PESTICIDE SUGGESTION AND CROP DISEASE CLASSIFICATION USING MACHINE LEARNING

A PROJECT REPORT

Submitted by

FARZANA.A.J [REGISTER NO:211417104062]

YAMINI.B [REGISTER NO:211417104308]

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

PANIMALAR ENGINEERING COLLEGE,
CHENNAI-600123.

ANNA UNIVERSITY: CHENNAI 600 025

MARCH 2021

BONAFIDE CERTIFICATE

Certified that this project report **"PESTICIDE SUGGESTION AND CROP DISEASE CLASSIFICATION USING MACHINE LEARNING"** is the bonafide work of "**AJ .Farzana (211417104062)**,**B .Yamini (211417104308)"** who carried out the project under my supervision.

SIGNATURE

Dr.S. MURUGAVALLI, M.E., Ph.D.,
HEAD OF THE DEPARTMENT

DEPARTMENT OF CSE
PANIMALAR ENGINEERING
COLLEGE,

NAZARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

SIGNATURE

R. SALINI M.tech.,
SUPERVISOR,
ASSISSTANT PROF.

DEPARTMENT OF CSE
PANIMALAR ENGINEERING
COLLEGE,

NAZARATHPETTAI,
POONAMALLEE,
CHENNAI-600 123.

Certified that the above candidate(s) was/ were examined in the Anna University Project Viva-Voce Examination held on.....

INTERNAL EXAMINER

EXTERNALEXAMINER

ACKNOWLEDGEMENT

We express our deep gratitude to our respected Secretary and Correspondent **Dr.P.CHINNADURAI, M.A., Ph.D**. for his kind words and enthusiastic motivation, which inspired us a lot in completing this project.

We would like express our heartfelt and sincere thanks to our Directors**Tmt.C.VIJAYARAJESWARI, Dr.C.SAKTHIKUMAR, M. E, Ph.D.,**and **Tmt.SARANYA SREE SAKTHI KUMAR B.E., M.B.A.**, for providing us with the necessary facilities for completion of this project.

We also express our gratitude to our Principal **Dr.K.Mani, M.E., Ph.D**. for his timely concern and encouragement provided to us throughout the course.

We thank the HOD of CSE Department, **Dr. S. MURUGAVALLI, M.E., Ph.D.,** for the support extended throughout the project.

We would like to thank my Project Guide **Mrs.Salini**, Assistant **Professor Cse Department** and all the faculty members of the Department of CSE for their advice and suggestions for the successful completion of the project.

NAME OF THE STUDENTS,

FARZANA.A. J

YAMINI.B

ABSTRACT

Crop cultivation plays an essential role in the agricultural field. Presently, the loss of food is mainly due to infected crops, which reflexively reduces the production rate. To identify the plant diseases at an untimely phase is not yet explored. The main challenge is to reduce the usage of pesticides in the agricultural field and to increase the quality and quantity of the production rate. Proposed system explores the leaf disease prediction at an untimely action. We propose an enhanced Machine Learning to predict the infected area of the leaves. A color-based segmentation model is defined to segment the infected region and placing it to its relevant classes. Experimental analyses were done on samples images in terms of time complexity and the area of infected region. Plant diseases can be detected by image processing technique. Disease detection involves steps like image acquisition, image pre-processing, image segmentation feature extraction and classification.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

CNN-CONVOLUTION NEURAL NETWORK
DNN-DEEP NEURAL NETWORK
HOG-HISTORY OF AN ORIENT GRADIENT
UML-UNIFIED MODELLING LANGUAGE

Chapter 1

INTRODUCTION

1.1 Overview:

India is a land of Agriculture. Agriculture plays an important role because of rapid growth of population and increase in demand for food. It is the basic foundation of economic development of the country. Plant diseases seriously affect the normal growth of plants, the yield and quality of agricultural products. In recent years, with the dramatic changes in climate, the natural environment of the plant growth has been damaged by pollution, frequent natural disasters. Normally, we can identify diseases by our naked eye but it is difficult to tell how the disease has occurred. In this proposed system, different image processing and machine-learning techniques called Support Vector Machine is used in the identification of rice plant diseases based on im- ages of disease infected rice plants.

One third of the population of India depends on the agriculture so we need more crop productivity for the entire population. But one of the major reasons for reduction in crop productivity is the plants gets effected to disease. The diseased should be identified and particular pesticides should be used for that particular disease to avoid unwanted soil pollution. The main objective is to detect rice plant diseases in advance and notify the name of the disease so that necessary precautions and pesticides can be used according to the disease type. So, for better identification of diseases, we can use machine learning techniques to ensure high accuracy.

Machine learning (ML) is the scientific study of algorithms and statistical

models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. It is seen as a subset of artificial intelligence. Machine learning build, known as" training data", in order to make predictions or decisions without Machine learning algorithms are used in a wide variety of applications, such as email filtering and computer vision, where it is difficult or infeasible to develop a conventional algorithm for effectively performing the task. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it learn for themselves.

Identifying the disease through naked eye is quite a difficult task that too at the early stage. As the technology has been developed now it is a simple task to solve these kinds of problems. So, to help the farmer to identify the diseases at the early stage, we have developed an interesting system that identifies and detects the disease with the help of images. By seeing the image, it displays the name of the disease for the rice plant. Thus, earlier plant damages caused by plants can be reduced easily by using appropriate pesticides and fertilizers.

1.2 Problem Statement:

First of all, we have collected the dataset from uci repository. The dataset comprises of various diseases namely Bacterial Leaf Blight, Brown Spot, Leaf Blight. Each category has 40 images each in that dataset. In this project we have developed a system to identify and detect the diseases using Image Processing and Machine Learning Techniques. Now the images are subjected to preprocessing steps like noisy data removal, feature extraction and improving image quality. Then the data is divided as training data and testing data. We have used SVM (Support Vector Machine) algorithm which has been trained separately using training dataset and test using test datasets.

Chapter 2

LITERATURE REVIEW

This chapter gives the overview of literature survey. This chapter represents some of the relevant work done by the researchers. Many existing techniques have been studied by the researchers on plant disease detection using Machine learning, we review some of them below.

The author in Xin Yang, Ting wei Guo presents a prototype system for detection and classification of rice diseases based on the images of infected rice plants. This prototype system is developed after detailed experimental analysis of various techniques used in image processing operations. They consider three rice plant diseases namely Bacterial leaf blight, Brown spot, and Leaf smut. They captured images of infected rice plants using a digital camera from a rice field. And empirically evaluated four techniques of background removal and three techniques of segmentation. To enable accurate extraction of features, they proposed centroid feeding-based K-means clustering for segmentation of disease portion from a leaf image. They enhanced the output of K-means clustering by removing green pixels in the disease portion and extracted various features under three categories: color, shape, and texture.

They used Support Vector Machine (SVM) for multi-class classification

and achieved 93.33 percentage accuracy on training dataset and 73.33 percentage accuracy on the test dataset and also performed 5 and 10-fold cross-validations, for which we achieve 83.80 percentage and 88.57 percentage accuracy, respectively.

In this work, the author Yang Lu, Shujuan Yi, presents a machine learning algorithm to find the symptoms of the disease in the rice plant. Automatic detection of plant disease is carried out using machine learning algorithm. Images of healthy and blast disease affected leaves are taken for the proposed system. The features are extracted for the healthy and disease affected parts of the rice leaf. The total data set consists of 300 images and divided for training and testing purposes. These images are processed with the proposed method and the leaf is categorized as either infected or healthy. The simulation results provide an accuracy of 99 percentage for the blast infected images and 100 percentage for the normal images during the training phase. The testing phase accuracy is found to be 90 percentage and 86 percentage for the infected and healthy images.

In this work, the author P. R. Rothe and R. V. Kshirsagar makes use of Random Forest in identifying between healthy and diseased leaf from the data sets created. This proposed paper includes various phases of implementation namely dataset creation, feature extraction, training the classifier and classification. The created datasets of diseased and healthy leaves are collectively trained under Random Forest to classify the diseased and healthy images. For extracting features of an image, they used Histogram of an Oriented Gradient (HOG). Overall, using machine learning to train the large data sets available publicly gives a clear way to detect the disease present in plants in a colossal scale.

In this work, author Aakanksha Rastogi, Ritika Arora and Shanu Sharma main goal was to propose improvement in existing classification techniques for plant leaf disease detection using machine learning and that is proved using KNN classifier over SVM classifier. They made past studies in which, 42 percentage of agricultural production is in loss and that too only because of the increasing rate of loss due to plant leaf diseases. To overcome this major issue, this plant leaf disease detection technique can be applied to detect a disease from the input images. This process involved steps like image preprocessing, image segmentation, feature extraction. Further K Nearest Neighbor (KNN) classification is applied on the outcome of these three stages. Proposed implementation has shown 98.56 per- centage of accuracy in predicting plant leaf diseases. It

also presents other information regarding a plant leaf disease that is Affected Area, Disease Name, Total Accuracy, Sensitivity and Elapsed Time.

In this work, the authors S. S. Sannakki and V. S. Rajpurohitactualizes a machine learning system to recognize the type of malady on various plant species where phases include dataset acquisition, feature extraction, training and classification. The datasets of both healthy and diseased leaves are trained using various machine learning classifiers. Feature extraction is done using Histogram of Oriented Gradient (HOG) where features of color, shape and texture are extracted. Then, the accuracy of all classifiers is com- pared to obtain the best classifier. Random Forests is found to be the best classifier for all variety of plants. Later, a test image of leaf is supplied to the classifier to classify whether the leaf is diseased or not and retrieve the disease type.

The authors in present an algorithm based on machine learning, which can self-learn, so that there is no requirement of a specific background or image quality.

Chapter 3

System Analysis

3.1 Existing System

The existing system consists of temperature, humidity, pH and soil moisture sensors for collecting data from soil based on variation in temperature, humidity, pH level of water and soil moisture content present plant leaves. The data collected consists of current environ- mental factors like temperature, humidity, pH level, and soil moisture content. The data collected from temperature, humidity, soil moisture and pH sensors are given to raspberry pi from which the information is communicated to the farmers using LCD. They used the DHT11 to sense the temperature on the surface of leaf to determine whether it is healthy or diseased sensors using raspberry interface. The output is displayed by means of LCD or a real application of IoT is used known as blynk application. The blynk application deals with data that is being displayed on the LCD. The sensor data that is displayed on the LCD screen is being displayed in this application.

3.2 Proposed System

The proposed system objective is to detect rice plant diseases in advance and notify the name of the disease so that necessary precautions and pesticides can be used according to the disease type. The initial step of the proposed system is collection of datasets. The dataset comprises of three diseases namely leaf smut, bacterial leaf blight and brown spot. Next the step is image preprocessing in which the noise is removed and the quality of the image is enhanced. All the libraries are installed in this

preprocessing step like cv2, NumPy, pandas, seaborn, jupyter etc., A classification is selected called SVM. From that dataset collected we train the machine using train dataset. This algorithm then classifies the disease type and gives the output of disease name.

Advantages

The proposed system is an advanced form of the existing once. The proposed system uses SVM algorithm to depict the disease name and notify the farmer about disease.

3.3 Requirement Analysis and Specifications

3.3.1 Feasibility Study

Feasibility Study is assessment of the proposed project plan or method. It tells that the project should be a burden and finds whether the project is feasible or not by using the some considerations such as:

1. Economic Feasibility.
2. Technical Feasibility.
3. Social Feasibility.

3.3.2 Economic Feasibility

This is completely a software-based project which does not require to buy any license or products. So that the project can be done within the given project budget.

3.3.3 Technical Feasibility

We use Machine Learning algorithm (SVM) in this project for better result i.e., it will classify the images according to its belonging class.

3.3.4 Social Feasibility

This project helps the farmers by predicting the disease name so that farmers can use required pesticides for different diseases.

3.4. System Specification

Hardware Requirements

- Processor - Intel Core i7
- RAM - 4 GB
- Hard Disk - 500 GB

Software Requirements

- Operating System - Windows 7/8/10
- Programming Language: Python, MySQL

3.4.1 Standards and Policies

Jupyter notebook

The Jupyter Notebook is an open-source web application that al- lows you to create and share documents that contain live code, conditions, representations, and content.

Standard Used: ISO/IES 6090

PYTHON

Python is the interpreted, high level programming language that is mostly used for machine learning problems. Coding part will be less so it is

considered as easiest language now in this project we used python version3.6.1.

Standard Used: ISO/IEC WD TR 24772-4

3.5 Software Specification

3.5.1 Machine learning

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people. Although machine learning is a field within computer science, it differs from traditional computational approaches.

In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs. Any technology user today has benefitted from machine learning. Facial recognition technology allows social media platforms to help users tag and share photos of friends. Optical character recognition (OCR) technology converts images of text into movable type. Recommendation engines, powered by machine learning, suggest what movies or television shows to watch next based on user preferences. Self-driving cars that rely on machine learning to navigate may soon be available to consumers. Machine learning is a continuously developing field. Because of this, there are some

considerations to keep in mind as you work with machine learning methodologies or analyze the impact of machine learning processes.

Here in this thesis, we are providing basic info of the common machine learning methods of supervised and unsupervised learning, and common algorithmic approaches in machine learning, including the k-nearest neighbor algorithm, decision tree learning, and deep learning**.**

3.5.2 SUPERVISED LEARNING

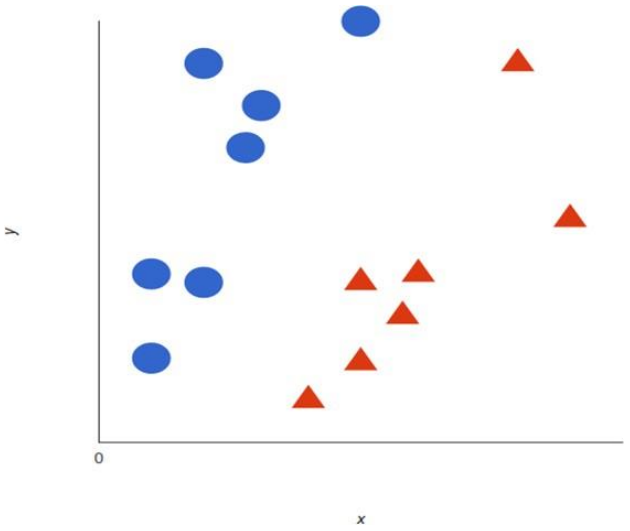
In machine learning, tasks are generally classified into broad categories. These categories are based on how learning is received or how feedback on the learning is given to the system developed. Two of the most widely adopted machine learning methods are supervised learning which trains algorithms based on example input and output data that is labeled by humans, and unsupervised learning which provides the algorithm with no labeled data in order to allow it to find structure within its input data. Let's explore these methods in more detail. The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output Y = f(X). The goal is to approximate the mapping function so well that when you have new input data. That you can predict the output variables (Y) for that data. Techniques of Supervised Machine Learning algorithms include linear and logistic regression, multi-class classification, Decision Trees and support vector machines.

Supervised learning requires that the data used to train the algorithm is already labeled with correct answers. For example, a classification algorithm

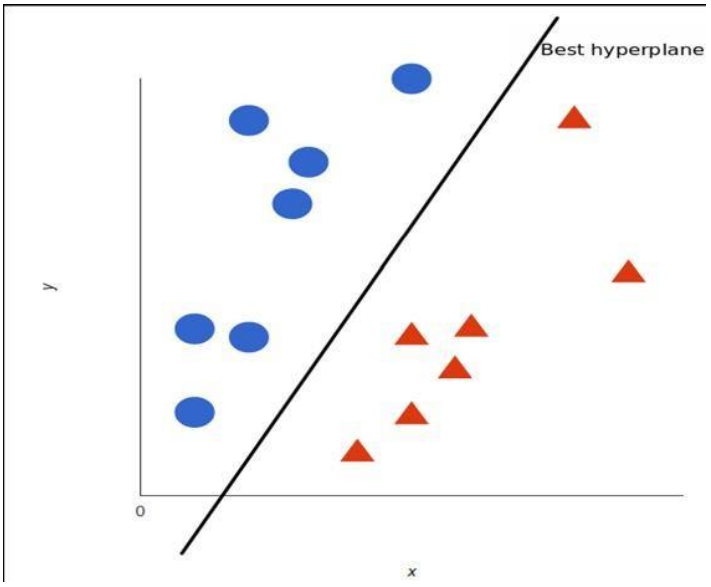
will learn to identify animals after being trained on a dataset of images that are properly labeled with the species of the animal and some identifying characteristics. Supervised learning problems can be further grouped into Regression and Classification problems. Both problems have as goal the construction of a succinct model that can predict the value of the dependent attribute from the attribute variables. The difference between the two tasks is the fact that the dependent attribute is numerical for regression and categorical for classification.

Support Vector Machine:

A support vector machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group classification problems. After giving an SVM model sets of labeled training data for each category, they're able to categorize new text. The basics of Support Vector Machines and how it works are best understood with a simple example. Let's imagine we have two tags: red and blue, and our data has two features: x and y. We want a classifier that, given a pair of (x, y) coordinates, outputs if it's either red or blue. We plot our already labeled training data on a plane:

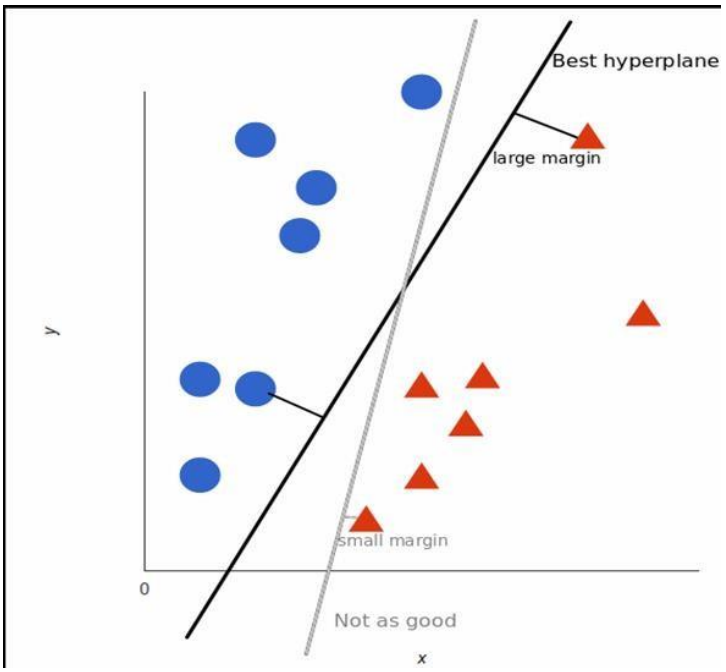


A support vector machine takes these data points and outputs the hyperplane (which in two dimensions it's simply a line) that best separates the tags. This line is the decision boundary: anything that falls to one side of it we will classify as blue, and anything that falls to the other as red.



In 2D, the best hyperplane is simply a lineBut, what exactly is the best hyperplane? For SVM, it's the one that maximizes the margins from both

tags. In other words: the hyperplane (remember it's a line in this case) whose distance to the nearest element of each tag is the largest.



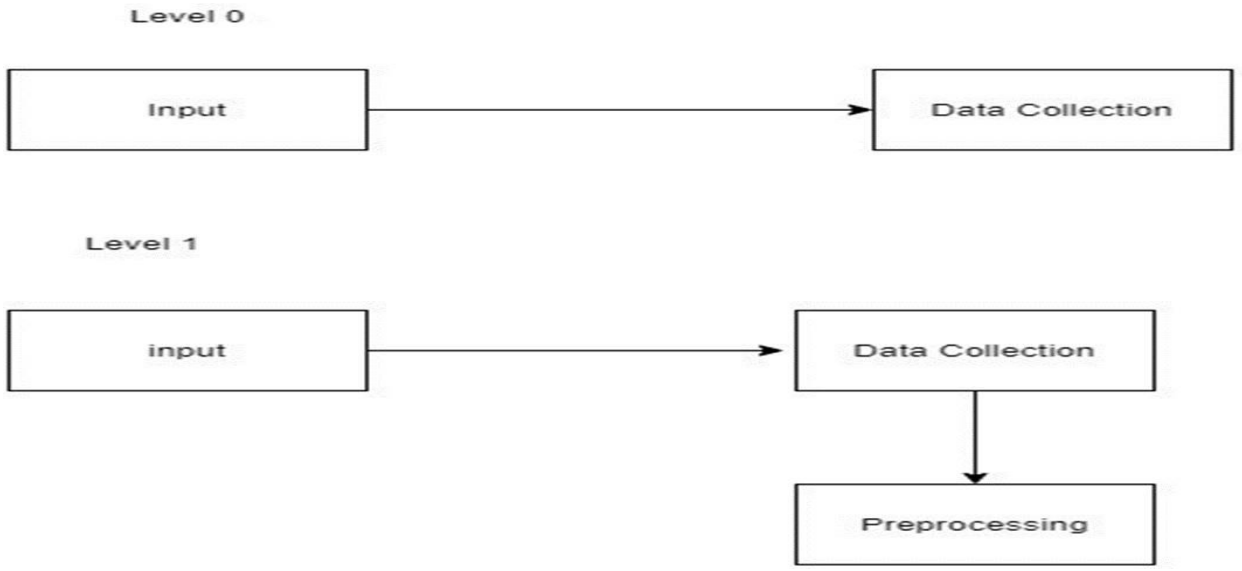
UNSUPERVISED LEARNING

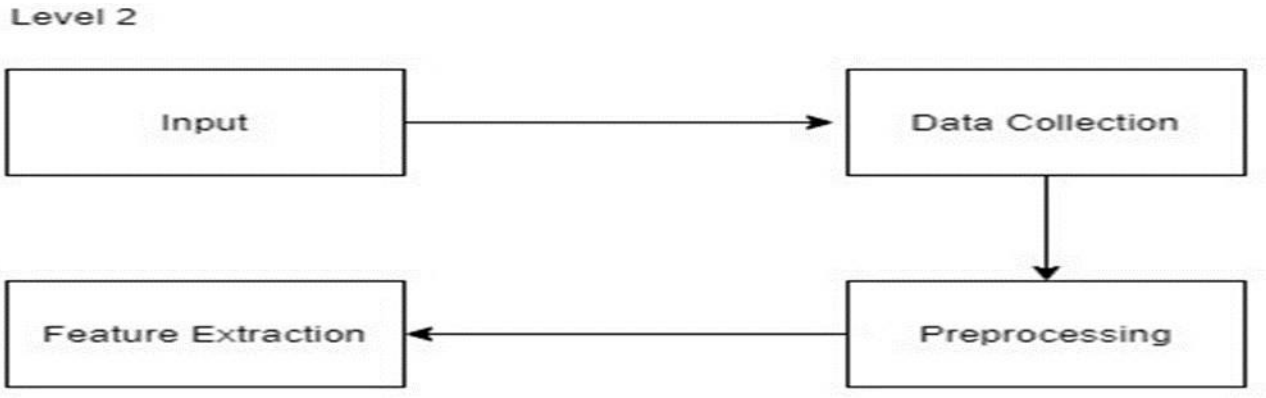
Unsupervised learning is where we only have input data (X) and no corresponding output variables. The goal for unsupervised learning is to model the underlying structure or distribution in the data in order to learn more about the data. These are called unsupervised learning because unlike supervised learning above there is no correct answer and there is no teacher. Algorithms are left to their own devises to discover and present the interesting structure in the data. Unsupervised learning problems can be further grouped into clustering and association problems.

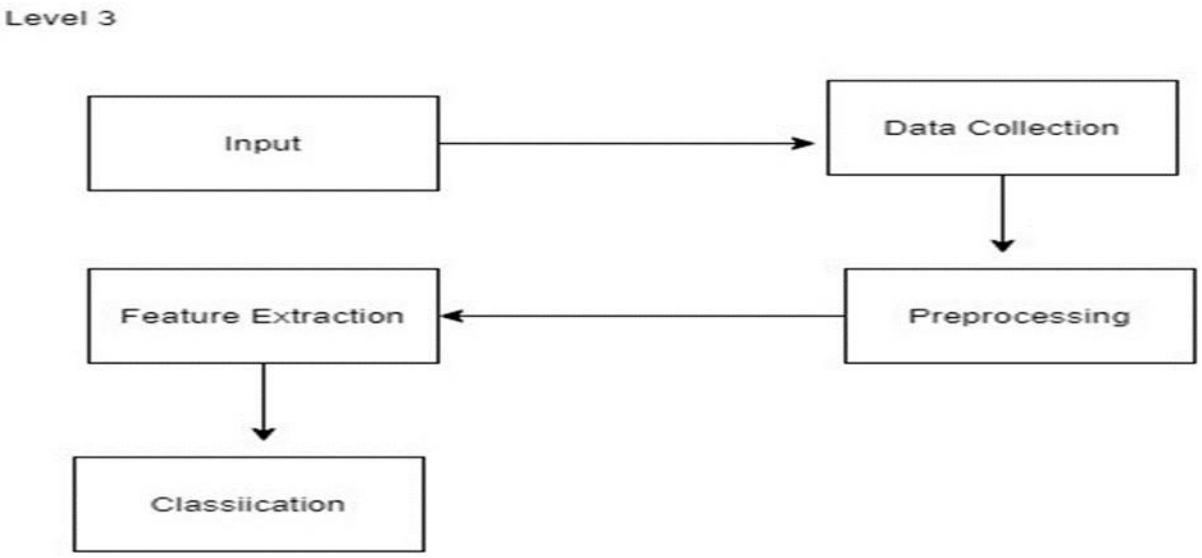
Chapter 4

System Design

4.1 Data FlowDiagram



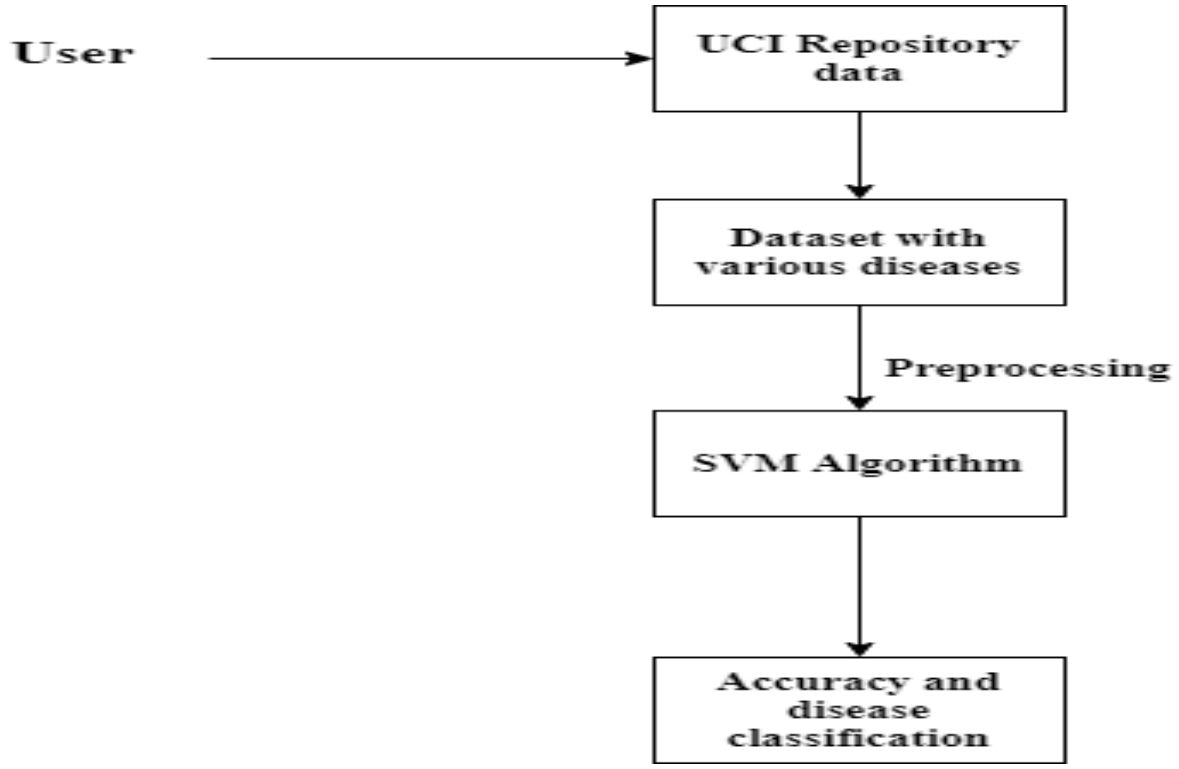




Description

The data set is splitted into training dataset and testing dataset. This training dataset undergoes some preprocessing steps which enhance the quality of the diseased part. An algorithm is used to train the data which will give an output of identification of the disease.

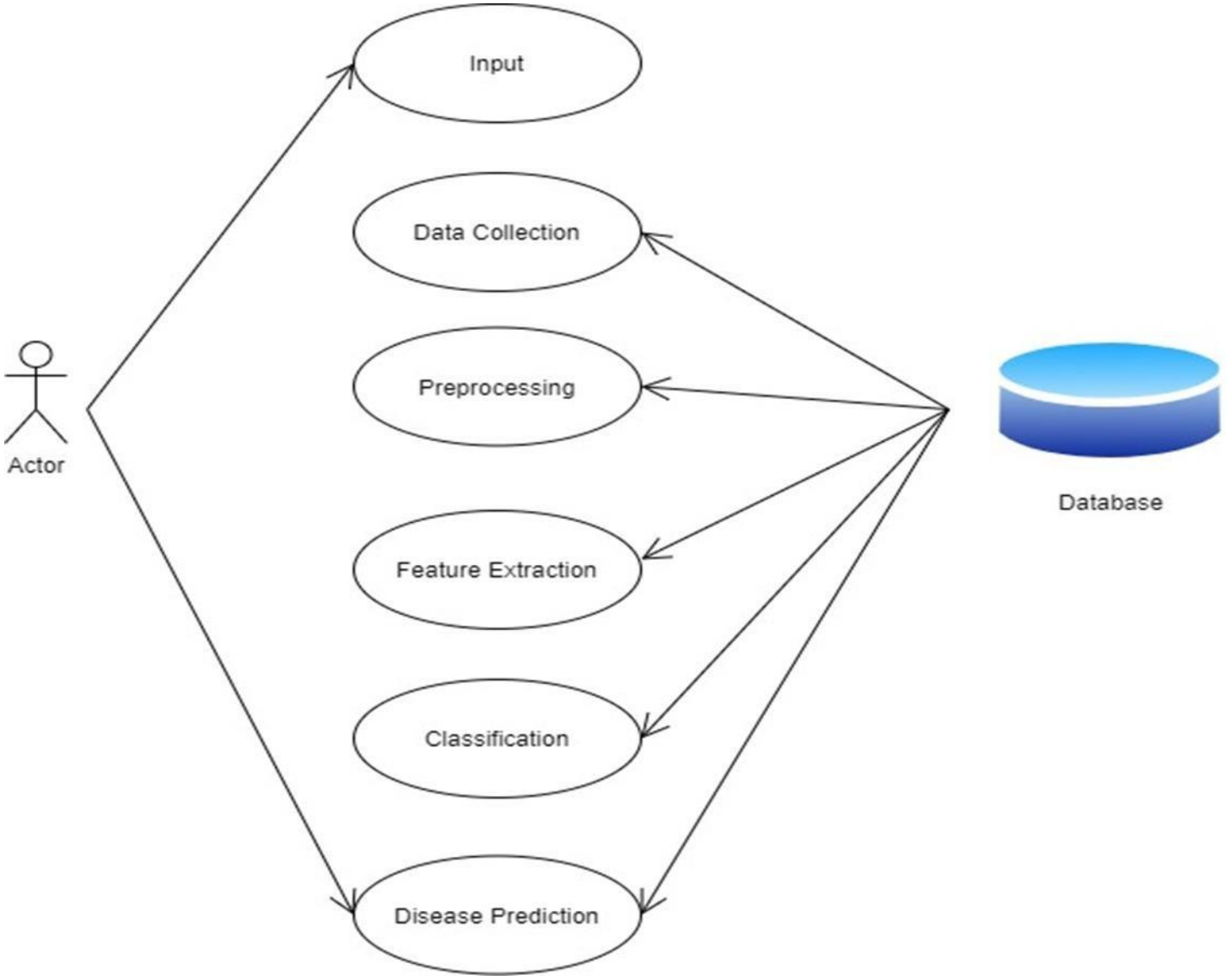
4.2 UML Diagram



Description

The user collects the data from uci repository i.e., rice plant dis- ease dataset which contains all various leaf disease data that are leaf smut, bacterial leaf blight, and brown spot. This dataset undergoes preprocessing steps and an algorithm is used to depict the name of the disease and accuracy.

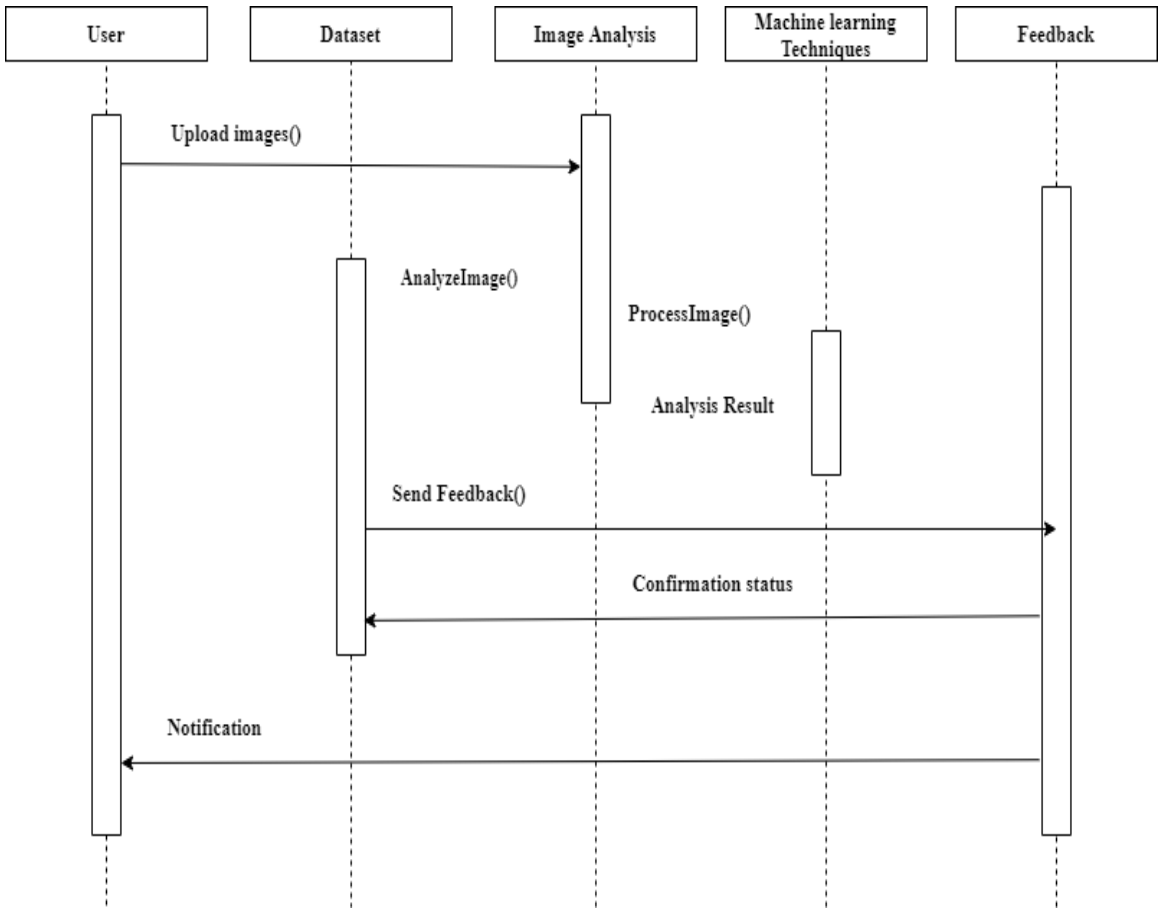
4.3 Use Case Diagram



Description

Here we are building a model for the user which is used to predict what kind of disease that has been occurred for a rice plant. User collects the dataset which have three disease names and each disease have 40 images each. System will extract features and depicts the output and accuracy of the model.

4.4 Sequence Diagram



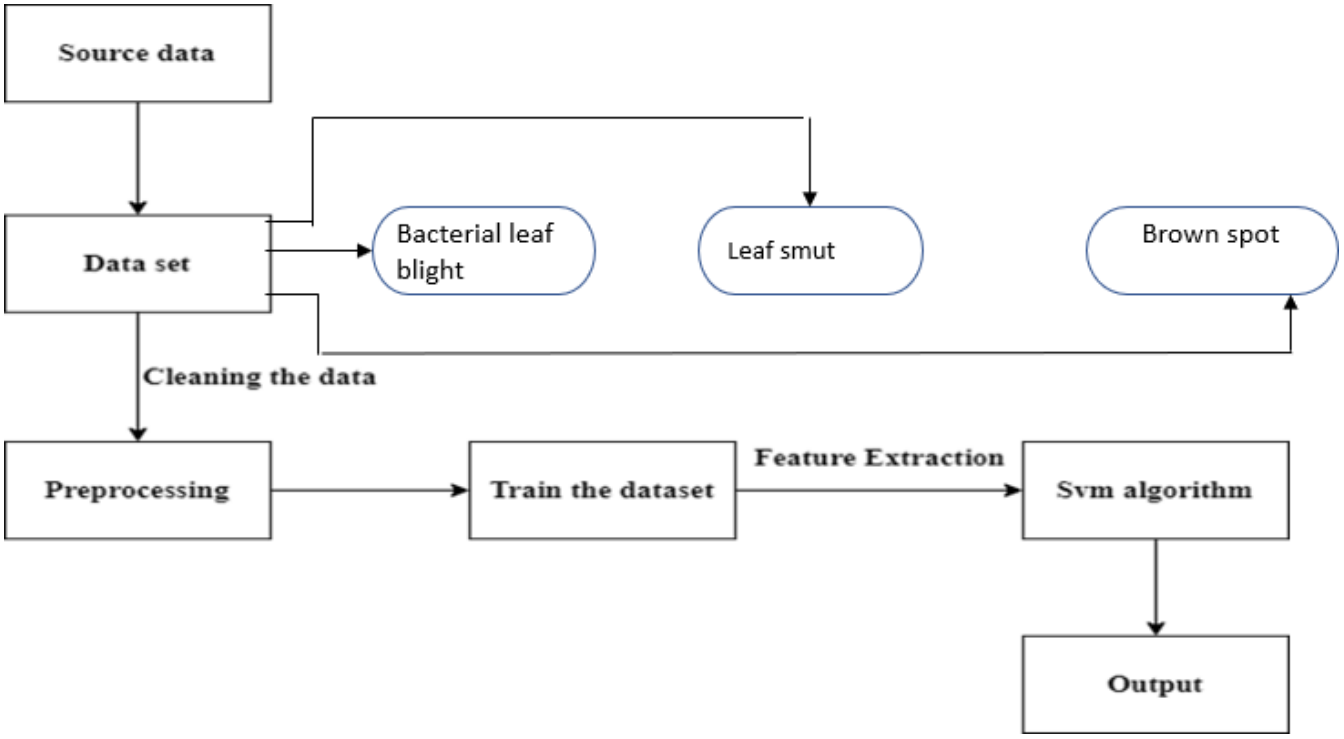
Description

The sequence diagram will give an idea of how the model is processing or trained step by step. The image data set is collected by user from uci repository and the images are analyzed and undergoes some image preprocessing techniques to increase the quality of the image. Algorithm is selected to train the data set and depicts the output i.e., as a feedback to user.

CHAPTER 5

System Architecture

5.1 Architecture overview



Description

In this diagram, we first collected the source data which is from the UCI machine learning repository. That source data has a data set which comprises of three diseased leaf images and that data set is subjected to the preprocessing steps like cleaning the data etc.

5.2 Module Description

Data Collection

- The data collection is done Kaggle. The data collected has been check one by one before taking it as input data. There are 2 classes named as affected plants and normal plants will be classified and each of the class contains images. From the 80% of images used for training and 20% of images used for validation and testing.

Data Augmentation

1) Data augmentation is a method by which you can virtually increase the number of samples in your dataset using data you already have.

2) For image augmentation, it can be achieved by performing geometric transformations, changes to color, brightness, contrast or by adding some noise. In order to let our model, adapt to a different situation that might occur noisy data such as when raining season.

Plant classification

After training for each model, the final two models were generated. One model was trained directly based on original images, while the other one was created using image data produced by data augmentation. Model history was used to record the training process. which displays the accuracy on the training set and testing set, it is described that the accuracy for the training set is having differences with 5% by with and without data augmentation. Besides, the performance of the model without data augmentation having lower accuracy on the validation set, while the model with data augmentation has a higher result in the testing set.

CHAPTER 6

IMPLEMENTATION

Implementation is the stage of the project when the theoretical design is turned out into a working system. In our project, we have to consider all the stages like downloading or arranging all the required materials, software, etc. We should Use the collected data set to train and test, train the entire machine learning model with the help of SVM algorithms, to perform the entire task. Generally, 80 percentage of the entire data set is considered as training part.

6.1 Input and Output

6.1.1 Input Design

The input design includes analyzing the data set that is collected from the UCI machine learning repository. Then, with the help of the dataset we prepared a machine learning model to perform the operations like training, testing the data. The algorithm which we used for training is Support Vector Machine.

6.1.2 Output Design

Test Data

The data set which was collected has been splitted into two parts they are training and testing data. Testing part is used for sample checking purposes whether the training model which we created is showing accurate results or not.

Train Data

The training data is used to develop the model and the data inside helps in training the algorithm.

6.1 Sample Code

```
from PIL import Image Tk
import PIL. Image
from tkinter import *
import tkinter as tk
from tkinter import file dialog
from skimage.io import imread
from skimage. transform import resize
import skimage, pickle
windo = Tk ()
windo. Configure(background='white')
windo. Title ("Rice Disease Prediction App")
# width = windo. winfo_screenwidth ()
```

```
# height = windo. winfo_screenheight ()

# windo. Geometry(f'{width}x{height}')

windo. Geometry('1120x820')

windo. iconbitmap('./images/rice.ico')

windo. Resizable(0,0)


#Size for displaying Image

w = 650; h = 270

size = (w, h)


def upload_ im ():

    try:

        global im, resized, cp, path, display, image Frame, dn1

        image Frame = tk. Frame(windo)

        imageFrame.place(x=415, y=160)

        path = file dialog. askopenfilename ()

        im = PIL.Image. open(path)

        resized = im. resize (size, PIL.Image. ANTIALIAS)
```

```
        tkimage = ImageTk.PhotoImage(resized)
        display = tk.Label(image Frame)
        display. imgtk = tkimage
        display. Configure(image=tkimage)
        display. Grid ()
        dn1 = tk.Label(windo, text='Original\ud83d\ude80 Image ', width=20, height=1, fg="white", bg="brown4",
                       font=('times', 22, ' bold '))
        dn1.place(x=570, y=120)
        cp = tk.Button(windo, text='Predict\ud83d\ude80 Disease', bg="brown4", fg="white", width=20,
                       height=1, font=('times', 22, 'italic bold '),command = prediction,activebackground = 'yellow')
        cp.place(x=570, y=440)
    except Exception as e:
        print(e)
        noti = tk.Label(windo, text = 'Please upload an Image\ud83d\ude80 File', width=33, height=2, fg="white", bg="brown4",
                        font=('times', 23, ' bold '))
```

```
        noti.place(x=454, y=540)
        windo.after(5000, destroy_widget, noti)


    def destroy_widget(widget):
        widget.destroy()


    def prediction():
        windo.after(2000, destroy_widget, cp)
        def load_image(im_file):
            dimension = (104, 104)
            flat_data = []
            img = skimage.io.imread(im_file)
            img_resized = resize(img, dimension, anti_aliasing=True,
mode='reflect')
            flat_data.append(img_resized.flatten())
            return flat_data
        img = load_image(path)
        try:
            with open('./model/rice_pred.pkl', 'rb') as f:
```

```
            clf = pickle.load(f)
            pred = clf.predict(img)
        except:
            noti = tk.Label(windo, text = 'Model not Found', width=33, height=2, fg="white", bg="brown4",
                            font=('times', 23, ' bold '))
            noti.place(x=454, y=580)
            windo.after(5000, destroy_widget, noti)
        labels = ['Bacterial leaf blight', 'Brown spot', 'Leaf smut']
        s = [str(i) for i in pred]
        a = int("".join(s))
        lab = str("Predicted Disease is "+ labels[a])
        pred = tk.Label(windo, text=lab, width=33, height=2, fg="white",
                        bg="dark green",
                        font=('times', 23, ' bold '))
        pred.place(x=454, y=540)
        windo.after(7000, destroy_widget, pred)
        windo.after(7000, destroy_widget, display)
        windo.after(7000, destroy_widget, imageFrame)
```

```
    windo.after(7000, destroy_widget, dn1)
    ri = PIL.Image.open('./images/rice.png')
    ri =ri.resize((351,263), PIL.Image.ANTIALIAS)
    sad_img = ImageTk.PhotoImage(ri)
    panel4 = Label(windo, image=sad_img,bg = 'white')
    panel4.pack()
    panel4.place(x=20, y=170)

    up = tk.Button(windo,text = 'Upload\ud83d\ude80 Image',bg="brown4",
fg="white", width=20,
                   height=1, font=('times', 22, 'italic bold '),command =
upload_im, activebackground = 'yellow')
    up.place(x=20, y=440)

    pred = tk.Label(windo, text="Rice Disease Prediction", width=30,
height=2, fg="white",bg="dark green",
                    font=('times', 25, ' bold '))
    pred.place(x=254, y=20)

    windo.mainloop()
```

Chapter 7

System Testing

7.1 Testing Techniques:

Testing is a process of executing a program with the intent of finding an error. A good test case is one that has a high probability of finding an as-yet – undiscovered error. A successful test is one that uncovers an as-yet-undiscovered error. System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. It verifies that the whole set of programs hang together. System testing requires a test consists of several key activities and steps for run program, string, system and is important in adopting a successful new system. This is the last chance to detect and correct errors before the system is installed for user acceptance testing. The software testing process commences once the program is created and the documentation and related data structures are designed. Software testing is essential for correcting errors. Otherwise, the program or the project is not said to be complete. Software testing is the critical element of software quality assurance and represents the ultimate the review of specification design and coding. Testing is the process of executing the program with the intent of finding the error. A good test case design is one that as a probability of finding a yet undiscovered error. A successful test is one that uncovers a yet undiscovered error. Any engineering product can be tested in one of the two ways:

WHITE BOX TESTING

This testing is also called as Glass box testing. In this testing, by

knowing the specific functions that a product has been design to perform test can be conducted that demonstrate each function is fully operational at the same time searching for errors in each function. It is a test case design method that uses the control structure of the procedural design to derive test cases. Basis path testing is a white box testing**.**

Basis path testing:

- Flow graph notation
- Kilometric complexity
- Deriving test cases
- Graph matrices Control

BLACK BOX TESTING:

In this testing by knowing the internal operation of a product, test can be conducted to ensure that “all gears mesh”, that is the internal operation performs according to specification and all internal components have been adequately exercised. It fundamentally focuses on the functional requirements of the software.

Basis Path:

- Graph based testing method
- Equivalence Partitioning
- Boundary value Analysis

- Comparison testing

7.2 SOFTWARE TESTING STRATEGIES:

A software testing strategy provides a road map for the software developer. Testing is a set activity that can be planned in advance and conducted systematically. For this reason, a template for software testing a set of steps into which we can place specific test case design methods should be strategy should have the following characteristics:

Testing begins at the module level and works “outward” toward the integration of the entire computer-based System.

INTEGRATION TESTING: defined as a type of testing where software modules are integrated logically and tested as a group. A typical software project consists of multiple software modules, coded by different programmers. The purpose of this level of testing is to expose defects in the interaction between these software modules when they are integrated.

Software Engineering defines variety of strategies to execute Integration testing, they are,

- Big Bang Approach:
- Incremental Approach: which is further divided into the following

1. Top-Down Approach
2. Bottom-Up Approach

- Sandwich Approach - Combination of Top Down and Bottom Up

VALIDATION TESTING:

We check whether the developed product is right. Validation is also known as dynamic testing. Validation includes testing like functional testing, system testing, integration, and User acceptance testing. It is a process of checking the software during or at the end of the development cycle to decide whether the software follow the specified business requirements. Quality control comes under validation testing. In the validation testing, we can find those bugs, which are not caught in the verification process. In this type of testing, we can validate that the user accepts the product or not.

Test case ID	Test Cases	Priority	Pre-condition	Input Test data	Steps to be executed	Expected result	Actual result	Failure Test case
1	Collect the data	A	User must have processed and generate the data	Collected images	Process and collect from the data	Dataset should be expected using the processed data	Dataset executed from the given data	No failure
2	Sample images	B	User should have Already processed data	images	Process according through the condition	Result is expected from the processed one.	Result executed successfully from the data	No failure
3	Restored sample	c	Process according through the data	Defected data	Process and suggest the result	Identified and suggested from data	Data is executed successfully	Over fitting

Test Case Id	Test Cases	Priority	Pre-condition	Input Test data	Steps to be executed	Expected result	Actual result	Failure Test case
4	Stored sample data	D	User must have proceesed	Images	Check the data and process based on the condition	Disease is detected and executed	Acquired from source as result	Unidentified Disease
5	System must have stored data	E	Data accessed from users condition	Collected data	Process and execute condition	Detected according through the condition and identified	Result is executed successfully	Pesticide cannot be determined.

Chapter 8
CONCLUSION AND FUTURE ENHANCEMENTS

8.1 Conclusion

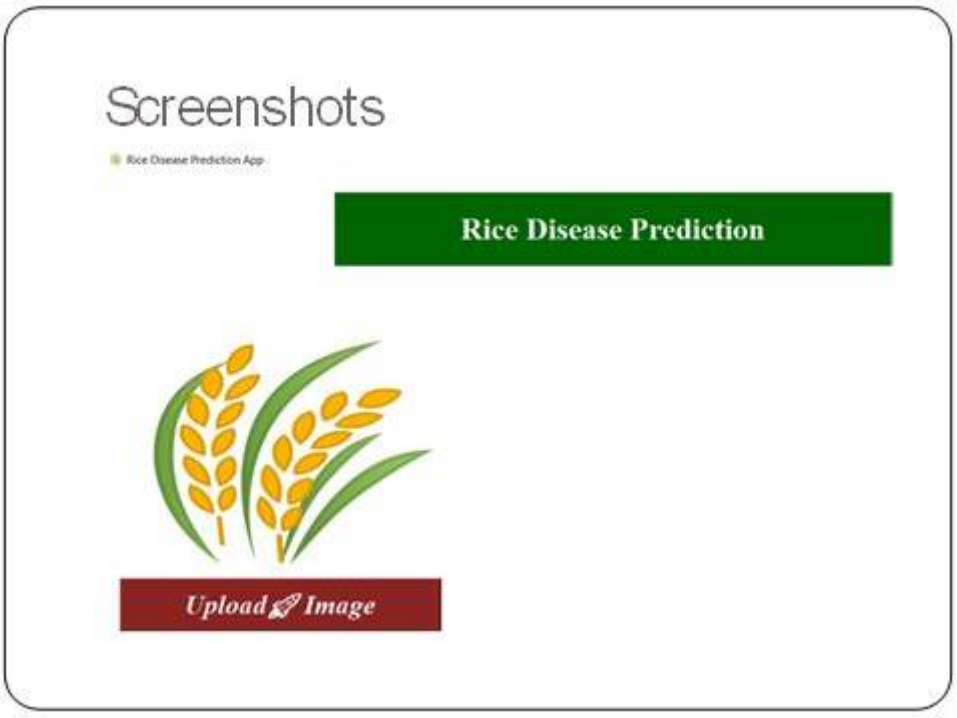
Rice plant diseases can make a big amount of loss in the agricultural domain. Due to climatic conditions and global warming issues the plants are deficient in terms of water content and necessary nutrients and easily get affected by diseases. Generally with the help of naked eye, farmers used to identify the disease, but they cannot identify what was the disease. In this project we developed a system to detect three rice plant diseases namely Bacterial Leaf Blight, Brown Spot, Leaf Blight. We have collected dataset of leaf images from uci repository. We had used Machine learning techniques to identify the diseases accordingly. So, it helps farmers to improve the quality of farming and increase the crop production.

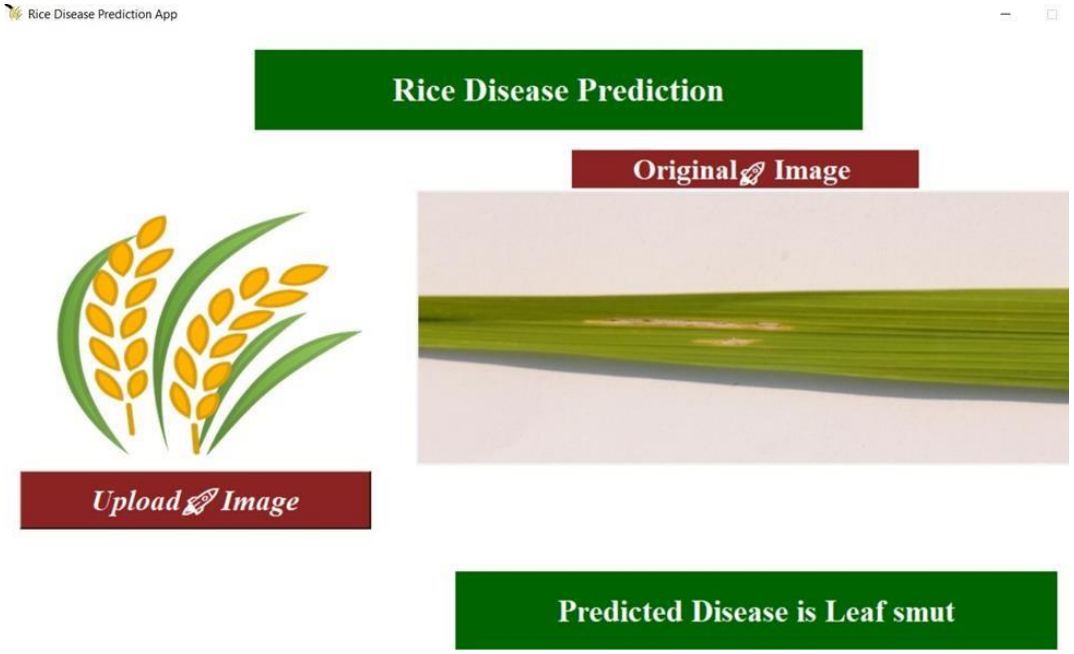
Future Enhancements

Here the work has been done with the help of one of the machine learning techniques. This system is developed for early and accurate detection of diseases. As the technology has been rapidly increasing these days, we have a lot of opportunities to solve any kind of problem. We can also further enhance this project with the help of deep learning and other techniques. In future we can go with other algorithms like random forest, logistic regression etc.., for better accuracy. Images in the datasets can be increased for proceeding with random forest and logistic regression which gives accuracy according to the number of images.

Appendixes:

A1. SCREENSHOTS:





Rice Disease Prediction App

Rice Disease Prediction

Upload Image

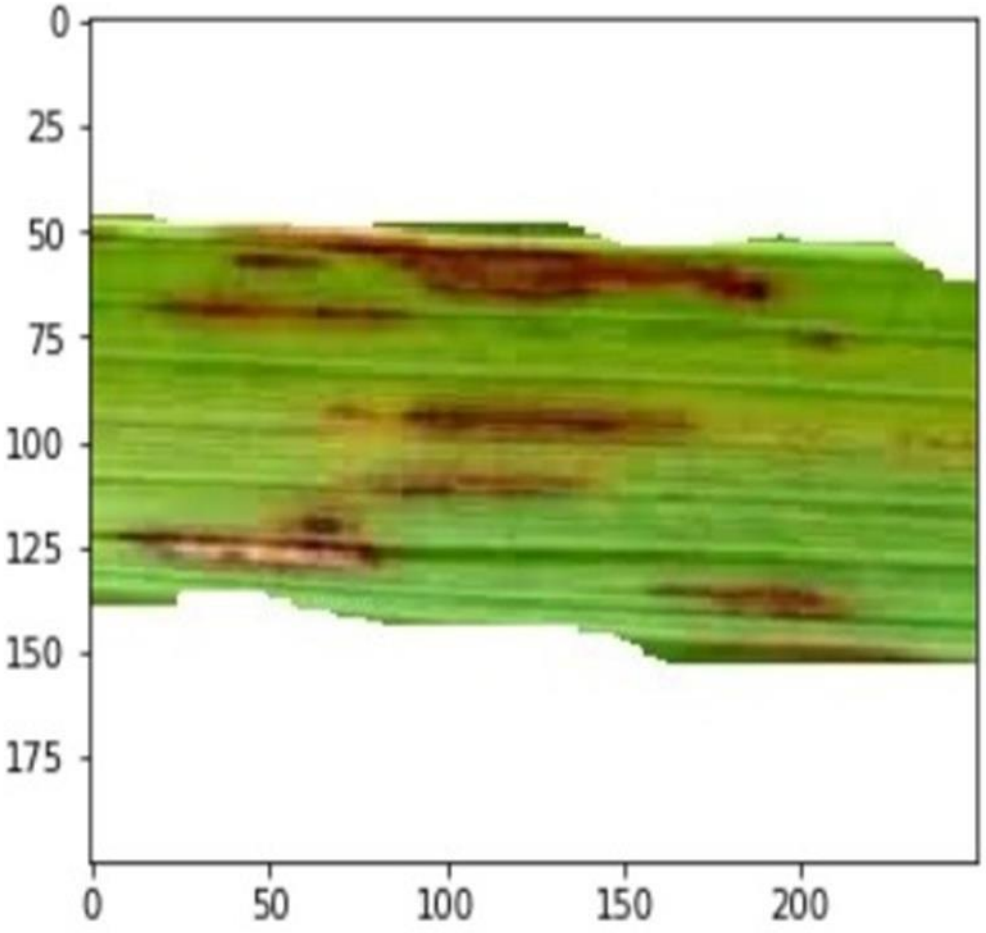
Rice Disease Prediction App

Rice Disease Prediction

Original Image

Upload Image

Predicted Disease is Leaf smut




```
Predicted Disease is  Leaf smut
```

A2.Publications

PESTICIDE SUGGESTION AND CROP DISEASE CLASSIFICATION USING MACHINE LEARNING

Abstract- Crop cultivation plays an essential role in the agricultural field. Presently, the loss of food is mainly due to infected crops, which reflexively reduces the production rate. To identify the plant diseases at an untimely phase is not yet explored. The main challenge is to reduce the usage of pesticides in the agricultural field and to increase the quality and quantity of the production rate. Proposed system explores the leaf disease prediction at an untimely action. We propose an enhanced Machine Learning to predict the infected area of the leaves. A colour-based segmentation model is defined to segment the infected region and placing it to its relevant classes. Experimental analyses were done on samples images in terms of time complexity and the area of infected region. Plant diseases can be detected by image processing technique. Disease detection involves steps like image acquisition, image pre-processing, image segmentation Feature extraction and classification.

I.INTRODUCTION

India is a land of Agriculture. Agriculture plays an important role because of rapid growth of population and increase in demand for food. It is the basic foundation of economic development of the country. Plant diseases seriously affect the normal growth of plants, the yield and quality of agricultural products. In recent years, with the dramatic changes in climate, the natural environment of the plant growth has been damaged by pollution, frequent natural disasters. Normally, we can identify diseases by our naked eye but it is difficult to tell how the disease has occurred. In this proposed system, different image pro- cessing and machine-learning techniques called Support Vector Ma- chine is used in the identification of rice plant diseases based on im- ages of disease infected rice plants. One third of the population of India depends on the agriculture so we need more crop productivity for the entire population. But one of the major reasons for reduction in crop productivity is the plants gets effected to disease. The diseased should be identified and particular pesticides should be used for that particular disease to avoid unwanted soil pollution. The main objective is to detect rice plant diseases in advance and notify the name of the disease so that necessary precautions and pesticides can be used according to the disease type. So, for better identification of diseases, we can use machine learning techniques to ensure high accuracy.

II.LITERATURE SURVEY

This chapter gives the overview of literature survey. This chapter represents some of the relevant work done by the researchers. Many existing techniques have been studied by the researchers on plant dis- ease detection using Machine learning, we review some of them below.

1.Auother: Xin Yang, Ting wei Guo, Title:" Machine learning in plant disease research", Year:2020, Method: K-means clustering used to segment the defected area Advantages: They create their own data set Disadvantage: The accuracy level is 92.2%. 2.Auother. Yang Lu,Shujuan Yi,Nianyin Zeng,Yurong Liu,Yong hang,Title of the paper:" Identification of rice diseases using deep convolutional neural networks"Year:2018,Method: we propose a novel rice diseases identification method based on deep convolutional neural networks (CNNs) techniques, Advantages: The simulation results for the identification of rice diseases show the feasibility and effectiveness Disadvantages: Less Number of dataset will be trained.3.Auother: P. R. Rothe and R. V. Kshirsagar,Title of the paper:" Cotton Leaf Disease Identification using Pattern Recognition Techniques",Year:2017, Method: Uses Snake segmentation, Hu's moments are taken as feature, Advantages: Active contour model used to minimize the energy inside the disease spot, BPNN solves the multiple class problems, average classification is found to be 85.52%.,Disadvantages: Snake segmentation is a very slow process.4.Auother: Aakanksha Rastogi, Ritika Arora and Shanu Sharma, Title of the paper:" Leaf Disease Detection and Grading using Computer Vision Technology &Fuzzy Logic",Year:2017,Method: K-means clustering used to segment the defected area; GLCM is used for the extraction of Texture features, Disease grading using fuzzy logic. Advantages: Severity of the disease is checked, Fast and highly efficient. Disadvantages: Low-level segmentation 5. Author's. S. Sannakki and

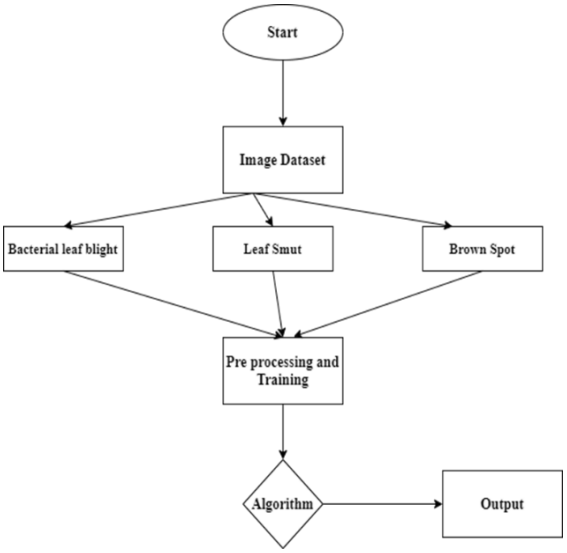
V. S. Rajpurohit,Title of the paper: "Classification of Pomegranate disease based on convolution Neural Network",Year:2015,
Method: K-means clustering used to segment the defected area, colour and texture are used as the features. Advantages: RGB image is converted to L*a*b to extract chromaticity layers of image, classification is found to be 97.30%. Disadvantage: Only applicable for limited crops.

III. PROPOSED SYSTEM

Our project is to detect the plant diseases and provide the solutions to recover from the leaf diseases. In our proposed system we are providing a solution to recover from the leaf diseases and also show the affected part of the leaf by image processing technique. The existing system can only identify the type of diseases which affects the leaf. It's not efficient. We will provide a result within fraction of seconds and guided you throughout the project. We briefly explain about the experimental analysis of our methodology. Samples of 1000 images are collected that comprised of different plant diseases like Alternaria Alternata, Anthracnose, Bacterial Blight, Cercospora leaf spot and Healthy Leaves. Different number of images is collected for each disease. The initial step of the proposed system is collection of datasets. The dataset comprises of three diseases namely leaf smut, bacterial leaf blight and brown spot. Next the step is image pre- processing in which the noise is removed and the quality of the image is enhanced. All the libraries are installed in this pre-processing step like cv2, NumPy, pandas, seaborn, jupyter etc.,. A classification is selected called SVM. From that dataset

collected we train the machine using train dataset. This algorithm then classifies the disease type and gives the output of disease name.

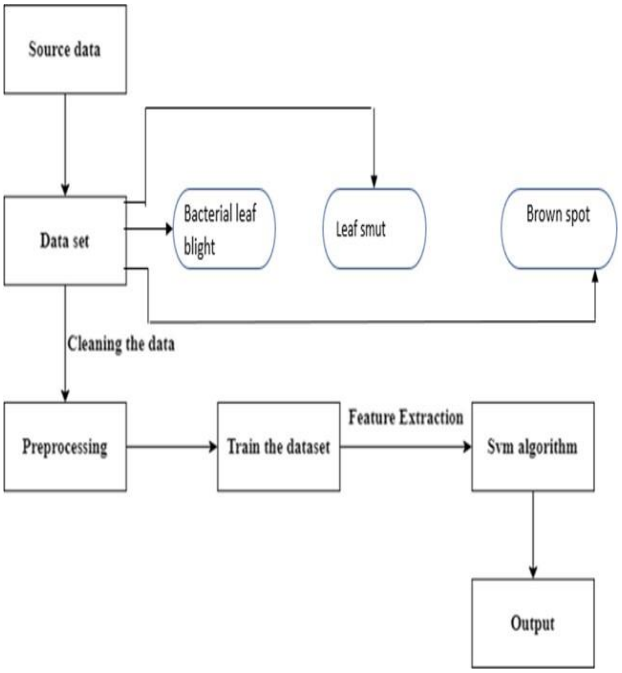
Diagram:



The input is nothing but the entire image data set which had been collected for the processing of the model.

IV. SYSTEM ARCHITECTURE

A system architecture is the conceptual model that defines the structure, behaviour, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviours of the system. A system architecture can consist of system components and the sub-systems developed, that will work together to implement the overall system. There have been efforts to formalize languages to describe system architecture, collectively these are called architecture description languages (ADLs).Diagram:



In this diagram, we first collected the source data which is from the UCI machine learning repository. That source data has a data set which comprises of three diseased leaf images and that data set is subjected to the pre-processing steps like cleaning the data etc. The input is nothing but the entire image data set which had been collected for the processing of the model. The output will be what disease that has been occurred for the plant and It also shows the accuracy of the model. The structure of components, their interrelationships, and the principles and guidelines governing their design and evolution over time.

V.MODULES

Data Collection: The data collection is done Kaggle. The data collected has been check one by one before taking it as input data. There are 2 classes named as affected plants and normal plants will be classified and each of the class contains images. From the 80% of images used for training and 20% of images used for validation and testing.

Data Augmentation:1) Data augmentation is a method by which you can virtually increase the number of samples in your dataset using data you already have.2) For image augmentation, it can be achieved by performing geometric transformations, changes to colour, brightness, contrast or by adding some noise. In order to let our model, adapt to a different situation that might occur noisy data such as when raining season.

Plant classification: After training for each model, the final two models were generated. One model was trained directly based on original images, while the other one was created using image data produced by data augmentation. Model history was used to record the training process. which displays the accuracy on the training set and testing set, it is described that the accuracy for the training set is having differences with 5% by with and without data augmentation. Besides, the performance of the model without data augmentation having lower accuracy on the validation set, while the model with data augmentation has a higher result in the testing set.

VI.IMPLEMENTATION

Hardware Requirements

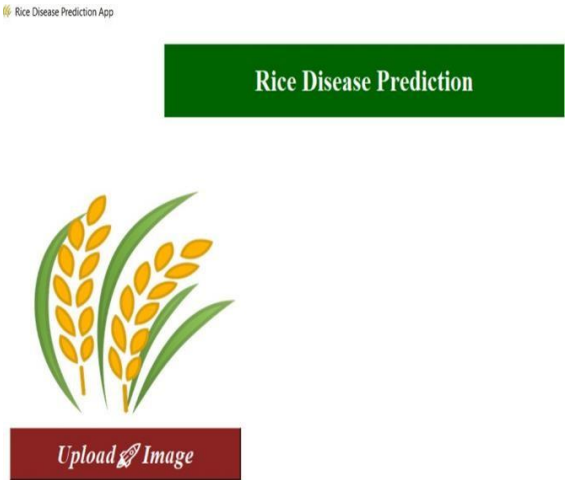
- Processor - Intel Core i7
- RAM - 4 GB
- Hard Disk - 500 GB

Software Requirements

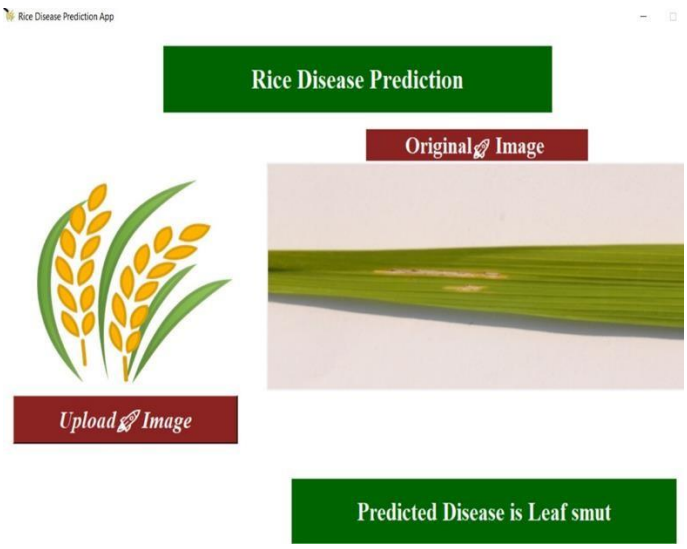
Operating System - Windows 7/8/10

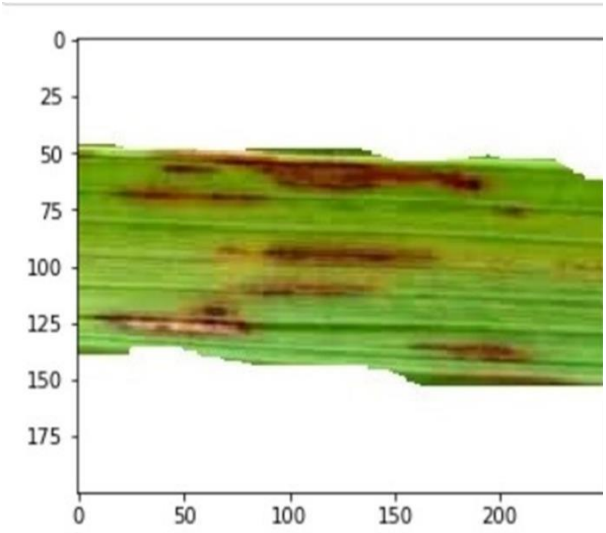
Programming Language: Python,

MySQL Screenshots:



Disease prediction:





Predicted Disease is Leaf smut

VII.CONCLUSION

Rice plant diseases can make a big amount of loss in the agricultural domain. Due to climatic conditions and global warming issues the plants are deficient in terms of water content and necessary nutrients and easily get affected by diseases. Generally, with the help of naked eye, farmers used to identify the disease, but they cannot identify what was the disease. In this project we developed a system to detect three rice plant diseases namely Bacterial Leaf Blight, Brown Spot, Leaf Blight. We have collected dataset of leaf images from uci repository. We had used Machine learning techniques to identify the diseases accordingly. So, it helps farmers to improve the quality of farming and increase the crop production.

REFERENCE

[1] Jitesh P. Shah, Harshad Kumar B. Prajapati, Vipul K. Dabhi, 2017, "Detection and Classification of Rice Plant Diseases", Journal: Intelligent Decision Technologies, IOS Press.

[2] S. Ramesh, Devaki, 2018, "Rice Blast Disease Detection and Classification using Machine Learning Algorithm", International Conference on Micro- Electronics and Telecommunication Engineering doi:10.1109/icmete.2018.00063.

[3] Shima Ramesh, Niveditha M, Pooja R, Prasad Bhat N, Shashank, N Ramachandra Hebbar, P.V. Vinod,2018, "Plant Disease Detection Using Machine Learning, International Conference on Design for 3Cs Compute Communicate Control, DOI 10.1109/ICDI3C.2018.00017.

[4] Amrita S. Tulsan, NaTasha Raul, 2019, "Plant Leaf Disease Detection using Machine Learning",10th International Conference on Computing, Communication and Networking Technologies (ICC- CNT), IEEE -45670.

[5] Parul Sharma, Yash Paul Singh Berwal, Wiqas Ghai, 2018, "KrishiMitr (Farmer's Friend): Using Machine Learning to Identify Diseases in Plants", IEEE International Conference on Internet of Things and Intelligence System (IoTaIS).

[6] Swetha V, Rekha Jayaram, 2019, "A Novel Method for Plant Leaf Malady Recognition using Machine Learning Classifier",3rd International conference on Electronics, Communication and Aerospace Technology (ICEC).

Reference

1.Jitesh P. Shah, Harshadkumar B. Prajapati, Vipul K. Dabhi, 2017, "Detection and Classification of Rice Plant Diseases", Journal: Intelligent Decision Technologies, IOS Press.

2.S. Ramesh, D.vydeki, 2018, "Rice Blast Disease Detection and Classification using Machine Learning Algorithm", International Conference on Micro-Electronics and Telecommunication Engineering doi:10.1109/icmete.2018.00063.

3.Shima Ramesh, Niveditha M, Pooja R, Prasad Bhat N, Shashank N, Ramachandra Hebbar,P.V.Vinod,2018,"Plant Disease Detection Using Machine Learning,International Conference on Design for 3Cs Compute Communicate Control, DOI 10.1109/ICDI3C.2018.00017.

4.Amrita S.Tulshan , Nataasha Raul, 2019, "Plant Leaf Disease Detection using Machine Learning",10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE -45670.

5.Parul Sharma, Yash Paul Singh Berwal, Wiqas Ghai, 2018, "KrishiMitr (Farmer's Friend): Using Machine Learning to Identify Diseases in Plants", IEEE International Conference on Internet of Things and Intelligence System(IoTaIS).

6.Swetha V, Rekha Jayaram, 2019, "A Novel Method for Plant Leaf Malady Recognition using Machine Learning Classifier",3rd International conference on Electronics, Communication and Aerospace Technology(ICEC)