PADDY LEAF DISEASE DETECTION USING MACHINE LEARNING TECHNIQUE

BY

MUSTAFIZUR RAHMAN KONIK

ID: 191-15-2626

AND

AZAZ AHMAD SWAPNIL

ID: 191-15-2426

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

Supervised By

DR. S.M. AMINUL HAQUE
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

AMIT CHAKRABORTY CHHOTON
Senior Lecturer
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

FEBRUARY 2023

APPROVAL

This Project titled **"Paddy Leaf Disease Detection Using Machine Learning Technique."**, submitted by Mustafizur Rahman Konik, ID No:191-15-2626, Azaz Ahmad Swapnil, ID No:191-15-2426 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *Sunday, 5th February, 2023*.

BOARD OF EXAMINERS

7/2/23

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Nazmun Nessa Moon
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Raja Tariqul Hasan Tusher
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Ahmed Wasif Reza
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that this research has been done by us under the supervision of **Dr. S.M. Aminul Haque, Associate Professor, Department of Computer Science and Engineering and** co-supervision of **Amit Chakraborty Chhoton, Senior Lecturer, Department of Computer Science and Engineering**, Faculty of Science and Information Technology, Daffodil International University. We also declare that neither this research nor any part of this research has been submitted elsewhere for the award of any degree.

Supervised by:

Dr. S.M Aminul Haque
Associate Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Amit Chakraborty Chhoton
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Mustafizur Rahman

Mustafizur Rahman Konik
ID: 191-15-2626
Department of CSE
Daffodil International University

Azaz Ahmad

Azaz Ahmad Swapnil
ID: 191-15-2426
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

At First we would like to express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year thesis successfully.

We really grateful and wish our profound our indebtedness to **Supervisor Dr. S.M. Aminul Haque,** Associate Professor, Department of CSE, Faculty of Science and Information Technology, Daffodil International University, Dhaka. His Deep Knowledge & keen interest with supportive instructions helped us in the field of machine learning based research, finally we completed our work on "Paddy Leaf Disease detection using machine learning technique". Their endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to **Dr. S.M. Aminul Haque**, Associate Professor, **Amit Chakraborty Chhoton**, Sr. Lecturer and **Dr. Touhid Bhuiyan**, Professor and Head, Department of Computer Science and Engineering, Faculty of Science and Information Technology, DIU, for his valuable support and advice to finish our project and also heartiest thanks to other faculty member and the staff of department of CSE, Daffodil International University.

At last, again we want to thank all the good wishers, friends, family, seniors for all the help and inspirations. This research is a result of hard work and all those inspirations and assistance.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

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

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Introduction

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

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

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

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

In the image preprocessing system, after the disease images are read and acquired then, Using the color conversion, Red Green Blue (RGB) photos are transformed into grayscale images. Various contrast-enhancing techniques, such as contrast adjustment, are employed to boost the contrast. And same types of disease images have been removed by the image preprocessing task. Once complete of the preprocessing, in segmentation, lowering the image's complexity and allowing each image fragment to be processed or analyzed further. And K-means clustering techniques has been proposed for segmentation. Lastly, Image classification has some model's techniques which is used for identify disease images. In machine learning, image classification has various types of models like K-means, Support vector classifier(SVC), Random forest, Decision tree and Logistic regression. Those classification have been usable for the measure of paddy leaf diseases.

1.2 Objectives

In the current era is one of technological transformation. Technology can be the solution to all problem. So, In agriculture sector many science have been used for grow up this sector. Plant disease has been the most vulnerability of this era of agriculture sector. Many diseases can be occurred in plants. In this research, mainly focused on paddy leaf disease detection using machine learning techniques. Our main goal is to classify the various disease with the accuracy to predict the disease leaf. There are different types class of disease like Brown spot, blast and blight with another one is healthy leaf. Using machine learning and image processing approaches, we aim to categorize the paddy leaf disease. Measuring the size of the disease's impact of segmentation of disease leaf using K-means techniques. Therefore, we may list these objectives as follows:

- To assist farmers and make a contribution to our agricultural economy.
- A paddy's leaves may be used to assess the condition of the plant.
- To use machine learning to detect and categorize the many types of sickness.
- To get information that is applicable to disease prediction.
- Gaining knowledge about all aspects of the machine learning industry.
- Putting different methods to use to enhance outcomes.

1.3 Motivation

We were interested in doing something for the farmers, as they are our main source of food production. Both me and my teammate belong to rural areas, and that's the reason behind our motivation to do something in the agriculture sector to help the farmers by the help of artificial intelligence (AI) and machine learning (ML). First, we thought what should we select for our paper topic but couldn't find out any idea that can fulfill our desires for the research so then we get consulted with one of our honorable teachers and he loves that we choose agriculture for our research topic and he suggest as paddy is our main crops so we should do something related to paddy. That's how we chose our topic, "**Paddy Leaf Disease Detection Using Machine Learning Technique"** In addition, we observe that the modern world is enhancing the agricultural sector with cutting-edge technologies and that researchers are conducting an increasing amount of research on agricultural sectors in order to enhance our agricultural sector. This piqued our curiosity about undertaking such research-based activity. The world we live in is intimately connected to machine learning methods.

1.4 Rationale of the study

There are undoubtedly thousands of studies in the fields of object classification and image processing. However, there are only a handful of completed works on "**Paddy Leaf Disease Detection Using Machine Learning Technique."** So our research is working with different approaches to algorithms and classifications. We have properly planned to complete our work, so we have developed our own model and classifier for this topic to maintain the flow of work.

Image compression, image enhancement, restoration, and measurement extraction are some of the numerous categories that make up the advanced method known as image processing. It helps to lessen how much memory is required to store a digital image. Images can have flaws. Images can have defects due to flaws in the digitization process. Image enhancement techniques can be used to fix damaged images. Additionally, machine learning classifications and algorithms are able to recognize them.

1.5 Research Question

For completing this study, there have been so much hard working and passion able for completing this work. To finish this work was really difficult for us. There have so many challenges to have an answer to the problem that is reasonable, effective, and correct. To communicate these thoughts and address this issue, the researchers want to put up the following questions:

- Are raw image data available for our machine learning research?
- Is the paddy disease raw image data labeled by an Agricultural Officer?
- Can we use a machine strategy to pre-process the raw data?
- Do these techniques and measures have the potential to enhance the agricultural sectors?
- How can farmers benefit from this work & approaches?

1.6 Expected Outcome

In this portion, there are some facts, assuming that those points were our minimum anticipated result. For the outcome of the results of the following research, there have been used different types of classification algorithm for prediction of actual result of the leaf disease. This research-based project's anticipated result is the creation of an algorithm or comprehensive, effective process that will classify paddy leaf disease in accordance with the created model of training and testing the dataset. Consequently, the following is a list of these expected outcome:

- We will show after seeing the paddy leaf whether it's infected or not.
- An improved knowledge of paddy leaf disease detection using ML.

- Detection of Leaves Malnutrition part.
- We anticipate comparing our results to earlier research using dated data of image.
- Identifying the machine learning technique that is most suited for the data at hand.

1.7 Layout of the Report

Chapter One has shown how to introduce the project by its determination, motivation, research questions, and estimated results of research. This part also defines the report's overall format.

In Chapter 2, it is discussed what has already been done in this field. The scope resulting from their limiting of this field is then demonstrated in the latter portion of this second chapter. The main hurdles or impediments to this research are discussed very last. There are some sectors of this chapter like: related works, research summary besides challenges which consume been handled to build up this project.

In Chapter 3, The theoretic analysis of this research project is described. This chapter elaborates on the statistical techniques used in this effort in order to speech the theoretical portion of the study. Additionally, this chapter illustrates the machine learning techniques' procedural approaches. Dataset collecting procedure and data preprocessing system all have been described in the following chapter. Additionally, confusion matrix analysis is provided in the latter serving of this subdivision to evaluate the model and toward display the classifier's accuracy tag and implementation analysis should take part in the valid accuracy using machine learning approaches .There are some sectors of this section like: Research Subject and Instrumentation, Workflow, Data Gathering Procedure, Data Processing , Proposed Model, Training the Mode and Implementation Requirements which have been faced to build up this project. Each algorithm includes a thorough explanation of all the machine learning methods and classifications that were employed in this research.

The experimental findings, performance assessment, and result discussion are existing in Chapter 4. This chapter includes a few experimental images to help the project come to life. After this chapter, result discussion follows the implementation of the machine learning techniques.

The summary of the study, next projects, and conclusion were covered in Chapter 5 & 6. To demonstrate how the entire project report complies with recommendations, this chapter is accountable. The chapter is concluded by highlighting the shortcomings of our efforts, which may have an impact on those people who wish to pursue careers in this area in the future Impact on Society, Environment and Sustainability.

.

CHAPTER 2

Background

2.1 Introduction

This part determination provides the research difficulties, relevant works, and research summary. In the section under "Related Works," we will review research papers by other authors and the ways in which their techniques and accuracy connect to our own. We shall talk about other research papers' works, methodologies, and accuracy that are connected to our study in the part on related works. We shall summarize our linked studies in the study summary unit. We resolve go through how we raised the accuracy layer in the challenges part and all the problems faced for doing this study. Everything has been discussed.

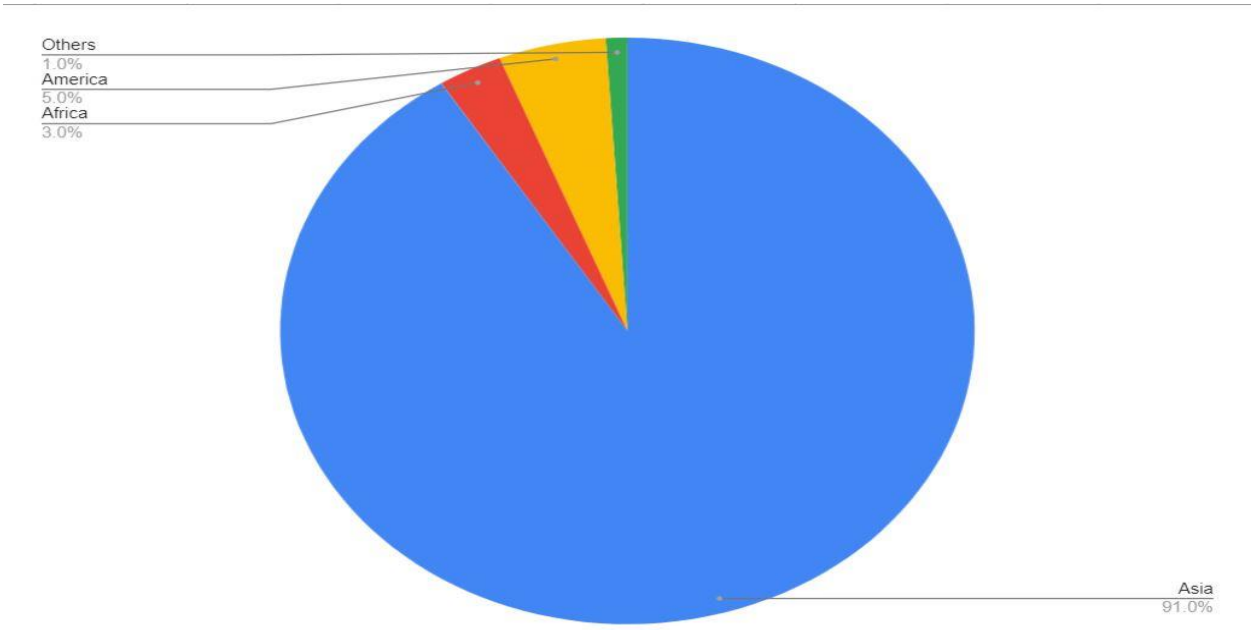

Fig 2.1: Production of Rice World (Source FAOSTAT)

The rural population intentionally uses rice as a significant food source, and it is also the second most widely farmed crop in the world. The two main subspecies of rice, Japonica and Indica, come from the Pinaceae family of plants. The most well-known, affordable, and nutrient-dense food in Asia is rice. Asia, Africa, America, Europe, and Oceania are the five regions of the globe where

rice is grown. Asian countries provide and devour 91.05% of the world's rice, according to the Food and Agriculture Organization of the United Nations' (FAOSTAT) summary. The residual rice is separated awake among various areas of the globe. For instance, 3.5% by Africa, 5.19% by America, 0.67 percent by Europe, and 0.15 percent by Oceania.

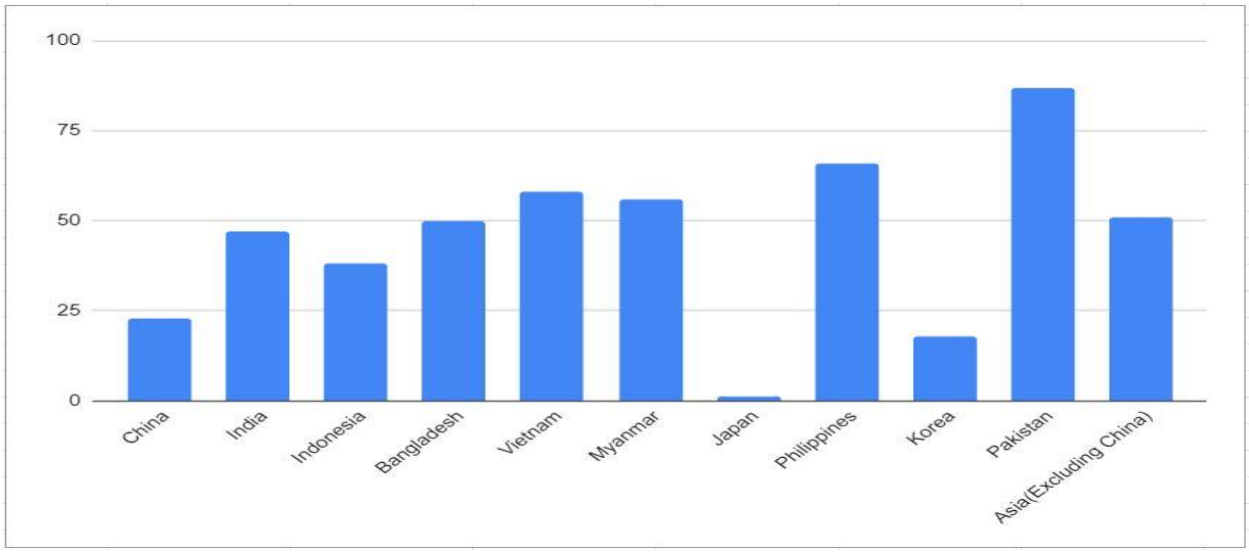

Fig 2.2: Projection of Rice Consumption in Percentage of Asian Countries, 1995-2025. (Source World Bank Population Projection, 1995-2025)

The projected demand for rice consumption, which is substantially higher than the growth rate of the population, is predicted by the World Bank to rise by 51% by the year 2025. Here, in Fig 2.2 is a projection of rice consumption for the main Asian nations between 1995 and 2015.

2.2 Related Works

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

paddy in order to finish their work. Different steps of segmentation using k-means clustering for locating bacterial leaf blight and brown spot in rice plant in their work.

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

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

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

Paddy illness detection using AI technology has received a lot of attention so far or is now receiving it.S.Pavithra et. al [5] focused on Paddy illness detection using AI technology has received a lot of attention so far or is now receiving it. With current image processing and machine learning techniques, how to identify and categorize paddy illness. Identification, Acquisition, Image preprocessing, feature extraction, and classification are the four phases they use to do their

task. SVM classifiers and KNN have been used for paddy ill detection and measurement. The main topics of this analysis are paddy blast and brown spot disease. They achieved the best accuracy from SVM 95% for classification.

2.3 Research summary

The many approaches that the community has to offer is the main focus of our project effort. Four distinct methods have been employed in total, and we have utilized other techniques on our own dataset. Our compiled dataset from the field and the internet has been used as the main data source in this instance. As we previously indicated, our dataset includes both data that has already been utilized and data that we have newly collected. We will be able to examine things like the effect of additional data that we added from the same source and determine how accurate the four algorithms we employed were. The new data is identical to the preceding dataset with which it was integrated. The same types and classes of labels exist, thus that's what it means. The primary language of choice was Python, and our feature extraction processes made use of ML approaches.

2.4 Challenges

The key challenge of this work is to collect besides process the dataset as they are image data and dealing with the data set was too hard. We had to visit several times on the fields to collect data as well as government agriculture office for gathering knowledge about this topic was really tough. To clean our dataset and normalize it, we used several steps and tools. Datasets was in huge amount and tool many layers with dissimilar scope of epoch that took diary time in our machine, so for the ending output we delayed so much with keeping patience. There were some other datasets regarding this domain, but they were not accurately representing our knowledge after we examined them several times and got knowledge on fields from government organizations, so we had to collect datasets from real fields. As we didn’t do any research related work, we had to work really hard to find best ways to do the work done and completed it with our efficiency.

CHAPTER 3

Research Methodology

3.1 Introduction

In this section, we are working to explain about the entire roadmap and approaches to classification of the paddy leaf disease of our study. Some important elements are the gathering of data, its processing, and the suggested model, which is also explained with the appropriate equation, graph, table, and description. In this research using segmentation and prediction with classification model of machine learning that give the best accuracy and this applied on our own raw dataset in our work. Giving an overview of the statistical theories used in our project and outlining the obvious requirements for execution serves to finish the chapter. In this research there have been 4 types of classes that have been used for our models. Mainly, paddy have different types of disease but in our research work we focused 3 disease class like Brown spot, Leaf Blast and leaf Blight with another class of paddy is Healthy leaf. All the models have been trained of all images by this following 4 types of classes in our research.

3.2 Research Subject and Instrumentation

A research theme is a field of study that consumes been examined and investigated to clarify concepts. not just for implementation, but also for model creation, data collection, performance, processing, and model training. Instrumentation is the area where we discuss the technologies and procedures we employ. We used the Windows platform and Python language with many packages like NumPy, SkLearn, OpenCV, etc. We have used the Google Colab platform for entirely the training and testing processes. Google Colab is an allowed platform for Python programming languages to implement code for data science and machine learning algorithms.

Libraries used:

- Matplotlib: A set of routines referred to as Pyplot is one of Matplotlib's plotting utilities. It enables you to perform tasks like locating lines in a plot and defining a plot's boundaries while creating forms, among other things.

- NumPy: The NumPy Python module is a well-liked option for manipulating arrays. Its scope includes matrix operations, linear algebra, and the Fourier transform. The NumPy library for Python provides a variety of tools and methods for working with arrays of different sizes. Arrays may be accomplished mathematically and logically with NumPy. NumPy is a Python library for numerical calculation, to put it simply. Also known as "numerical Python," this word is utilized.

- Sklearn: Sklearn is a simple-to-use and efficient application for studying prognostic data. Free for everyone to use and customize to their own needs Developed with NumPy, SciPy, and matplotlib.

- Seaborn: This seaborn is well-known as a Python tool for data visualization that is compatible with matplotlib. It is a user-friendly tool for producing intelligent and aesthetically pleasing data representations.

- CV2: A collection of Python bindings called OpenCV-Python was created to address computer vision issues. It allows one to process photos and movies to recognize items, people, and even human handwriting etc.

- Joblib: Joblib offers a more efficient means of avoiding repeatedly calculating the same function, which may save a lot of time and computational expense.

- H5py: An interface for the HDF5 binary data format in Python is provided by the h5py module. Huge quantities of numerical data may be simply stored in HDF5 and edited using NumPy.

- Os: Developers can communicate with the Operating system they are working on using the different capabilities provided by the Python OS module.

3.3 Workflow

In this research there are some potential things or working flow for determining the analytics of data. Data gathering, data processing, data resizing and augmentation, model selection, etc. are some of the workflow phases for this study.

Stage-1: - **Data Collection:** We gathered information from the internet and then processed it to produce our own data collection. There isn't a large classful dataset available in this sector because data collection for the particular paddy sickness and class dataset classification was so difficult.

Stage-2: **Data Processing:** Following the acquisition of all data from multiple sources and raw data, each kind of data was processed separately. There are a lot of data with inaccuracies and noise. We physically process those data initially, and then we apply the selective dataset to the following phase.

Stage-3: **Data Resize and Augmentation:** Data have remained expanded and resized after being processed class by class. We had to resize and supplement the data for training purposes. We only performed the most significant and minimal augmentations because augmented data can lead to some overfitting.

Stage-4: **Model Selection:** We select a model and train and test it against our data to improve accuracy. Machine learning uses a variety of classifiers. One model was ultimately chosen to estimate the data accuracy after several models were used to improve accuracy with our machine setup.

Stage-5: **Performance Evaluation:** All of the outcomes are covered in this unit. After training and testing, such procedures offered us a limited degree of accuracy for the next 4 classes. Additionally, the confusion matrix and a table displaying the accuracy, recall, and f1 measure were calculated.

Stage-6: **Conclusion and Future Work:** A summary and a roadmap for future development will be included in this section.

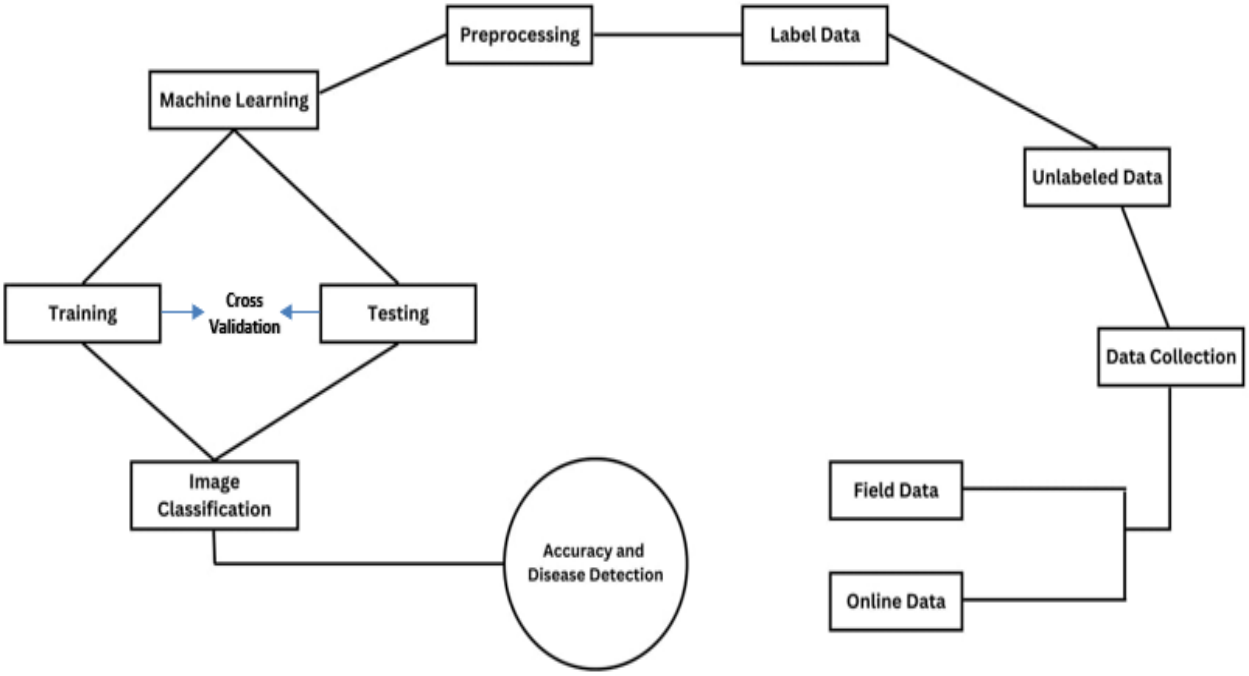

Fig 3.1: Proposed model for whole research work.

Fig:3.1 shows that the basic model has been used for the classification of leaf disease. Firstly, collection of the raw data from field of paddy and online data from web both are merged for creating a dataset. After that, image have been labeled and classified by agriculture officer and resize, augmented of the data. Then the machine may then be supplied with this data after that. We may utilize this data to train and test our intended ML techniques on new datasets. The algorithm's precision will then enable us to achieve our objective of detecting paddy disease detection accuracy with the highest accuracy of ML models.

3.4 Data Collection Procedure

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

We took 80% images of the dataset for training which means 480 images and 20% of images of the dataset which means 120 images for testing. In the training dataset, each class contains approximately 120 images, and in the test dataset, apiece covers approximately 30 images.

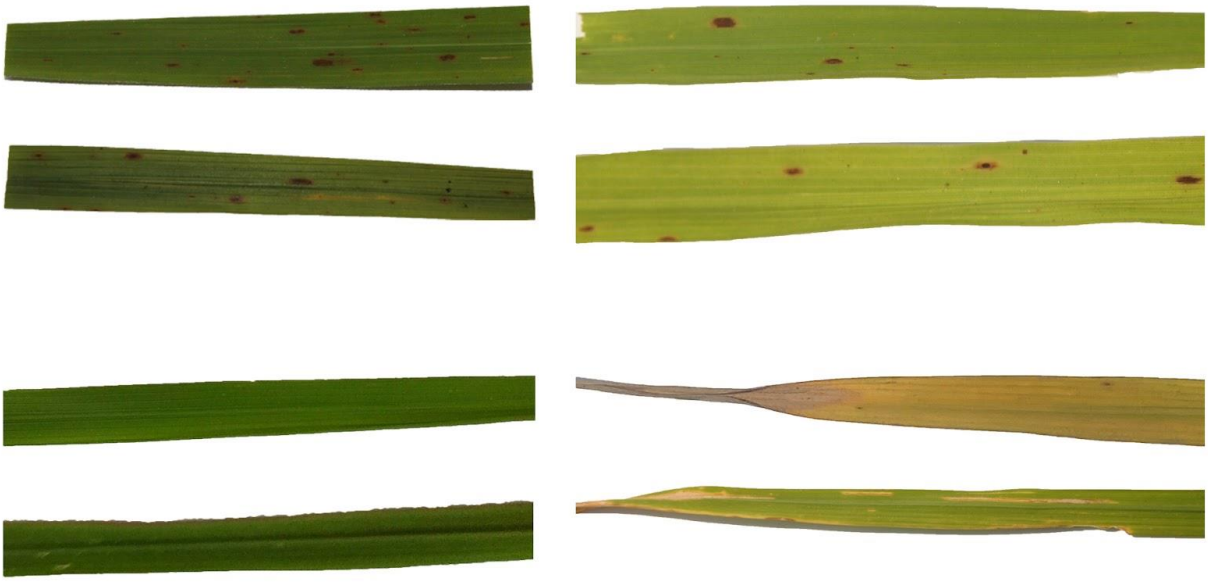

Fig 3.2: A small part of our Dataset

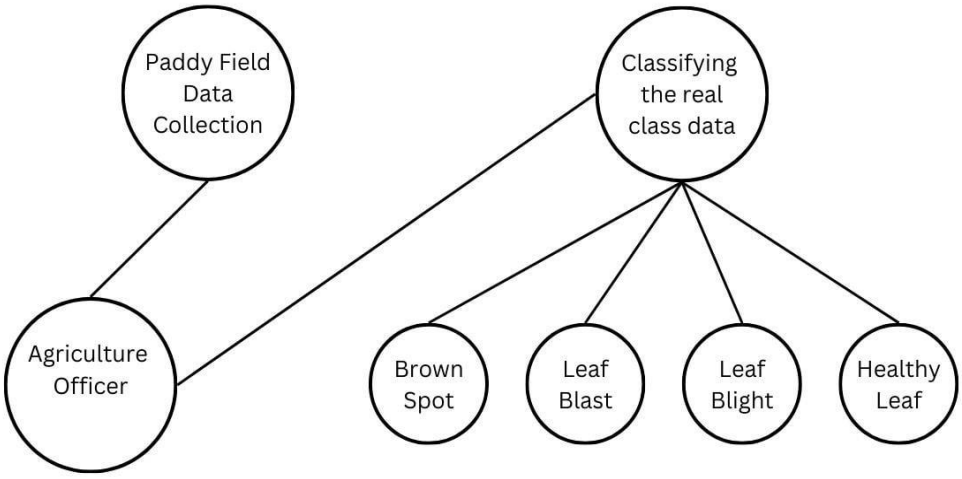

Fig 3.3: Raw data collection and labeling work Flow

Table 3.1: Dataset Table

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

Label Encoder:

In the field of machine learning, we frequently deal with datasets that include a large number of labels, either in a single column or multiple columns. For these identifiers, words or numerals may be utilized. In order to make the training data more accessible to humans or more understandable, words are frequently used to label the data.

Label encoding is the process of converting labels into a language that computers can understand. The labels must be changed into a numerical format as part of this procedure. The ultimate decision on the appropriate usage of such labels may be made by ML algorithms. This stage of pre-processing the structured dataset in supervised learning is crucial.

3.5 Statistical Analytics

3.5.1 Data processing

Data processing is a main portion of a data. A data set's processing of the data is crucial. Preprocessed data plays a major role in our ability to successfully work with row data. In this research, we are going in paddy field for capturing disease leaf row data and with some online resources data have been merged for creating a healthy dataset of 4 classes. For adjusting dataset, the achievement typically depends on the preprocessing of data. Pre-processing data more effectively will result in more accurate results. The two stages of a data processing system, two of them are data preparation and data augmentation. It is the initial difficulty for this species of research-based job, to put it simply.

I. **Data augmentation:** We use some online dataset to prepare significant data for prediction. Some data has been scaling, rotating left or right with some degrees and flips of the same category of dataset. That the way to combining all the dataset of the different class label. The dataset's volume of pertinent data is increased as a result.

II. **Data preparation & Build:** All the images of our dataset which have been collected from the field raw data and some online resources data with different heights and widths. Our data is resized to 500 × 500 pixels since our model needs a fixed pixel for each image with using of resize code to resize image for fixed size (500 × 500) pixels. And all the images have been preprocessed in our model with 'jpg' extension. We have too changed the images into grayscale, RGB with HSV after that image have been segmented which is ready for classification. So, in order to train the model, we used segmented image total dataset.

- Fixed size images with code.
- jpg format conversion.
- Delete error images.
- Duplicate images deleted.

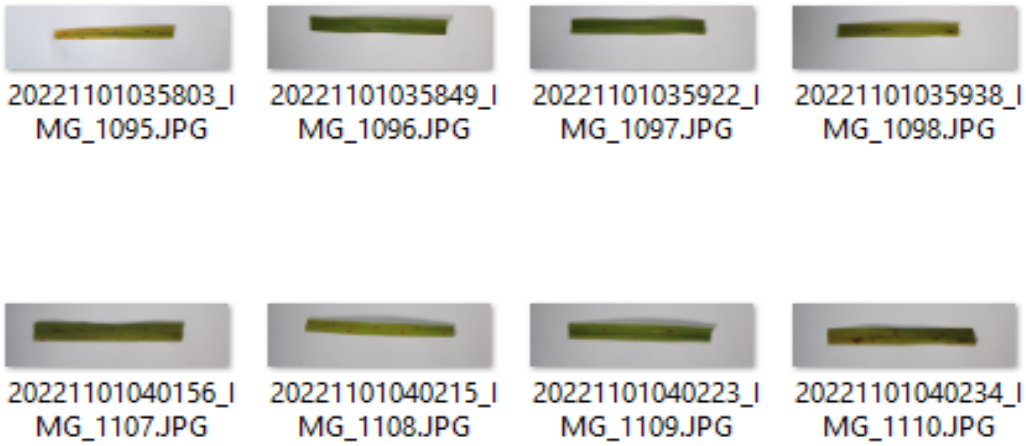

Fig 3.4: Unclassified and unlabeled data.

After labeling the un labeling classless dataset there have been 4 classes of major disease have been labeled by agriculture officer. Agriculture officer labeled them as 4 different class like: Brown spot, leaf blast, leaf blight and healthy leaf.

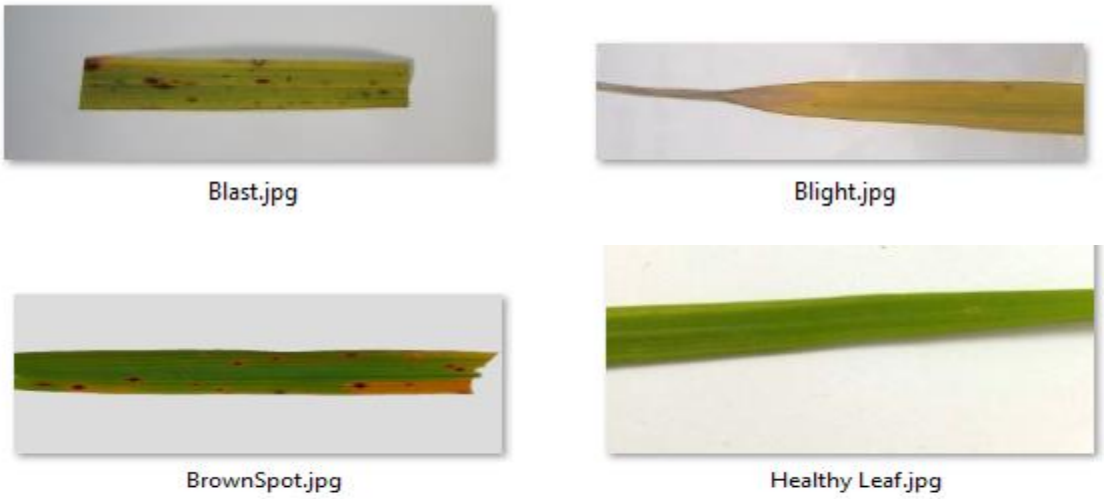

Fig 3.5: Label & classes of data Blast, Blight, Brown Spot and Healthy Leaf.

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

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

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

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

Fig 3.6: Classes of 4 paddy leaf's.

3.5.2 Image segmentation

In our research 4 types of classes have been used for image processing techniques. For image processing techniques, K means algorithm have been used for creating the segment of images of those classes. K means clustering is a widespread machine learning algorithm. K-means clustering can be used to create groups of observations having related properties.

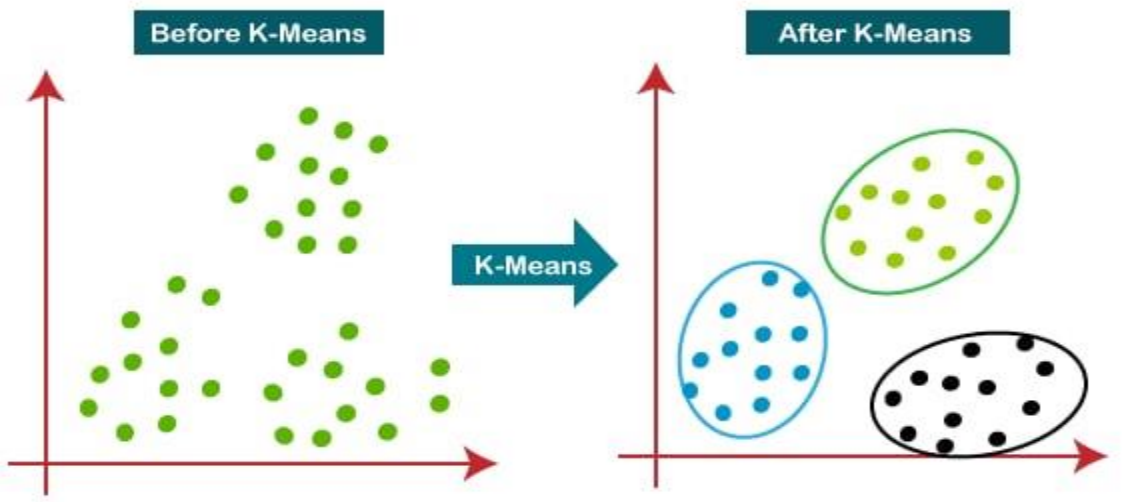

Fig 3.7: K-means cluster

In ML techniques, K -means always create significant impact for segment or classify the image disease detection. The sick image always be able to see the infected part of the image using ML techniques. There has been some scaling for the image data of disease that should be converted into an array. The sounds in a picture that degrade its quality are found during image segmentation. K means algorithm is used to eliminate noise and extraneous spots k- means algorithm converted enhanced image which always produces significant quality images for the detection of paddy leaf disease in Fig:3.8.

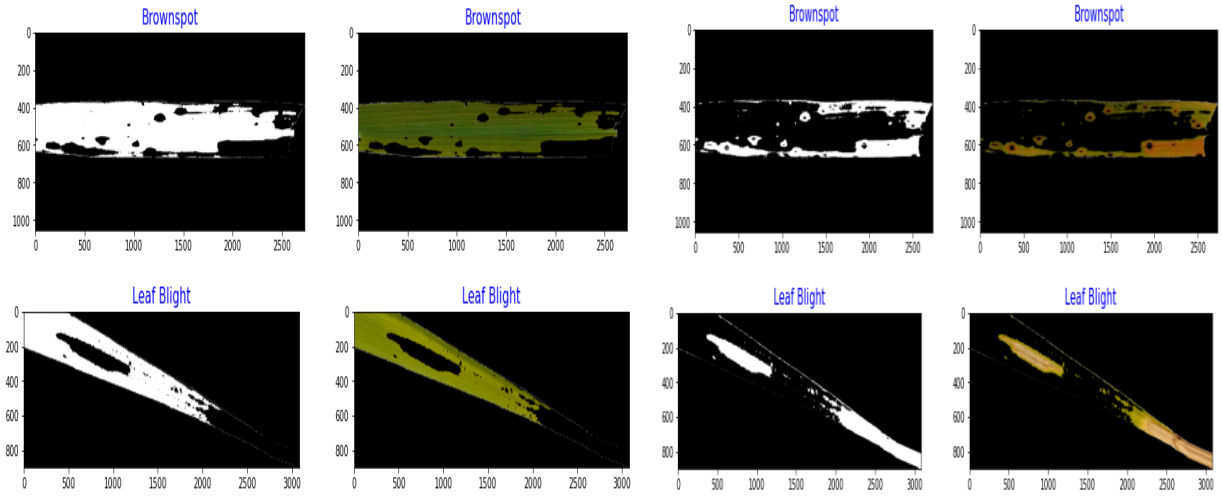

Fig:3.8: Segmentations of Brown spot & Leaf Blight.

3.5.3 Feature extraction

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

3.5.4 Training and Testing data

Researching and developing algorithms that can learn from data and provide predictions based on that data is one of the most common activities in the zone of machine learning. As a means of accomplishing their goals, these algorithms first use the input data to build a mathematical model, which they then employ to make inferences or assessments. These kinds of inputs are often partitioned into many data sets before they are utilized to build a model. The model-building process typically employs three distinct data sets: training, validation, and test. We trained with 80% of the training dataset and tested with 20%. With a total of 600 photos in the dataset, we used 480 for training and 120 for testing.

3.5.5 Classification model

1. **Random Forest:** Use the tree-based strategy of the RF Classifier for both classification and regression. It is an algorithm that produces a tree-based hierarchy in the context of machine learning. It's a method used by AI that builds "decision trees" in a hierarchy. Random Forest Classifier creates a large number of decision trees as an ensemble method before averaging them. This helps to explain the overfitting issues. Machine learning is one of the most talked-about topics in business today since it can be used in any circumstance where there is a lot of data. There are several benefits that RF Classifier offers over other algorithms, which makes it a very popular machine learning technique. The technique was developed to handle big datasets and was originally published in1997.

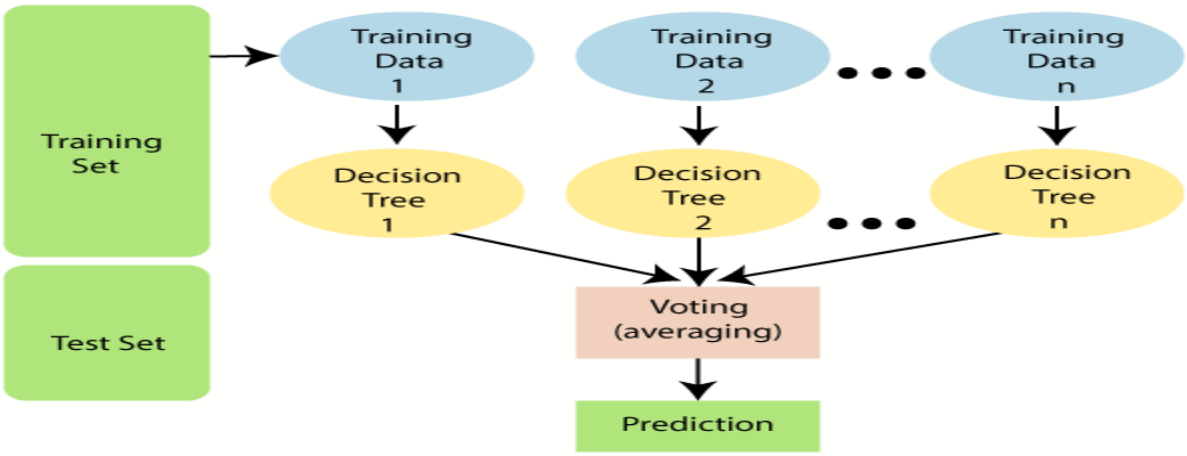

Fig 3.9: Random Forest classifier work flow

As long as there are some training data sets with labels for each potential category and each set includes at least two samples, it may be used to solve any classification-related problem. Because it requires training sets provided by a supervisor, Random Forest Classifier is a supervised classifier. Once there are sufficient datasets, this technique uses selection, splitting, and out-of-bag estimation to produce predictions based on the data.

2. **Logistic Regression:** One of the supervised machine learning approaches, logistic regression, is frequently used aimed at binary class classification matters, where the results are static and be contingent on the incidence of a certain sort of data. For values between 0 and 1, the basic workings of logistic regression are represented by a sigmoidal function.

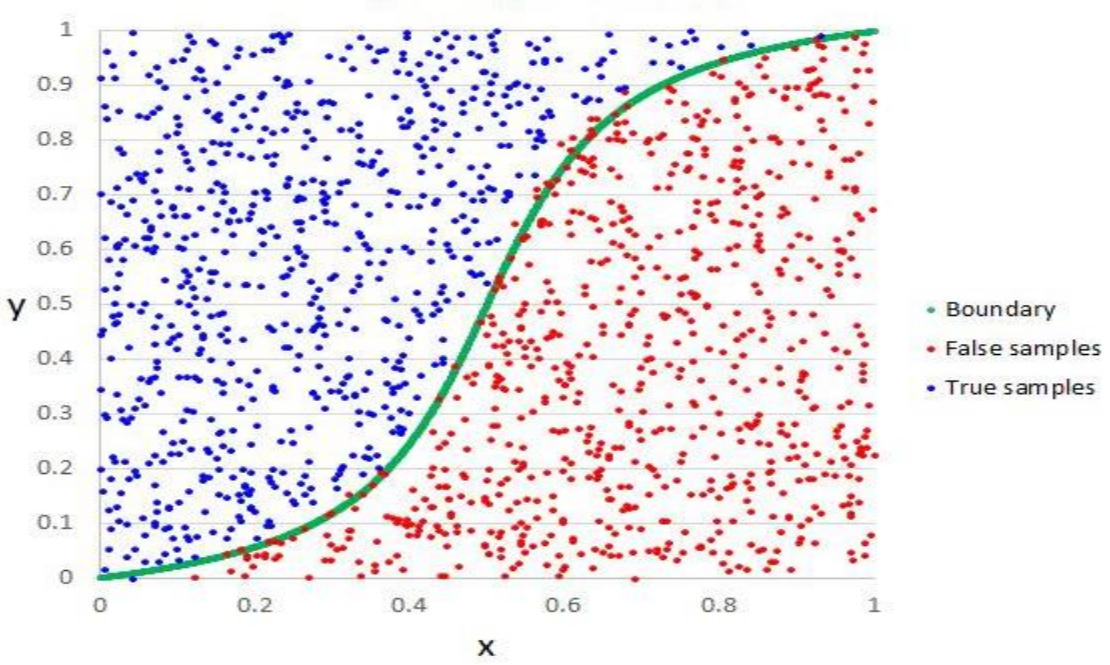

Fig 3.10: Logistic Regression example

Logistic regression is a sort of classification method. According to one definition, a binary result is one in which there are only two possible outcomes: Either it will happen (1) or it won't happen at all (0). Simply explained, independent variables are anything other than the dependent variable that might have an impact on a study's conclusion (or dependent variable). Consequently, a logistic regression is the proper approach of analysis when dealing with binary data. You are dealing with binary data when the result or determining factor is of a binary or categorical nature; alternatively, you will know you are working with binary data if it falls into one of two categories.

3. **Decision Tree:** One of the techniques for supervised machine learning is the decision tree. Regression and classification issues may be solved with this approach. It is a tree-structured classifier, where internal nodes stand in for the dataset's characteristics, branches for the rules of classification, and each leaf node for the result. Two nodes the Decision Node and the Leaf Node make up a decision tree. A choice is made using a Decision node, which has several branches, whereas a Leaf node represents the result of that decision and does not have any more branches. The characteristics of the provided dataset are used to execute the tests or make the judgments.

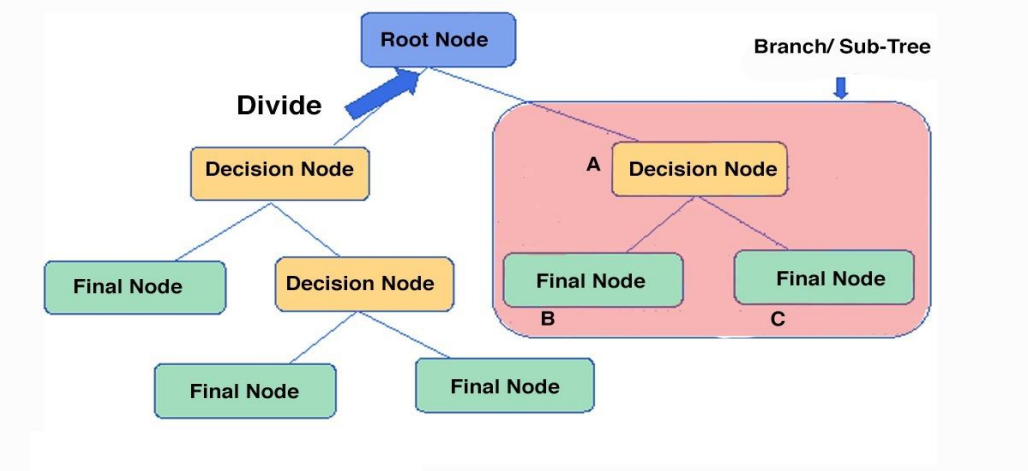

Fig 3.11: Decision Tree work flow

Decision trees are typically simple to grasp since they reflect how people think while making decisions. Because it displays a structure like a tree, the decision tree's reasoning is simple to comprehend.

4. **SVC:** It's called Support Vector Classifier also called SVM.SVC is a nonparametric clustering approach by definition since it makes no assumptions about the underlying structure of the data, such as the number of clusters or their relative sizes. As a result, if the dimensionality of your data is high, a preprocessing step, such as doing the principal component analysis, is typically required. Previous research has shown that it functions best with low-dimensional data. A number of modifications have been made to the original strategy, and these modifications offer specialized methods for computing the clusters by merely processing a portion of the edges in the adjacency matrix. Many of these changes have been put out.

3.6 Implementation

After all of the following works, we need to apply the data set for accuracy. For implementation, we separated our effort into five key sections. These are the measures that must be taken for our project to succeed.

- Image Acquisition.
- Image pre-processing.
- Segmentation for disease.
- Feature extraction.
- Algorithm Implementation
- Image Classification
- Result & accuracy Discussion.

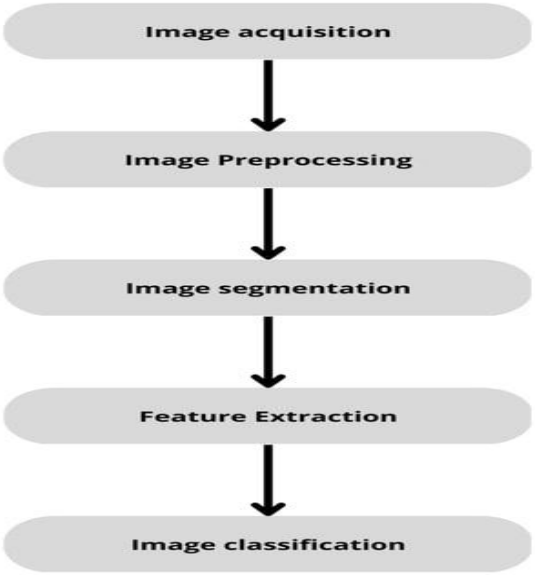

Fig 3.12: Proposed Model of work.

For data collection we need to go in the paddy field, for collecting brown spot, leaf blast, leaf blight and healthy leaf with online images both has been combined by a fine dataset. Following that, we got to work on the data preprocess. Here, we eliminated all the extraneous components of our data, such as noise and error pictures, unscaled images etc. We also use features extraction for the increased time of data train and test.

The first step in our algorithm implementation required us to work on the algorithm's code. We employed four total algorithms and tested their accuracy. After finishing the algorithm, we evaluated the accuracy. We compared the accuracy and determined which would be most

beneficial for our task. Which has been so much predictable for the leaf disease. After the proper study on all necessary numerical or theoretic ideas and methods, a list of requirements takes been created that must be compulsory for such an effort of image Classification. The possible necessary effects are:

1.Hardware/Software Requirements

- Operating System (Windows 7 or above)
- Hard Disk (minimum 1 TB)
- Ram (Minimum 4 GB)

2.Developing Tools

- Python Environment
- PyCharm.
- Google Colab.

CHAPTER 4

Experiment Results and Discussion

4.1 Introduction

This section contains a description of how the classification model for paddy leaf disease was created. The model's entire process was broken down into a few parts, including the acquisition of pictures, data preparation, data augmentation, data resizing, suggested model description, and training with model accuracy. Our experimental outcomes are presented and discussed throughout this chapter.

4.2 Experimental Result

Paddy leaf disease detection has been predicted by various types of algorithms. Because of this, we employed several techniques throughout the procedure. To find the optimum strategy for the experiment, we tested a variety of approaches and analyzed a variety of approaches. We experimented with many methods to enhance the quality of our work. We employed two distinct datasets, one of which was obtained from the internet and the other by our own raw data collection which have been labeled by agriculture officer for 4 classes like: Brown Spot, leaf blast, leaf blight and healthy.

We made use of the existing Python libraries, content classification methods, and dictionaries. For analyzing the dataset of disease algorithm makes the dataset a significant result or accuracy of the disease detection.

4.3 Descriptive Analysis with Best Classification Model

Depending on the categorization techniques we employed, we obtained various results. We have been used 4 machine learning algorithms for the accuracy of paddy disease. For the accuracy measurement we applied K-fold cross validation techniques for the means of all the model's algorithm after that a prediction result have been given by K-fold cross validation techniques. We used Decision tree, Random Forest, Linear regression, Support vector Machine (SVM) with classifier and we applied it on K-fold cross validation techniques. Each models used the same dataset, which included both previously accessible data and our own dataset that we had collected

from the internet and raw data, after labeling the final dataset. After completing the dataset procedure, we utilized Python and its prebuilt libraries to evaluate the algorithms accuracy.

Table 4.1: Accuracy Table

Classifier	**Accuracy Score (AUC)**
Decision Tree	85.83%
Random Forest	94.16%
Logistic Regression	85%
SVM (classifier)	82.50%

Performance of several classifiers is displayed in this section. CoLab and PyCharm, two pieces of open-source tools, were employed throughout the entire process. Four classifiers in all were used: Support vector classifier, Decision tree, Random Forest, and linear regression.

	precision	recall	f1-score	support
0	0.92	0.92	0.92	39
1	1.00	0.96	0.98	27
2	0.96	0.88	0.92	25
3	0.91	1.00	0.95	29
accuracy			0.94	120
macro avg	0.95	0.94	0.94	120
weighted avg	0.94	0.94	0.94	120

Fig 4.1: Classification Report of Random Forest

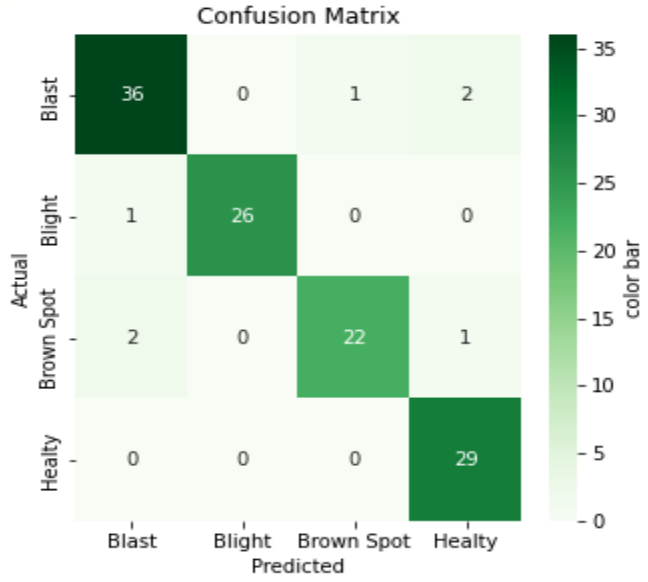

Fig 4.2: Confusion matrix of Random Forest.

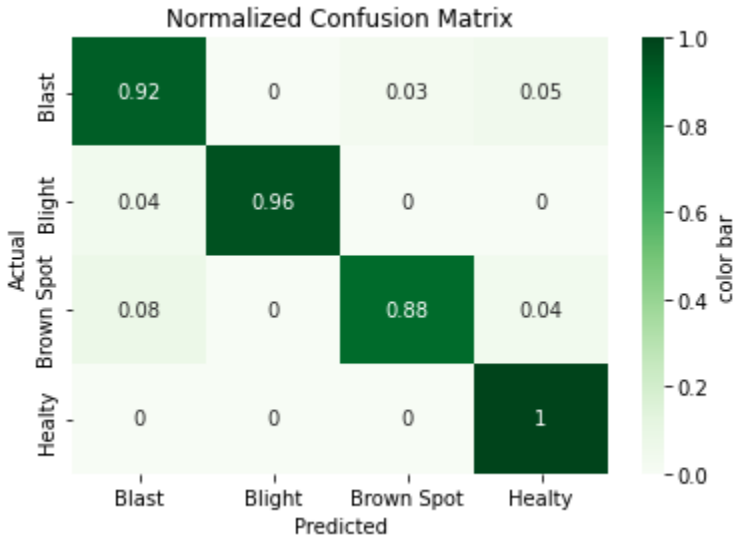

Fig 4.3: Normalization Confusion matrix of Random Forest.

We are showing only Random Forest classification all report and normalization matrix for the achieving highest accuracy. Models Performance measurements result of three Disease leaf classes as well as healthy leaf below Table: 4.2,4.3,4.4,4.5 with the value of K=10 in cross validation techniques.

Table:4.2 Performance Measurements of Blast.

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

Table:4.3 Performance Measurements of Blight.

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

Table:4.4 Performance Measurements of Brown Spot.

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

Table:4.5 Performance Measurements of Healthy Leaf.

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

Table:4.6 Average Disease Leaf Accuracy followed by models.

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

Here above Table:4.6 shows the 4 classes of paddy disease leaf accuracy according to classification models.

Table:4.7 Comparison of Some Previous Research on Paddy leaf Diseases.

References	Diseases	Tools and classifier	Accuracy
Xiaochun Mai et al. (2016)	Brown spot, bacterial blight, leaf scald, leaf blast	Random forest classifier	dataset2: 80%
Prabira Kumar Sethy et al. [2]	Brown spot, bacterial blight, leaf scald, leaf blast	Fuzzy logic, computational intelligence, RF, SVM, K-means	86.35%
Mohd adzhar abdul kahar et al [17]	LBD, BSD, BLB	Neuro-Fuzzy expert system	74.21%
Suman T1, Dhruvakumar T2 [18]	rice blast diseases, narrow brown spot, BLB, brown spot,	SVM with classifier	70%
Suresha M et al. (2017)	Various disease	k-Nearest neighbor (k-NN) classifier	76.59%
In this Work	**Brown Spot, Leaf Blast, Leaf Blight, Healthy Leaf.**	**Random Forest**	**94.16%**

In above Table:4.7 discuss about different methods with techniques for identify leaf disease of rice with using of classification. In past, research have been occurred for detection various typed of disease. Our model's accuracy gave better result for perspectives of ML in previous research of Table: 4.6. Fig: 4.4 will show the comparison of average cross validation accuracy.

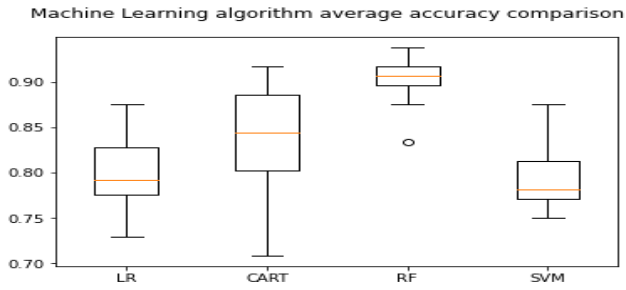

Fig: 4.4 Comparison of average cross validation accuracy.

4.4 Discussion

In our research, we will apply ML models for the predicting of paddy disease. In every sector research, every term should be very plays an important role for the classification. Our goal for this research from the start is to detect the paddy leaf disease using image classification. There has also segmentation of the following classes of dataset using K-means algorithm. One of the most crucial components of every experiment is the data. Depending on the available data, an experiment of the same type might provide extremely different outcomes. We were certain that the findings that other researchers would obtain utilizing one of our experiment datasets that was already accessible would differ since we employed a blend of two datasets. Because we utilized additional data, our accuracy may have improved or decreased.

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

After using K-fold cross validation techniques for the following 4 models' classifier, our most accuracy was in Random Forest where we got 94.16% of accuracy score. It is most significant for

our dataset after getting the highest accuracy with 4 classes of leaf's accuracy mentioned above Table: 4.6.

Precision: Precision, or the accuracy of a successful prediction produced by the model, is one measure of the model's performance. By dividing the total number of positive predictions through the number of real positives, we may measure precision.

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

Recall: The proportion of pertinent examples that have been recovered ended all pertinent instances is known as recall. A high recall rate indicates that the majority of pertinent results were returned by an algorithm.

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

F1-Score: Using both precision and recall, the f-score calculates a test's accuracy. Both recall and precision are a harmonic average.

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

Accuracy: The familiarity of the measured worth to a known cost is referred to as accuracy.

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

Cross Validation: Following the creation of a model, cross-validation is essential. We used stratified K-fold cross-validation as a result. Each time, the data set was divided into 10 equal folds. Nine folds were used to create a model, and one-fold was used to assess it. As a result, after ten rounds, the average value has been compared to the model's accuracy. The Stratified K-fold is recommended because it equally distributes the target attribute's classes across all folds. We have experimented with several different methods. We go into great detail on how we succeeded in achieving our objectives.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Utilizing raw and integrated internet data, machine learning techniques are applied to the diagnosis of paddy leaf disease. Considering that it could have a big impact on society. These are a few possible social repercussions of these techniques:

1. **Improve agriculture sector:** One of the biggest potential effects of machine learning methods is the ability to detect paddy leaf disease and improve the industry with better results. Early detection of a leafs condition will assist farmers in bringing an instant fix to the field to increase crop production. In the end, this might result in increased agricultural output and a more effective way to combat disease.

2. **Economic benefits:** Society may profit economically from the deployment of machine learning techniques to identify paddy leaf disease. For instance, it might aid in increasing crop production, as this form of sickness in the agricultural sector is causing damage to crops. because treatment that is more intensive and expensive may not be necessary if the illness is treated early on. Additionally, it can result in better farming productivity and financial results. Farmers who are having this kind of issue will grow fewer crops than other farmers.

3. **Social benefits:** The application of machine learning strategies to the detection of leaf disease may provide societal advantages in addition to potential economic gains. In the case of farmers who are dealing with disease issues on their crops, crop damage and a reduction in production will significantly alter their economic benefits or situation. For instance, it could help to improve farmers; economic circumstances and create a society free from poverty, which is frequently a sign of a better society. We must coexist as a society, and because Bangladesh is an agricultural nation, any issues faced by farmers would have an immediate impact on society as a whole.

Overall, machine learning methods used to identify leaf disease will have a big impact on society. Economic gains and social benefits may be obtained by reducing paddy leaf disease and the agricultural sector, which would help both the well-being and standard of the agricultural sector as a whole as well as society.

5.2 Impact on Environment

There is little chance that applying machine learning techniques to identify paddy leaf illness from disease leaf data would have a significant negative effect on the environment. However, it's feasible that the creation and use of these approaches might have an indirect effect on the environment by increasing resource usage and emissions. Here are some possible effects on the environment that might result from using these methods:

1. **Energy consumption:** It can take a lot of resources to create and implement machine learning algorithms, especially if they need a lot of processing power. Increased energy demand might be the outcome, and the generation of greenhouse gas emissions could have an adverse effect on the environment. It is crucial to take measures to reduce energy usage whenever practical in order to reduce these effects. This includes thinking about how energy-efficient the hardware and algorithms being utilized are.

2. **Data storage:** Large volumes of data are necessary for the training and testing of machine learning algorithms. Storage of this data necessitates the usage of resources, including energy and materials, which may have an adverse effect on the environment. It's critical to think about how data storage will affect the environment and take action to reduce resource usage where it can be to reduce these effects.

3. **Transportation:** It's feasible that the employment of machine learning techniques might lead to higher transportation emissions if the techniques are utilized in a distributed fashion and data and algorithms are accessible from various places across the world. It is crucial to take into account how transportation affects the environment and to do all in your power to reduce emissions in order to lessen these effects.

Overall, the use of machine learning techniques for detecting paddy leaf diseases using raw field and online resources data is not likely to have a direct impact on the environment. However, it's crucial to take into account any indirect environmental effects that the creation and application of these approaches may have and to do all in your power to reduce them. This might involve cutting back on resource use and emissions by using energy-efficient hardware and algorithms, storing data in a way that uses the fewest resources feasible, and, where practical, decreasing emissions from transportation.

5.3 Ethical Aspects

Using field data gathering and machine learning approaches, rice leaf disease can be identified. Some of the most important ethical considerations are listed below:

1. **Crops Safety:** Naturally, it takes a lot of work and money to grow crops, and farms are where or how farmers make a living. We need to address the farmer; worries about gathering data from their fields and talk to them about their existing position before we can take any action on this matter. Based on the information gathered from the field, it is crucial to take problems and risks into account.

2. **Bias:** Another ethical concern is the potential for bias in the machine learning algorithms. For example, if the algorithms are trained on data that is not representative of the population, it could result in biased predictions. It is important to ensure that the data used to train the algorithms is diverse and representative of the crops disease, in order to reduce the potential for bias.

3. **Access:** Access to machine learning methods for detecting paddy leaf dying is another moral dilemma to take into account. Inequities in the ability to detect and treat disease may emerge if the procedures are only available to specific groups or individuals. It is crucial to make sure that the approaches are widely used and accessible to everyone who could gain from them.

4. **Responsibility**: The topic of who is accountable for decisions made in light of the output of machine learning algorithms is left unanswered. It is crucial to think about who is accountable for making sure the algorithms are utilized ethically and properly and for ensuring that farmers who are afflicted by paddy disease are connected to the right resources and help.

Overall, the use of machine learning techniques for detecting paddy leaf disease raises a number of ethical concerns that must be carefully considered and addressed. It is important to ensure that the techniques are developed and used in an ethical and responsible manner, in order to maximize their potential benefits while minimizing any negative impacts.

5.4 Sustainability Plan

In order to make sure that the use of these approaches is sustainable over time, a sustainability strategy for applying machine learning techniques for diagnosing paddy leaf diseases using raw field and online resource data should take a number of aspects into account. A sustainability strategy must to have the following important components:

1. **Data privacy and security:** For machine learning approaches for identifying paddy leaf disease to remain viable, social media data privacy and security must be guaranteed. It's crucial to take into account the potential privacy implications of identifying someone as being at risk for leaf disease based on their field's raw data collected by the farmer.

2. **Accuracy and reliability:** The sustainability of the approaches also depends on ensuring the precision and dependability of the machine learning algorithms. This involves continuously examining and upgrading the algorithms as necessary, as well as extensively testing and verifying the algorithms to guarantee their correctness.

3. **Responsibility:** For machine learning techniques to be sustainable, it is essential to clearly define the roles and responsibilities of persons involved in their development and application. This might entail laying down precise standards for the moral use of the

methods as well as outlining the duties and responsibilities of individuals in charge of making sure that they are applied morally and correctly.

In order to make sure that the use of these approaches is sustainable in the long run, a sustainability strategy for the use of machine learning techniques for detection and prediction utilizing raw and online resource data should take a variety of aspects into account.

CHAPTER 6

Conclusion and Future Research

6.1 Summary of the Study

We've learned a lot about this subject thanks to this investigation. Paddy's lead illness is still a delicate subject right now. This is a significant factor in the yearly decline in crop productivity. This is the reason we utilized ML to identify roughly three illnesses that pose a significant hazard.

As we've already indicated, to get the most data on this subject for our study, we combined data from the web with field data. With the use of this data, we were able to train and educate our algorithms on the illness pattern, which improved our ability to identify diseases. A few issues were resolved at the start.

The objective we set out to accomplish was accomplished. Different algorithms produce different results for humans. In the section after that, we went into further detail.

6.2 Conclusion

According to our work, the study findings and procedures we employed are excellent. After this study is complete, we believe and hope that it will spur further research in this area. We will have a lot of ideas for increasing our work thanks to this study. While working, we came across a few mistakes. We discovered additional directions in which this investigation may go. It will enable us to address any faults or other issues that come up while we work on this project. Additionally, we are considering how to integrate algorithms in the future to provide more effective solutions to the issues raised by this study. Through this research, we will be able to learn more about our area of study. We anticipate that it will further the development of paddy-lead disease detection and innovative approaches to technology that enable us to contribute to agriculture. We hope to propose a new system for paddy lead disease detection based on this study.

6.3 Possible impacts

We think that the results of our work vary depending on where they are seen. Due to this, we took extra care to do our tasks correctly. We took care to ensure the integrity of our work. We choose to utilize a dataset of our own for this reason. Our work may be used in a variety of sectors, including the advancement of agriculture and computer science. Farmers, especially those working in the agriculture industry, would benefit from it. It may also simplify our lives and protect agriculture. It could support agricultural yield.

6.4 Implications of Further Study

There are many possibilities for further study in this field, including in our own work. We found different ways to make our work better. As we mentioned, we have also found some errors, and these errors are ways to make our study better. We will try to fix this error by implementing our work and increasing prediction results with better opportunities to catch good accuracy. We have plans to fix those errors we get.

We also have other goals in mind. We would like to make our prediction better. That is why we can use an algorithm to work with an application that people can use to predict paddy leaf disease, and we will add features like disease detection types and other tools to help the user get the most out of this.

We think we can use this kind of work to make the agriculture sector more technology-based and increase the production rate of crops. We might help the people grow more crops.

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 paddy leaf disease detection," International Journal of Advanced Computer Science and Applications, 12(1), 2021.

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

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

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

[6] R. Deshmukh, and M. Deshmukh, "Detection of paddy 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 paddy leaf disease detection, classification and diagnosis in crop health monitoring unit," In Computer Aided Intervention and Diagnostics in Clinical and Medical Images (pp. 281-291),2019.

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

[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.

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

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

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

Paddy leaf disease detection using machine learning technique

ORIGINALITY REPORT

17%	14%	7%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	10%
2	Submitted to Daffodil International University Student Paper	1%
3	Md. Ashiqul Islam, Md. Nymur, Muhammad Shamsojjaman, Shazid Hasan, Md. Shahadat, Tania Khatun. "An Automated Convolutional Neural Network Based Approach for Paddy Leaf Disease Detection", International Journal of Advanced Computer Science and Applications, 2021 Publication	1%
4	www.researchgate.net Internet Source	<1%
5	Submitted to Visvesvaraya Technological University, Belagavi Student Paper	<1%
6	tnsroindia.org.in Internet Source	<1%
7	Submitted to University of Westminster	