“PSCS69 INTEGRATED CROP PROTECTION SYSTEM”

A PROJECT REPORT

Submitted by,

VAISHNAVI M - 20211CSE0801
KUSUMA B - 20211CSE0822
YAMUNA A K - 20221LCS0020

Under the guidance of,

Dr. SERIN V SIMPSON

in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING

at

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

PRESIDENCY UNIVERSITY

BENGALURU

JANUARY 2025

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

CERTIFICATE

This is to certify that the Project report **"PSCS69 INTEGRATED CROP PROTECTION SYSTEM"** being submitted by **Kusuma B, Yamuna A K, Vaishnavi M** bearing roll number **20211CSE0822, 20221LCS0020, 20211CSE0801** in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a Bonafide work carried out under my supervision.

Dr. Serin V Simpson
Assistant Professor
School of CSE&IS
Presidency University

Dr. Asif Mohammed H.B
Professor & HoD
School of CSE&IS
Presidency University

Dr. L. SHAKKEERA
Associate Dean
School of CSE
Presidency University

Dr. MYDHILI NAIR
Associate Dean
School of CSE
Presidency University

Dr. SAMEERUDDIN KHAN
Pro-VC School of Engineering
Dean-School of CSE&IS
Presidency University

PRESIDENCY UNIVERSITY

SCHOOL OF COMPUTER SCIENCE ENGINEERING

DECLARATION

We hereby declare that the work, which is being presented in the project report entitled **INTEGRATED CROP PROTECTION SYSTEM** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of **DR. SERIN V SIMPSON, Assistant Professor, School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Name	**Roll Number**	**Signatures**
VAISHNAVI M	20211CSE0801	
KUSUMA B	20211CSE0822	
YAMUNA A K	20221LCS0020	

ABSTRACT

This project focuses on the development of a machine-learning model to predict plant diseases and enable better crop selection by farmers. Built in Python, this model scans thousands of plant images, along with integration of different environmental factors such as weather, soil quality, and fertilizer usage, for more accurate predictions. This should enable farmers to make more informed decisions that will eventually increase crop yield and overall productivity.

Combining traditional farm wisdom and the newest techniques of data science allows real-time insight as to the type of crops better suited for prevailing conditions or detect possible disease. Based on some sophisticated machine learning algorithms particularly Hybrid Ensemble Learning, improved predictive accuracy in predictions from the system was noted. This approach can, therefore, be particularly useful in countries such as India, where agricultural products are a source of living for a significant portion of the population, but is lagging behind in other technologies.

It's the ultimate aim that closes the gap between traditional agriculture and modern technology, providing farmers with tools leading to better decision-making, higher productivity, and more sustainable farming practices. The project now uses data mining and machine learning that has given farmers useful insights for optimizing their crops with resultant improvements in their outcomes for their livelihoods and overall, the agricultural industry.

ACKNOWLEDGEMENT

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**,

Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair,** School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Asif Mohammed H.B** Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Dr. SERIN V SIMPSON, Assistant Professor** and Reviewer **Dr. Prasad P S, Assistant Professor**, School of Computer Science Engineering & Information Science, Presidency University for his inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar A and Mr. Md Zia Ur Rahman,** department Project Coordinators **Mr. Amarnath J.L & Dr. Jayanthi K.** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

VAISHNAVI M
KUSUMA B
YAMUNA A K

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER-1
INTRODUCTION

1.1 Detection of Disease

1.1.1 Motivation:

Farmers support the agriculture system. Agriculture, as everyone knows, forms 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 that Indian farmers face is they don't choose the right crop for their soil. Among the most popular issues that Indian farmers suffer from is failing to give protection to their crops/plants in time from diseases. [1].

1.1.2 Project Objective:

The first and foremost objective of the detection of plant diseases is the accurate and rapid disease diagnosis. This will prevent further spreading of the disease and also stop further crop loss. Early detection of the plant diseases would mean that there would not be a wide spread of the disease and the loss in the crop would be at a minimum [2]. Reliability of detection of the disease by Machine Learning Algorithms and resultant results must also be ensured.

1.1.3 Plant Leaf Disease:

A stress in the stock condition of a herb that kills or modifies vital. All kinds of herbs, broken and educated, are susceptible to disease. Each tribe is prone to some diseases, but each of these is relatively rare. The incidence and prevalence of plant disease vary from prime to prime, depending on the company of pathogens, soil conditions, and the availability and types that are cultivated. Some species of plants are susceptible to epidemics while others may be resistant. [1].

- **Definitions of plant disease:**

Whenever they initiate abnormal physiological processes which interfere with normal structure, growth & function, plants will readily be infected with disease-causing pathogens.

Infective agents include fungi, bacteria, mycoplasma, viruses, viroids, nematodes, or parasitic flowering plants in case of infectious plant diseases. Infective organisms can reproduce and be spread from host to host when contained either within or outside the host [2]. Not all growth conditions enhance growth but are composed of the following: extreme temperatures; adverse ratios of moisture oxygen, pollutants in soil and air, or an over- and under-availability of crucial minerals. All these are a basis for non-infectious plant diseases.

1.1.4 Importance of Plant Disease Detection:

Proper identification of disease control programs can prevent them from becoming a waste of time and resources. Other plant losses may be caused by disease control measures that are insufficient in controlling the disease-causing agent. Infectious parasites like nematodes, fungi, oomycetes, viruses, and bacteria are the most common cause of plant diseases. Since a diversity of organisms causes a variety of symptoms (Figure 1), proper identification of the pathogen is a must to create a management plan [3]. Injury vs. Disease It is important to know what the difference between a plant injury and a disease is. A sudden injury occurs due to an external factor within a short period of time.

1.1.5 Methods for Disease Detection:

Machine Learning Techniques

K-Nearest Neighbour (KNN) Algorithm for Machine Learning

• One of the easiest Machine Learning algorithms, K-Nearest Neighbor is based on the Supervised Learning technique.

• The K-NN algorithm assumes the new case and the data are comparable to already existing cases, and thereby puts the new instance in the category most similar to such cases.

• The K-NN algorithm retains every piece of information that is there and classifies other data items based on similarity. Therefore, with the use of the K-NN technique, new data can be classified relatively very fast and almost accurately [3].

• The K-NN algorithm can be used for classification problems or even in solving regression. The K-NN algorithm is more used in solving classification problems.

• K-NN is a non-parametric method. Therefore, it does not even have any assumption about the parameters of the underlying data.

• It is also known as lazy learning algorithm since it memorizes the training dataset rather than learning from it all at once. Instead, it uses the dataset to take some action while classifying data.

• Example: for training, KNN algorithm stores only the dataset; when it has new data, it classifies that data into a category that is very similar to the new data.

Support Vector Machine Algorithm

Most predominantly used supervised learning algorithm is Support Vector Machine or SVM that addresses classification and regression problems, however being used mostly for machine learning classification problems.

The SVM technique is targeted towards finding the best decision boundary or line that may classify the n-dimensional space, and allows easy classification of other data points in the future. Such an optimal boundary is known as a hyperplane [1]. To assist in the creation of this hyperplane, SVM has selected extreme vectors and points.

These extreme cases can be represented by the support vectors that make up the basic idea of SVM methodology. Look at how in the graph below, the choice classifies two different groups.

Types of SVM

There are two kinds of SVM:

- Linear SVM: The term "linearly separable data" is that the data which can be segregated into two categories using only one straight line. This kind of data is classified using Linear SVM and the classifier used is termed as the Linear SVM classifier.
- Non-linear SVM: The Non-linear data set is said to be a set whose the same cannot be classified with the help of a straight line. In such cases, the technique of classification that is used in this set is called as a Non-linear SVM classifier [2].

Random Forest Algorithm

Favored machine learning algorithm Random Forest falls under the class of supervised learning strategy. It could therefore be applied onto ML problems that require regression and classification.

It is based on the ensemble learning concept, which is a technique to integrate several classifiers useful for solving complex problems and enhances the model performance by averaging many decision trees that have been applied to varied subsets of the given data.

Instead of using a single decision tree, the random forest collects forecasts from all decision trees and makes a prediction by majority vote among the projections [3].

1.2 Crop and Fertilizer Recommendation

Agriculture has been one of the vital constituents of human society, first for survival and then as an important source of growth in the economy, especially in a country like India. Agriculture is a significant source of employment generation for the majority section of the population while being an important source for food security in the country. However, the high demands of agriculture products around the world have encouraged a great dependence on modern technology by using hybrid crop varieties [4]. Such crop varieties, while beneficial for large-scale production, do not provide nutrients in naturally grown crops to the plants and eventually result in soil degradation and environmental pollution.

Current hardships to farmers have now been affected by climate change and low yields. This is because the old mode of farming through experience will not adopt new realities. For instance, the stage at which farmers cannot hence predict what crops should be farmed and which fertilizers would give maximum yield.

Data science and machine learning can reduce this gap between traditional practices and technology. In this project, a machine learning-based platform needs to be designed, which will help farmers with better decisions with respect to recommended crops and fertilizers on the basis of soil composition. These technologies will improve productivity among Indian farmers, lower losses as well as profitability and hence move the sector towards a robust agriculture economy [4].

1.2.1 Problem Statement

There are so many farmers in the Indians. Most of them plant the same varieties, and fertilizer usage is absolutely random because nobody knows what the soil needs to be fertilized. Crop yield has gone down, and fertility of the soil has also declined, which causes acidic soil. We shall propose for this problem a system design using a machine learning algorithm to suggest to farmers what better to grow in a certain place and what more useful it is when talking about fertilizer.

It will suggest some specific crops according to the type of soil and considering climatic conditions. It will also notify farmers about various types of fertilizers and their quantities so that farmers choose the appropriate ways and get maximum benefits [4].

1.2.2 Objective

The main goals of this project are to:

- Recommend the most suitable crops for farmers to plant based on several factors to help them make informed decisions.
- Recommend the best fertilizers for improving soil and crop health.

The project will be implemented as a website that includes two main applications:

i. **Crop Recommendation**: Farmers can input soil data, and the app will predict the best crop to plant.

ii. **Fertilizer Recommendation**: Farmers can provide soil data and the type of crop they plan to plant, and the app will recommend fertilizers based on soil nutrient deficiencies or excesses.

1.2.3 Motivation

Agriculture is one of the biggest sectors in India and accounts for nearly 70% of small and medium enterprises. Even though the farmers are also embracing the modernities, there is still a knowledge gap on what crops to take and which fertilizer to use in agriculture. Crop and fertilizer recommendation systems help farmers in taking better decisions about what to grow and how to maximize yield for improved economic returns, thereby contributing to the growth of Indian agriculture [3].

1.2.4 Dataset Description

We are using two datasets for this project:

i **Plant Village Dataset**: This dataset contains over 38,000 images of plants from 15 species, labeled with 38 different plant diseases. It is useful for developing models that can accurately identify plant diseases and recommend appropriate treatments [2].

ii **Agricultural Production in India Dataset**: This dataset provides 46 years of crop production data in India, including information on cultivated areas, crop yields, and demographic details. It helps in understanding agricultural trends and making informed decisions to improve productivity and food security [1].

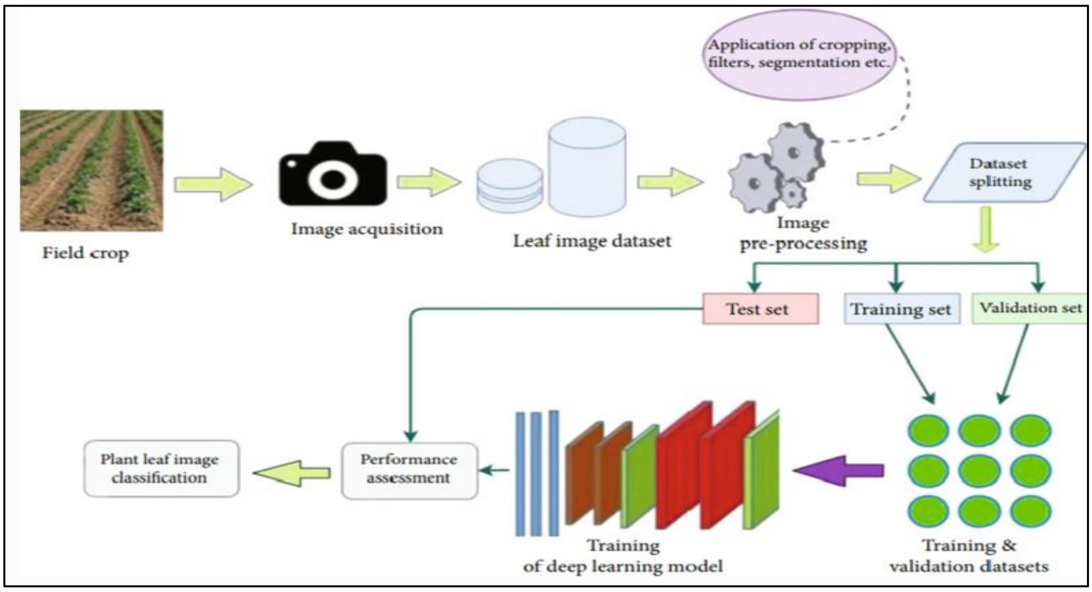


Fig 1.1 Flowchart

Here is a brief description of the flowchart:

1.Image acquisition: A variety of sources are used to gather crop photos.

2.Image pre-processing: The photos are subjected to pre-processing methods like noise reduction, trimming, and image improvement.

3.Disease identification: Using pre-processed images, a deep learning model built on a convolutional neural network (CNN) is taught to detect diseases in crops.

4.Crop recommendation: Based on soil and temperature factors, a machine learning algorithm is created to suggest suitable crops.

5.Fertilizer recommendation: Based on the crop variety and soil nutrient level, another machine learning model is created to suggest the right fertilizer.

6.Integration: A singular system incorporates disease detection, crop suggestion, and fertilizer recommendation algorithms.

7.Deployment: The combined system is set up in the field so that farmers can use it to spot agricultural diseases, get suggestions for suitable products, and get specialized fertilizer advice.

CHAPTER - 2

LITERATURE SURVEY

2.1 AIM:

The aim of our project is to concentrate on one of the major problems in agricultural land, that is, disease prediction. Disease in crop plants affects agricultural production, so a model is proposed to automate a method for the prediction of disease in the plants [4].

2.2 Literature Survey:

1. Unsupervised image translation using adversarial networks for Improved plant disease recognition.

Reference: Nazki et al.

Dataset: 2789 tomato plant disease images

The techniques used: Generative Adversarial Network and Deep CNN

Output: Accuracy= 86.1%

Advantages:

Better demonstration of information appropriation (pictures more honed and more cleared).

GANs can prepare any sort of generator organization.

Disadvantages:

Difficult to prepare, unstable training process.

Require many guidelines to obtain Satisfying results. Mode Collapse issue.

2. Cucumber leaf disease identification with global pooling dilated convolutional neural network.

Reference: Zhang et al.

Dataset: Acquisition of 600 cucumber sick leaves of 6 regular cucumber leaf infected

Technique used: GPDCNN

Output: Accuracy = 94.65%

Advantages:

GPDCNN is more robust than different strategies.

Disadvantages:

The completely associated layer has such a large number of parameters which decreases the Speed of preparing (training) and effectively bringing about over-fitting

3. Multilayered Convolution neural network for the Classification of Mango leaves infected by Anthracnose Disease.

Reference: SINGH CHOUHA N et al.

Dataset: Captured images at SMVDU, Katra

A technique used: Multilayer convolutional neural network (MCNN)

Output: Accuracy = 97.13%

Advantages:

The essential advantage of MCNN diverging from its paradigms is that it therefore Perceives the critical features with no human administration.

Disadvantages:

MCNN has a few layers then the training process takes a ton of time if the PC doesn't Comprise of a good CPU.

4. Sunflower leaf disease detection using Image Segmentation based On Particle swarm optimization.

Reference: Vijai Singh

Dataset: Capture Sunflowers leaves.

A technique used: Particle Swarm Optimization Algorithm.

Output: Accuracy = 98%

Advantages:

The upsides of PSO are that PSO is easy to implement and there are scarcely any Boundaries to change.

PSO performs in a way that is better than the GA for computational efficiency.

Disadvantages:

PSO is one of the well-known techniques, however, its application for the issue isn't Confounded because of its simple characteristics.

5. Deep Convolutional neural network-based detection system for real-time corn plant disease recognition.

Reference: Mishra et al.

Dataset: Plant Village dataset.

Technique used: Deep Convolution Neural Network

Output: Accuracy = 88.46%

Advantages:

With little dependence on pre-processing, this algorithm requires less human effort.

It is a self-learner, which makes the preprocessing phase, easier.

Disadvantages:

It requires an enormous dataset to process and train the neural organization.

6. Performance analysis of deep learning CNN models for disease detection in plants using image segmentation

Reference: Sharma et al.

Dataset: Tomato healthy and infected leaves images

Technique used: Convolution Neural Network

Output: Accuracy = 98.6%

Advantages:

Perhaps the greatest favorable position of CNN is the programmed extraction of highlights by handling straightforwardly the crude pictures.

Disadvantages:

CNNs don't have arranged outlines which are a fundamental component of human vision.

7. Tomato Leaf Disease Detection using Convolution Neural Network.

Reference: Agarwal et al.

Dataset: Images taken from Plant Village dataset.

A technique used: Images taken from Plant Village dataset.

Output: Classification Accuracy = 76% to 100%,

Average Accuracy for disease = 91.2%

Advantages:

The Storage space needed by the proposed model was of the order of 1.5MB where as pre-prepared models had additional room needs of around 100MB appropriately demonstrating the upside of the proposed model over pre-trained models.

Disadvantages:

A CNN is essentially slower because of an activity, for example, pooling.

8. Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network

Reference: Khamparia et al.

Dataset: Plant Village Dataset

Technique used: Deep Convolution Encoder Network

Output: Accuracy = 97.50%

Advantages:

Softmax classifier is used at the output layer. It returns the probabilities of each class if there should arise an occurrence of a multi-order model, and the target class should have a high probability.

Disadvantages:

This method lacks a mechanism to map deep layer feature maps to input dimensions.

9. Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification

Reference: Sladojevic et al.

Dataset: Capture images by agricultural experts.

Technique used: Deep Convolution Neural Network

Output: Accuracy= 96.3%

Advantages:

DCNNs included image and object classification, face detection, and image segmentation.

DCNN have more hidden layers especially more than 5, which increases the accuracy.

Disadvantages:

CNN don't encode the position and direction of an object.

Absence of ability to be spatially invariant to the input information.

10. A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification

Reference: MAJJI V APPLALANAIDU, G. KUMARAVELAN.

Dataset: plant village dataset

Technique used: Color Co-occurrence Matrix (CCM), Gray Level Co-occurrence Matrix (GLCM), Minimum Enclosing Rectangle (MER), Color Co-occurrence Matrix (CCM), CCM, GLCM, Discrete Wavelet Transform (DWT) Scale Invariant Feature Transform (SIFT)

Objective: This review provides a comparative analysis of various state of-the-art ML and DL algorithms to identify and categorize plant leaf diseases.

11. Research on machine learning framework based on random forest algorithm

Reference: Qiong Ren, Hui Cheng and Hai Han.

Technique used: Random Forest algorithm17

Objective: This article examines and analyses the machine learning framework based on the random forest algorithm with the goal of enhancing the random forest algorithm's current restrictions [4]. It also creates and implements a number of machine-learning frameworks.

12. Random Forest with Adaptive Local Template for Pedestrian Detection

Reference: Tao Xiang, Tao Li, Mao Ye, and Zijian Liu.

Dataset: TUD Pedestrians, INRIA pedestrians

Technique used: Random Forest

Output: Accuracy= 90.8%

Objective: Pedestrian detection in cluttered environments. Our main idea is to combine a set of adaptive local templates in a weak classification specified by a Random Forest and build the forest iteratively, layer by layer. The adaptive local templates learn the splitting functions in the forest; when they all have the same depth, they produce a weak classifier. This means that sample weights are updated by minimizing a global loss and adding each new weak classifier [5]. The proposed technique achieves the state-of-the-art or competitive performance according to the final experimental findings on two challenging pedestrian datasets.

13. Improving the Random Forest Algorithm by Randomly Varying the Size of the Bootstrap Samples

Reference: Md Nasim Adnan.

Technique used: Random Forest

Objective: For example, applying the technique of Random Subspace to the high-dimensional data set through the algorithm of Random Forest allows one over bootstrap sample. This way the choice of available basic classifiers namely, decision trees come fairly diversified. Further, increased diversification of decision trees could be very pretty helpful to ensure an increase in the accuracy of the ensemble. The size that doesn't change with the generation of decision trees is fixed-size, and not the one that changes with the generation of the decision trees since the basis classifier in Random Forest [5], maintains random bootstrap samples of. Also to even more improve forest accuracy, the size of bootstraps needs to be altered randomly within any given range, we propose it. We tested our approach on an extensive study on datasets for the UCI Machine Learning Repository. The paper presents results that reflect the immense promise of this approach.

14. An Ensemble Random Forest Algorithm for Insurance Big Data Analysis

Reference: Ziming Wu, Weiwei Lin, Zilong Zhang and Angzhan Wen.

Technique used: Random Forest

Objective: This paper presents the imbalance distribution of insurance business data analysis, pre-processing algorithms of the imbalance dataset, and plans an ensemble random forest algorithm using Apache Spark that may be used in large scaled imbalanced classification tasks on insurance business data. Based on the results of the experiment, it indicates that using a random forest ensemble algorithm is more appropriate for this type of insurance product recommendation or potential customer analysis than the traditional approach of Preprocessing algorithms suited for underbalanced classifications possibly adapted with the suggested bootstrap under-sampling approach in conjunction with KNN [3].

15. Leaf and skin disease detection using image processing

Reference: Manjunath Badiger, Varuna kumara, Sachin CN Shetty, Sudhir Poojary

Technique used: K-means algorithm and SVM classifier

Output: Accuracy = 96.3%

Advantages:

Easy to understand and implement. Can handle large datasets well. Disadvantages of KMeans Sensitive to number of clusters/centroids chosen.

Disadvantages:

It requires to specify the number of clusters (k) in advance.

It cannot handle noisy data and outliers.

It is not suitable to identify clusters with non-convex shapes.

16. Plant disease detection using machine learning

Reference: Niveditha M, pooja R, prasad Bhat N, Shashank N

Technique used: HOG, Random Forest

Output: Accuracy = 70%

Advantages:

Work well for small resolutions.

Typically does detection via classification, i., e uses a Binary classifier.

Disadvantages:

More time consuming to construct than a frequency polygon.

17. Plant disease detection using CNN

Reference: Nishant Shelar ,Suraj shinde ,Shubham sawant ,Shreyas dhumal

Technique used: CNN

Output: Accuracy = 96%

Advantages:

Local spatial coherence in the input (often images), which allows them to have fewer weights as some parameters are shared. This process, taking the form of convolution makes them especially well suited to extract relevant information at a low computational cost.

Disadvantages:

Classification of Images with different Positions, Adversarial examples, Coordinate Frame, and Other minor disadvantages like performance.

18. Pest detection in crop using video and Image processing

Reference: Madhuri Devi Chodey, Dr.Noorilla Shariff c, Gauravi Shetty

Technique used: K-means algorithm and SVM classifier

Output: Accuracy = 96.3%

Advantages:

SVM works relatively well when there is a clear margin of separation between classes.

SVM is more effective in high-dimensional spaces.

Disadvantages:

The SVM algorithm is not suitable for large data sets. SVM does not perform very well when the data set has more noise i.e. target classes are overlapping.

19. Image Classification Using Resnet-50 Deep Learning Model

Reference: Aryan Garg Dataset: STL-10 Technique used: Resnet-50

Output: Accuracy= 76.229%

Advantages:

Networks with large numbers (even thousands) of layers can be trained easily without Increasing the training error percentage.

Disadvantages:

High computational complexity - Residual neural networks can often require Significant processing power and may not be suitable for certain tasks.

20. Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer

Reference: Devvi Sarwinda, Radifa Hilya Paradisa , Alhadi Bustamam, Pinkie Anggia

Dataset: Warwick-QU

Technique used: Deep Residual Network (ResNet)

Output: Accuracy= 73% -88%

Advantages:

ResNets help in tackling the vanishing gradient problem using identity mapping.

Disadvantages:

High memory requirements - Residual networks require large amounts of memory to store the necessary parameters and weights.

2.3 SCOPE:

- To predict the precise plant disease that suits the factor of the picture.
- To analyze the data to draw inferences.
- To analyze and compare the cleaned dataset with different machine learning algorithms.
- With the dataset, frame the hybrid ensemble model.

CHAPTER 3
RESEARCH GAPS OF EXISTING METHODS

I. Disease Detection

- **No data:** All the models built for disease detection are based on large data that is not available for most crops or even in a region. It prevents them from generalizing if the plants and diseases are less represented.
- **Environmental factors:** Most of the techniques developed to date are not environmentally friendly and take into consideration factors like temperature, humidity, and quality of soil that play a very crucial role in disease formation and detection.
- **Real-time Detection:** Most of the models are not optimized for real-time deployment, especially in low-resource environments [6]. What is required are lightweight models that could do reasonably well on mobile devices or edge computing systems.
- **Early-Stage Detection:** Early-stage detection is very hard because the manifestation of a disease on crops is very subtle. Existing models are failing at that point, and hence research is required regarding the methods to improve early-stage detection [6].
- **Generalization Across Varieties:** Most of the models are working on a particular crop variety, ultimately failing to generalize across different species, let alone different varieties in the same species.
- **Explainability and Interpretability:** The deep learning model is non-explainable by nature, and hence farmers will not understand the diagnosis process and take appropriate action. Disease detection using interpretable AI may fill this gap.
- **Multimodal Approaches:** Most of the models focus on visual data such as leaf or plant images, and the multimodal approaches that involve a combination of visual, environmental, and textual data are not much explored.
- **Disease Co-occurrence:** Different diseases may attack crops at the same time; however, models available so far are single disease-based, which somehow limits practical utility.

II. Crop Recommendation Systems

- **Lack of personalization**: Crop recommendation models in general are based on data that is not fully particularized for the unique needs of an individual farmer, yet more relevant, specific conditions under which a given field stands, such as prior yields or soil health.
- Most of the existing models have not considered changing climate patterns that affect crop suitability over time.
- **Dynamic Data Integration:** Most systems treat static data and are not updated in real-time by novel weather events, soil properties, or market conditions. Systems that update their suggestions dynamically according to the prevalent data are required [4].
- **Sustainable Agriculture Practices:** The overall attention that sustainable agriculture practices, like crop rotation and conservation tillage, have generally been less while designing recommendation systems tend to result in long-term environmental illness.
- **Economic Factors:** Most of the models contain basic economic factors. However, much more research needs to be conducted on crop recommendation systems in the areas of market fluctuations, input costs, and subsidies so that it is feasible for the farmers financially.
- **Integration with IoT:** Crop recommendation systems can be made more efficient with the help of IoT devices which are monitoring field conditions [7]. Such information gives a real-time perspective, which will increase the precision of the recommendation.
- **Locality-based recommendations:** Most models only give general recommendations that are not hyperlocal, such as climatic and soil characteristics that carry out precision agriculture.

III. Fertilizer Recommendation Systems

- **Nutrient Profile Limitations:** Current models mainly recommend the three major nutrients N, P, and K and almost completely ignore the other nutrients, calcium and magnesium and the micronutrients, which are as important as these three major nutrients for plant health.
- **Soils Degradation:** Many of the above models fail to account for the period-dependent degradations that the soil undertakes [3]. So, the answers presented may be unsound and not sustainable. Models that demonstrate health and the degradation of soils must be made.
- These fertilizers may give general recommendations about the regions or crop type, rather than specific recommendations according to the profiles of the soils and crops in a given region.

- There are very few models which explain what might be the results of over-fertilization. For example, run off nutrients and soil acidification could be named as a few known impacts. More needs to be carried on the formulation of models. Optimizing yield sets equal focus on environment sustainability while on fertilizer application.
- It incorporates crop growth stage information into legacy systems. As these typically do not change fertilizer applications according to growing stages, so often, important precision fertilizing opportunities that create the most additional growth at various crop life stages points are missed.
- **Temporal Variation in Soil Fertility:** fertilizer recommendations based on any once off soil test may be misleading because soil fertility status varies with growing season or even within the growing season. Models of such temporal variation need to be further investigated [5].
- **AI Fertilizer Optimization:** In the current model, it more or less leans on fixed formulae and simpler data analysis. This hidden potential lies in using its continuous ability to learn its optimization of an AI-driven recommendation for fertilizers and its use with environmental inputs.
- **Combination with Organic Farming:** Typically, such models rely more heavily on chemical-based fertilizers with minimal research directed into organic fertilizer and its other applications for scenarios involving other crops.

CHAPTER - 4
PROPOSED METHODOLOGY

I. Disease Detection

4.1.1 EXISTING SYSTEM:

This system proposed combines fuzzy modeling with expert knowledge to predict which crop would most likely be well suited to a piece of farmland. Fuzzy sets present the existing knowledge about conditions of land, weather, quality of air, and agricultural practice, finally used in arriving at the decision-making rules regarding crop recommendations [1]. The crop prediction about optimal utilization is processed on multiple state variables using a fuzzy model.

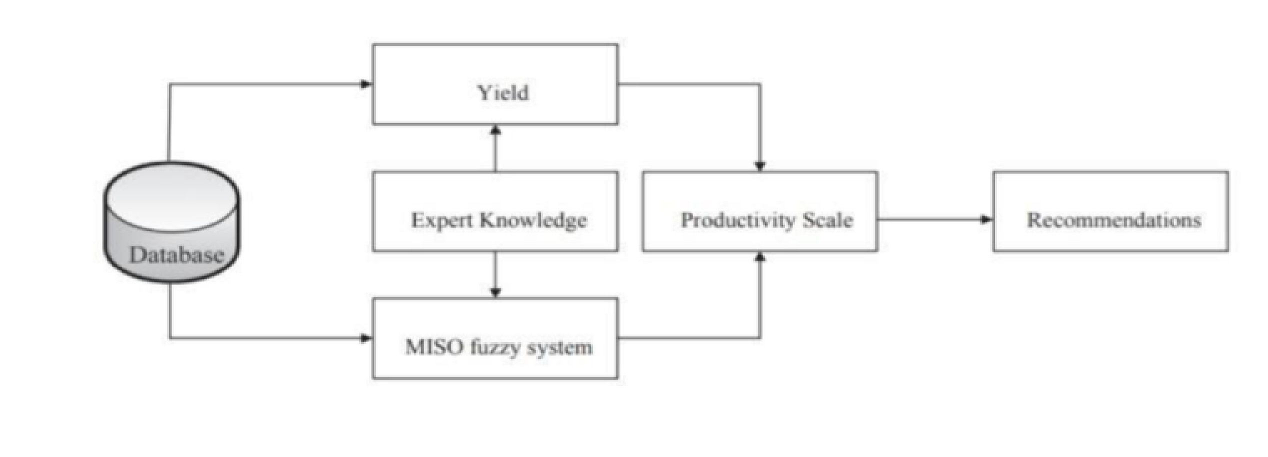


Fig.4.1: EXISTING SYSTEM

4.1.2 MAIN DISADVANTAGES OF THE EXISTING SYSTEM:

Much hardware testing is required to validate and verify a fuzzy knowledge-based system. Fuzzy rules and membership functions cannot easily be generated without making some errors. Fuzzy time logic along with fuzzy probability theory abound, and these words are jargon for the most part. Rules of precision in a control system through use of fuzzy logic need to be updated quite frequently. In fact, an exact estimation yield of a system is rather a challenging task for a system. The entire development of logic for the fuzzy expert system is time-consuming and demanding [2].

4.1.3 PROPOSED SYSTEM:

In the proposed methodology, we will predict crop yields that consist of optimum crop results. The System will be applied on NPK values to produce soil fertility by Figure 3.2. More importantly, a number of parameters such as pH, temperature, average rainfall and humidity have an influence on the best favorable conditions of good crop growth. It will, according to these parameters, suggest the best crop for that soil condition.

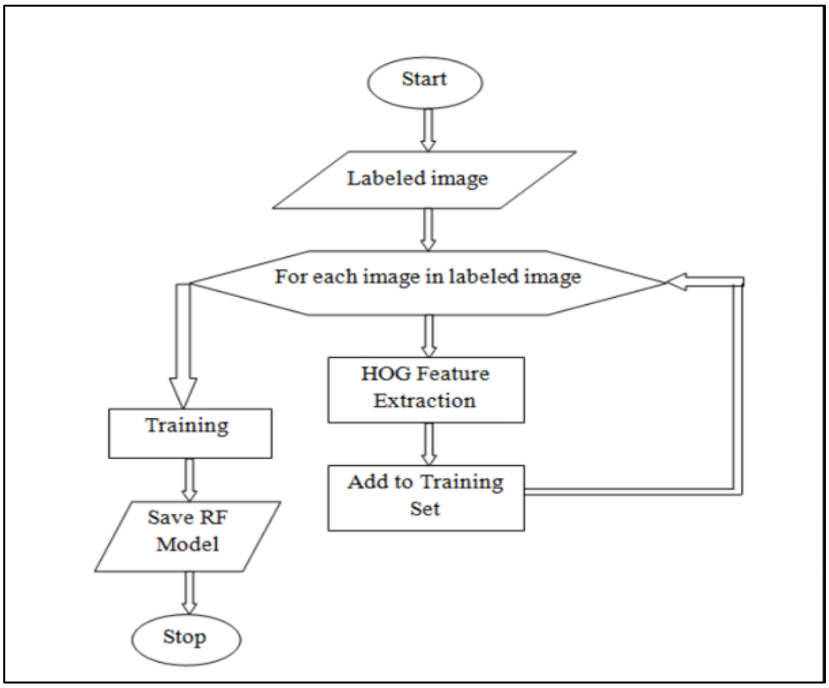


Fig.4.2: Proposed System

We have compared some of the most popular classification algorithms to predict the best crop for the soil in this project. Algorithms used in this analysis are Logistic Regression, Decision Tree, Random Forest Classification, Support Vector Machine (SVM), and Naive Bayes Classification.

4.1.4 EXPLORATORY DATA ANALYSIS:

While exploratory data analysis, that is part of every data science project, often gets underrated or left untouched [5], at the heart lies advanced mathematical methods in machine learning

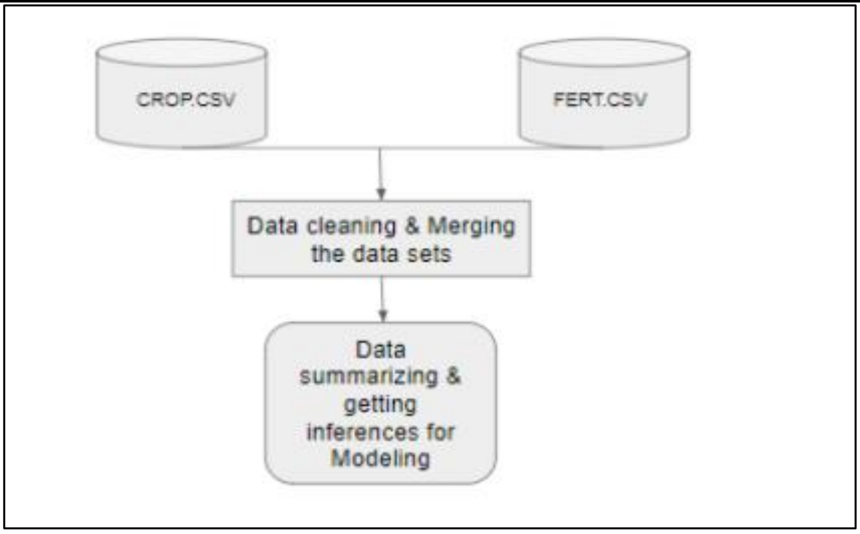


Fig.4.3: Diagram for EDA

This can be very direct in comparing the exploratory methods to be created with all the data sets. Data understanding phase: Understand the data given, highlight missing values, outliers, etc., and narrate the story based on both graphical and numerical tools. Have an idea about what the next logical step for the project regarding specific issues or aspects to be explored further. The steps of EDA are shown in Figure 3.3: identification of variable and type, correlation analysis, simple metrics analysis, transformation of variables, and handling missing values. Let us assume that the data is separable into two classes. If so, it would mean that we can classify a two-class classification problem completely, and let the class label yi − 1,1.

II. Crop Recommendation

Crop recommendation systems are designed to inform the farmer which crop is more ideal for planting in a specific farm in a given environment and agronomic conditions. That is, from soil conditions, climate, history of crops grown, and so many more, so the best crop recommendations will be produced that gives optimum yields.

4.2.1 Data Collection:

A good crop recommendation system always begins with the collection of a very wide variety of data influencing crop performance:

- **Soil Properties:** Soil pH Level Soil pH is significant, which determines nutrient availability and crop health. Different pH preferences are available for the crop; for example, wheat thrives well between slightly acidic to neutral pH range of 6-7.

- **Nutrient content:** The levels of N, P, and K, along with other minor nutrients like Ca and Mg, determine which crops can be produced. For instance, corn is one crop that places high demands on nitrogen, while beans, which are legumes, require far less because it has the intrinsic ability to fix its own N.
- **Moisture Content:** The moisture content in the soil determines the crops that can be cultivated. Sandy soils need drought-resistant crops, whereas loamy soils retain moisture better, and hence more diverse crops can be grown.
- **Soil Texture:** This includes the percentage of sand, silt, and clay in the soil. It is what determines the water retention drainage and aeration and determines what crop may be grown there. Rice prefers more water-retaining clay soils while root crops such as carrots grow well in a well-drained, loamy soil.

4.2.2 Climate Data:

- **Temperature:** The temperature levels support different forms of crops. For example, crops such as barley and wheat can grow better in temperature but rice sugarcane Tropical crop requires temperate for warmth.
- **Rainfall:** Places where rainfall is either moderately or highly appearing during the time of cropping also play important roles. For instance, maize, sorghum, and millet crops should be grown where moderately to high rainfall is noticed whereas, for barley and sunflower, they can be grown in semi-arid areas.
- **Humidity and Wind:** This microclimate, comprising humidity and wind, will determine crop health, the population of insects, and the loss of water.

4.2.3 Topography and Geography:

- **Altitude:** This might affect temperature and oxygen. Tea and coffee, for instance, are usually cultivated at high altitudes where temperatures are low.
- **Slope:** Steeper slopes will suffer more from erosion, and the scope of what can be farmed will be restricted.
- **Historical Crop Data:** Records fine-tuning the past performance of crops in a given area. If crops have been good under similar conditions in the past, then they are probably recommended.

Feature Engineering and Data Preprocessing

After collecting the required data, that data needs to be preprocessed and formatted into a suitable structure for machine learning models. This includes:

- **Normalization and Standardization:** Normalize the pH, nutrient levels, temperature, etc., of soil/climate features to bring everything onto a comparable scale so no feature dominates learning.
- **Handling Missing Data:** Fill missing or partially available records within climate or soil datasets using imputation techniques, either by average values or nearest-neighbor values filling gaps.
- **Encoding Categorical Variables:** Type of soil, crop category, or region encoded in some way that might be better suited for machine learning algorithms to digest.

4.2.4 Model selection

Once all the data is readied, all sorts of types of machine learning models are deployed to predict which crops would ideally be best considering the chosen input features or variables, such as soil or climate. There are so many of those that get employed extremely widely.

- **Random Forest Classifier:** For agricultural data, random forests are particularly effective because they can capture the complex interactions between multiple variables in a model. The model learns how to best predict the appropriate crops after training a series of decision trees on a random subset of data.
- **Decision Trees:** These models lend themselves to some very strong features for feature selection and interpretability. The basic idea is to split the data across different features, such as soil pH and temperature, to predict the most appropriate crop for a region.
- **XGBoost or Extreme Gradient Boosting**: is an improvement over decision tree-based ensemble learning to big-size datasets. It generates one after the other sequentially, for each such tree, which minimizes the error that has occurred with the previous one, thereby leading to much more accurate and generalized predictions with lesser overfitting.
- **KNN:** It can be used if adequate history of data exists. It identifies the crops grown in similar soil and climate conditions. It recommends such crops.
- **SVM:** SVM can be used so as to classify the different crops by using soil and environmental factors. It is more suitable for the classification problem, where the suitability of the crop is sharply differentiated.

The crop recommendation model is trained on historical data. Features are the characteristics of soil, climate, and other appropriate agronomic variables; the target variable is a type of crop. Training of the model involves feeding the data into the model and adjusting its hyperparameters-whether that is changing the depth of decision trees or the number of trees in a random forest-for optimization of its prediction capability.

Such techniques could be optimized while training with the help of techniques like grid search with cross-validation; therefore, that alone reduces overfitting but in turn increases the performance of a model by an important percentage. Cross-validation implies making training and testing datasets from any one dataset, as that would show whether the current model generalizes well to some other new or unseen data or not. As such, a good model will generate more accurate and better recommendations to farmers.

- **Precision and Recall:** This will now evaluate how well the model can suggest the right crops to be grown without simultaneously suggesting the wrong crops.
- **Confusion Matrix:** This will now point out where the model's performance is bad or good regarding the crops the model classified correctly or wrongly.

With such metrics, further enhancement of the model is performed using adjusted hyperparameters or through retraining.

4.2.5 Recommendation Output

When the farmer inputs data about his or her field-whether it would be soil test results, climate in the locale, or some other relevant piece of information model analyzes those inputted data and provides a ranked list of available crops for which the farmer might choose. The output will be included:

- **Yield Estimation:** It even forecasts the yield of each recommended crop by using previous data, soil quality, and prevailing climatic conditions.
- **Planting Schedule:** It even gives a calendar along with the suitable time for sowing, growth, and harvest of every crop according to the prevailing weather of that region.
- **Profits Analysis:** The models may even include economic factors like market prices, estimated costs of cultivation, and expected returns to help farmers determine which crop would be the most profitable.

For example, if the data input by the semi-arid farmer with loam sandy soils indicated a high potassium content but low nitrogen, then it will advise the model to provide drought-resistant crops like sorghum or millet and, at the same time, nitrogen-fixing legumes such as chickpeas to be farmed that improve the condition of the soil over time.

4.2.6 Benefits of Crop Recommendation Systems

- **Personalized Recommendations:** It gives personalized recommendations instead of generic recommendations based on specific soil and climatic conditions, thus guiding the farmers better while planning crops.
- **Better Yield:** "When crops are planted based on suitability, the yields will surely be better, and thus crop failure is reduced," according to Klavans.
- **Sustainability:** This will encourage sustainable agriculture by proposing the type of crops that need fewer inputs- water and fertilizers - suggested by soil health and regional climate in order to minimize the degradation of the environment.
- **Resource Optimization:** It will help the farmer to utilize his water, fertilizers, and other resources in totality not to waste them.
- **Economic benefits:** With an analysis of both agronomic as well as economic factors, the system has made sure farmers will benefit to their maximum profitable value by producing crops that come in higher demand in the market as well as high profitability too.

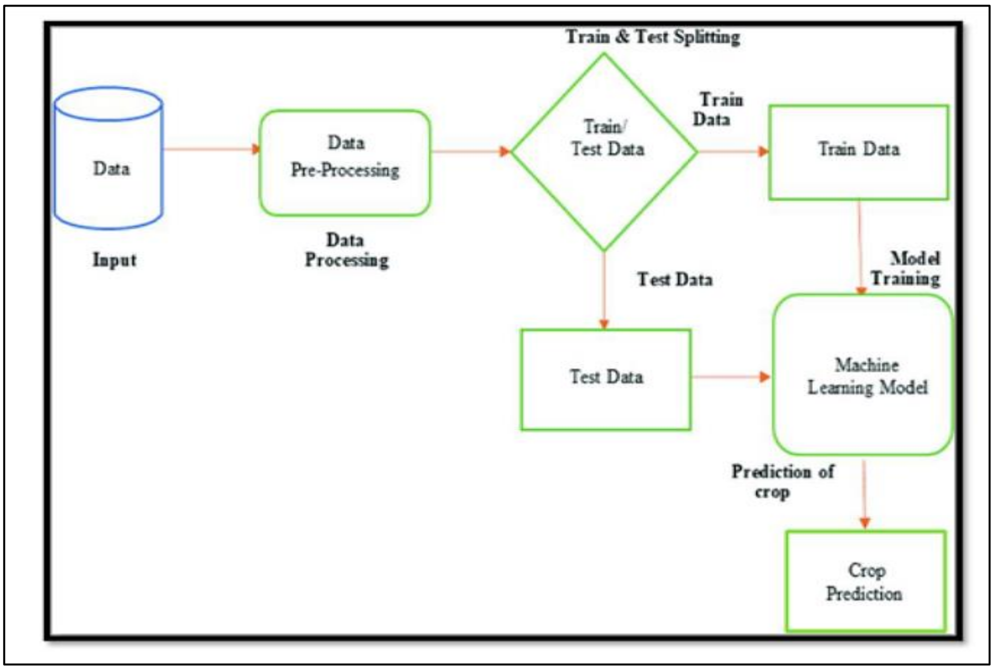


Fig 4.4: Crop Recommendation Architecture

III. Fertilizer Recommendation

The proper fertilizer recommendation is one of the key aspects in maintaining soil fertility and therefore, appropriate availability of nutrients to crops at the right time. This method gives fertilizer recommendations as a function of the level of nutrients available in the soil, the requirement of crops, and other field conditions.

Data collection is the very first step of the process. Testing of the soil is carried out regarding Nitrogen (N), Phosphorus (P), and Potassium (K). Other factors like soil pH, moisture content, and soil type, whether sandy, loamy, or clay, are also recorded. All this collected information is important for making proper fertilizer recommendations.

Nutrient levels and soil properties are preprocessed on collected data. Different crops require the identification of deficiencies or excess of nutrient, with the use of threshold values for determination. In the case where nitrogen in soil is below the threshold value, the system shall advise the farmer on fertilizers with nitrogen content.

The recommendation system can be rule-based or machine learning-based. A rule-based system would be based on agricultural guidelines and expert knowledge specifying what kind of fertilizer, in what amount, to apply given the specific nutrient deficiency.

The model was trained on past soil test data and the types and amounts of fertilizers applied. In the training data, the model learned how best to recommend the type and dosage of fertilizer for any given set of current soil conditions and types of crops.

Once trained, the model advises on the type of fertilizer to be applied; it could either be Nitrogen-rich, Phosphorus-rich, Potassium-rich, or organic fertilizers. The system also advises the amount of fertilizer needed and the method of application. That is, the app advises foliar spraying or soil application, among others. The app will also give time suggestions to make sure the fertilizer is applied during the most critical growth stages of the crop.

By implementing such recommendations, soil health is enhanced and the crops will flourish while at the same time-wasting fewer fertilizers, hence sustainable farming and more savings in farming.

4.3.1 Dataset

We have considered 2 datasets. One helps recommendation of crops, and second dataset helps in prediction or recommendation of fertilizer.

- **Dataset for crop recommendation**

As we know good crop production or good yield of crop depends on many factors, so in this data set, various factors are presented that are included in the crop production process. With the use of this data set, a crop recommendation model can be developed.

Dataset for crop recommendation have following data fields

a) N: talks about the ratio of nitrogen

b) P: talks about the ratio of Phosphorous

c) K: talks about the ratio of Potassium

d) Temperature: in Celsius

e) Humidity: relative humidity in %

f) Ph: tells either soil is acidic or basic

g) Rainfall: in mm

	A	B	C	D	E	F	G	H
1	N	P	K	temperatı	humidity	ph	rainfall	label
2	90	42	43	20.87974	82.00274	6.502985	202.9355	rice
3	85	58	41	21.77046	80.31964	7.038096	226.6555	rice
4	60	55	44	23.00446	82.32076	7.840207	263.9642	rice
5	74	35	40	26.4911	80.15836	6.980401	242.864	rice
6	78	42	42	20.13017	81.60487	7.628473	262.7173	rice
7	69	37	42	23.05805	83.37012	7.073454	251.055	rice
8	69	55	38	22.70884	82.63941	5.700806	271.3249	rice
9	94	53	40	20.27774	82.89409	5.718627	241.9742	rice
10	89	54	38	24.51588	83.53522	6.685346	230.4462	rice
11	68	58	38	23.22397	83.03323	6.336254	221.2092	rice
12	91	53	40	26.52724	81.41754	5.386168	264.6149	rice
13	90	46	42	23.97898	81.45062	7.502834	250.0832	rice
14	78	58	44	26.8008	80.88685	5.108682	284.4365	rice
15	93	56	36	24.01498	82.05687	6.984354	185.2773	rice
16	94	50	37	25.66585	80.66385	6.94802	209.587	rice
17	60	48	39	24.28209	80.30026	7.042299	231.0863	rice
18	85	38	41	21.58712	82.78837	6.249051	276.6552	rice
19	91	35	39	23.79392	80.41818	6.97086	206.2612	rice
20	77	38	36	21.86525	80.1923	5.953933	224.555	rice
21	88	35	40	23.57944	83.5876	5.853932	291.2987	rice

crop_recommendation

Fig 4.5: Dataset for crop prediction

4.3.2 Dataset for fertilizer recommendation

Only finding the right crop to grow is not enough for good yield or good yield production we must also find what fertilizer must be used for crop care.

Dataset for fertilizer recommendation have following data fields

a) N: talks about the ratio of nitrogen

b) P: talks about the ratio of Phosphorous

c) K: talks about the ratio of Potassium

d) Ph

e) soil moisture

f) crop

1		Crop	N	P	K	pH	soil_moisture
2	0	rice	80	40	40	5.5	30
3	3	maize	80	40	20	5.5	50
4	5	chickpea	40	60	80	5.5	60
5	12	kidneybeans	20	60	20	5.5	45
6	13	pigeonpeas	20	60	20	5.5	45
7	14	mothbeans	20	40	20	5.5	30
8	15	mungbean	20	40	20	5.5	80
9	18	blackgram	40	60	20	5	60
10	24	lentil	20	60	20	5.5	90
11	60	pomegranate	20	10	40	5.5	30
12	61	banana	100	75	50	6.5	40
13	62	mango	20	20	30	5	15
14	63	grapes	20	125	200	4	60
15	66	watermelon	100	10	50	5.5	70
16	67	muskmelon	100	10	50	5.5	30
17	69	apple	20	125	200	6.5	50
18	74	orange	20	10	10	4	60
19	75	papaya	50	50	50	6	20

Fig 4.6: Dataset for Fertilizer Recommendation

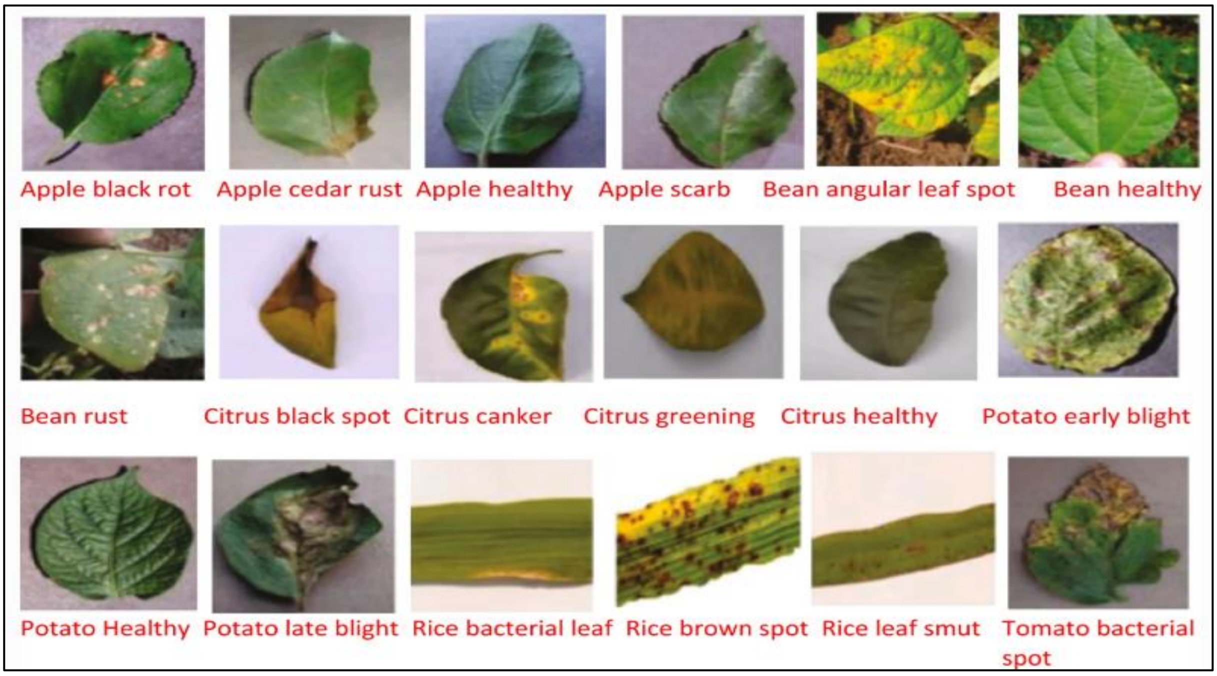


Fig 4.7: Dataset for Disease Detection

CHAPTER-5
OBJECTIVES

Crop Disease Detection

Objective:

Explore the usage of computer vision for the early detection and identification of crop's diseases by applying deep learning. It will help the farmers analyze the image, and take an appropriate measure on time, which decreases gigantic loss from crops, thus enhancing the output level.

Key Objectives:

- Disease Detection in Time: Use real-time crop disease detection. The same can be achieved with image-capturing devices such as smartphones, drones, and cameras in the fields.
- Extensive Disease Coverage: Develop a robust system that can identify multiple diseases across different crop types so that farmers get critical information regardless of the crops they grow.
- Actionable Insights: Provide the farmer with actionable insights on disease management, which involves treatment methods and preventive measures that can control its spread.
- Prediction of Disease Spread Severity: Provide an estimate of the severity of disease spread so that farmers can give priority to appropriate treatment efforts.
- Sustainability: Suggest the judicious use of chemical pesticides; do not use excessive chemical pesticides in agriculture. Counseling on pesticide application should be according to proper identification of disease.

Benefits:

- Healthier Crops: The diseases will be identified beforehand, and thus the farmer will act as early as possible to reduce crop losses.
- Economical: Early diagnosis will not require heavy and expensive treatments; hence, the resources will be saved, and the profit from the farm will be maximized.
- Accessible technology Technology that is available, a feature that easily provides access via mobile apps. The farmer promptly uploads images with a diagnosis being returned on the same screen.

Crop Recommendation

Objective:

Crop recommendations should be based on facts, and a farmer should be able to decide which crop he needs to grow based on the information available. The local environment is taken into consideration to optimize land use and increase crop yield accordingly.

Key Objectives:

- Localized Recommendations: Crop recommendations following the local environment, incorporating considerations about soil type, pH levels, and other climatic conditions such as temperature, rainfall, and humidity.
- Productivity Maximization: The productive crop be handed over to the farmer who is most suitable for that soil and climate. Then they shall produce crops and earn revenues
- Risk Reduction: Support the farmers in planting sturdy crops against localized pest attacks, harsh climatic conditions, and poor market trends. Thus, the loss of crop is reduced
- Seasonal Basis: The decision is taken and recommendation made that will help the farmer by considering either whether it is currently or the upcoming cropping season.
- Crop Rotation Advice: Advise the farmer on crop rotation to ensure proper soil health, avoid soil degradation, and minimize disease outbreaks.
- Nutritional Value Orientation: Suggest crops that answer market needs and nutritional gaps in the region, enhancing food security.

Benefits:

- High Yield: The right crop in the right environment assures high yields and better-quality produce.
- Optimized Resources: Crop recommendations ensure that water, labor, and fertilizers are used in the most efficient way possible. It makes farming economical.
- Environmental Sustainability: It promotes sustainable farming by suggesting crops suitable to natural conditions. It decreases artificial inputs and environmental impact.
- Market Aligned: The crop recommendation also suggests crops aligned with the demand of the market, which can be grown by farmers with a higher probability of selling them at a profit-generating price.

Fertilizer Recommendations

Goal:

Assist farmers to get the right fertilizer recommendations from data so that the appropriate nutrient is applied at the right time and in the right quantity to help crops thrive. This will enable farmers to ensure best practice in the management of nutrients in a more sustainable manner.

Key Results:

- Tailor-made Fertilizer Programs: Design fertilizers according to soil type, crop kind, and its growth stages to avoid over-fertilization in crops.
- Nutrient Management: Over-fertilization or under-fertilization of crops are avoided by ensuring the right type of fertilizer and amounts for use among farmers.
- Organic as well as Eco-friendly Options: Use organic or eco-friendly fertilizers whenever available. Encourage the sustainable farming so that the soil and water becomes less chemically impure.
- Testing for Soil Health: Regular testing of soil health should be conducted and frequent changes in the fertility of the soil by organic or chemical fertilizer should be advised to the farmer.
- Nutrient Deficiency Detection: Nutrient deficiency symptoms of the crop such as yellow color in the leaves should be presented and information about how to overcome this problem should be conveyed to the farmer.
- Application of Balanced Nutrition: Amounts of required nutrients, like quantities of nitrogen, phosphorus, and potassium. This is an obvious strategic removal of overuse of fertilizers to avoid degradations in soils and poisonous contents in the environment

Benefits:

- Balanced Growth for Proper Crop: Precise recommendations regarding fertilizers are used for better production with proper cultivation that gives excellent results.
- Save the Money: Does not waste unnecessary, saves money in the hands of farmers. Money conservation is possible only through proper fertilizer recommendation that stops the usage of excessive or incorrect fertilizers
- Natural sustainability discourages the unnecessary excessive chemical fertilizers with minor ill effects such as soil pollution, water contamination.
- Soil Conservation Balanced nutrient applications continue to preserve the health of the soil avoiding long-term degradation

CHAPTER-6
SYSTEM DESIGN & IMPLEMENTATION

1. System Overview

A mobile or web application this interacts with farmers, gathering data from the them and processing it using the Machine Learning models to yield actionable insights as regards disease diagnosis, crop suggestions, and fertilizer recommendation. It can be cloud hosted for scalability and to secure data.

2. Component of the System:

- **User Interface (UI)**
  - Mobile app or Web app (Cross-platform framework like Flutter for mobile, React for web)
  - Friendly dashboard for the farmer to interface with the system
  - Capture image of plants, input soil conditions, and fetching recommendations
  - Push notifications and alerts for disease outbreaks and seasonal crop recommendations
- **Backend**
  - REST APIs that manage data flow between frontend and backend
  - Machine Learning Models for predicting diseases, and recommending crops and fertilizers.
  - User Details Database, Raw Input, Crop Historical Data, etc.
  - **Data Storage**
  - Image storage: Cloud Storage, such as AWS.
  - Crop and Soil Information Storage: Relational such as PostgreSQL
  - Historical data storage: This will be to track User recommendations and disease outbreaks
- **Machine Learning Models**
  - **Disease Detection Model:** Deep learning-based CNN with TensorFlow for image classification
  - **Crop Recommendation Model:** Utilizing soil conditions, climate, location, and season by making use of a machine learning model such as a Random Forest.
  - **Fertilizer Recommendation Model:** Use the regression model that gives the best fertilizer recommendation for data on nutrient contents in soil, crop type, and weather
- **Integration with External APIs**
  - Weather API-For extracting real-time weather data from any weather API sources like OpenWeatherMap or WeatherStack
  - Api for Soil/Nutrient For pulling data regarding soil properties with a location

- Geo-API: Pull info about that location that will be suggested, for instance, Google Maps API

- **Notification System**
  - Send push notifications about the diseases identified, which crops are to be suggested, and fertilizers
  - Link to services, like Firebase Cloud Messaging (FCM)

3. System Design Architecture

i Frontend (Web App)

- **Web application development using React:**

- Capture images to identify diseases
- Forms to provide soil information and crop requirement
- Crop and fertilizer recommendations
- Optional Chatbot feature to guide the farmer

ii API & Processing

- **API Layer:**

- API Gateway (Node.js) to process requests
- Authentication (JWT-based for secure login)
- Routing for different services (disease detection, crop/fertilizer recommendation)

Data Processing: Image Preprocessing (for disease detection): Resize, normalize, and augment input images

ML Model Inference:

- Disease Detection Model (trained on plant image datasets such as Plant Village)
- Crop Recommendation (soil input parameters, location, and climate)
- Fertilizer Recommendation depending on crop type and soil nutrients

Database:

- Relational Database (PostgreSQL) to handle history for users and crops
- Cloud Storage (e.g., AWS S3) for images of plants
- NoSQL Database (MongoDB), only in case unstructured data needs to be stored

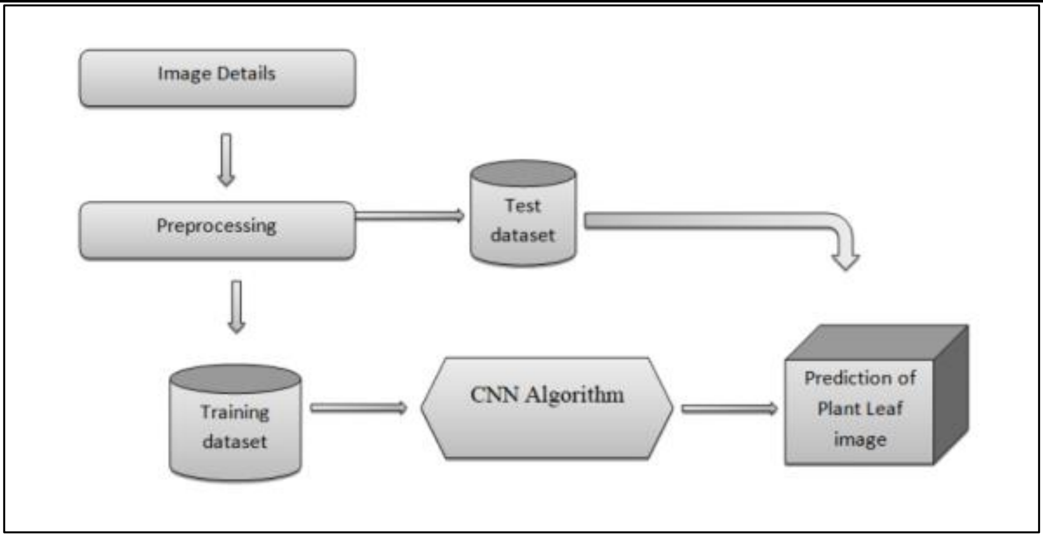


Fig 6.1: Overview of the System

- ## Data Flow Diagram

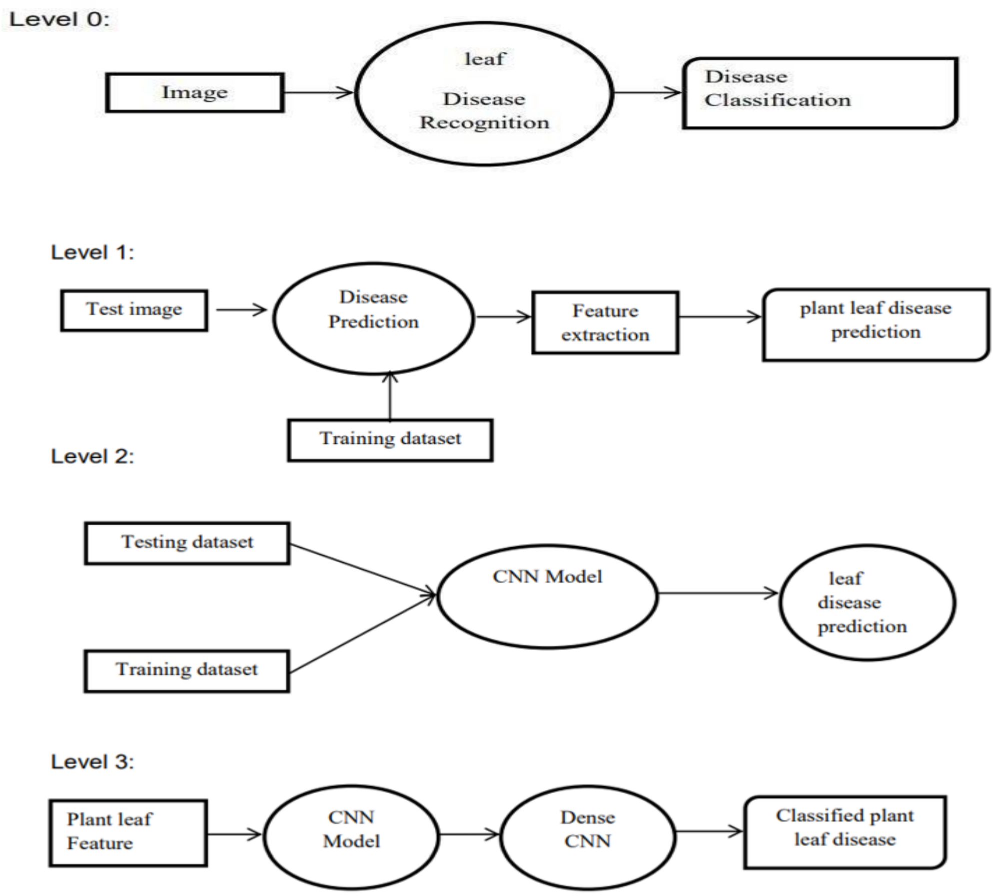


Fig 6.2: Data Flow Diagram for disease prediction

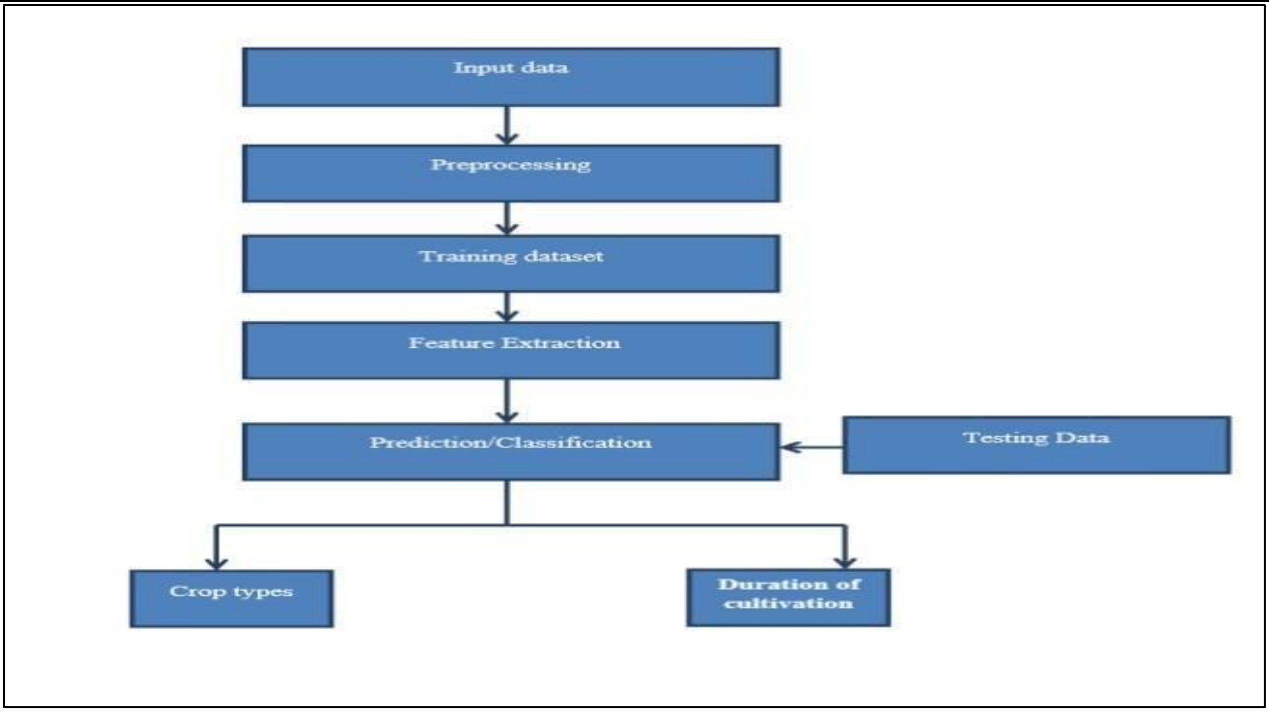


Fig 6.3: Data Flow Diagram for Crop Recommendation

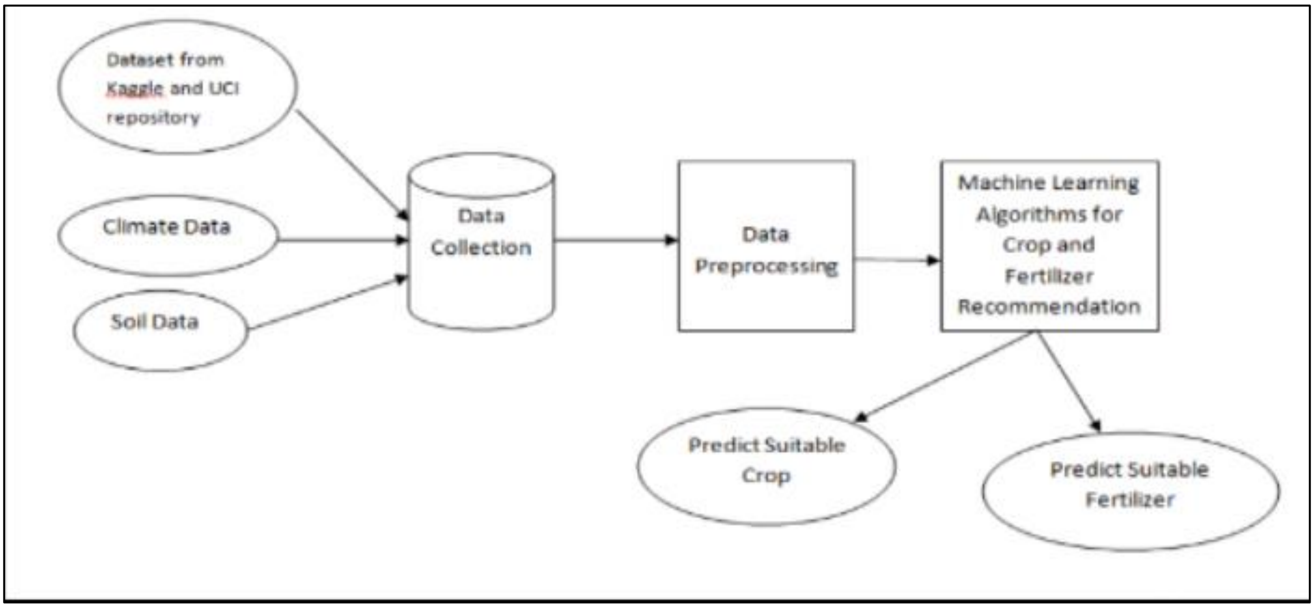


Fig 6.4: System Architecture for crop and fertilizer recommendation

4. Flowcharts:

- Fertilizer Recommendation

The user has to input the values of Nitrogen, Phosphorus, Potassium along with crop Name. A POST request is sent to the flask API. In this application, the fertilizer recommendation classifier is hosted. An HTTP response is sent to the front-end and on the front-end the user is getting a recommendation for the fertilizer.

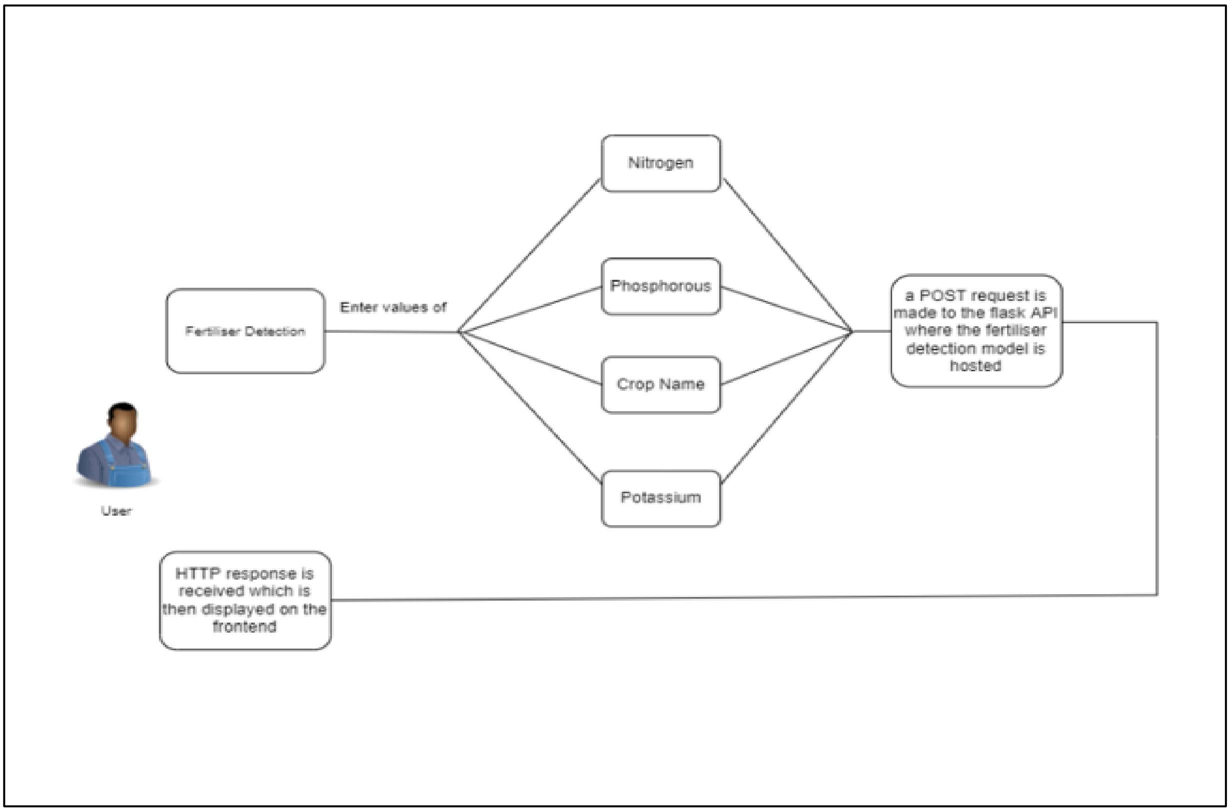


Fig 6.5: Flow Diagram for Fertilizer Recommendation

- **Disease Detection**

In the disease-detecting functionality, the user has to choose an image to click or load it directly into the model. The back-end image gets processed by the model, and its output sends back to the front-end in HTTP response. There, it delivers the disease plant is suffering from and the appropriate remedies for its cure.

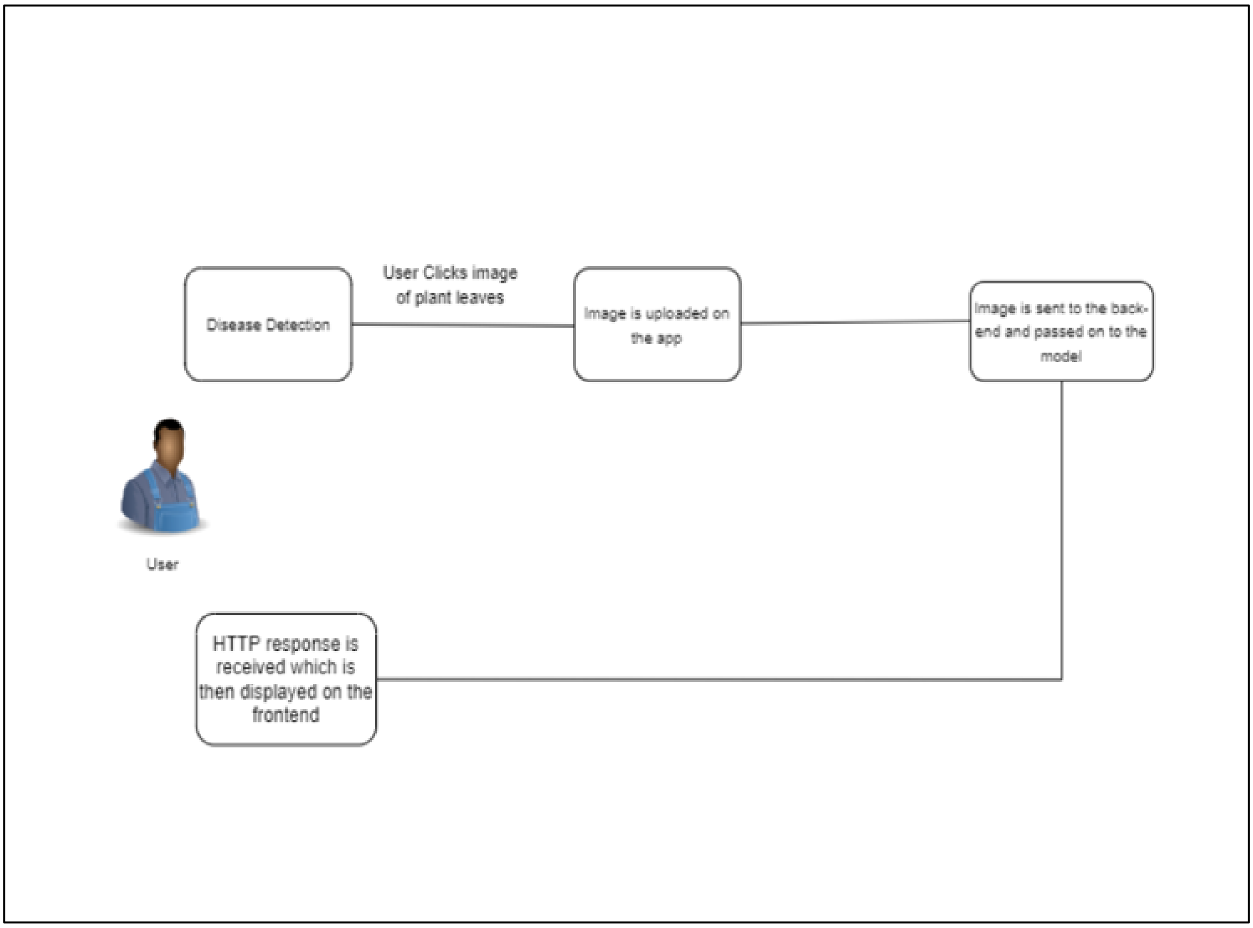


Fig 6.6: Flow Diagram for Plant Disease Detection System

- **Crop Recommendation**

Once values of Nitrogen, phosphorus, and Potassium are submitted, a post request is made to the flask API. The model runs on the data submitted, and the HTTP response sent back to the front-end explains which crop the farmer should cultivate in that soil to get the best crop out of the land given.

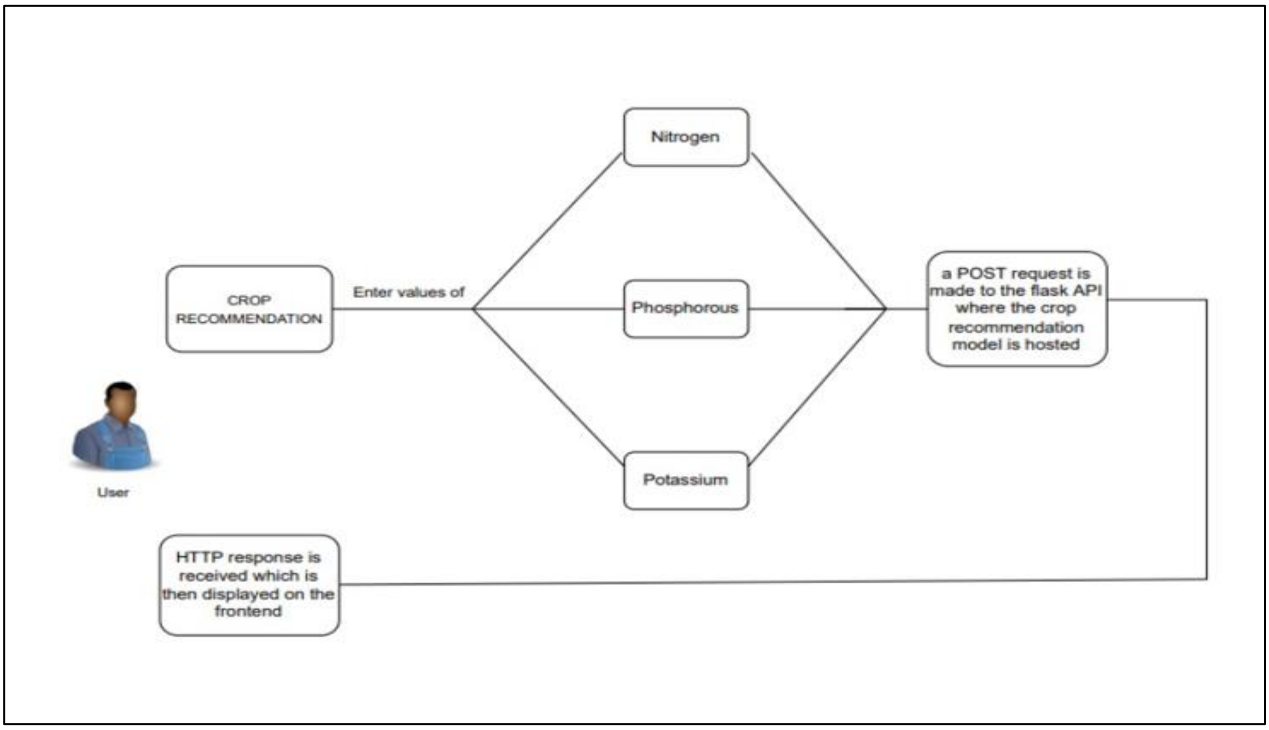


Fig 6.7: Flow Diagram for Crop Recommendation System

.

CHAPTER-7

TIMELINE FOR EXECUTION OF PROJECT (GANTT CHART)

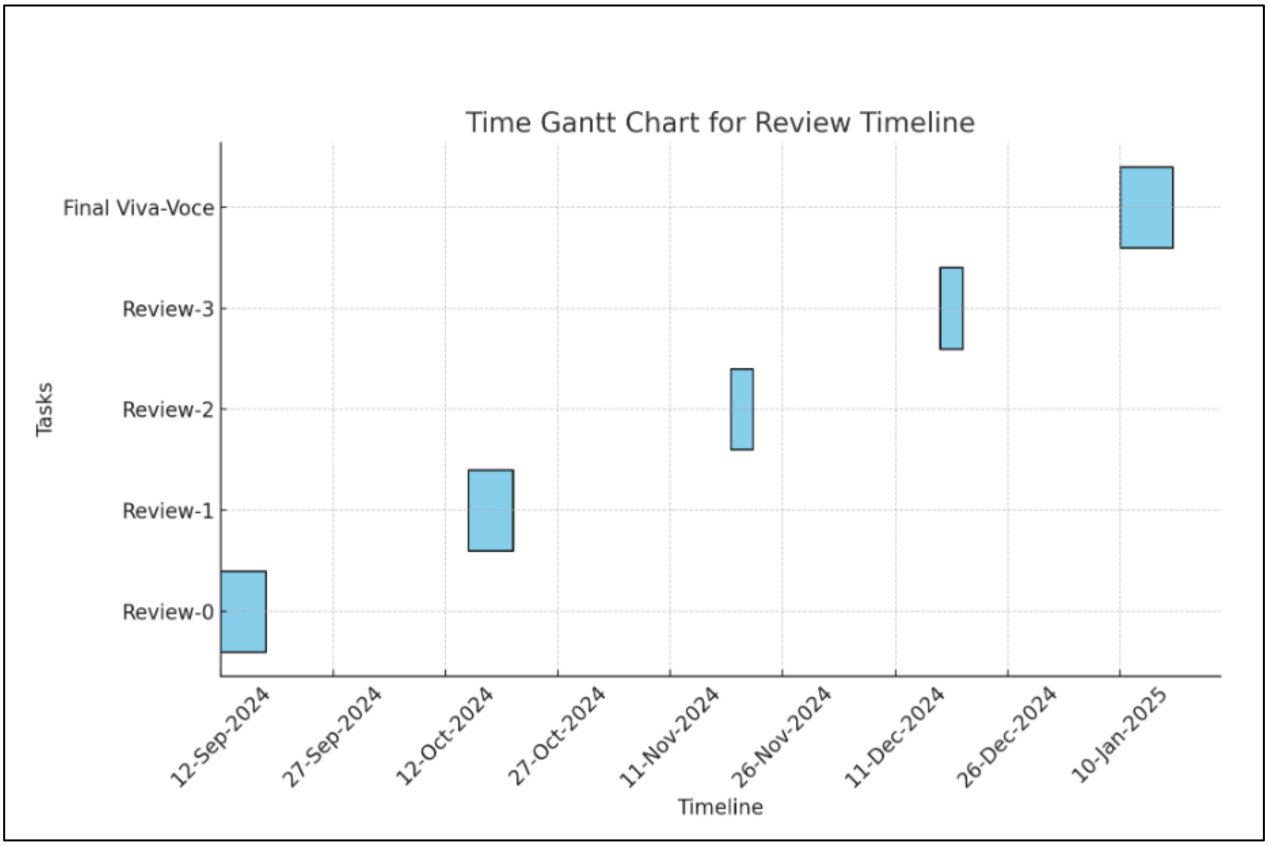


Fig 7.1: Gantt Chart

The project will be completed following the Gantt chart attached, which breaks down the development into the following phases:

Phase	TimeLine
Planning and Requirement Gathering	Sept 09 - Sept 15
UI/UX Design	Sept 16 - Sept 22
System Architecture	Sept 23 - Sept 29
Frontend Development	Sept 30 - Oct 19
Backend Development	Oct 19 - Nov 9
AI/ML Model Development	Nov 10 - Nov 30
Testing	Dec 1 - Dec 7

Table 7.1: Phases

CHAPTER - 8

OUTCOMES

8.1 Crop Recommendation

The results for our crop recommendation experiments are summarized in the Table 8.1. Fig 8.1 also plots these scores on a bar-chart for easy comparison. From the table we can see that the Random Forest and Naïve Bayes models performed the best while the XGBoost model ranked at the third position. Generally, one would expect the boosting (Random Forest) and bagging (XGBoost) models to perform better and generalize better than most non-ensemble methods.

Model Type	*5-Fold Cross-Val Accuracy*
Decision Tree	0.914
Naive Bayes	0.995
SVM	0.983
Logistic Regression	0.955
Random Forest	0.995
XGBoost	0.992

Table 8.1: Accuracy Comparison of Crop Recommendation Models

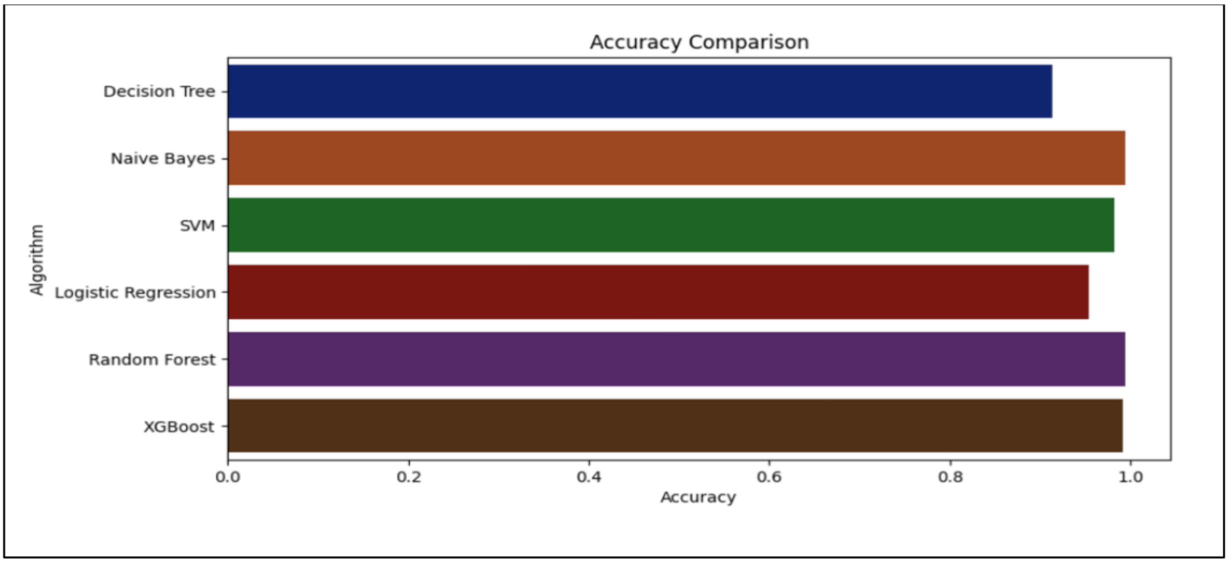


Fig 8.1: Accuracy Comparison of Crop Recommendation Models

8.2 Disease Detection

Advanced deep learning techniques for the detection of diseases in images of crops provide an output showing potential diseases present in the image. The actual disease that may be affecting the crop, including bacterial blight, powdery mildew, or rust, is precisely identified. Output Outcome: Output Outcome includes farmers with information related to the specific disease detected along with its severity. Related fungicides and biological controls to be adopted in treating the crop will be recommended. This will enable the farmer to take action prior to more damage of the crops and loss of yield.

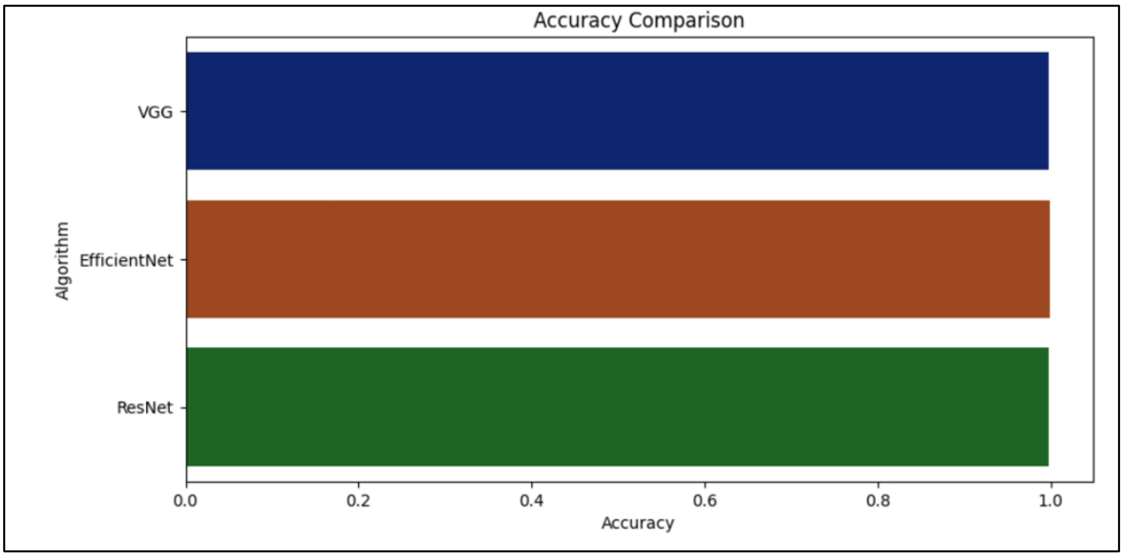


Fig 8.2: Accuracy Comparison of Plant Disease Detection Models

8.3 Fertilizer Recommendation

The fertilizers recommendation tool will provide appropriate fertilizers based on the requirement of nutrients in the soil, the type of crop and stages of growth of the crop. Therefore, the algorithm shall calculate how much and what kind of fertilizers should be put up at the appropriate dosage so the growth of the plants can become faster without applying too much of it since, overfertilization does huge damage to the crops and also to the surroundings. The outcome gives a farmer more detailed fertilizing schedules in terms of organic and chemical fertilizer options and therefore ensures proper soil health is balanced with good crop yields.

CHAPTER-9

RESULTS AND DISCUSSIONS

9.1 Home Page: This is the home page of a program that will help farmers implement well-thought decisions regarding the farming strategy. Solutions will be provided on some key questions of what to grow, which fertilizer to apply, and what type of diseases the crop may get infected with. The layout has a navigation bar with sections such as "Home," "Crop," "Fertilizer," and "Disease," showing the holistic approach the platform will take to managing agriculture. The background image of a farmer in a field at sunrise or sunset would be a visual connection to farming. Overall, the platform will help farmers by giving them data-driven insights that enhance productivity and sustainability in agriculture.

Fig 9.1: Home page

9.2 Platform Features: This can be proven by three major characteristics of the "AGRICULT" platform: this is environmental intelligence, the nature of which can improve a farming practice. The first is "Crop," which indicates advice on what crops should be planted under certain conditions in the environment and soil, and therefore makes easier optimization of the cultivation. The second is "Fertilizer," which advises what fertilizer is best used for the type of soil in question and recommended crop, keeping nutrient management in balance. The third feature, "Crop Disease," predicts crop diseases by showing the causes and treatment options.

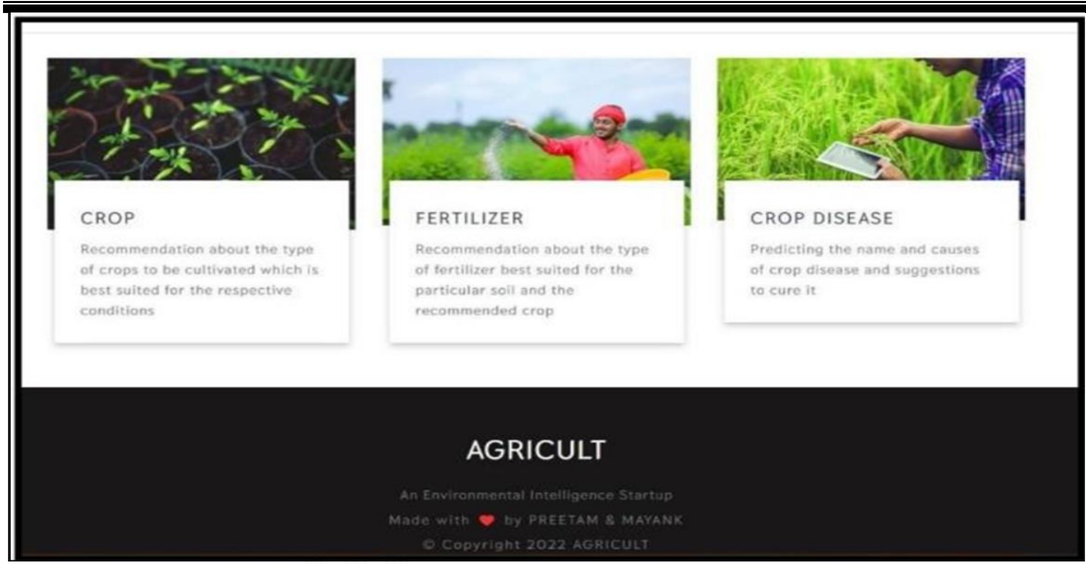


Fig 9.2: Platform Features

9.3 Crop prediction with input: The attached image is a user interface for Crop Recommendation System. It is supposed to decide the crop that has to be grown by the farmers by considering various agricultural inputs. User will be allowed to input several key parameters relating soil and environment, including Nitrogen, phosphorus, Potassium levels, pH level, Rainfall (in mm), and location details such as State and City. Once the user has filled in these values, they can click on the Predict button, which probably uses a machine-learning model to suggest the optimal crop for their farm. This system would support farmers in making data-driven decisions on soil health and local weather conditions.

Fig 9.3: Crop Recommendation System with input

9.4 Crop prediction system with output: Below is an image representing a user interface of a crop recommendation system. The system must, therefore likely guide farmers to pick crops for their environment. From the application of this case, the system had learned of the user's farm environment and went ahead to let him know it would be appropriate for him to grow coffee since his environment was rich in climate and other environmental features that are relevant for coffee crops. The idea of the system is to maximize crop selection so that there will be a good yield and farm productivity.

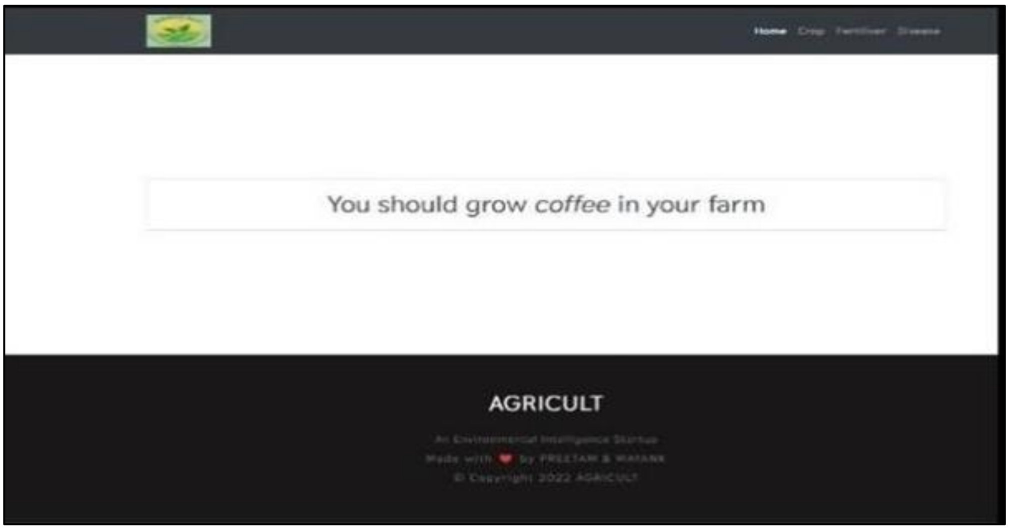


Fig 9.4: Crop Recommendation System giving output

9.5 Disease Detection system: This is the interface of a plant disease detection system. It looks like one that could be used to identify diseases affecting the plants. A user would upload an image of the diseased plant, and it most probably employs image recognition and machine learning algorithms to determine the specific disease affecting the plant. Results are presented before the user with possible treatment options or recommendations. Such a tool can be very useful to the farmers and the gardeners too, as one comes to know promptly which problem the plants are facing, and accordingly acts to salvage the crops.

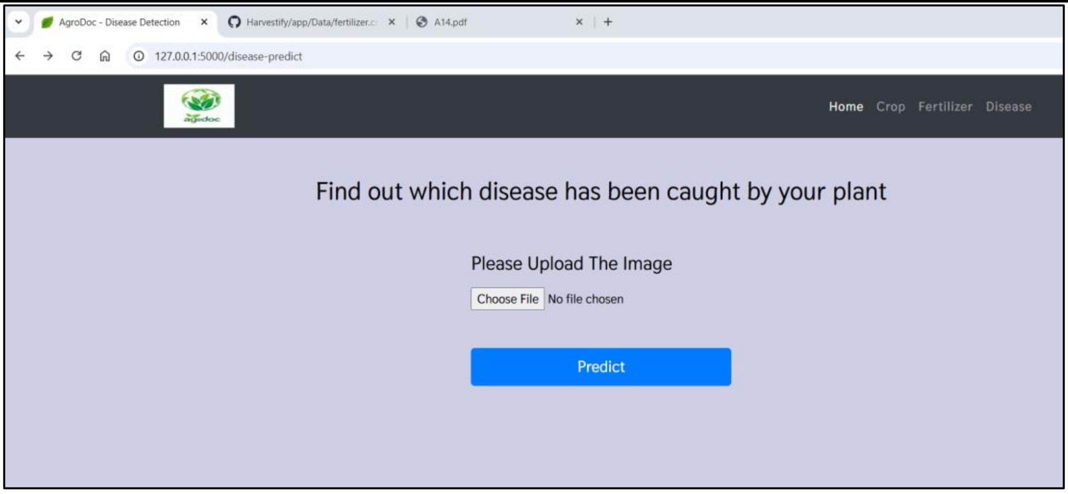


Fig 9.5: Disease Detection System

9.6 Disease Detection with output: It is actually a plant disease-detecting system, whereby the output image you have shown was from a result page. It scans the uploaded photo of the infested plant to determine that its problem is in fact Leaf Scorch in the Strawberry plant. Some of the information about the disease would appear on the result page, referring to the cause for which it is believed to be due to infection by a fungus called Diplocarpon earliana. It even takes it further to describe a control and treatment measure by proper garden sanitation, removing infected debris, and avoiding waterlogged soil. Users can reap this information to manage the disease in plants.

Fig 9.6: Disease Detection System giving output

CHAPTER-10

CONCLUSION

We propose, in this paper, a friendly user-based web application system based on machine learning and web scraping that we named 'Farmer's Assistant'. Our system has successfully applied to all features above: crop recommendation using Random Forest, fertilizer recommendation through a rule-based classification system, and detection of disease from the leaf image through EfficientNet. Our UI has forms as input, that returns the almost real-time output. We're also using a method called LIME interpretability to explain on this disease detection image what our model predicts, therefore maybe understanding the reason why the model predicts just that, then improve the corresponding datasets and model.

While our application runs quite smooth, still are too many areas through which we can make our application shining. For instance, during the provision of crop recommendations along with fertilizer recommendation, we might display the availability over most favorite shopping sites by allowing the consumer to shop for the same crop and fertilizers in the application itself.

Another dimension that might be exploited using the fertilizer recommendation is to seek other information referring to different brands and their various products regarding N, P, K values. We only present today six kinds of recommendations, but in the future, we shall have a possibility to work with highly advanced systems of machine learning, which will be really detailed.

Now, we are aware that the dataset we are working with here for disease classification is not exhaustive.

That means our model will perform only on those images that lie in the class for which the model has information. That will not be able to identify the class for out-of-domain data and that has to be addressed at a later point of time and that can be done in two ways. One way might be to hunt for other datasets similar in scale pertaining to other crops and/or diseases or generatively model/scale those datasets to add to our training set. In this way, our model is likely to be better generalized. The alternative option would be to let users input their images by creating an annotation portal for our web application. It has also proved that LIME explanations, by themselves, are not always reliable,

as they are only local explanations of the examples and not information about what it is the model focuses at a global level. We can therefore apply some other techniques such as GradCAM, Integrated Gradients, etc. or other training schemes like sparse-linear layers with LIME to give better explanations to our models' predictions.

Finally, we also have to give detailed segmentations of the diseased portion of the dataset.

Not available currently because we don’t have it.

In our application, we can include an addition of the segmentation annotation tool where users will help us cover the deficiency. Also, we will make use of some unsupervised algorithms where it can point out the disease-prone area of the image. We are going to feature that in our future work where we close these gaps.

REFERENCES

1. D.Jayanarayana Reddy, Dr. M. Rudra Kumar, “Crop Yield Prediction using Machine Learning Algorithm”, 2021.
2. R, Harish D, Priya B, “A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation”, 26th August 2022.
3. Suruliandi, G. Mariammal & S.P. Raha, “Crop prediction based on soil and environmental characteristics using feature selection techniques”, 12 Mar 2021.
4. Nishu Bali & Anshu Singal, “Deep Learning Based Wheat Crop Yield Prediction Model in Punjab Region of North India”, 16 Sep 2021.
5. Reyana, Sandeep Kautish, P. M. Sharan Karthik, Ibrahim Ahmed Al-Baltah, Muhammed Basheer Jasser, (Member, IEEE), And Ali Wagdy Mohamed,” Accelerating Crop Yield: Multisensor Data Fusion and Machine Learning for Agriculture Text Classification”, 27 February 2023.
6. Alexandros Oikonomidis, Cagatay Catal & Ayalew Kassahun, “Deep learning for crop yield prediction: a systematic literature review”, 06 Feb 2022.
7. Sujatha M., Jaidhar C.D, “Machine learning-based approaches to enhance the soil fertility”, 14 November 2023.
8. P K Devan, Swetha B, Uma Sruthi P, Varshini, “Crop Yield Prediction and Fertilizer Recommendation System Using Hybrid Machine Learning Algorithms”, 14th March 2023.
9. Mamatha, J.C. Kavitha c, “Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques”,
10. 10 January 2023. 10. Chetan R, D.V. Ashoka, Ajay Prakash B, “IMLAPC: Interfaced Machine Learning Approach for Prediction of Crops”, 11 January 2022.

11. Simraneet Kaur, Getanajali Babar, Navneet Sandhu, Dr. Gagan Jindal, „Various Plant disease detection using Image Processing Methods", June 2019.
12. Er. Varinderjit Kaur,Dr. Ashish Oberoi , „Wheat disease detection using svm classifier‟ , Aug 2018.
13. Barbedo, „Plant Disease Identification from individual lesions spots using deep learning‟, Apr 2019.
14. K. Naga Subramanian, A.K. Singh, A. Singh, S. Sarkar and Ganpath Subramanian, "Usefulness of interpretability methods to explain deep learning-based plant stress phenotyping", July 2020.
15. W. Yang, C. Yang, Z. Hao, C. Xie, and M. Li, „„„Diagnosis of plant cold damage based on hyperspectral imaging and convolutional neural network,‟‟‟ IEEE Access, vol. 7, pp. 118239– 118248, 2019.
16. K. Nagasubramanian, S. Jones, A. K. Singh, S. Sarkar, A. Singh, and B. Ganapathy subramanian, „„„Plant disease identification using explainable 3D deep learning on hyperspectral images".
17. N. Zhang, G. Yang, Y. Pan, X. Yang, L. Chen, and C. Zhao, „„„A review of advanced technologies and development for hyperspectral-based plant disease detection in the past three decades,‟‟‟ Remote Sens., vol. 12, no. 19, Sep. 2020, Art. no. 3188.
18. J. Chen, J. Chen, D. Zhang, Y. Sun, and Y. A. Nanehkaran, „„„Using deep transfer learning for image-based plant disease identification,‟‟‟ Comput. Electron. Agricult., vol. 173, Jun. 2020, Art. no. 105393.
19. M. Agarwal, A. Singh, S. Ajaria, A. Sinha, and S. Gupta, „„„ToLeD: Tomato leaf disease detection using convolution neural network,‟‟‟ Procedia Comput. Sci., vol. 167, pp. 293–301, Jan. 2020.
20. G. Hu, H. Wu, Y. Zhang, and M. Wan, „„„A low shot learning method for tea leaf's disease identification,‟‟‟ Comput. Electron. Agricult., vol. 163, Aug. 2019, Art. no. 104852.

APPENDIX-A

PSEUDOCODE

1. Disease Detection for Plants (Using CNN)

```
# Load Pre-trained CNN Model (e.g., ResNet, MobileNet)
MODEL = load_pretrained_model('PlantDiseaseModel')
# Preprocess the image
function preprocess_image(image):
# Resize the image to the required input size for the model (e.g., 224x224)
resized image = resize (image, (224, 224))
# Normalize pixel values
normalized_image = normalize(resized_image)
return normalized_image
# Predict disease from image
function predict_disease(image):
# Preprocess the image
input_image = preprocess_image(image)
# Expand dimensions to match model input (batch size, height, width, channels)
input_image = expand_dims(input_image, axis=0)
# Make prediction using the model
prediction = MODEL.predict(input_image)
   # Find the class label with the highest probability
   disease_class = argmax(prediction)
   # Map the class index to the corresponding disease name
   disease_name = map_class_to_disease(disease_class)
   return disease_name
  # Main function
   function detect_plant_disease(image):
   disease = predict_disease(image)
   return "Detected Disease: " + disease
```

2. Crop Recommendation

```
# Load pre-trained Crop Recommendation Model (e.g., Random Forest)
MODEL = load_model('CropRecommendationModel')
# Define a function to collect soil and environmental inputs
function collect_input_data():
# Collect soil pH
soil_ph = get_input("Enter soil pH: ")
# Collect nutrient values (Nitrogen, Phosphorus, Potassium)
nitrogen = get_input("Enter Nitrogen level (mg/kg): ")
phosphorus = get_input("Enter Phosphorus level (mg/kg): ")
potassium = get_input("Enter Potassium level (mg/kg): ")
# Collect climate and location data
location = get_location() # e.g., latitude, longitude
season = get_season()      # e.g., 'Summer', 'Winter'
temperature = get_temperature(location)
rainfall = get_rainfall(location)
# Return the collected data as a feature vector
return [soil_ph, nitrogen, phosphorus, potassium, season, temperature, rainfall]
# Predict the optimal crop based on input data
function recommend_crops():
# Collect input data
input_data = collect_input_data()
# Convert input data to model-compatible format
input_vector = preprocess_input(input_data)
# Get predictions from the model
crop_recommendations = MODEL.predict(input_vector)
# Return the top recommended crops (e.g., top 3)
top_crops = get_top_n_crops(crop_recommendations, 3)
return top_crops
# Main function to recommend crops
function crop_recommendation_system():
```

```
    recommended_crops = recommend_crops()
    return "Recommended Crops: " + join(recommended_crops, ', '
```

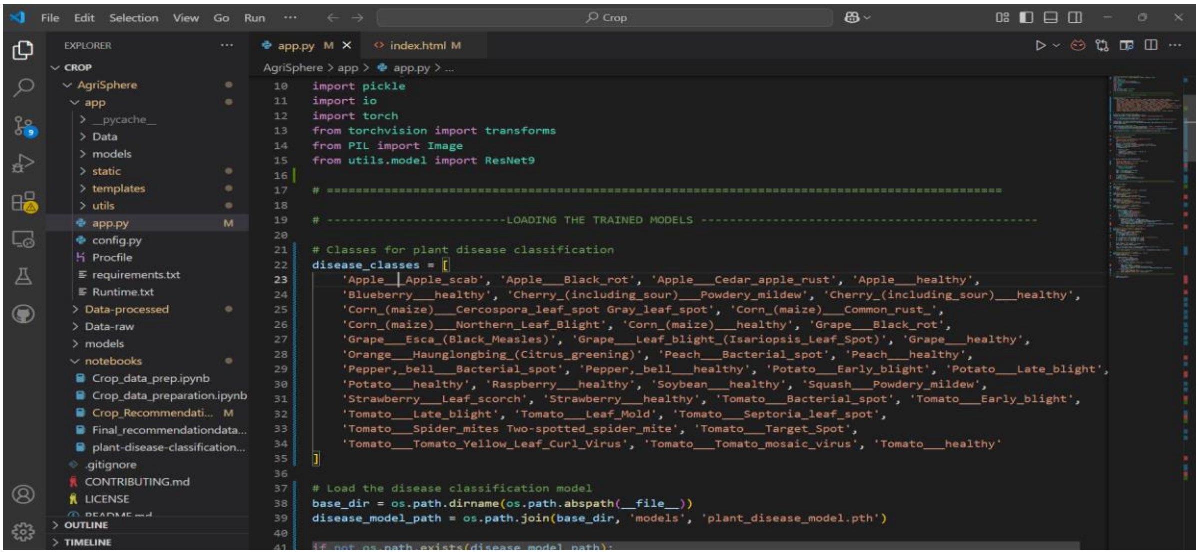
3. Fertilizer Recommendation

```
# Load pre-trained Fertilizer Recommendation Model (e.g., Regression or Decision Tree)
MODEL = load_model('FertilizerRecommendationModel')

# Collect soil and crop data
function collect_fertilizer_input():
# Collect soil nutrient levels
    nitrogen = get_input("Enter Nitrogen level (mg/kg): ")
    phosphorus = get_input("Enter Phosphorus level (mg/kg): ")
    potassium = get_input("Enter Potassium level (mg/kg): ")

 # Get current crop being grown
    crop_type = get_input("Enter crop type (e.g., wheat, rice): ")

 # Get real-time weather data
    location = get_location()  # e.g., latitude, longitude
    temperature = get_temperature(location)
    rainfall = get_rainfall(location)
    return [nitrogen, phosphorus, potassium, crop_type, temperature, rainfall]

# Recommend the appropriate fertilizer
function recommend_fertilizer():
# Collect input data
    input_data = collect_fertilizer_input()

# Convert the input data to a feature vector
    input_vector = preprocess_input(input_data)

# Get fertilizer recommendation from the model
    fertilizer_recommendation = MODEL.predict(input_vector)

# Extract NPK ratio and suggested quantity
    npk_ratio = fertilizer_recommendation['npk_ratio']
```

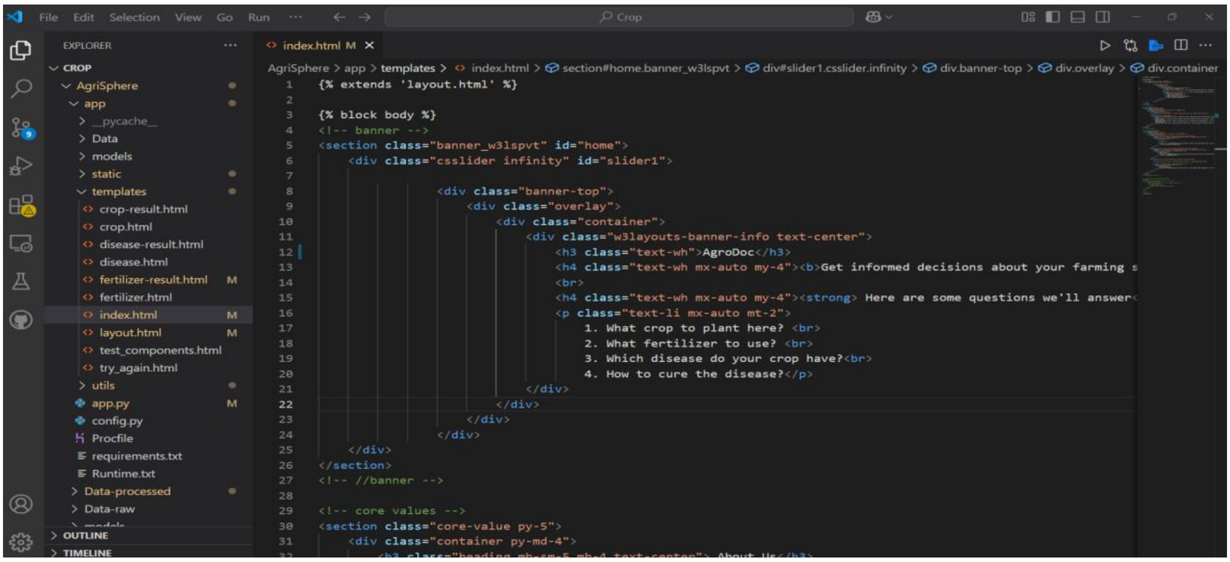
```
    fertilizer_quantity = fertilizer_recommendation['quantity']
      return "Recommended NPK Ratio: " + npk_ratio + ", Quantity: " + fertilizer_quantity + "
    kg/hectare"

# Main function for fertilizer recommendation
function fertilizer_recommendation_system():
    recommendation = recommend_fertilizer()
    return recommendation
```

APPENDIX-B

SCREENSHOTS



Screenshot 1: Workflow – Backend



Screenshot 2: Workflow – Frontend

APPENDIX – C

ENCLOSURES

1. Journal Publication Paper Presented Certificates of all students.

IJSREM e-Journal

ISSN: 2582-3930

Impact Factor: 8.448

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

VAISHNAVI M

in recognization to the publication of paper titled

INTEGRATED CROP PROTECTION SYSTEM

published in IJSREM Journal on Volume 09 Issue 01 January, 2025

Editor-in-Chief
IJSREM Journal

www.ijsrem.com

e-mail: editor@ijsrem.com

ISSN: 2582-3930

Impact Factor: 8.448

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

KUSUMA B

in recognization to the publication of paper titled

INTEGRATED CROP PROTECTION SYSTEM

published in IJSREM Journal on Volume 09 Issue 01 January, 2025

Editor-in-Chief
IJSREM Journal

www.ijsrem.com

e-mail: editor@ijsrem.com

ISSN: 2582-3930

Impact Factor: 8.448

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING & MANAGEMENT

An Open Access Scholarly Journal || Index in major Databases & Metadata

CERTIFICATE OF PUBLICATION

International Journal of Scientific Research in Engineering & Management is hereby awarding this certificate to

YAMUNA A K

in recognization to the publication of paper titled

INTEGRATED CROP PROTECTION SYSTEM

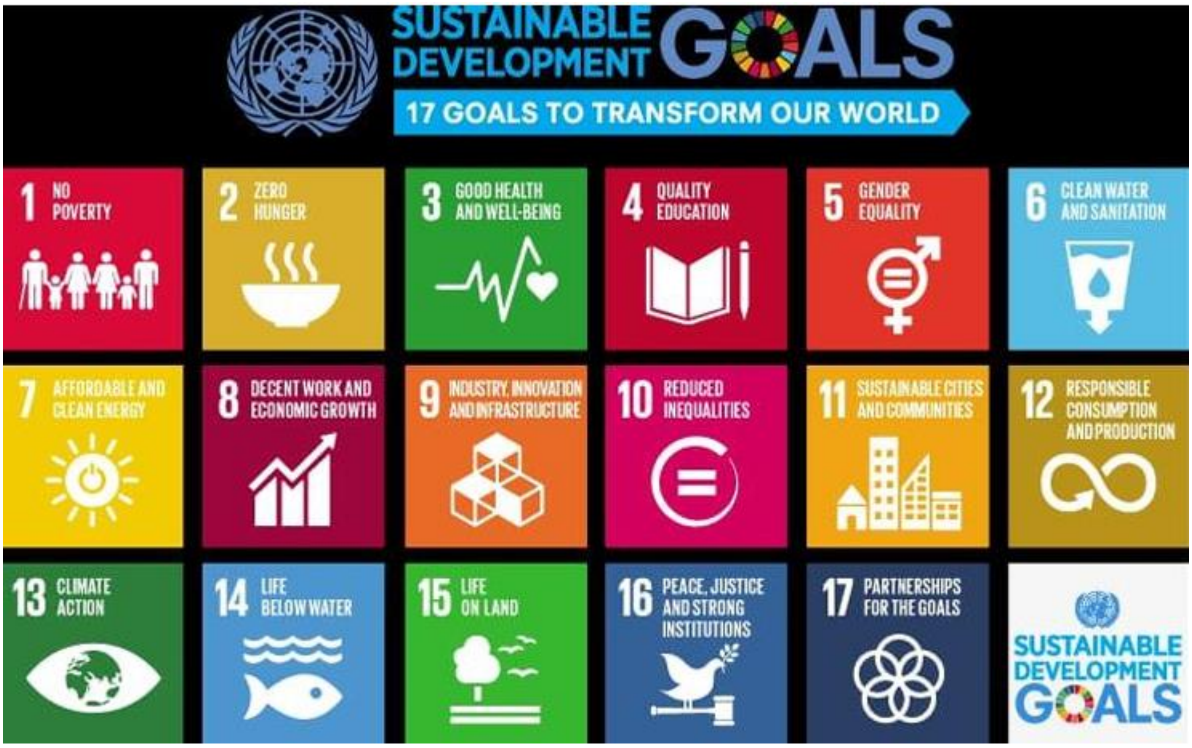
published in IJSREM Journal on Volume 09 Issue 01 January, 2025

Editor-in-Chief
IJSREM Journal

www.ijsrem.com

e-mail: editor@ijsrem.com

2. Details of mapping the project with the Sustainable Development Goals (SDGs).



- **SDG 2: Zero Hunger**

AgroDoc contributes to ending hunger by improving crop yields through optimized farming practices, disease detection, and resource-efficient fertilizer use. It helps ensure food security by reducing crop losses and promoting sustainable agriculture.

- **SDG 12: Responsible Consumption and Production**

The app promotes efficient use of natural resources, such as water, soil, and fertilizers, by offering data-driven recommendations. This reduces waste and minimizes the environmental impact of farming, supporting sustainable consumption and production practices.

- **SDG 13: Climate Action**

AgroDoc can assist farmers in adapting to climate change by integrating real-time weather data and offering guidance on how to respond to changing conditions. This enables farmers to mitigate climate-related risks.

- **SDG 9: Industry, Innovation, and Infrastructure**

By integrating deep learning and advanced technologies into agriculture, AgroDoc fosters innovation in farming techniques, contributing to the development of sustainable infrastructure in agriculture.

- ### SDG 15: Life on Land

AgroDoc helps protect ecosystems by promoting sustainable land management practices. By reducing overuse of fertilizers and improving crop health, the app supports the conservation of biodiversity.

erin V Simpson - ABSTRACT

ORIGINALITY REPORT

18%	17%	7%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ece.anits.edu.in Internet Source	14%
2	ijireeice.com Internet Source	2%
3	ebin.pub Internet Source	<1%
4	Submitted to Presidency University Student Paper	<1%
5	Submitted to University of Hertfordshire Student Paper	<1%
6	www.hindawi.com Internet Source	<1%
7	Submitted to The Robert Gordon University Student Paper	<1%
8	Sakshi Raina, Abhishek Gupta. "A Study on Various Techniques for Plant Leaf Disease Detection Using Leaf Image", 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), 2021 Publication	<1%

9	Submitted to Unity College Student Paper		<1%
10	Submitted to Republic of the Maldives Student Paper		<1%
11	www.coursehero.com Internet Source		<1%
12	Weiwei Lin, Ziming Wu, Longxin Lin, Angzhan Wen, Jin Li. "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis", IEEE Access, 2017 Publication		<1%
13	dokumen.tips Internet Source		<1%
14	www.researchgate.net Internet Source		<1%
15	edepot.wur.nl Internet Source		<1%
16	pdfs.semanticscholar.org Internet Source		<1%
17	tnsroindia.org.in Internet Source		<1%
18	www.ijitee.org Internet Source		<1%
19	www.ijraset.com Internet Source		

		<1%
20	www.kluniversity.in Internet Source	<1%
21	www.mdpi.com Internet Source	<1%
22	Ziming Wu, Weiwei Lin, Zilong Zhang, Angzhan Wen, Longxin Lin. "An Ensemble Random Forest Algorithm for Insurance Big Data Analysis", 22017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2017 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		