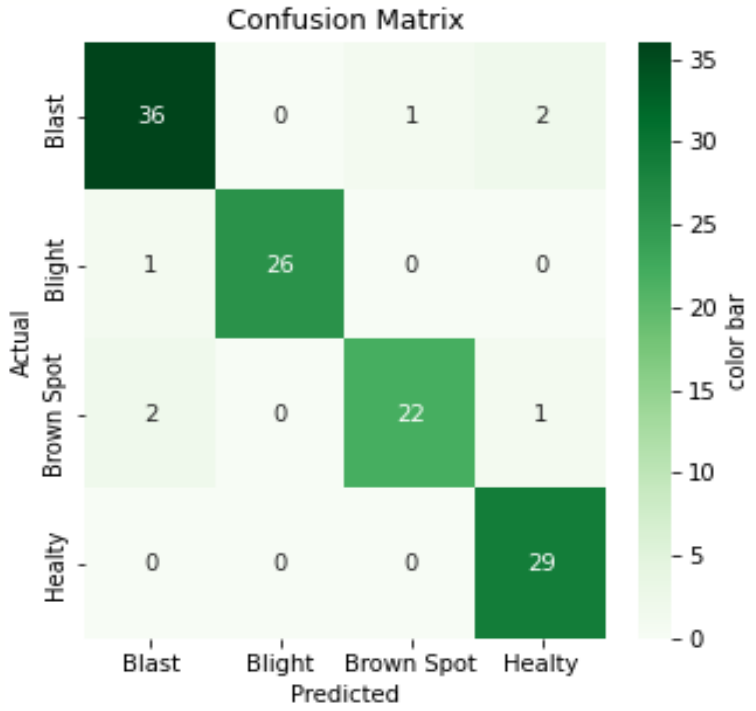


Fig 4.2: RF (Random forest's) confusion matrix.

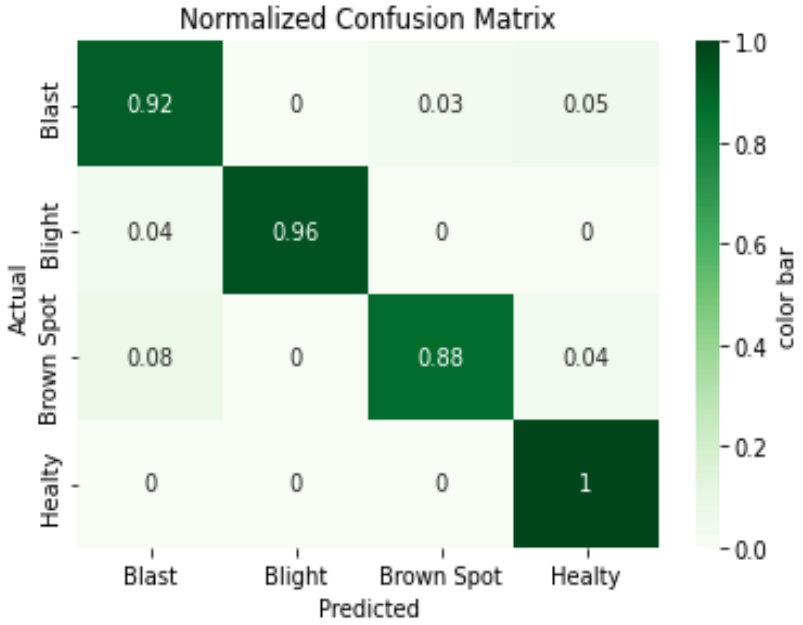


Fig 4.3: RF Normalization Confusion Matrix.

Simply showing the whole RF classification report & normalizing will get the greatest accuracy.

The procedure for determining a specific rice leaf classification and using the CNN InceptionV3 emulate to create a prototype web-based application is shown in Fig. 4.4 below. Figure 4.6, 4.7, 4.8, and 4.9 show the three forms of rice disease 4 class that were correctly predicted by a web application using InceptionV3. As is typically the case, deep learning is among the least supervised techniques for object detection or prediction.

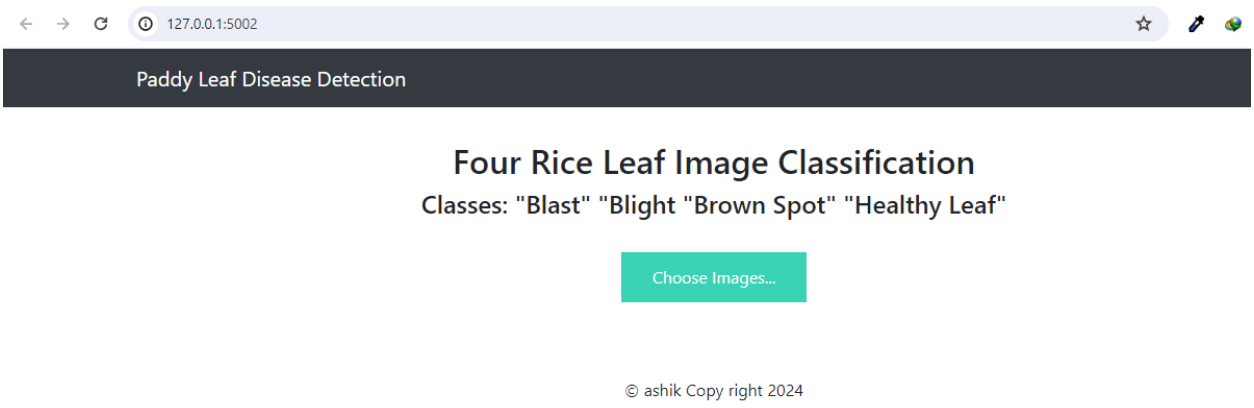


Fig 4.4: Web prototype for rice leaf classification.

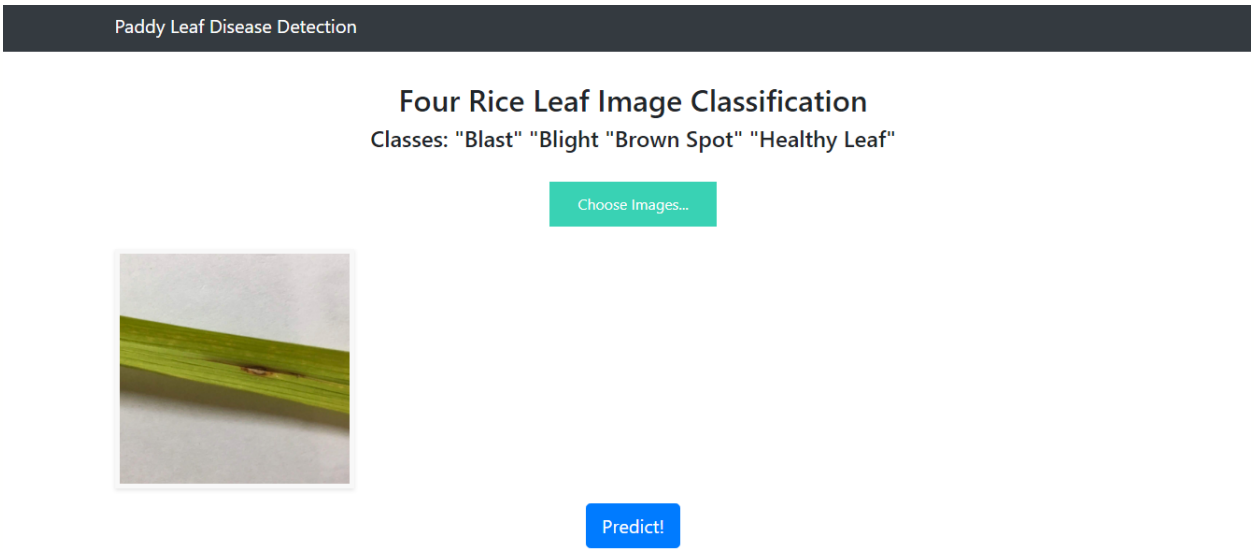


Fig 4.5: Selecting images of Web prototype.

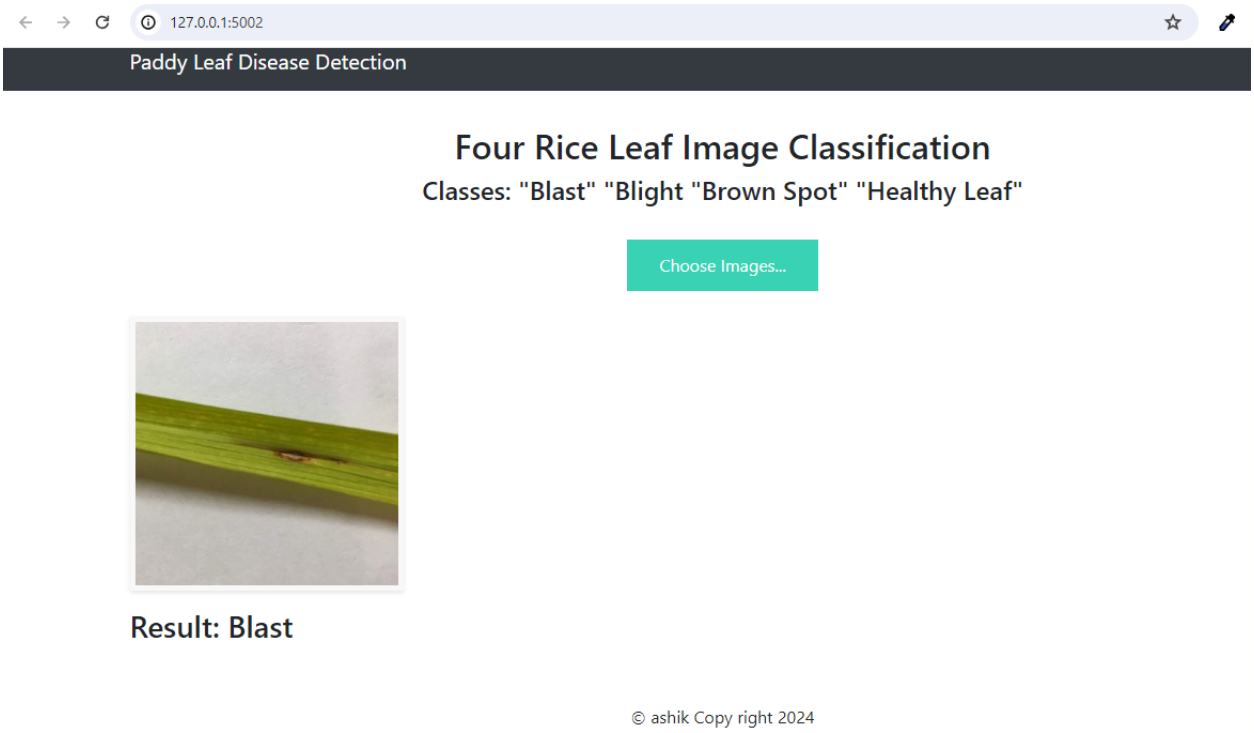


Fig 4.6: Leaf Blast classification using web application prototype.

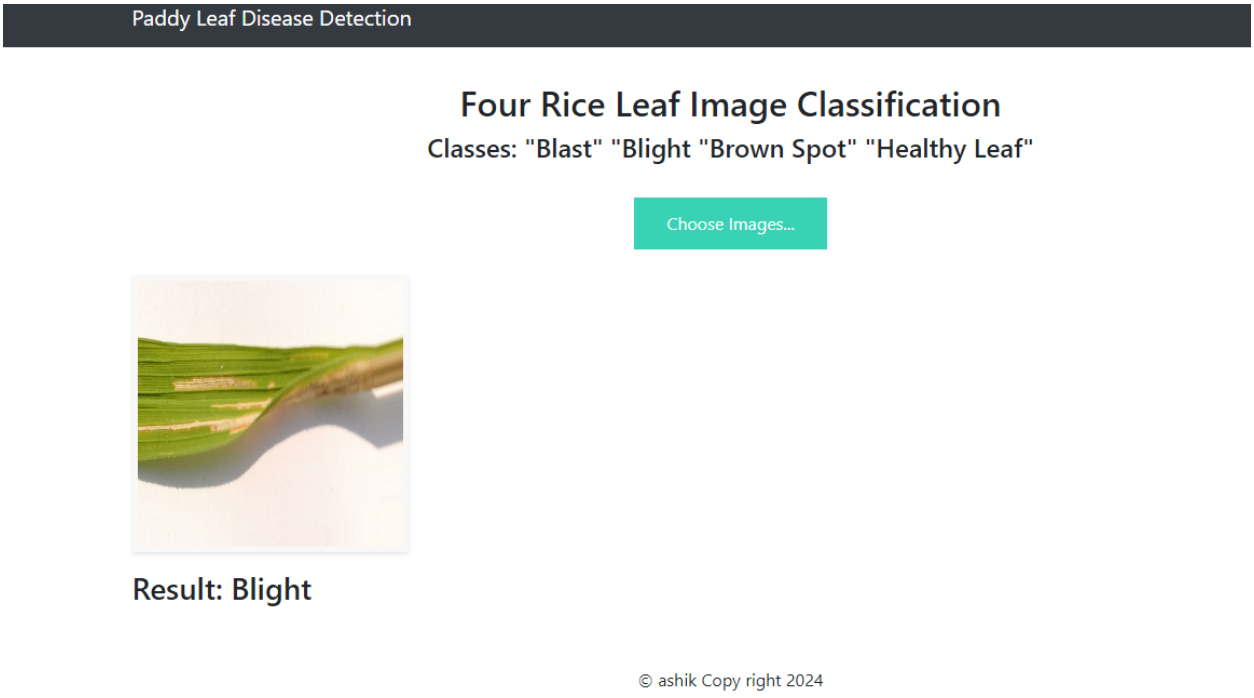


Fig 4.7: Leaf Blight classification using web application prototype.

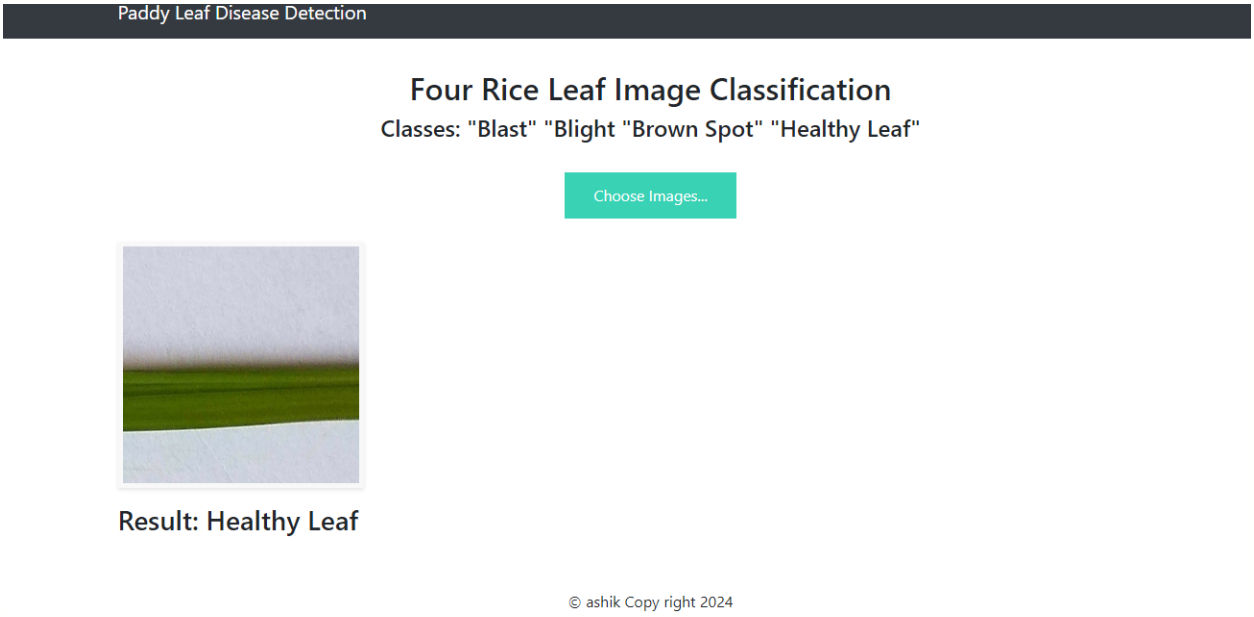


Fig 4.8: Healthy Leaf classification using web application prototype.

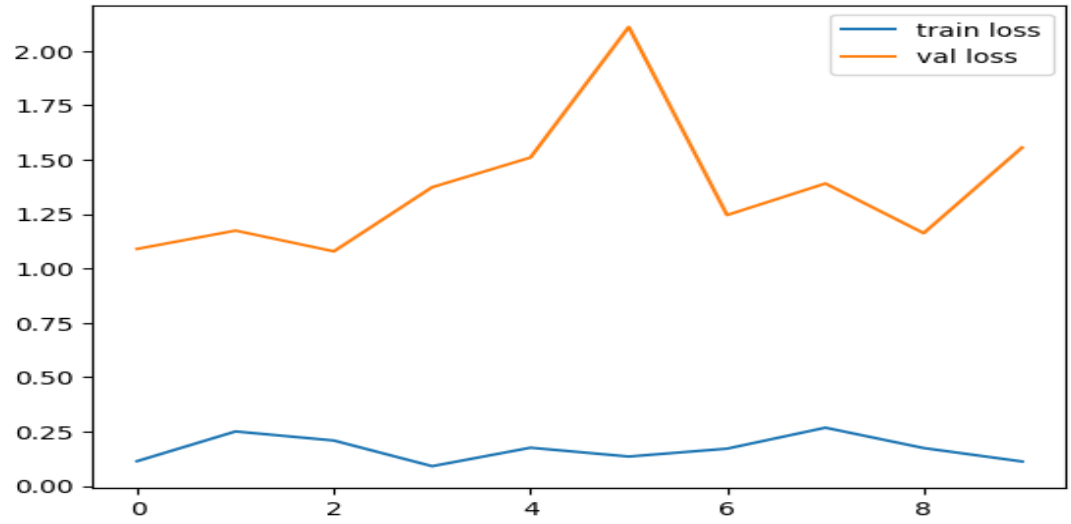


Fig 4.9: CNN validation loss and training for classifying web prototypes.

4.4 Discussion

In my research, I'll employ machine learning methods to predict rice disease. In any field of study, each word should be extremely important to the classification process. The goal of my study has always been to recognize the rice leaf disease by picture classification. The datasets have also been divided into succeeding groups using the K-means approach. One of the most crucial components of any study is the data. A similar experiment might yield very different findings depending on the available data. Since the results were a mix of two datasets, I was certain that other researchers would get different findings if they used one of the two previously accessible datasets. Given that I utilized more data, I may have proved more or less accurate.

Using a range of machine learning techniques and cross-validation to achieve the K-fold scale for the accuracy and average score, I was able to accomplish my aim. I employed a total of four distinct algorithms for this project. I had to search for several things before I could begin the current assignment. I did start working on the algorithm after selecting it. I then ascertained the accuracy of each algorithm. I requested an agriculture officer present my dataset, as I previously indicated.

I was able to obtain the greatest accuracy of 94.16% in the Random Forest model, the classifier for the following 4 categories, by utilizing both of these strategies. Given that my dataset produced the best accuracy out of the four-leaf accuracy classes, it is quite notable.

Precision: Exactness, or the accuracy that defines the accurate forecasting the model makes, is a commonly used metric to assess the efficacy of the model. By taking the total number of correct predictions and multiplying it by the entire number of true positives, I can determine accuracy.

$$\text{precision} = \frac{TP}{TP+FP}$$

Recall: Regardless of all relevant instances, retrieval is the proportion of relevant examples that were eventually located and retrieved. When a technique has a high recall rate, it is considered to have produced the most relevant findings.

$$\text{recall} = \frac{TP}{TP+FN}$$

F1-Score: When assessing a test's validity, the f score takes accuracy and memory into consideration. Accuracy and recall are harmonically averaged.

$$\text{F1} = \frac{2 \times precision \times recall}{precision + recall}$$

Accuracy: The relationship between an assessed value and a recognized cost is known as reliability.

$$\text{accuracy} = \frac{TP+TN}{TP + FN + TN + FP}$$

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Society Impact

Using ml techniques on both basic and aggregated web information, Rice leaf disease is identified. considering that it may significantly impact society. The following are a few possible social implications of these methods:

- **Boost the agricultural sector**: The ability of machine learning algorithms to detect problems in rice leaves and improve the sector with better results is one of the most important potential advantages. By quickly diagnosing a leaf issue early on and treating it in the field, farmers may crop output.

- **Benefits of economy**: The application of ml techniques for rice leaf disease might benefit society economically. For instance, it may increase crop output because this specific sickness is negatively affecting harvests in the agricultural sector. because expensive and intensive therapy not be necessary if the illness is handled early. Additional advantages include increased farming output and positive economic impacts. Farmers that face this of issue will plant than others.

- **Benefits of society:** Using ml techniques for identify diseases that have positive social effects in addition to potential financial gains. Farmers that battle crop disease issues will find crop and a decrease in output their financial situation or benefits. For instance, it may help alleviate poverty in society, which is often an indication of a better environment, and improve farmers' financial situation. Living as a community is crucial.

In general, the use of ML approaches to catch leaf disease will have a significant influence on society. The financial and benefits of eradicating the rice disease of the plant's leaves and the farming industry might improve the standard of living in both the agricultural sector and society at large.

5.2 Impact on Environment

Detecting rice leaf illness from diseased leaf data utilizing ml methods is unlikely to have a significant negative environmental impact. However, there's a chance that the creation and implementation of these tactics would result in higher resource usage and carbon dioxide carbon dioxide emissions, which would have an unintended negative effect on the environment. These methods might potentially have the following negative consequences on the environment:

- **Energy consumption**: Creating and putting into practice machine learning algorithms might need a lot of resources, if they require a various power. One possible outcome might be a rise in energy usage, and the release of greenhouse gas emissions could harm. To lessen these consequences, efforts must be taken to reduce energy use whenever it is practicable. This means accounting for the energy-efficiency of the parts and processes that are being used.

- **Information storage:** Big data required to the development for evaluation of ML algorithms. All data must be stored, requiring the usage of material and energy resources that might be hazardous to the environment. In order to mitigate these effects, it is critical to take into account how data storage affects whenever feasible.

- **Shipping:** This is feasible that the use of ML methods, particularly if they are dispersed and data and algorithms are obtainable from several global locations, may lead to a rise in emissions from the transportation sector. To lessen these effects, It is essential to consider the environmental impact of transportation along with try our hardest to reduce pollution.

Machine learning algorithms for rice leaf disease diagnosis that use raw field asset data are typically not anticipated to have direct adverse effect. It is imperative to consider and implement the necessary measures in light of the potential unexpected indirect environmental impacts. Using environmentally conscious characteristics and calculations to reduce emissions and resource consumption, maintaining data in a form that requires less resources when possible, and, where practicable, reducing gasses in comparison to mass transit are some examples of these safeguards.

5.3 Ethical Aspects

Field data gathering and machine learning approaches may be used to identify diseases of rice of the leaf. A list of some of the most important moral factors to examine is provided below:

- **Agricultural protection**: Since farms are the source of revenue for farmers, growing crops demands a lot of time and resources. Prior to taking any action on this matter, I need to address the farmers' concerns on the information gathering from their fields and speak with them regarding their present circumstances.

- **Accessibility:** The availability of methods for detecting rice leaf death presents another moral dilemma. Disparities in the ability to recognize and treat sickness may occur if some persons or groups are the sole group with access to these techniques. Making sure that a lot of people apply the strategies and available to who might benefit from you is essential.

CHAPTER 6

Conclusion and Future Research

6.1 An overview of the research

This inquiry has taught us a great deal about this topic. Rice lead poisoning situation is a sensitive matter. This makes a substantial yearly contribution to the decline in agricultural productivity. Because of this, I was able to use machine learning to diagnose about three major disorders.

As we've already discussed, I combined field data with online data to gather as much data as possible for my research. I was able to get better at identifying diseases by using this data to train my software algorithms about the pattern of sickness. At first, a few problems were fixed. I was able to accomplish my desired goal. For various people, different algorithms provide different outcomes. I went into further detail in the next section.

6.2 Conclusion

This works shows that the way and research results The ones I work with are top-notch. I think and hope that this study will spur additional research in this field once it concludes. This investigation has provided us with several ideas for growing my body of work. As I worked, I noticed a few mistakes. I discovered other possible directions to pursue with this investigation. It will allow us to carry on with the existing project while resolving any bugs or other problems that may arise. Furthermore, I have ideas on how algorithms may be applied in future research to offer more robust answers to the problems that this investigation has shown. I'm sure I'll be enabled to learn further about other aspects of my chosen topic of study because of this examination. In addition to helping identify paddy-lead sickness, I think it will further research and development of innovative technical solutions that enable us to assist agriculture. Using DL to identify rice leaf disease, I aim to present a unique approach for rice lead identification of diseases based on this study.

6.3 Possible impacts

I think that depending on where my work is viewed, different results are obtained. As a result, I took additional care to do my duties accurately. I used caution to ensure that my work was truthful. I choose to utilize my own dataset as a result. My study may find use in several fields, such as computing and agribusiness. Farmers would gain from this, particularly those employed in the agricultural sector. It may also safeguard agriculture and make my life easier. It might boost the output of agriculture.

6.4 Implications of Further Study

There are several chances to conduct more study in this field, including as part of my own job. I devised many tactics to enhance my work. As said before, I have also found certain errors; these mistakes provide me the chance to make improvements to my study. I will try to fix this error by putting my research into practice and enhancing prediction results with more possibilities of capturing exceptional accuracy. I wish to address the errors that I come across. Further goals are on our minds. My goal is to outperform my prediction. To help consumers get the most out of this, I'll add features to the application that will allow me to predict rice leaf illness using an algorithm, such as the ability to recognize different disease types. This kind of work, in my opinion, might increase agricultural output and move the agriculture sector closer to utilizing technology. Maybe I could help the people grow more food.

Reference

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

[7] 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.

[8] 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.

[9] 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.

[10] A.D. Nidhis, C.N.V. Pardhu, K.C. Reddy,and K. Deepa," Cluster based rice 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.

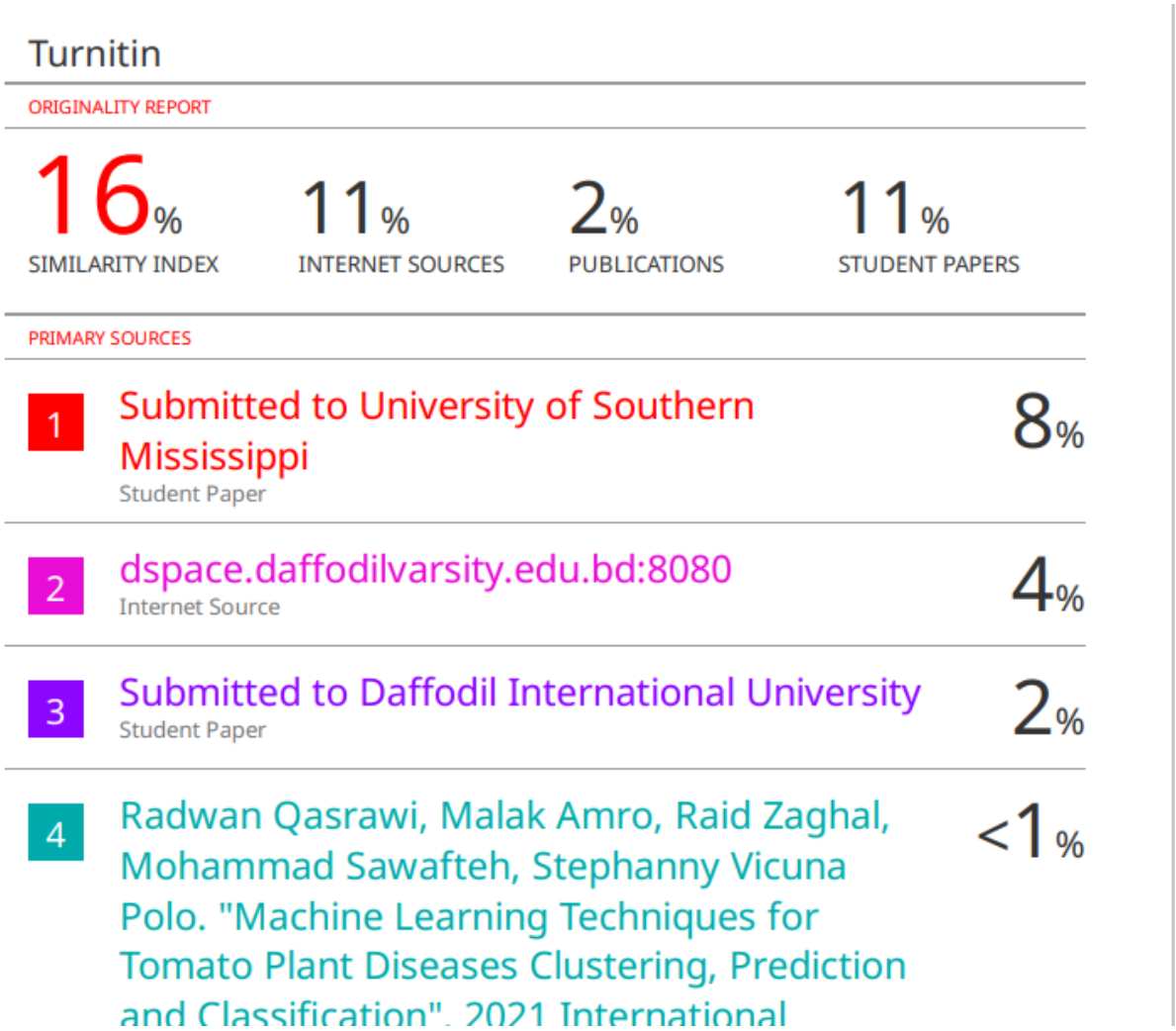
[11] S. Ramesh, and D. Vydeki, "Recognition and classification of rice leaf diseases using Optimized Deep Neural network with Jaya algorithm," Information processing in agriculture, 7(2), pp.249-260, 2020.

[12] 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.

[13] 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.

[14] 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.

[15] 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.



Turnitin

ORIGINALITY REPORT

16%	11%	2%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Southern Mississippi Student Paper	8%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	4%
3	Submitted to Daffodil International University Student Paper	2%
4	Radwan Qasrawi, Malak Amro, Raid Zaghal, Mohammad Sawafteh, Stephanny Vicuna Polo. "Machine Learning Techniques for Tomato Plant Diseases Clustering, Prediction and Classification", 2021 International	<1%