ABSTRACT

This work presents a machine learning model, in the form of a Jupyter Notebook.The business logic in Python uses Hybrid ensemble Learning techniques in order to predict the disease present in the plant or in its leaves. The proposed system will integrate the data obtained from clicking 2k photos and by applying machine learning algorithms and a prediction of most suitable crops according to current environmental conditions is made. This provides young farmers with a variety of options of crops that can be cultivated. Machine learning is concerned with the development of computer programs that can access data and learn on their own. Plant Disease prediction solves one of precision agriculture's most difficult issues, and several models have been proposed and validated so far. Since plant diseases are affected by a variety of factors such as climate, weather, soil, fertilizer usage, and seed variety, this issue necessitates the use of multiple datasets. This suggests that predicting plant disease is not a simple task; rather, it entails a series of complex steps. Farmers are the core of the agricultural system. Agriculture is an important part of a country's development because, as everyone knows, a country's gross domestic product must be equitable. Farmers are critical to the agricultural system's success because crops must be planted and processed for it to work. A prototype for a real-time plant disease prediction algorithm in Python using Hybrid Machine Learning and Data Analytics was created to assist these farmers. Machine learning is a subset of artificial intelligence (AI) that enables computers to learn and evolve without being explicitly programmed..

CONTENTS

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
EDA	Exploratory Data Analysis
RF	Random Forest
SVM	Support Vector Machine

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 MOTIVATION :

The agriculture system is supported by farmers. Agriculture, as is well known, is an integral part of a country's development. Agriculture is very important in India's economy and job market. One of the most common problems faced by Indian farmers is that they do not choose the appropriate crop for their soil. One of the most popular issues that Indian farmers face is failing to protect their crops/plants in time from diseases.

1.2 PROBLEM STATEMENT:

Plant diseases have turned into a dilemma as it can cause significant reduction in both quality and quantity of agricultural products. Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. The proposed system is a software solution for automatic detection and classification of plant leaf diseases. The scheme consists of four main steps, first a color transformation structure for the input RGB image is created, then the green pixels are masked and removed using specific threshold value followed by a segmentation process, the texture statistics are computed for the useful segments, finally the extracted features are passed through the classifier.

The method for detection and classification of leaf diseases is based on masking and removing of green pixels, applying a specific threshold to extract the infected region and computing the texture statistics to evaluate the diseases using MATLAB.

1.3 OUTLINE OF THE PROJECT:

Many studies are being conducted in order to develop an accurate and effective crop prediction model. Ensembling is one of the techniques used in these types of studies. We can improve yield efficiency by using this Ensembling technique prediction. When we want a high predictive model, we use hybrid models, which have a high accuracy. Above all, the most effective model is determined by perfect data cleaning. We used Python for both the backend and the frontend in this project.

1.4 DATA COLLECTION:

Kaggle data set collection provides the dataset for this project. Description. In this work, we introduce a field dataset to diagnose and monitor plant symptoms called DiaMOS Plant—an extended dataset analyzed in [5]. DiaMOS Plant is a pilot dataset containing images of an entire growing season of a pear tree, from February to July, in order to build a representative sample that covers the main cultural aspects of this plant. The dataset is suitable for performing machine and deep learning methods in classification and detection tasks. A total of 3505 images were collected, including 499 fruit images and 3006 leaves images, respectively. The fruit is portrayed in the following four phases: fruit set, nut fruit, fruit growth and ripening. Similarly, biotic and abiotic stresses fall into four categories: leaf spot, leaf curl, slug damage and healthy leaf.

CHAPTER 2
AIM AND SCOPE

2.1 AIM:

The aim of our project is to focus on one of the main challenges in agricultural land i.e., disease prediction. The disease in crop plants affects agricultural production, so a model is proposed to automate a method for the prediction of disease in the plants.

2.2 LITERATURE SURVEY:

[1] Prabakaran,G et al., has presented the “Fuzzy decision support system for improving the crop productivity and efficient use of fertilizers' ' states that the process of reducing the fertilizer consumption and improving the crop productivity using the fuzzy logic systems. This system comprises two parts; land report based expert knowledge to stimulate the yield potential through appropriate organic lacking minerals in soil. This research work is to improve productivity with minimum consumption of fertilizer. This study has been carried out to assess the fertilizer consumption in both the ACZ (Agro Climatic Zone) with an exhaustive daily field measurements and lab analysis of a duration of three years to determine exact fertilizer needs for every individual land. The data was analysed in MATLAB to establish feasibility rules for decision support systems for the crops to get the targeted output.A major drawback of Fuzzy Logic control systems is that they are completely dependent on human knowledge and expertise. You have to regularly update the rules of a Fuzzy Logic control system. These systems cannot recognize the accurate yield prediction. And also it is quite difficult to write the logic for the whole Fuzzy expert system.

[5] Nishit jain et al., has developed “Crop Selection Method Based on Various Environmental Factors Using Machine Learning”,The Proposed method is used to determine the maximum yield by summing up the analysis of all the affecting parameters using a feedback propagation algorithm.

[6] Kirtan Jha et al., has reviewed the “A comprehensive review on automation in agriculture using artificial intelligence”, This paper talks about use of different automation practices like IOT, Machine learning and Artificial Intelligence in agriculture systems.

2.3 SCOPE:

- To predict the appropriate plant disease based on the image factor.
- To analyze the data to get the inference.
- To analyze and compare the cleaned dataset with various machine learning algorithms.
- To frame the hybrid ensemble model along with dataset.