Crop Disease Prediction

A major project report submitted in partial fulfilment of the requirement for the award of a degree of

Bachelor of Technology

in

Computer Science & Engineering / Information Technology

Submitted by

Anant Rao (201160)

Medhavi Singh (201119)

Under the guidance & supervision of

Dr. Amit Kumar Jakhar

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology, Waknaghat, Solan - 173234 (India)

CERTIFICATE

This is to certify that the work which is being presented in the project report titled "Crop Disease Prediction" in partial fulfillment of the requirements for the award of the degree of B.Tech in Information Technology and submitted to the Department of Computer Science & Engineering And Information Technology, Jaypee University of Information Technology, Waknaghat is an authentic record of work carried out by "Medhavi Singh (201119) and Anant Rao(201160)" during the period from August 2023 to May 2024 under the supervision of Dr. Amit Kumar Jakhar, Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat.

Medhavi Singh 201119 Anant Rao 201160

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Dr. Amit Kumar Jakhar
Assistant Professor (SG)
Computer Science and Engineering/Information Technology Department
Dated:

Candidate's Declaration

We hereby declare that the work presented in this report entitled **'Crop Disease Prediction'** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science & Engineering / Information Technology** submitted in the Department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology, Waknaghat is an authentic record of my own work carried out over a period from August 2023 to May 2024 under the supervision of **Dr. Amit Kumar Jakhar** (Assistant Professor(SG), Department of Computer Science & Engineering and Information Technology).

The matter embodied in the report has not been submitted for the award of any other degree or diploma.

(Student Signature with Date)
Student Name: Medhavi Singh
Roll No.: 201119

(Student Signature with Date)
Student Name: Anant Rao
Roll No.: 201160

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

(Supervisor Signature with Date)
Supervisor Name: Dr. Amit Kumar Jakhar
Designation: Assistant Professor (SG)
Department: Computer Science and Engineering
Dated:

ACKNOWLEDGEMENT

Before we begin, we would like to express our heartfelt gratitude to almighty God for his heavenly grace, which enabled us to successfully complete the project work.

We are so grateful and wish to express our deep gratitude to Supervisor **Dr. Amit Kumar Jakhar, Assistant Professor (Senior Grade), Department** of CSE&IT Jaypee University of Information Technology, Waknaghat. His never-ending patience, intellectual direction, persistent encouragement, constant and vigorous supervision, constructive criticism, helpful suggestions, and reading numerous poor versions and revising them at all stages allowed this project to be completed. We would like to express our heartiest gratitude to Dr. Amit Kumar Jakhar**,** Department of CSE&IT, for his kind help to finish our project.

We would also like to express our gratitude to everyone who has assisted us in making this project a success, whether directly or indirectly. In this unusual scenario, we did like to express our gratitude to the different staff members, both teaching and non-teaching, who have provided us with valuable assistance and assisted our project. Finally, we must express our gratitude for our parents' unwavering support and patience.

Medhavi Singh 201119 Anant Rao 201160

TABLE OF CONTENT

CONTENT **PAGE NO.**

LIST OF TABLES

LIST OF GRAPHS

CONTENT **PAGE NO.**

LIST OF FIGURES

ABSTRACT

Globally it is necessary to address a critical matter on the provision of food in order to sustain a growing population. In that sense, our research project shines as a vanguard developing an unprecedented path to forecasting crop diseases by using data science, information technology, machine learning techniques, and advanced technologies. Through this project we have witnessed how critical transformations in agriculture as a basis for our society have strengthened our farmers.

The emergence of data science has introduced new navigation techniques in the ocean of agricultural data and consequently unearth hidden beneficial information for agricultural researchers and farmers. We utilise complex data analytical approaches in analysing our project datasets consisting of various variables including historic crops, climatic conditions, as well as diseases. Our predictive model is based on this comprehensive analysis, which provides greater insight into the dynamic relationships between these factors.

Our framework features cutting-edge machine learning tools that are used to decimate leaf images in the most precise manner possible. Such a thorough investigation enables even small but distinct patterns which are indicators of many crop diseases to be revealed. The human-taught model after being trained on various datasets becomes an advanced diagnostic tool that can detect symptoms of most diseases which are usually not detected by the human eye. In turn, such a conceptual breakthrough has revolutionised the way of diagnosing plant diseases thus resulting in the notable reduction of yield losses caused by plant diseases. This project's reliance on technology goes beyond just diagnostic purposes, as it also helps to develop new treatment options. Researchers take a close look at the intricate details that cause the diseases while the machine-learning model attempts to decipher the intricate nuances that have an impact on disease patterns. Greater understanding of diseases leads to more specific and efficient treatments that change the disease management paradigm.

By focusing strongly on the importance of farmers' accessibility and usability of the prediction model in our project approach, we take great consideration to satisfy different agriculturally interested areas. The core aim herein is enabling farmers to obtain helpful ideas prior to making the right choices. To this end, we have designed user-friendly interfaces to ensure that benefits of our study are available in real time and comprehensible. Significantly, we emphasize on democratization of the technology ensuring even big and small farms access our predictions. Lastly, we have ensured that we prioritized our systems' safety. Our system allows access and contribution for permitted people only by means of several layers of authentication. By strictly complying with laws on protecting data privacy and an assuring atmosphere, we provide data privacy and security for farmers' data. We hope to build trust and reliability by merging stern security provisions in our solution. This effort will ensure widespread acceptance of our predictive models in agriculture for effective disease management.

Finally, this project is a holistic and revolutionary one with the intersection line of agriculture, data science, machine learning, and technology. In addition, it proves that an effective modern technology can change tomorrow’s agriculture. This research lays a foundation for technologically driven agriculture that is robust and forward-looking for tackling future issues in this new age.

CHAPTER-1 INTRODUCTION

1.1 INTRODUCTION

Modern age agriculture is underpinned by a new era wherein machine learning tools and data sciences are fully integrated with Information technology. This is just another major stride that seeks to deploy technology in the fight with the stubborn crop disease challenge. We adopt a data-driven approach that meticulously analyses data sets covering previous crop reports, weather features and disease cases. Adding machine learning into the equation helps dig deeper into the detail of each image for the analysis of specific disease markers. This innovative use of machine learning enhances accuracy and provides new research angles into disease identification. The detailed analysis of these insights forms a basis on which new therapeutic options could be built on. The project is multidimensional and it predicts more than predictions providing practical information for farmers for shift towards anticipatory disease management. In studying how technological advancement can help make crop resistant to disease and provide a more viable environment for future farmers, this paper seeks to address these issues in order to come up with solutions that will make things better for the global farming society in the near future.

Modern age agriculture is underpinned by a new era wherein machine learning tools and data sciences are fully integrated with Information technology. This is just another major stride that seeks to deploy technology in the fight with the stubborn crop disease challenge. We adopt a data-driven approach that meticulously analyzes data sets covering previous crop reports, weather features and disease cases. Adding machine learning into the equation helps dig deeper into the detail of each image for the analysis of specific disease markers. This innovative use of machine learning enhances accuracy and provides new research angles into disease identification. The detailed analysis of these insights forms a basis on which new therapeutic options could be built on. The project is multidimensional and it predicts more than predictions providing practical information for farmers to shift towards anticipatory disease management. In studying how technological advancement can help make crops resistant to disease and provide a more viable environment for future farmers, this paper seeks to address these issues in order to come up with solutions that will make things better for the global farming society in the near future.

This course takes the viewers through the intricacies of this case unravelling how the predictions can be changed using a noncontaminating method. By utilizing an innovative approach to data analyses and state of the art machine learning algorithms, we hope to set the pace in building the way forward. However, diagnosis of health beginning herein should go beyond detection and response because they should have been prevented beforehand. A possible scenario is one where farmers have easy access to timely and correct information needed for informed decision making, reducing the impact on plant diseases. Our goal is to explore this paradigm shift in forecasting future crop diseases and how it contributes towards not only strengthening individual farms, but also toward overall sustainability of our agricultural systems, so as our very foundation remains resilient even during these times.

A holistic, long-term solution is required in tackling crop diseases. The study exploits the revolutionary ability of machine leaning algorithms and data analytics in disentangling intricate connections among historical crops' data, environmental variability, and disease outbreaks. Through this process, we wish not only to forecast the emergence of agricultural diseases, but for our actions to be useful and pragmatic for the farmer. This innovative approach attempts to put arms into hands of farmers so that they could make knowledge-based decisions on uncertain agricultural environment.
Modern Agriculture continues to be challenged with changing scenarios, and this research is largely driven at these shifting challenges. The environment where crops grow changes for reasons of climate change uncertainty, as well as due to globalization of trade. Under such circumstances, it is imperative to establish an all-encompassing crop disease forecast system that transcends common preventive measures. Finally, this expedition aims at strengthening our agriculture, preparing us for impending dangers while fostering long-term stability for worldwide food production.

1.2) PROBLEM STATEMENT

Considering agriculture represents the most important occupation in the Indian economy, it brings aboard a slew of issues that can be disastrous for farmers. Addressing these concerns becomes critical, and utilizing modern equipment and approaches to relieve the problems faced by farmers is critical. Recognizing the critical importance of agriculture in the Indian economy requires not only acknowledging the difficulties but also actively engaging in innovative solutions that may empower and improve the agricultural community.

At this critical moment, a number of severe challenges are hindering crop disease prediction, with far-reaching implications for data privacy, security, forecast accuracy, and agricultural sustainability. The massive use of pesticides, fueled by incorrect predictions, poses a major threat to the environment and the agricultural industry's long-term viability. Errors in present prediction models exacerbate the situation, resulting in inaccurate diagnoses, wasteful pesticide treatments, and yield losses that severely reduce agricultural output.

The problem intrudes on the domain of data exchange and privacy. Farmers are hesitant to disclose essential information due to major privacy concerns, making it difficult to gather the massive and diverse datasets required to construct effective prediction models. This reluctance breeds ingrained concerns about data security and integrity, which are exacerbated by agricultural data's sensitivity to breaches. The sensitive nature of agricultural data, which includes crucial information on crop health and disease, raises the possibility of privacy infractions and generates major concern for farmers and other stakeholders.

Crop disease prediction encompasses a variety of interconnected challenges that must be addressed comprehensively in order to increase accuracy, safeguard data privacy, and boost data security standards. Maintaining agricultural data confidentiality and integrity while creating robust and long-lasting prediction models requires finding a compromise between these variables.

1.3) OBJECTIVE

This research project is taking shape with a multidimensional set of goals based on the combination of data science, machine learning technologies, and cutting-edge technology. Our major goal is to transform crop disease prediction by using the wealth of information contained in large datasets. We hope to uncover the subtle links between historical crop data, environmental variables, and disease patterns through meticulous data analysis. Our goal goes beyond traditional methodologies, diving into the area of leaf image dissection to detect subtle illness markers using the power of machine learning.

1. **Create a Prediction Model:** One, this is a complex and reliable project aimed at generating an agricultural disease predication model. The promise of modern machine learning methods is realized by using sophisticated algorithms and approaches within this model. Therefore, it primarily aims at true positive identification and detection of the specific diseases that affect agriculture. The computer will consider many types of information, such as crop pictures, climate elements, and past illness patterns through machine learning.

2. **Enhance Data Security:** This calls for a well-developed security architecture in the process to secure all agricultural data from any threat or breach. This is the primary goal. The installation of state-of-the art security processes is incorporated in this project to safeguard highly confidential agribusiness information. Protecting against unauthorized access and data breaches forms the core objective of the system's role in enforcement of access controls. The design of the enhanced data security architecture is carefully constructed to ensure the privacy and confidentiality of the stakeholder's and farmer's data.

3. **Reduce Pesticide Usage:** Reducing the use of pesticides as well as promoting a sustainability and minimization of pesticide environmental footprint. This is associated with the utilization of credible disease prediction models that give accurate information on the state of crops to farmers. This makes predictions of actual diseases that are likely to occur and allows them to apply a minimal amount of pesticides. It involves supplying comprehensive data concerning the frequency and intensity of rural illnesses and permitting farmers to use insecticides upon need and at their places of occupation. During this transition to focused or specific application of pesticides, there are numerous significant benefits.

4. **Measure Economic Impact:** Economic assessment of secure crop disease prediction system involves detailed study of the system's contribution in boosting productivity and profits. This study aims at ascertaining whether there can be monetary gains accrued from setting the model, particularly in their respective farms. The purpose here

is to evaluate the economic value addition from the assured plant disease prediction framework with regard to higher output, fewer losses, and general farming revenues.

1.4) SIGNIFICANCE AND MOTIVATION

SIGNIFICANCE:

This undertaking has a far-reaching effect on various significant fields. It has now emerged as an important element in enhancing agricultural sustainability and productivity. Accurate predictive modeling of diseases is essential in supporting agriculture as a major component of every economy in providing healthy crops. In this respect, the main objective of this project seeks to identify and control the spread of crop diseases with an aim of preventing huge agricultural losses and sustaining food security which constitutes the essential element of these countries.

Reducing economic risk to farmers is among the primary goals of this project. The crop diseases often lead to huge economic losses which puts at stake the viability of farm operations. Through such advance warning and precautionary measures, this program aims at minimizing economic exposures that give rise to a more robust and effective agricultural terrain.

Furthermore, the proactive method of diagnosing and managing crop diseases contributes greatly to the maintenance of a stable food production ecology. This research plays a critical role in guaranteeing stable food supplies by preventing potential crop ailments, hence lowering the dangers of food scarcity caused by disease-induced crop failures.

Finally, the application of advanced agricultural technology such as machine learning, image processing, and data analysis represents a significant movement toward innovation in farming operations. This integration, in addition to disease prediction, sets the door for further technical improvements, providing an atmosphere suitable for continual progress and innovation in the agricultural industry.

MOTIVATION:

This project's motivation has been diverse and impactful in several critical areas. To begin, it intends to boost the productivity of crops by enabling early disease identification and empowering farmers to take preventive measures quickly. This strategy supports long-term agricultural output, ensuring healthy crop yields that are critical to the world food supply.
Similarly, the study aims to help farmers make informed decisions. It enables farmers to make well-informed decisions about crop management, resource allocation, and disease control measures by offering predictive insights on crop diseases, hence optimizing

agricultural practices. This campaign embraces agricultural technological breakthroughs, recruiting tech-savvy persons to farming methods. Modern technology integration not only modernizes the agricultural sector, but it also has the ability to improve overall farming processes through innovation and efficiency.

The project has the potential to have a beneficial socioeconomic impact. It improves farmers' socioeconomic conditions by decreasing crop losses caused by diseases, and lowering financial vulnerabilities while promoting improved livelihoods within farming communities. Another critical factor is environmental sustainability. The research encourages sustainable farming practices by using targeted pesticide applications based on disease prediction, considerably decreasing environmental impact while ensuring effective disease management.

Finally, the project's concentration on crop disease prediction research development enhances the knowledge base in this domain. This need for further research adds to a better understanding of crop diseases, driving innovation in disease prediction and management approaches, and thereby helping the agricultural sector as a whole.

In summary, the project's importance stems from its potential to transform agricultural practices, improve food security, improve farmers' socioeconomic conditions, and pave the path for sustainable and technologically sophisticated farming methods.

1.5) Organization of Project Report

1. **Project Report Introduction:**
    - Goals and Objectives.
    - Importance and inspiration

2. **Content and Setting:**
    - Analysis of Agricultural Landscapes
    - Current Crop Disease Management Challenges

3. **Project Overview:**
    - Project Methodology
    - Utilised Tools and Technologies for User Interface Design

4. **Problem Synopsis:**
    - Importance of Crop Disease Identification
    - Issues Handled Effect on Farm Productivity

5. **Important and Motivating Elements of the Project:**
    - Democratising Predictive Models

- Empower Farmers via Technology
- Security Protocols

6. **Chapter Findings:**
    - Review of the Literature and Comparative Study
    - Gathering and preparing data; training and assessing models
    - Evaluation of Security and Privacy

Every segment will offer comprehensive analyses, strategies utilised, obstacles encountered, and the results attained during the specific phases of the undertaking. The goal of the paper is to provide a thorough overview of the research, development, and use of the crop disease prediction system, highlighting its importance, novel features, and added value to farming practices.

CHAPTER-2 LITERATURE SURVEY

The primary goal of this project is to create an advanced and resilient system dedicated to predicting and managing crop diseases, ultimately contributing to the substantial improvement of agricultural productivity. Through the fusion of cutting-edge machine learning methodologies and sophisticated image processing techniques, the project endeavours to establish a comprehensive framework capable of accurately detecting and diagnosing diseases affecting various crops. Additionally, the system will not only identify these diseases but will also propose targeted solutions and recommendations, empowering farmers with actionable insights for efficient disease control and crop management.

2.1) OVERVIEW OF RELEVANT LITERATURE

1. **“A Convolutional Neural Network Model for Wheat Crop Disease Prediction" by Mahmood Ashraf et al. focuses on addressing the issue of wheat crop diseases and their impact on production [1]:**

    It introduces a modified CNN architecture for effective disease detection, using datasets collected from wheat fields in Azad Kashmir, Pakistan. The proposed model is lightweight, designed to potentially operate on smartphones, and aims to provide real-time disease prediction. The study examines three different dataset variants to evaluate the model's performance, achieving a 93% accuracy rate with the most extensive training set.

2. **"DeepCrop: Deep learning-based crop disease prediction with web application." by Md. Manowarul Islam [2]:**

    The paper addresses the significance of agriculture in Bangladesh and focuses on the cultivation of various crops like corn, peach, grape, potato, and strawberry. It highlights the challenges faced by farmers due to plant diseases and pests, causing significant financial losses. The introduction emphasizes the prevalence of diseases specific to each crop and their impact on production rates.
    The proposed methodology involves using image processing techniques and Convolutional Neural Network (CNN) models to detect and classify leaf diseases in crops. The dataset used for training and testing consists of thousands of images for different diseases and healthy leaves of each crop.

3. **"Machine Learning Technique for Crop Disease Prediction Through Crop Leaf Image" by S. Nandhini and K. Ashokkumar explores the application of deep learning techniques for identifying crop diseases through leaf images [3]:**

The paper addresses the critical issue of crop diseases threatening food security and livelihoods, especially for small-scale farmers. It introduces the use of smartphone technology and machine learning, particularly convolutional neural networks (CNNs), to detect diseases in crops based on leaf images. The study trained a CNN on a dataset comprising 64,412 images of damaged and normal leaf tissue, achieving a high accuracy of 99.35% in detecting 16 crop species and 25 diseases. The research aims to enable widespread, accessible plant disease detection via smartphones, leveraging the ubiquitous presence of high-definition cameras and processing power in these devices.

4. **The paper titled "Crop Disease Prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata" by Kiran Sree Pokkuluri and SSSN Usha Devi Nedunuri [4]:**

   It addresses the critical concern of food security by focusing on predicting crop diseases and yield. The work introduces two novel classifiers, CNN-CA-I and CNN-CA-W, leveraging deep learning techniques to process images and weather data, respectively, for disease and yield prediction across prominent crops like rice, wheat, barley, sugarcane, cotton, and oilseeds.

5. **The paper titled "CROP DISEASE PREDICTION AND SOLUTION" by Nikita Yadav, Shivani Kasar [5]:**

   It focuses on the critical issue of crop disease detection and management, vital for sustaining agricultural productivity. It emphasizes the impact of diseases on crop quality and quantity, proposing a research framework that integrates machine learning (ML) and image processing techniques for accurate disease detection and subsequent solutions.

6. **The paper titled "Crop Disease Prediction Using Deep Learning Techniques" by Gargi Sharma and Gourav Shrivastava [6]:**

   It explores the potential of deep learning models in conjunction with smartphone technology to aid in the rapid identification of crop diseases. It leverages the increasing global penetration of smartphones, their high-resolution cameras, and recent advancements in computer vision powered by deep learning algorithms. Using a dataset of 54,306 images encompassing 14 crop species and 26 diseases, the study demonstrates the feasibility of training deep convolutional neural networks to identify these diseases or the absence thereof, achieving an impressive accuracy of 99.35% on a test set.

2.2) KEY GAPS IN THE LITERATURE

1. **Data Accessibility and Quality:**

    - **Limited Datasets**: The review emphasizes the scarcity of extensive and validated datasets of damaged and healthy crops, highlighting the PlantVillage dataset as a key resource. Lack of larger datasets or publicly available comprehensive datasets restricts the effectiveness of training models for accurate disease classification.

    - **Biased Data Collection:** There's a concern about biases in collected datasets due to standardized data collection processes, potentially affecting the model's performance when applied to real-world scenarios.

2. **Model Generalization and Real-World Application:**

    - **Limited Real-World Application:** While the CNN achieved high accuracy during testing on the provided dataset, there's a gap in its applicability to diverse real-world conditions. The review mentions limitations in the model's ability to recognize diseases beyond leaves or in varied environmental settings, affecting its practical use.

    - **Need for Diverse Perspectives:** The study primarily focuses on classifying individual leaves against a uniform background. However, many diseases affect different parts of the crop beyond the leaves, indicating a need for diverse image perspectives and more comprehensive data collection.

3. **Reliability and Scalability:**

    - **Robustness in Variable Conditions**: The review highlights challenges in the model's reliability when tested on images collected under different settings. While the accuracy remains significantly higher than random classification, a more diversified and expansive training dataset is essential for improved reliability in varied conditions.

    - **Scope for Improvement:** Despite achieving high accuracy with the chosen classes (number of crops and diseases), there's room for improvement by potentially narrowing the classification task or focusing solely on disease states, which could enhance practical usability.

4. **Future Directions and Methodology:**

   - **Realistic Data Collection:** The study suggests the importance of gathering images from various perspectives and environments to mimic real-world scenarios accurately.

   - **Enhanced Model Training:** Future research should focus on expanding datasets, reducing biases, and refining models to cater to real-world complexities, ultimately improving the accuracy and applicability of disease detection in crops.

Addressing these gaps would enhance the reliability, applicability, and scalability of machine learning models in detecting crop diseases through leaf images, potentially offering valuable support to farmers and agricultural practices globally.

Table 1: Related literature on plant disease using deep learning models.

Object	Total no. of images and classes used	DL frames	Accuracy (%)
Citrus leaf	609 images with	Inception-v3, VGG-19	Highest accuracy
	5 classifications	and VGG-16	for VGG16–89.5%
Apple leaf	2462 images with 6 classifications	DenseNet-121	93.71%
Potato leaf	2152 images with 3 classifications	VGG19	97.8%
Tomato leaf	736 images with 4 classifications	CNN based approach	98.12%
Tomato leaf	7500 images with 9 classifications	CNN based approach	91.2%
Paddy leaf	120 images with 2 classifications	DNN-CSA	96.96%
Citrus leaf	598 images with	Two-stage deep	94.37%
	4 classifications	CNN model	

CHAPTER-3 SYSTEM DEVELOPMENT

3.1) REQUIREMENTS AND ANALYSIS

Creating a robust crop disease prediction system necessitates the fulfilment of several critical requirements. To begin with, an extensive and diversified dataset including multiple crop types, diseases, and weather patterns needs to be collected. This dataset is used to train machine learning models like Random Forests and Neural Networks, ensuring their adaptability to various scenarios. Using modern image processing techniques that include noise reduction and feature extraction also improves the accuracy of disease identification from crop images.

Security is of the utmost importance. To protect sensitive agricultural data, the system must employ severe security features such as secure data administration, and strong user authentication. Simultaneously, a user-friendly interface that farmers and stakeholders can access is vital. This interface should allow for quick access to disease information, estimates, and solutions, allowing for more informed crop management decisions.

Another critical issue is optimizing pesticide use through precise disease predictions. This has the advantage of minimizing environmental impact while efficiently managing crop diseases. Extensive testing and validation are required to determine the prediction model's accuracy, scalability, and dependability, as well as to ensure its robustness across varied agricultural circumstances.

Critical steps include analyzing the dataset for correctness and diversity, evaluating model performance based on precision and recall metrics, reviewing security risks, and testing the user interface for accessibility and functioning. Furthermore, determining the system's performance and benefits requires assessing its economic, environmental, and social impact through field testing and user input. Lastly, assessing scalability, optimization efficiency, and long-term sustainability guarantees that the system remains effective and pertinent in the agricultural sector.

FUNCTIONAL REQUIREMENTS:

1. **Collection of Data and Processing:**

   - **Data Collection:** Gather data regarding various crop diseases, and environmental conditions.
   - **Image Processing:** Create algorithms to process crop photos in order to detect and classify diseases.

- **Extraction of relevant features** from images or data to train predictive algorithms.

2. **Prediction and recommendation:**

    - **Disease Prediction**: Create machine learning models that accurately predict crop diseases based on data and photos collected.
    - **Recommendation Engine:** Based on illness identification, suggest relevant remedies or solutions (pesticides).

3. **User Interface (UI):**

    - **Dashboard:** Construct a user-friendly interface for farmers or users to interact with the system.
    - **Input Mechanism:** Make it simple for users to enter data or images for disease prediction.
    - **Output Display:** Clearly and thoroughly present the forecast results and recommended solutions.

4. **Model Training and Development:**

    - **Model Training:** Using obtained data, train machine learning models and continuously improve them over time.
    - **Model Validation:** Put in place tools to test model accuracy and performance.

5. **System Maintenance:**

    - **System Upkeep:** Ensure that the system can be updated with fresh disease data, environmental changes, or enhanced algorithms on a regular basis.

NON-FUNCTIONAL REQUIREMENTS:

1. **Performance:**

   - **Accuracy:** Aim for high accuracy (>90% accuracy) in disease prognosis and recommendation.
   - **Response Time:** Ensure speedy response times for predictions and recommendations in order to provide farmers with real-time help.

2. **Scalability:**

   - **Data Scalability:** Make the system capable of handling a huge volume of data as the dataset grows over time.
   - **User Scalability:** Allow several users to access the system at the same time without affecting performance.

3. **Security:**

   - **Data Privacy:** Maintain privacy standards while ensuring the confidentiality and integrity of user data.
   - **Authentication:** Implement secure login techniques to allow farmers or users allowed access.

4. **Reliability:**

   - **System Availability:** Ensure that the system is always available and operable for users.

5. **Usability:**

   - **Intuitiveness:** Make the user interface straightforward and simple to use for people with varied levels of technical experience.
   - **Compatibility:** To ensure extensive usage, ensure compatibility across many devices and platforms.

The effectively stated requirements serve as the pillars that underlie the crop disease prediction system, providing its efficacy, optimal performance, robust security, and user-friendly interface. These standards, with a delicate balance, adapt to the dynamic demands of farmers seeking accurate insights into crop management and system administrators overseeing its functioning. Functionality guarantees accurate forecasting, performance assures operational excellence, security protects against threats, and usability ensures accessibility. This comprehensive approach not only strengthens the system's technological prowess but also ensures its usability, flawlessly matching the different needs of farmers and administrators alike in the complex domain of crop disease prediction.

3.2) PROJECT DESIGN AND ARCHITECTURE

1. **Data Collection and Preprocessing:** The first step is to collect a broad dataset of crop leaf photos damaged by various illnesses. Following that, preprocessing procedures are carried out in order to prepare the images for in-depth analysis.

2. **Feature Extraction and Model Development:** Relevant features are extracted from preprocessed pictures, opening the road for machine learning models to be developed. Based on these extracted data, several models such as CNNs, Decision Trees, and Random Forests are used to categorize and predict crop diseases.

3. **Backend Development:** The trained models are integrated into the system, which is enhanced further by the creation of a recommendation system. Based on the diagnosed diseases, this system recommends appropriate pesticides or management measures, boosting the platform's practical utility.

4. **User Interface Development:** To ensure accessibility, a user-friendly interface is designed using a web portal or application. Secure login procedures are used to preserve data privacy, offering a seamless and secure user experience.

5. **Testing and Integration:** The user interface is seamlessly integrated to the backend prediction and recommendation engines. Extensive testing is carried out to evaluate the system's operation and dependability in real-world circumstances.

6. **Scalability and Deployment:** The system has been precisely built to be deployed locally on dedicated servers. Its architecture is designed to allow for scalability within the local area while also providing adaptation to changing needs.

7. **Architecture:** The architecture is made up of a current web technology-driven frontend user interface, a Python-powered backend system for model integration and algorithm execution, and secure databases that store image metadata, disease information, and user data.

8. **Monitoring and maintenance:** To ensure peak performance, monitoring tools track both system performance and user interactions. Regular updates, bug repairs, and enhancements are prioritized based on user feedback and technology advances, assuring the system's long-term success.

Proposed work-flow diagram

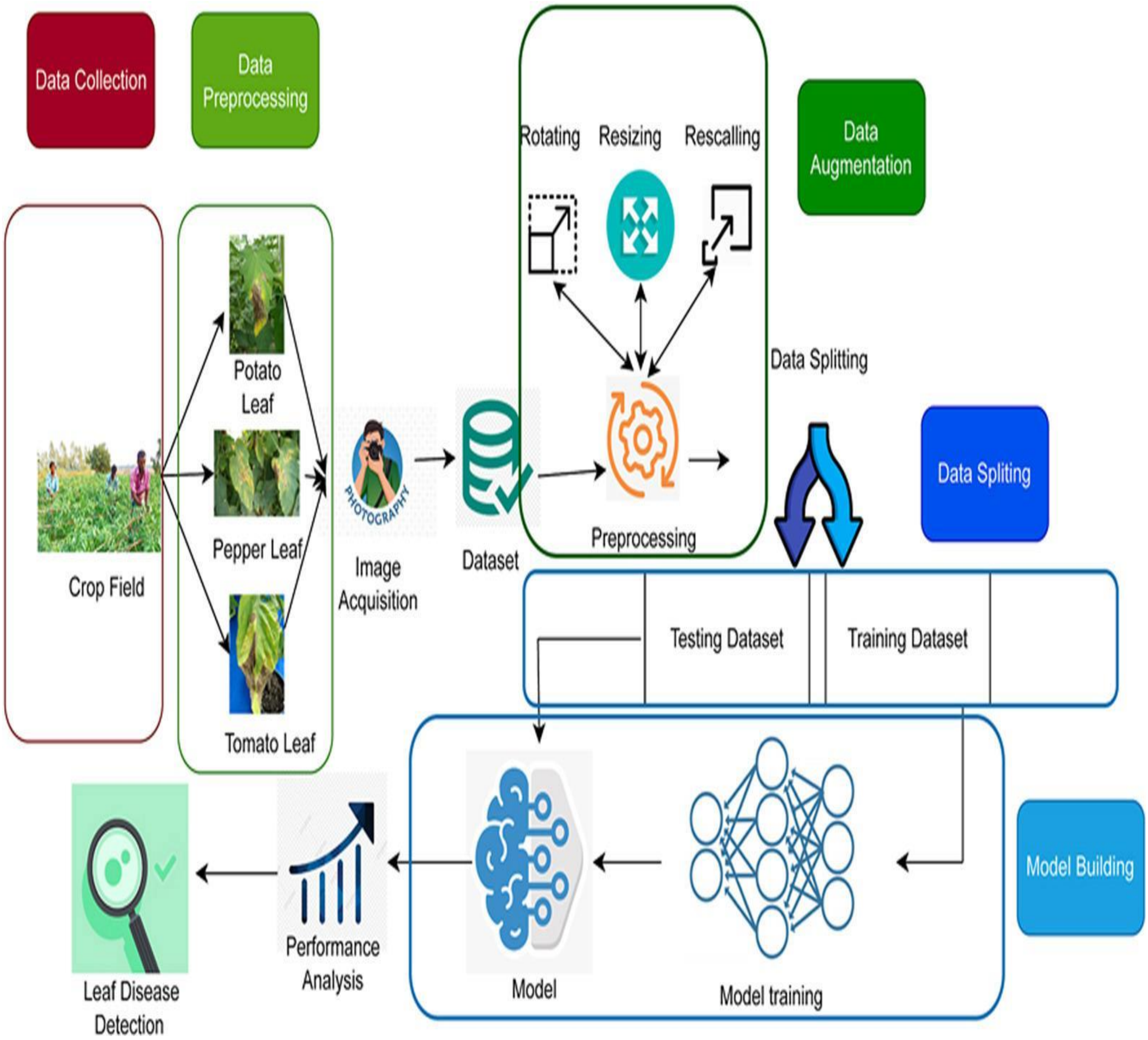


Figure 1: Proposed work-flow diagram.

SYSTEM ARCHITECTURE

1. **Dataset:** Collection of the input images representative of different leaf types including potatoes, tomatoes, and pepper, among others, will define the proposal's onset. To gather these images, one may use a handheld phone cam, or a live camera. The model was also used in testing where it was trained to a publicly available dataset for our deployment purposes.

2. **Preprocessing:** There may be certain types of noises in the raw images obtained via the data set and these need to be corrected before they are trained for the learning module. Preprocessing entails rotating, resizing, and shearing of the image in order to prepare it for processing.

3. **Training and building the model:** The second phase comprises two steps. In the first phase, this model is trained by using a set of the training images.
   In the final stage, the architecture is tested withheld-out performance evaluation images.

4. **Model construction:**
   To build the predictive model, we apply the following steps:
   - Collecting images from the dataset.
   - Ensure that all picture data is processed before use by implementing the steps of resizing and rotating of images.
   - Forming Convolved Features linked into fully connected layers. It is often flattened to 1D array and joined to fully connect layers for instance.
   - Similarly, lastly pick features for the various inputs class.

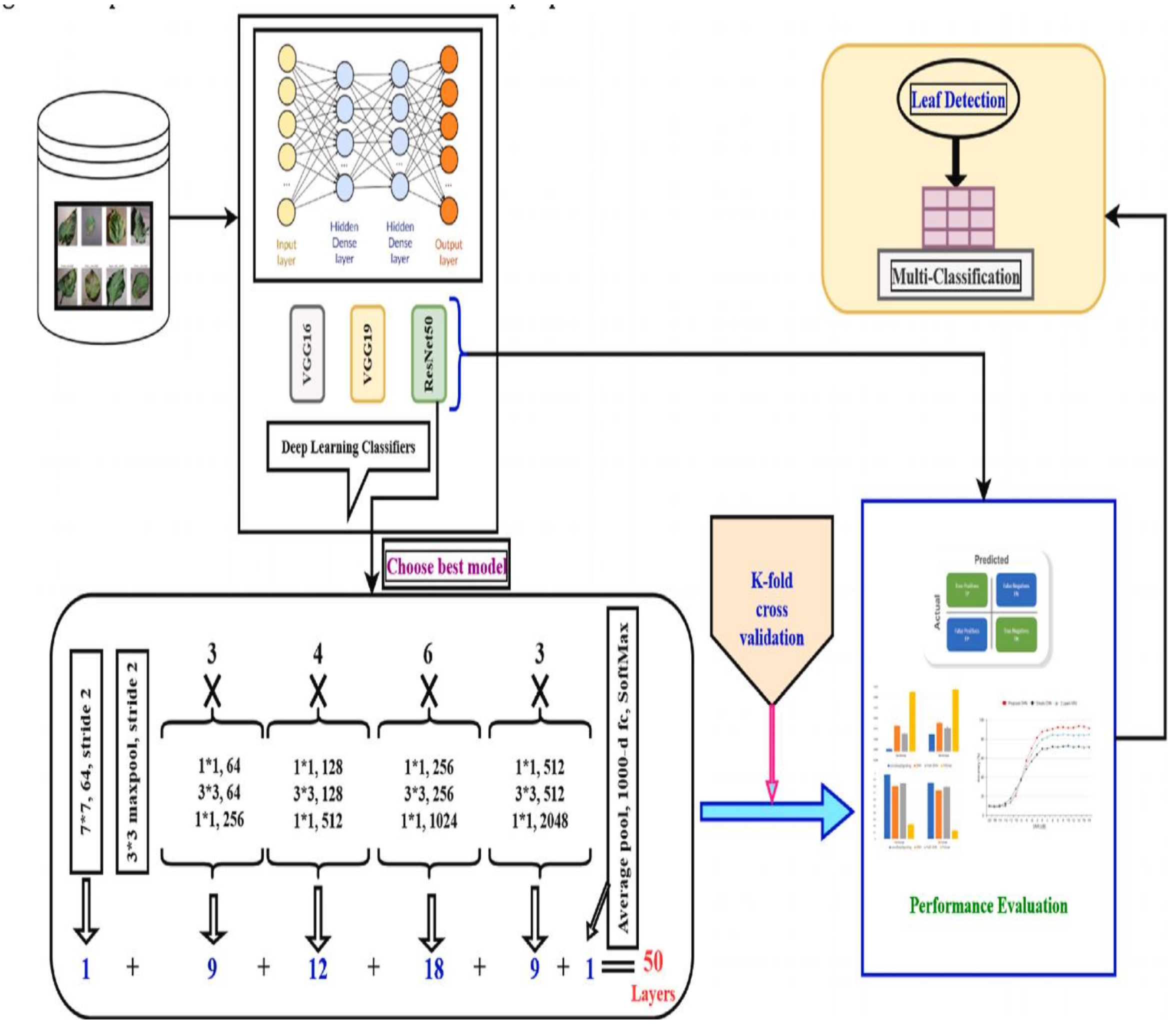


Figure 2: Deep learning model- ResNet50 is finalized for selection; the layer-wise architecture details of ResNet50 are included.

Backend Development

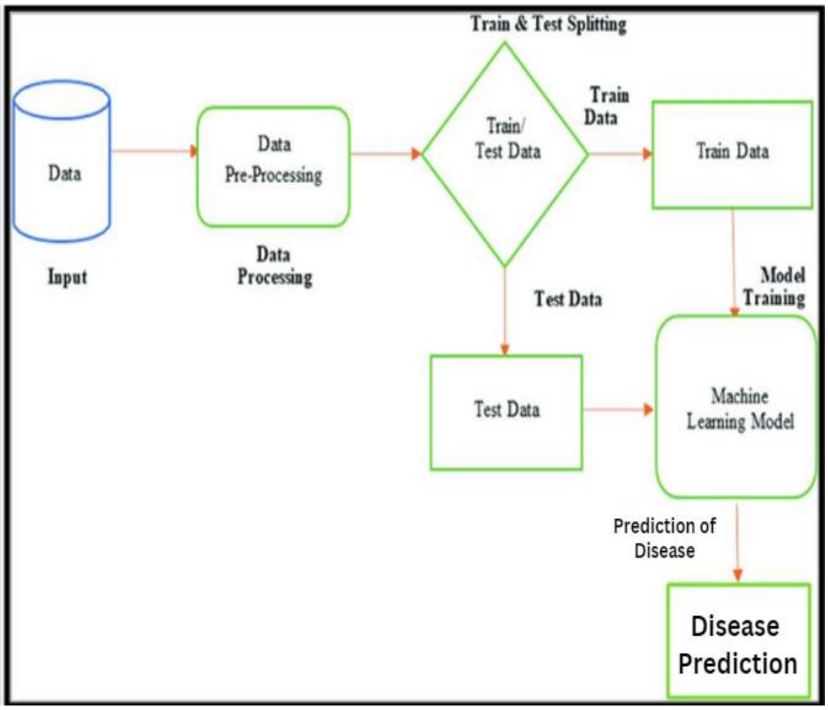
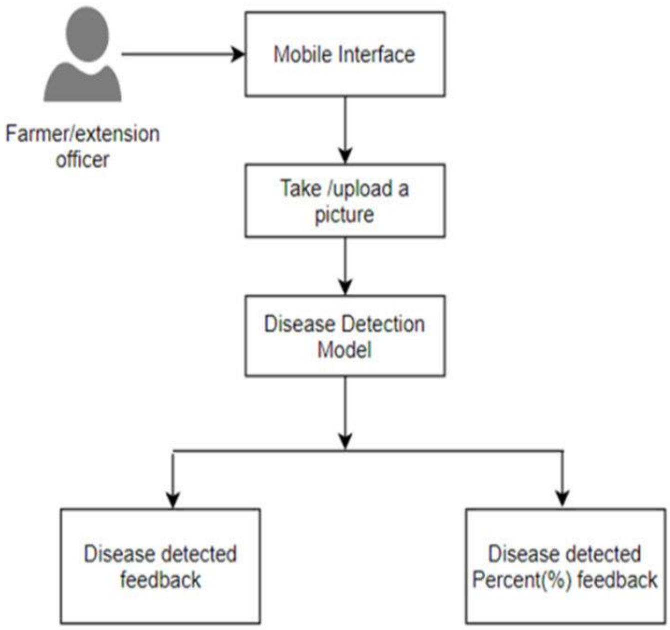


Figure 3: Crop Disease Architecture

User Interface Development



Graph 1: Crop Disease Architecture

3.3) DATA PREPARATION

In the course of the foundational phase of our project, we paid close attention to the comprehensive data management approach. The first requirement is the collection of diversified datasets that have been rigorously curated to include photographs of both healthy and damaged crops. This compilation, which ensures a representative cross-section of numerous crop kinds, illnesses, and environmental conditions, serves as the foundation for our forecasting model. Following the data collecting phase, a strict data cleaning strategy is undertaken to remove duplicates, inconsistencies, and superfluous data. Any missing values or flaws that may jeopardize the accuracy of our model are rigorously addressed in order to refine the dataset.

With a rectified dataset in hand, we turn our attention to data augmentation, a revolutionary step that improves the model's robustness. This is creating differences in existing images using techniques including rotations, flips, and colour modifications. Furthermore, the dataset is balanced to guarantee that different classes are represented fairly, striking a harmonious balance between pictures of healthy and ill crops.

The next important step is feature extraction, which involves extracting relevant information from images using sophisticated techniques such as edge detection, colour analysis, and texture analysis. The photos are converted into numerical or feature-based representations during this process, setting the framework for smooth incorporation into our machine-learning algorithms.

As the dataset approaches the end of its preparation, it is meticulously separated into training, validation, and testing sets. This separation is critical for good model training and evaluation since it ensures that prediction capabilities generalize far beyond the training data. In addition, all photos are shrunk to a uniform dimension in preparation for model compatibility, fostering consistency across the dataset. This rigorous data preprocessing methodology establishes the groundwork for the project's succeeding phases, presenting our dataset as a refined and standardized foundation upon which our revolutionary crop disease prediction model will be created.

Data preparation for a crop disease prediction system will consist of:

1. **Data Collection:**

    - Gather diverse datasets containing images of healthy and diseased crops.
    - Ensure the dataset covers various types of crops, diseases, and environmental conditions.

2. **Data Cleaning:**

   - Remove duplicates, inconsistencies, or irrelevant data from the dataset.
   - Handle missing values or errors that might affect the accuracy of the model.

3. **Data Augmentation:**

   - Augment the dataset by introducing variations in existing images (e.g., rotations, flips, colour adjustments) to diversify the training set.
   - Balancing the dataset by ensuring equal representation of different classes (healthy vs. diseased crops).

4. **Feature Extraction:**

   - Extract meaningful features from the images using techniques like edge detection, colour analysis, or texture analysis.
   - Convert images into numerical or feature-based representations that can be fed into machine learning models.

5. **Dataset Splitting:**

   - Divide the dataset into training, validation, and testing sets to train and evaluate the model effectively.

6. **Preprocessing for Model Compatibility:**

   - Resize images to a uniform size to ensure consistency across the dataset.

Dataset for Crop Disease Prediction have the following data fields:

➢ **Data Source:** Self Prepared
➢ **Data Description:** The data has 70,295 image data
➢ **Image Format:** .JPG

Table 2: Different types of crops and the data instances.

S.No.	Plant Disease	Sample Instances
1.	Apple Apple Scab	2015
2.	Apple Black Rot	1986

3.	Apple Cedar Apple Rust	1762
4.	Apple Healthy	2005
5.	Blueberry Healthy	1814
6.	Cherry Healthy	1726
7.	Cherry Powdery Mildew	1683
8.	Corn Cercospora Leaf Spot	1642
9.	Corn Common Rust	1907
10.	Corn Healthy	1859
11.	Corn Northern Leaf Blight	1908
12.	Grape Black Rot	1888
13.	Grape Esca	1920
14.	Grape Healthy	1692
15.	Grape Leaf Blight	1722
16.	Orange Haunglongbing	2010
17.	Peach Bacterial Spot	1838
18.	Peach Healthy	1728
19.	Pepper Bacterial Spot	1913
20.	Pepper Healthy	1988
21.	Potato Early Blight	1939
22.	Potato Healthy	1824
23.	Potato Late Blight	1939
24.	Raspberry Healthy	1781
25.	Soybean Healthy	2022
26.	Squash Powdery Mildew	1736
27.	Strawberry Healthy	1824
28.	Strawberry Leaf Scorch	1774
29.	Tomato Bacterial Spot	1702
30.	Tomato Early Blight	1920

31.	Tomato Healthy	1926
32.	Tomato Late Blight	1851
33.	Tomato Leaf Mold	1882
34.	Tomato Septoria Leaf Spot	1745
35.	Tomato Spider mites	1741
36.	Tomato Target Spot	1827
37.	Tomato mosaic	1790
38.	Tomato Yellow Leaf Curl	1961

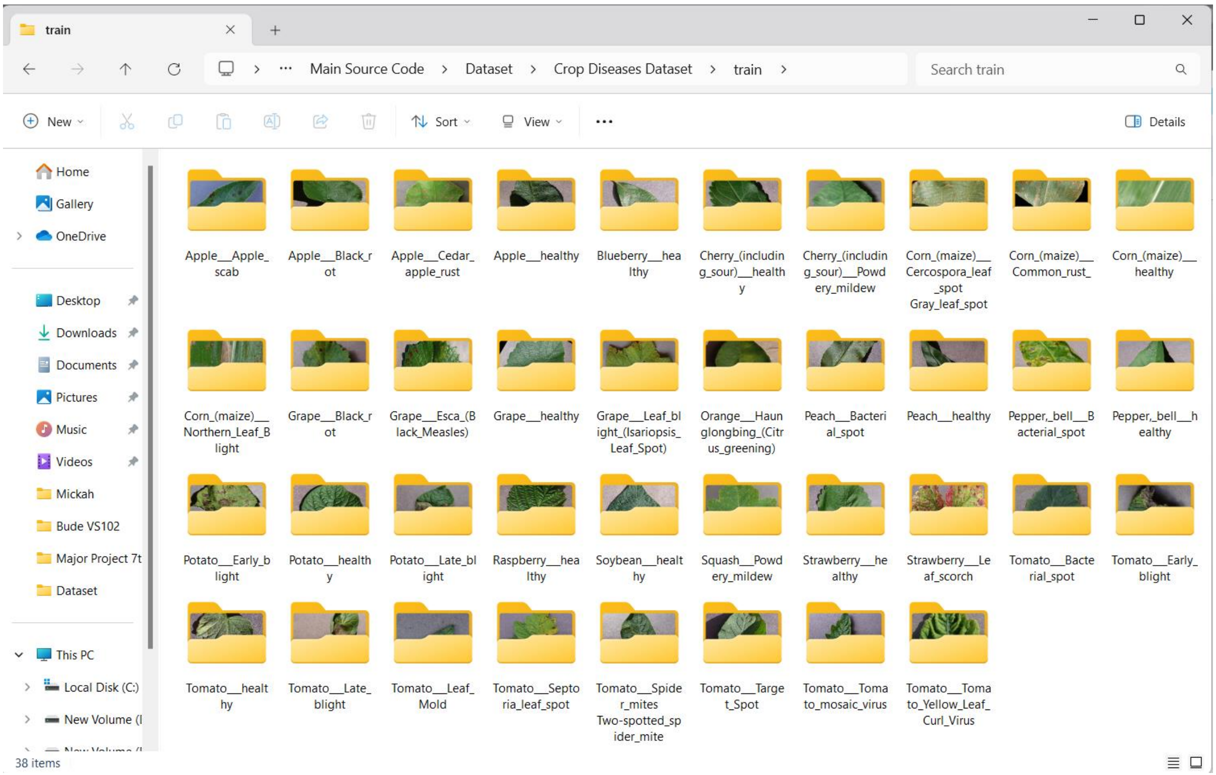


Figure 4: Representation of collected dataset.

Figure 4 displays the intricate manner in which our data is organized into files, providing a thorough understanding of the framework that underpins our dataset. Table 2 supplements this visual representation by providing extensive information on the specific features, formats, and rules of organization controlling data storage. These visual and tabular pieces work together to provide a clear overview of the rigorous file management approach used to handle our dataset.

The test dataset utilized in this study serves as an essential litmus test for the efficacy of our crop disease prediction algorithm, with 33 distinguished images chosen as individual data instances. These images illustrate the range of issues encountered in real-world agricultural scenarios, including a variety of crop varieties and diseases. Each image in this collection represents a distinct snapshot of the complex relationship between environmental conditions and crop health, reverberating the complexity faced by farmers in the field.

Table 3: Description of train dataset.

 (i)	 (ii)	 (iii)	 (iv)	 (v)	 (vi)
 (vii)	 (viii)	 (ix)	 (x)	 (xi)	 (xii)
 (xiii)	 (xiv)	 (xv)	 (xvi)	 (xvii)	 (xviii)
 (xix)	 (xx)	 (xxi)	 (xxii)	 (xxiii)	 (xxiv)
 (xxv)	 (xxvi)	 (xxvii)	 (xxviii)	 (xxix)	 (xxx)

 (xxxi)	 (xxxii)	 (xxxiii)			

3.4) IMPLEMENTATION

It was ensured that major steps were taken to ensure a highly-reliable and user-friendly disease prediction model; starting from collecting the datasets and pre-processing the crop images. The backend was constructed via Python with predefined functions to make disease detection and suggestions. The frontend was organized through the use of HTML and CSS, and JavaScript was similarly used which offered users extra features to allow them to upload images and receive predictions. Integration and testing to ensure effective communication between frontend and backend. The successful program helps farmers in early detection of diseases, promotes good crop management and increases agricultural productivity.

Code Snippets

Code for frontend of the website:

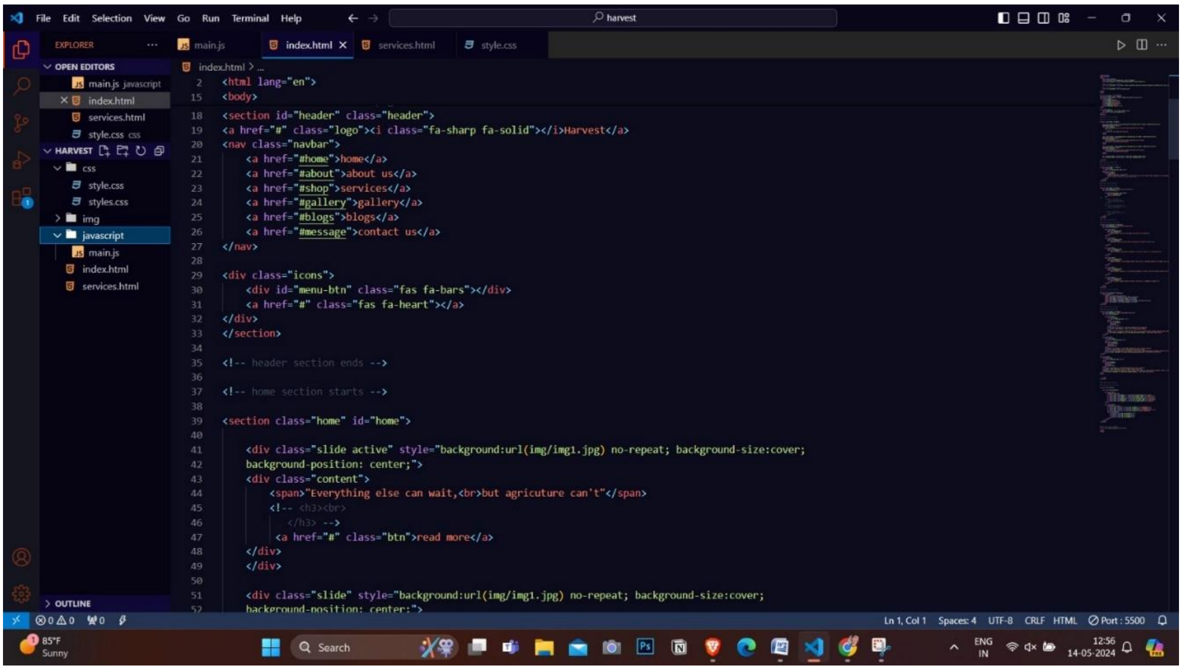


Figure 5: This snippet represents HTML code

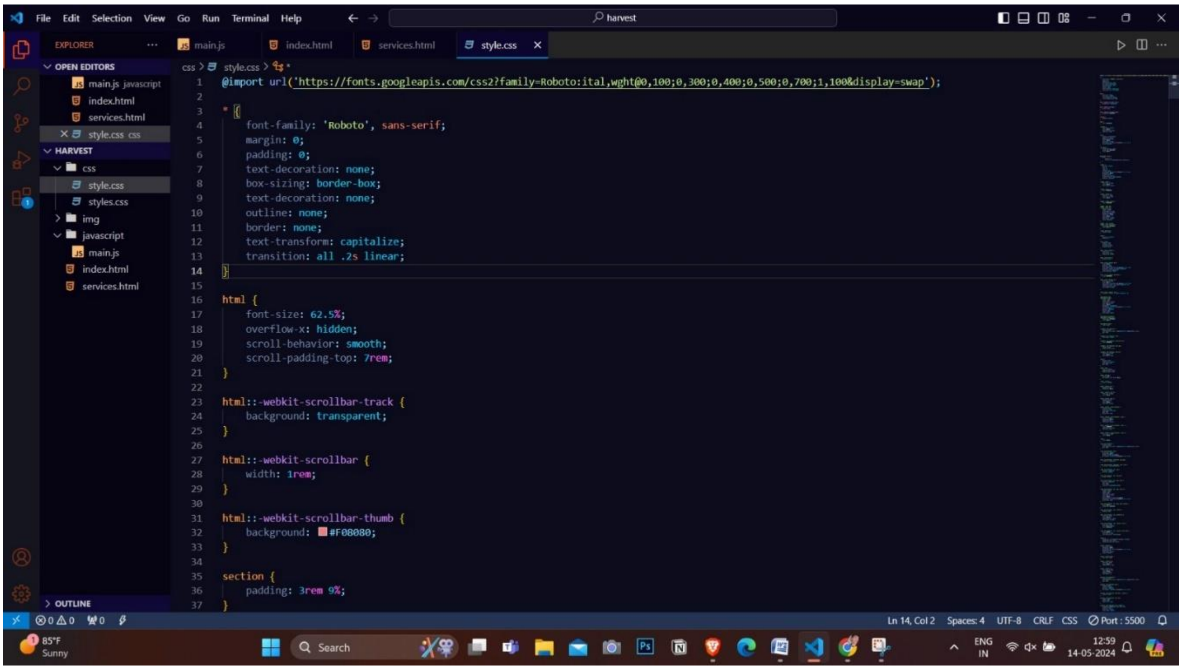


Figure 6: This snippet represents CSS code

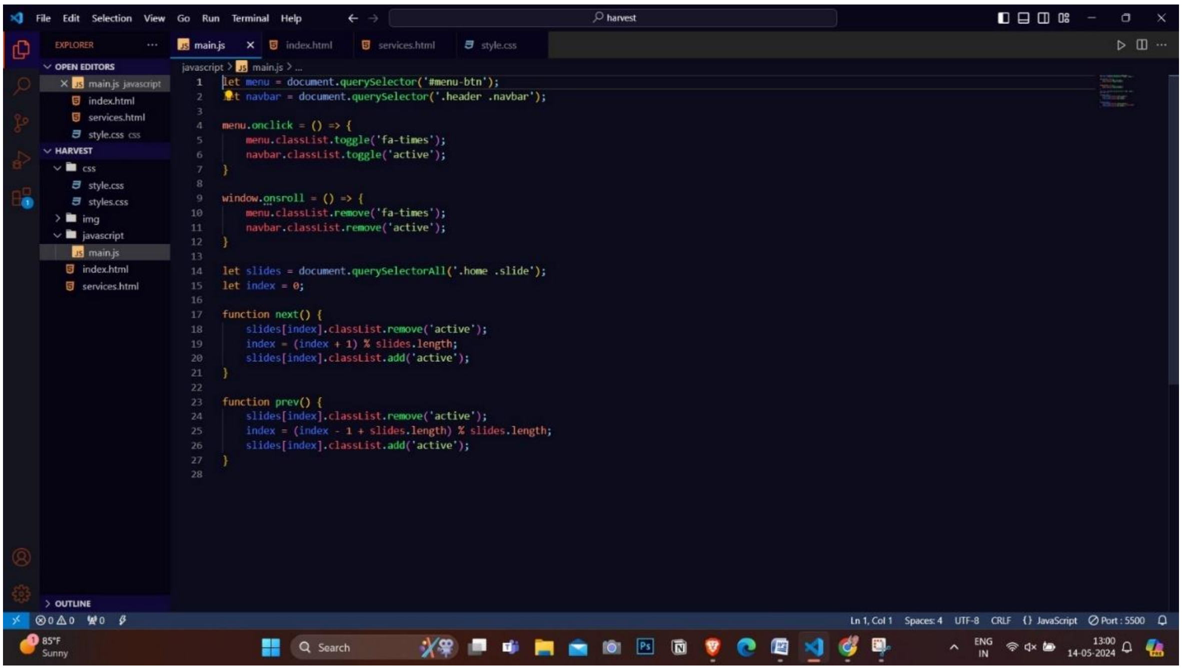


Figure 7: This snippet represents JavaScript code

Code of the backend of the website:

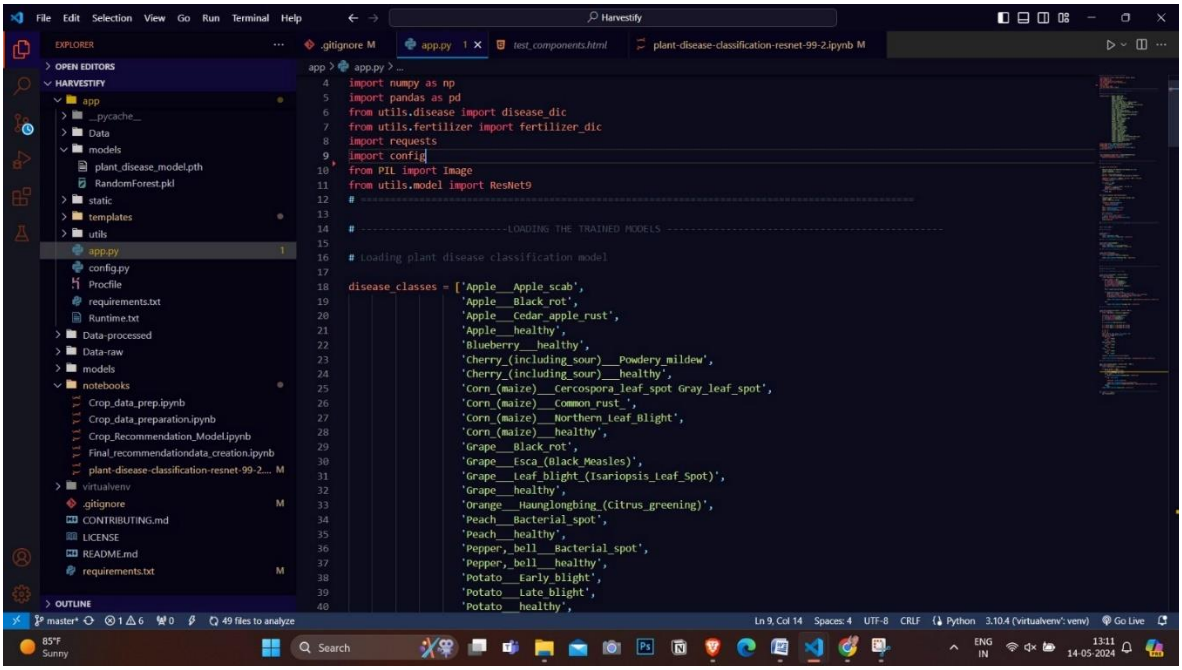


Figure 8: Snippet representing imported libraries

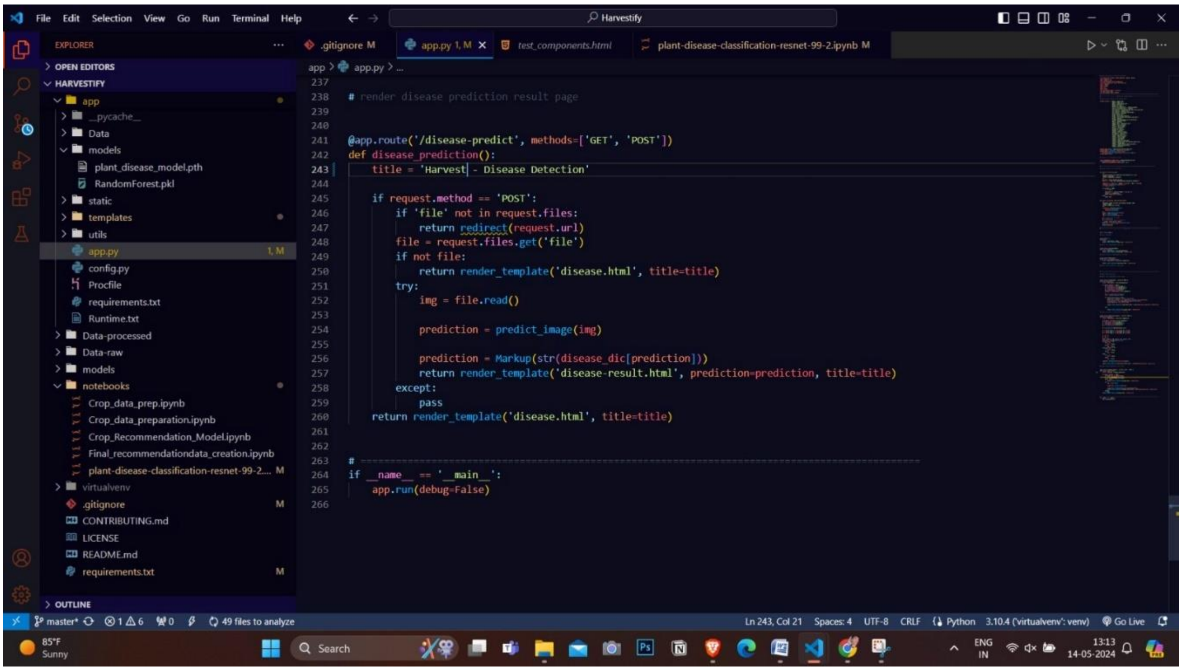


Figure 9: Snippet representing code for disease prediction

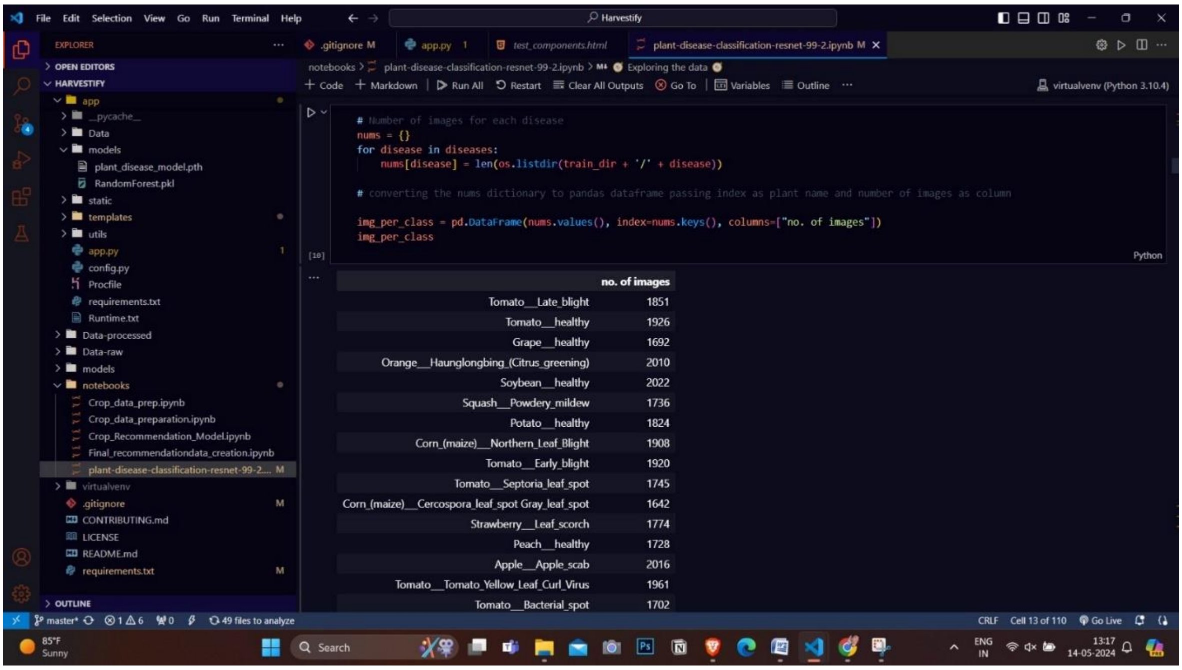


Figure 10: Snippet for dataset

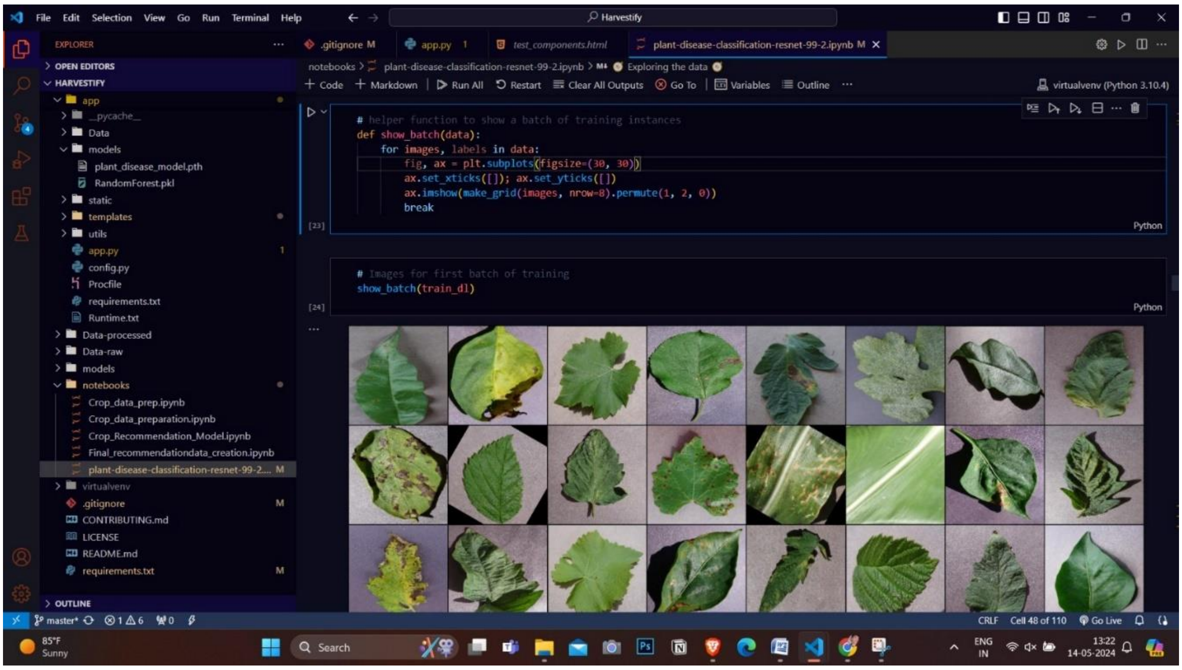


Figure 11: Snippet showing batch of the training data

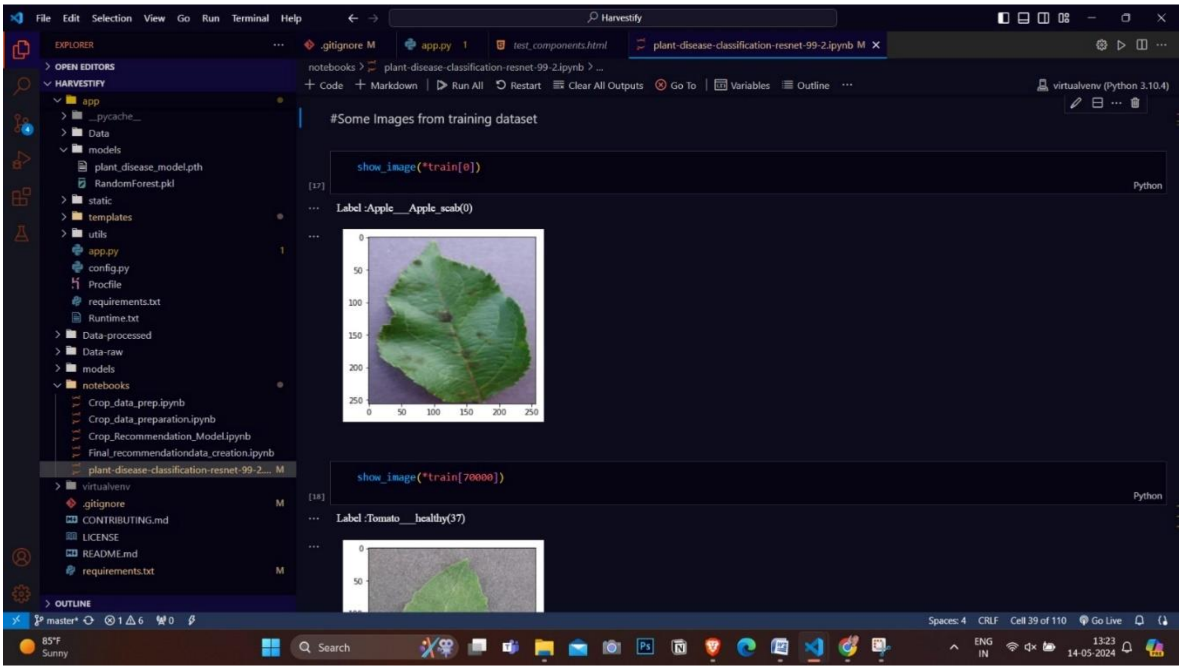


Figure 12: Snippet showing images from training dataset

Algorithms of machine learning that can be used

1. Naive Bayes

Naive bayes is a classifier based on probability that uses machine learning techniques to process data. It relies on bayes's theorem by stating that the presence or absence of one feature should not influence the occurrence of other attributes in the same class.

In naive Bayes, one considers the probability that a datum belongs to any given class having observed its feature values. Next, the data point is assigned to the class with maximal probability.

"Naively" is an assumption of no interdependence among the features that may, in reality, happen. However, even with that simplistic assumption, it has been observed that Naive Bayes is effective for many classification problems such as text classification and spam filtering.

2. Random Forest

One of the machine learning systems which are able to predict the yields in the agriculture sector is random forest. An ensemble learning approach employing multiple decision trees to enhance prediction accuracy and stability.

In crop prediction, random forest can be used to predict crop production, growth, and quality based on weather, soil, and historical data. This method constructs many decision trees using different sub-sets of the features, combines all these trees' results and gets one forecast.

Random Forest as a predictor for crops has one major advantage – it works well with large and complex data sets with many characteristics. Thus, it could pick out important aspects for form predictions, that will be very useful in the farmers' and academics' practice.

Despite this, it should be highlighted that Random Forest is not the ultimate answer that can predict true results all the time. Therefore, one must select and preprocess the input data well, adjust the model parameters correctly and evaluate it in an adequate way.

3. Decision Tree

A machine-based decision tree model can be used to accurately predict agricultural output. It is a straightforward but effective formulation that constructs a tree model representing the sequence of choices and results dependent upon supplied information.

Thus, it is possible to use decision tree in crop prediction in order to anticipate the outcome (amount produced), growth of a crop, or its quality in respect to weather, types of soil, and history. This involves dividing the original input area into smaller subsidiaries based on different choices/outcomes. The target variable is characterized by either acquiring information in high volume or minimizing impurity of that target variable thus each division is selected depending on individual strengths of the target variable.

Using a decision tree for crop prediction will help one understand which factors are worth looking at when conducting projections. Farmers and researchers can observe the resulting tree which readily gives out its inherent patterns and relationships within the results obtained.

Yet it should be emphasized that a Decision Tree suffers from an overfit which occurs if it becomes too complex to remember but not to infer new data. It is essential to select and prepare the source data correctly, trim the tree to reduce its level of complexity, and evaluate the model with suitable assessment rules.

4. Supervised Vector Machine (SVM)

The Support Vector Machine (SVM) is a strong machine-learning method that may be used to estimate agricultural productivity. It is a supervised learning approach that creates a hyperplane or group of hyperplanes in a high-dimensional space for classification or regression problems.

SVM may be used in agricultural yield prediction to predict crop production based on parameters such as weather, soil conditions, and historical data. The method finds the hyperplane with the greatest gap between the different classes or goal values while minimizing the amount of misclassified data points.

One of the primary benefits of utilizing SVM for agricultural production prediction is that it can handle high-dimensional, nonlinear data and can be utilized for both regression and classification tasks. It may also be trained on a limited number of samples, making it appropriate for small-scale farming operations.

SVM, on the other hand, may be computationally costly and may need careful selection and adjustment of the kernel function and regularisation parameter. To guarantee accuracy and generalizability, it is also necessary to thoroughly preprocess the input data and evaluate the model using appropriate evaluation criteria.

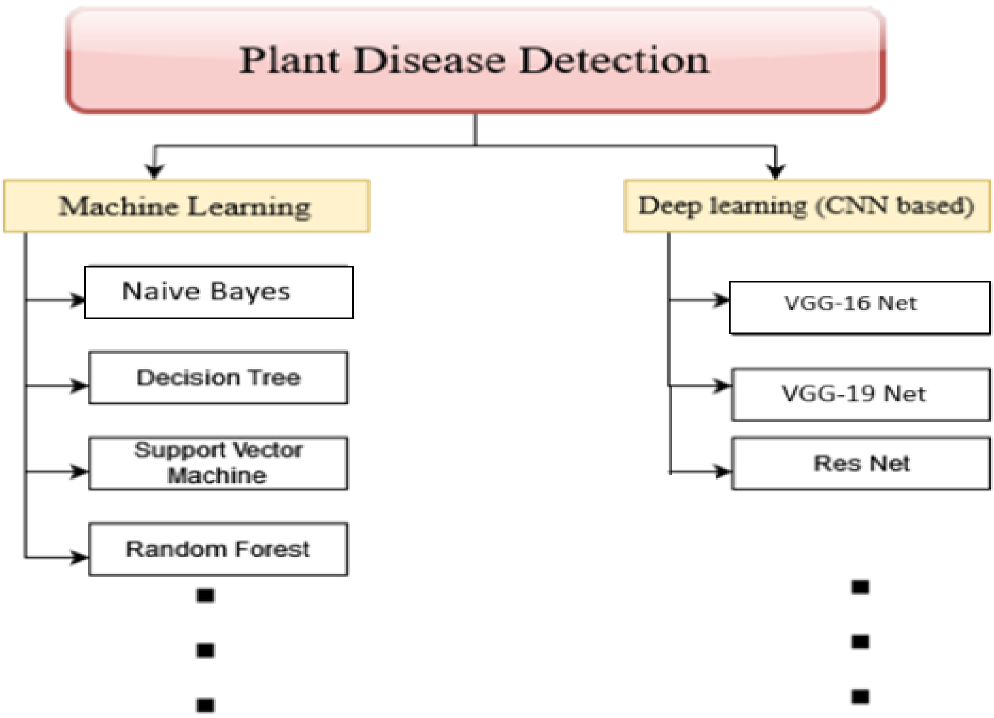


Figure 13: Approach for crop disease detection.

Tools and Libraries used

1. Python:

We opted for Python Language since it has a range of built libraries including pandas, NumPy, SciPy among others, with a plenty of incorporated features that facilitate both data science and machine learning methods implementation purposes for effective designing of our model. In doing so for this project we utilized assortment of Python packages to undertake several actions.
The language offers different prospects for writing programs in Python. It has broad capacities that can handle certain software products such as meta-programming and meta-objects. Power typing, reference calculation and trash collection for wastes are also used by Python. The word processing or late binding also happens in the process, which defines how words change.

Why should you use Python?

Python is being used because it runs on a variety of systems. Python is a stage-free language. Python is as easy to learn as English. Python has various libraries and a linguistic framework that is close to English, whereas Java and C++ have complex codes. Python programs have fewer lines than programs created in other languages. That is why we use Python for artificial intelligence, consciousness, and dealing with massive amounts of data. Python is a programming language that is article-oriented. Python has notions such as classes, objects, polymorphism, exemplification, legacy, and reflection.

2. NumPy Library:

NumPy is a Python program library that provides support for huge, multi-dimensional collections and matrices, as well as a vast library of mathematical functions designed to interact with these components.

The usage of NumPy in Python is similar to that of MATLAB in that they both translate and allow customers to construct projects more quickly as long as many jobs are focused on clusters or networks rather than scales. There are other alternatives, in addition to these important ones:

- NumPy is a Python scientific computing toolkit that supports massive, multi-dimensional arrays and matrices, as well as a variety of mathematical functions that work on these arrays. It is a critical component of the scientific Python environment, and it is widely used in domains such as physics, engineering, finance, and data science.

Overall, NumPy is a robust Python tool for scientific computing that is an essential part of many scientific and data analysis operations.

3. Pandas' Library:

It is a Python software package for decrypting and analyzing data. It manages number tables and time series using data structures and functions. The word is derived from "panel data," an econometrics term for data sets that have insight into a large number of identical persons.

4. Matplotlib Library:

Matplotlib Python library was created for creating graphs, charts, and high-quality statistics. It is fantastic that the library can update relatively little knowledge about mathematics. Matplotlib's primary principles and actions include:

Picture

Every picture is referred to as an image, and each image is an axis. Drawing may be thought of as a method of drawing several episodes.

Structure

The first item that should be drawn on a graph is data. It is possible to declare a keyword dictionary with keys and values such as x and y values. Following that, scatter (), bar (), and pie () can be used to create a structure as well as a variety of other functions.

Axis

The number and axis acquired from the sub-sections () can be adjusted. To change the x- and y-axis characteristics, use the set () function.

5. Scikit Learn Library:

A machine learning library called Skitlearn written in Python is the most successful. A lot of useful machine learning tool and mathematical modeling techniques such as division, deceleration, integration and size reduction are available in the Scikit learn library.eltemperaturennya: Sklearn employs machine learning models. Scikit-Learn is too pricey for data reading, playing, game theory, or storytelling and should be avoided. Others come here to help you interpret the spread.

3.5) KEY CHALLENGES

Here are some key challenges typically faced during the development of a crop disease prediction system and ways to address them:

1. **Limited and Diverse Data**:
   - **Challenge:** Obtaining a wide-ranging dataset encompassing multiple crop varieties, diseases, and environmental scenarios proves challenging due to limited accessibility and diversity of available data sources.

   - **Addressing:**
     - Dataset Augmentation: Augmenting the existing dataset by employing techniques like oversampling, or data augmentation can enhance diversity.

     - Collaboration with Research Institutions: Partnering with agricultural research organizations and institutions to access their extensive and diverse datasets broadens the scope and diversity of available information.

   By implementing these strategies, the project aims to enhance the dataset's variety, incorporating multiple crop types, diseases, and environmental factors. This approach facilitates a more comprehensive and representative dataset for effective machine learning model training and accurate disease prediction.

2. **Complexity in Image Processing**:

   - **Challenge:** The intricacy lies in extracting pertinent features from images and transforming them into formats suitable for model ingestion. This complexity arises due to the diverse nature of plant diseases, leaf variations, and the need to capture subtle but crucial visual cues for accurate disease identification.

   - **Addressing:**
     - Advanced CNN-based Feature Extraction: Leveraging Convolutional Neural Networks (CNNs) for feature extraction streamlines the handling of complex image data. CNNs excel in automatically learning hierarchical representations of images, and recognizing intricate patterns, textures, and shapes within the visual data.

     - Transfer Learning: Employing pre-trained CNN models or transfer learning techniques can assist in utilizing knowledge gained from previously trained models on large datasets. This method can be adapted for specific tasks, reducing the need for training from scratch and handling complexities more efficiently.

  - ○ Image Preprocessing Techniques: Implementing various preprocessing steps like normalization, noise reduction, and edge enhancement aids in simplifying image data and preparing it for effective feature extraction by the CNN architecture.

By utilizing these techniques, the project aims to tackle the challenges associated with the complexity of image processing, ensuring accurate and meaningful feature extraction from the visual data for disease identification and classification.

3. **Data Privacy and Security**:

- **Challenge:** Safeguarding sensitive agricultural data collected from farms and institutions is a pivotal concern. Protecting this data against unauthorized access, breaches, or misuse is crucial for maintaining the integrity and privacy of agricultural information.

- **Addressing:**
  - ○ OTP-Based Login Systems: Implementing One-Time Password (OTP) authentication systems adds a robust layer of security. OTPs are dynamic and time-sensitive codes sent to the user's registered mobile number or email upon login attempts. These codes are valid for a single login session and significantly reduce the risk of unauthorized access, as they require possession of the user's authorized device to complete the login process.

  - ○ Regular Security Audits: Periodic security audits/assessment gives insights on any gaps or weaknesses, as well ensuring the company follows up-to-date regulations and procedures. Continuous evaluations enable to take actions on time in case of discovering a security breach, a loophole, or any such threat.

By integrating OTP-based login systems along with other security measures like regular audits, the project aims to ensure the confidentiality, integrity, and security of agricultural data, mitigating potential risks and unauthorized access.

4. **Continual Model Improvement**:

- **Challenge:** Maintaining the model’s viability over time becomes difficult, especially within fluid agro-ecological environments that frequently introduce diverse diseases every now and then. It is important to keep current with changing situations so that the model is still accurate.

- **Addressing the Challenge:**

    - Active Learning Strategies: Utilize efficient ways of selecting and annotating significant new data through employing active learning strategies. The model makes active queries on the most informative data points for labeling. This is an approach that helps maximize use of resources towards bettering their models by focusing on their data instances with the highest levels of informative value to the knowledge of the models.

    - Monitoring and Evaluation: Regularly check on how well the model performs based on the metric values throughout a period. Evaluation of this involves measuring its accuracy, precision, recall and F1 score among other metrics. It is advisable to undertake regular evaluation since this assist in the detection of performance declines and areas that require adjustments, leading to immediate retraining or tuning up.

    - Collaboration with Domain Experts: Consider hiring domain experts, agronomists, or agriculture scientists who can shed light on the emergence of new disease and change in the environment, respectively. These experts could be used to pinpoint essential data points and attributes to improve a model.

By implementing active learning strategies, and engaging with domain experts, the model can continually evolve and adapt to new disease variations or environmental factors. This iterative approach ensures the model's sustainability and effectiveness in detecting crop diseases as agricultural landscapes change and new challenges emerge.

CHAPTER-4 TESTING

4.1) TESTING STRATEGY

As such, in this phase, we get the best model based on the numerous experiments done. Our approach involved employing accuracy, precision, recall, f1 score, training accuracy, training loss, validation accuracy, and validation loss. This will allow us to create a smart web-based application with intelligent deep learning.

DEEP LEARNING MODELS

We use the following deep TL methods for the prediction for the disease:

1. VGG-16

VGGnet-16 consists of sixteen convolutional layers with an identical arrangement among them, which is one of the most widely used models for diagnosis through image categorization. Firstly, VGG-16 shows that in certain situations increasing the number of neurons could boost system's efficiency. The VGG-16 transfer learning involves three components that are convolution, fully connected layer and pooling layer. The second layer is the convolution layer (Conv) that uses filters on images to reveal facts, these two essential parameters being kernel and stride size. The poling layer reduces the network's spatial area of coverage and ensuing computations.

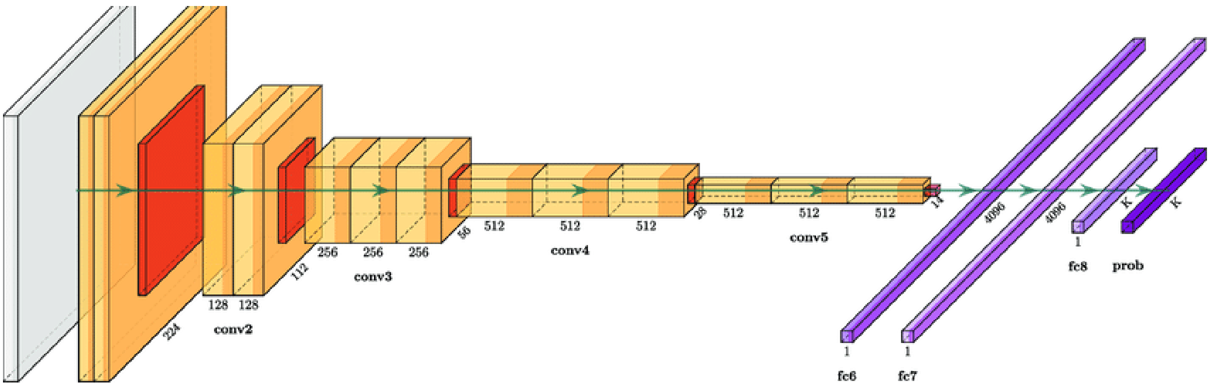


Figure 14: Representation of VGG-16 Architecture.

2. VGG-19

The CNN VGG-19 includes 3 dense layers, 16 convolutional layers and a total of 19 layers for classifying images with one thousand different types. It comprises three FC layers, two Conv 1 max pools, four Conv 1 max pools, and four Conv 1 max pools. Each convolutional neural network has numerous 3 × 3 filters, and hence, it is among the most used current photo prediction models.

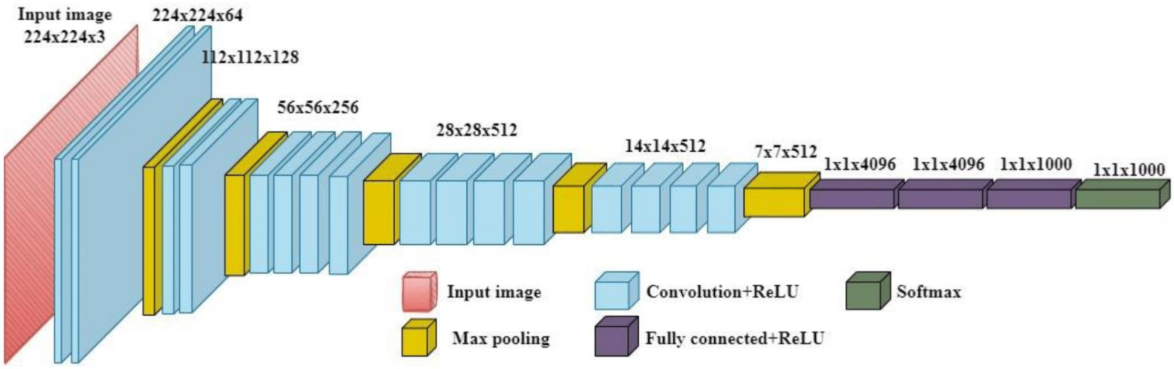


Figure 15: Representation of VGG-19 Architecture.

3. ResNet50

The structure consists of four convolutional layers of 128 units each which are then followed by one average pooling layer and one maxpooling layer before leading to the resnet50 model that forms a part of the ResNet family which was introduced by He et al., in It is preferred for classifying pictures among others. In this ResNET50 architecture, there are four distinct phases. Three layers comprise the first convolution stage: the first one is a 1*1, 64 kernel followed by a 3*3, 64 kernel, finalizing with a 1*1, 256 kernel. During this stage, these three layers have made copies thrice that makes up a sum of nine layers. After that, the kernel of 1*1, 128 appears, followed by the kernel of 3*3, 128 and then the kernel of 1*1, 512. This process was repeated four times for 12 layers. The subsequent kernel is 1 x 1,256, followed by two other kernels of 3 x 3, 256, and 1 x 1, 1024; this process is repeated six times, resulting in eighteen layers altogether. A 1*1,512 kernel followed, and then there were two more kernels of 3*3,512 and 1*1,2048. We were able to make four iterations of this method, which resulted in nine layers.

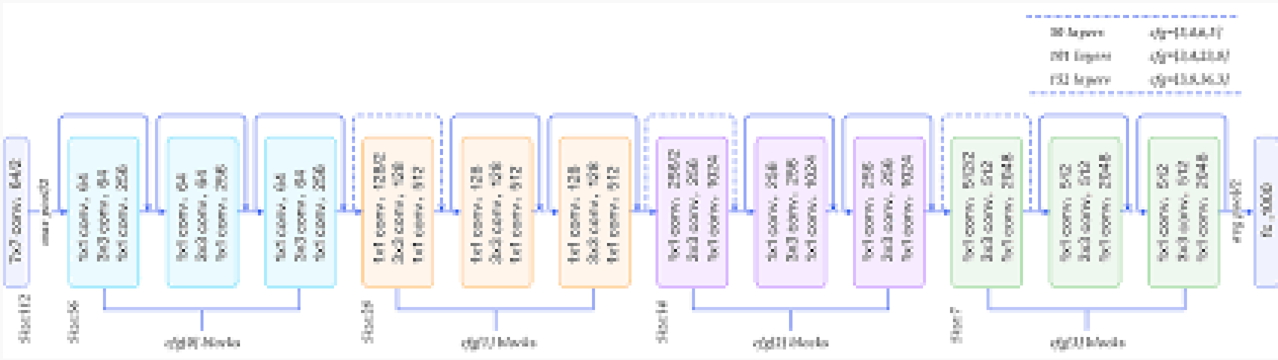


Figure 16: Representation of ResNet Architecture.

4.2) TEST CASES AND OUTCOMES

ARCHITECTURAL COMPARISON

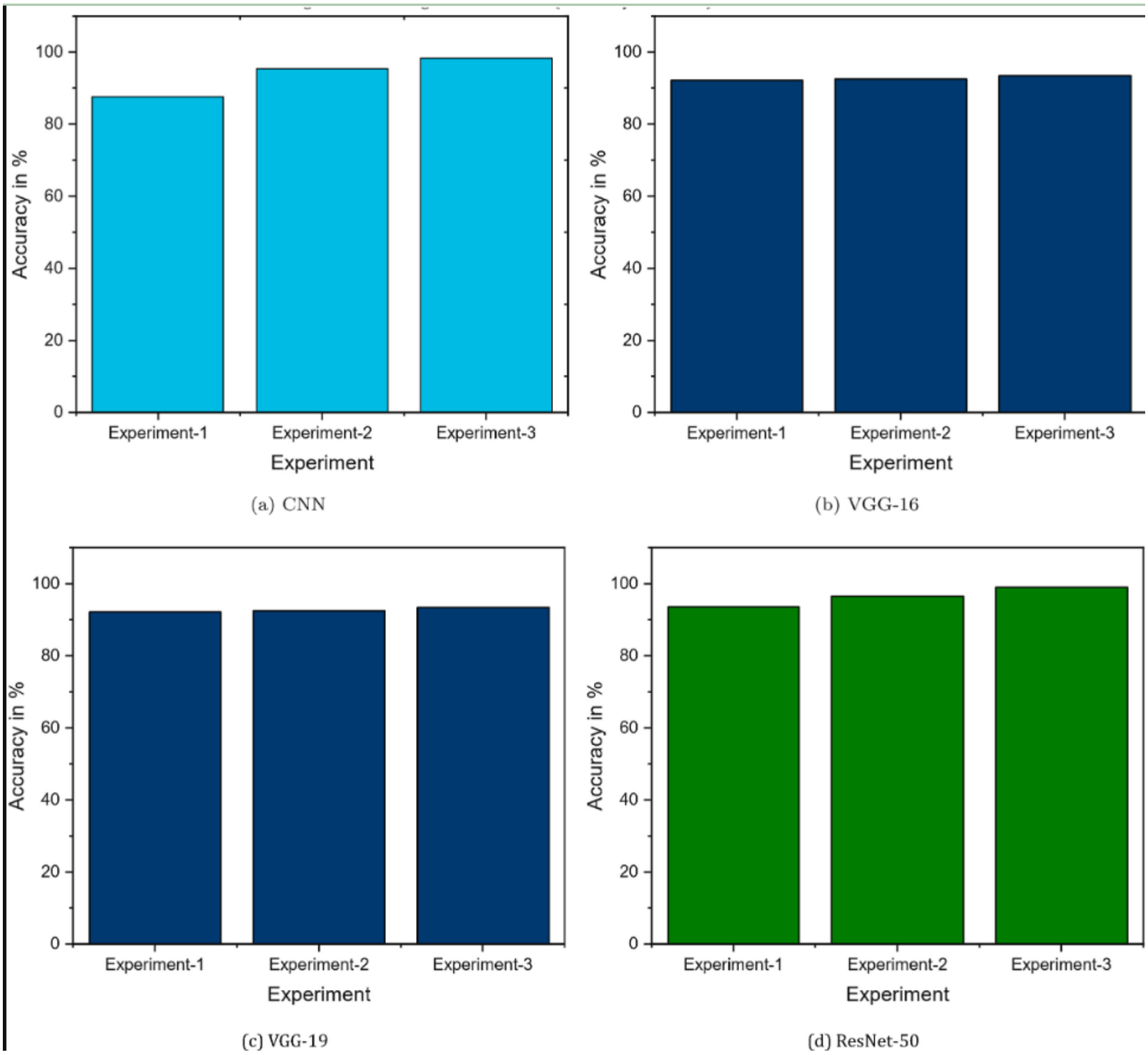
VGG-16, VGG-19, and ResNet50 are some of the most commonly used CNN transfer learning architectures. Each of these transfer Learning algorithms were trained using ImageNet dataset consisting of 1,000 unique image classes. Some of these factors that influence the accurateness and training error comprise the dataset, number of pictures and their dimensions, figure of convolution and swimming pool layers as well as, number of epoch and the number of batches. VGG-16 is a very deep CNN, which shares with VGG-19 the only property: the difference in depth. VGG-16 consist of 16 layers whereas the other one has 19. In most cases, accuracy provided by VGG-19 is somewhat higher because of its deeper architecture. These two models can be applied on image classification but they had “vanishing point” problems.! On the contrary, ResNet50 has higher precision resulting from more layers. In a nutshell, the more layers that are constructed on a pyramid, the more precise such representation of data will be and the higher its classification authority. On its part, it offers more accurate results than the other two models with ResNet50 being easier to optimize and to address the vanishing gradient issue.

Table 4: Architectural comparison of VGG-16, VGG-19 and ResNet50.

SL	Properties	VGG-16	VGG-19	ResNet-50
1	image	224*244*3	224*244*3	224*244*3
2	weight	imagenet	imagenet	imagenet
3	Model size	533MB	574MB	102MB
4	Total layers	16	19	50
5	Convolution layer	13	16	48
6	Max pool	5	5	1
7	Activation function	Softmax	softmax	Softmax
8	Total parameters	138.3 million	143.7 million	25.6 million

Advantages/Limitation

VGG-16	• It is painfully slow to train. • Weights are quite large. • Suffers from vanishing gradient.
VGG-19	• Deeper than VGG-16. • Weights are quite large. • Suffers from the vanishing gradient.
ResNet-50	• Faster than VGG-16 and VGG-19. • Low-power models parameterized to meet the resource constraints. • It can handle the vanishing gradient problem.



Graph 2: The accuracy results of different models.

The following figure shows the performance in terms of accuracy for every experiment on the suggested model. The accuracy results of any models are growing with the increasing number of images and we found the accuracy of ResNet50 model.

CHAPTER-5 RESULTS AND EVALUATION

5.1) RESULTS

Different deep learning models are explored and evaluated with respect to their performance based on the following metrics using the confusion matrix shown in

Table 5: Confusion Matrix

	Actual positive	Actual negative
Predicted positive	TP	FP
Predicted negative	FN	TN

This TP indicates a true positive; the model predicts the occurrence of one leaf with this disease and another leaf which really suffered from that. This TN reflects true negative; where it is established by a certain model that one among this instance is affected and the other one does not have this symptom.

- Accuracy: This may be expressed as the fraction of test cases predicted correctly.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{1}$$

- Precision: Precision refers to the ratio of correctly predicted disease-affected leaves to all positively predicted leaves by the model, which can be defined as:

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

- Recall: for this reason, recall is understood as the ratio between real and correct affirmative predictive results to the number of positive points of the study instance;

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

- F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$F1 - score = 2 * \frac{Precision*recall}{Precision+recall} \tag{4}$$

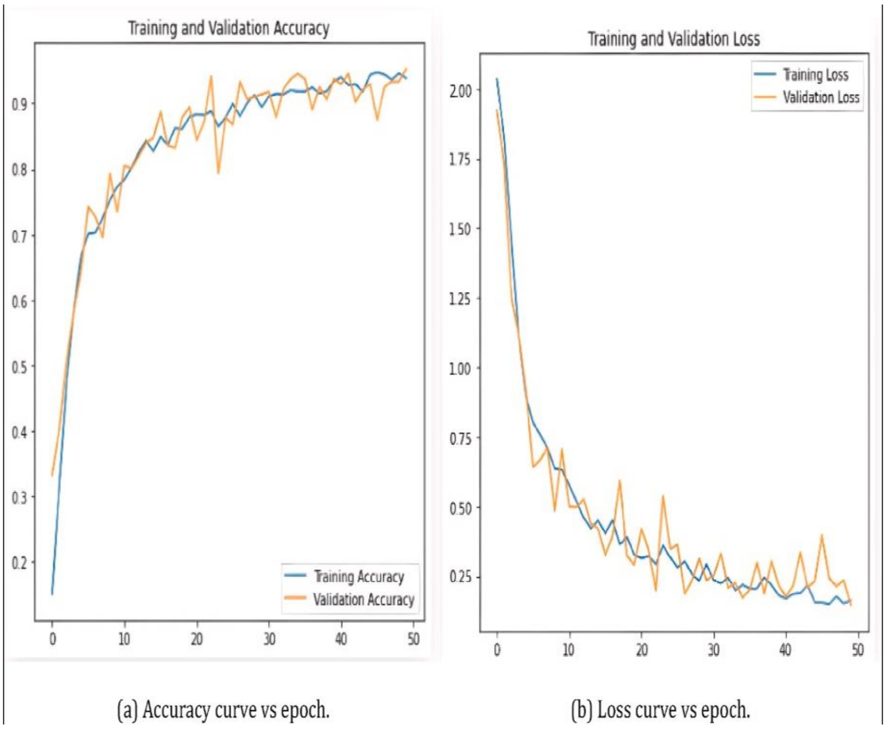
RESULT ANALYSIS

Table 6: Accuracy results.

Model Name	Training Loss	Training Accuracy	Validation loss	validation Accuracy
Sequential	0.0409	98.44%	0.1477	98.21%
VGG16	0.1218	96.03%	0.3484	91.91%
VGG19	0.0871	97.13%	0.1439	93.47%
ResNet-50	0.0711	95.84%	0.2143	96.51%

The table below provides the mean training and validation accuracy and loss for various models.

Any deep learning process has its learning curve that reflects the way of learning by means of the training process on increasing data set incrementally. The higher the number of epochs, the higher the interpreted accuracies for the training datasets. Moreover, the hold-out validation dataset is used to predict the generalizability of the model whose validity accuracy will be measured henceforth.



Graph 3: Training and validation (accuracy and loss) ResNet-50.

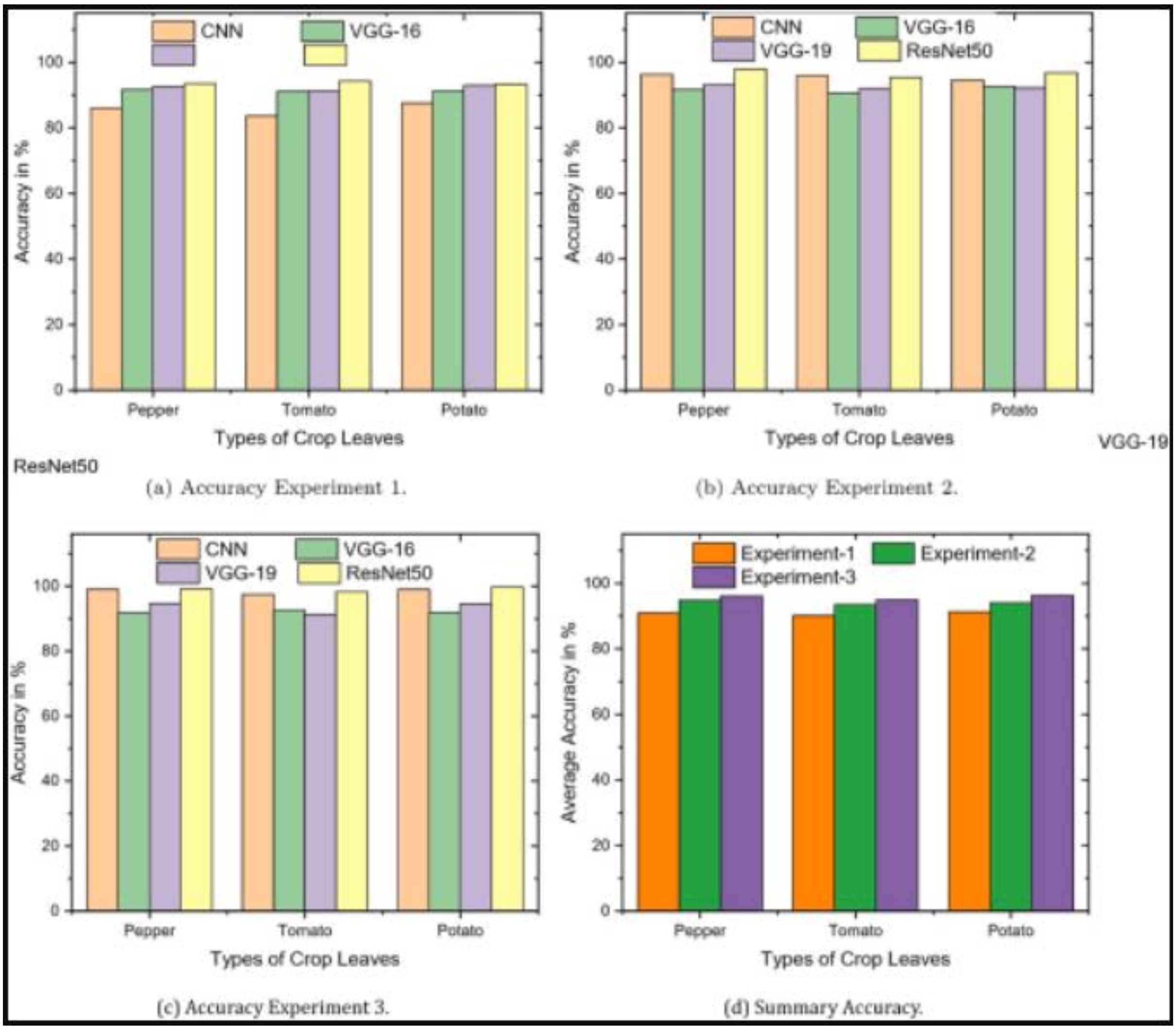
The trend on the loss curve indicates reduction of validation and training losses with minimal interval between each other. However, there were minor variations noticed in the validation tests.

It specifies the training's as well as the validation accuracy and loss. Chart four depicts validation accuracy and training accuracy; thus, this means the accuracy is improving with the training time, although there are some variations in the validation curve. The second issue is that the loss graph shows good fitting due to the close resemblance between the trained and validated loss graphed curves.

Table 7: Testing results summary.

Model	Testing Accuracy	Precision	Recall	f1score
Sequential	98.60%	96.83%	98.70%	97.68%
VGG-16	92.39%	93.55%	94.51%	95.71%
VGG-19	96.15%	96.42%	97.06%	96.60%
ResNet-50	98.99%	98.96%	99.05%	98.98%

The final consideration table below takes into account the results of accuracy, precision, f-measure, and recall. The table also reveals that out of the different models ResNet50 scores more than others for any model.



Graph 4: Accuracy results.

This figure shows the summary of the accuracy performance of each of the model. From the graphs, it is clear that the detection accuracy of ResNet50 is higher than any models for any experiments.

5.2) COMPARISON WITH EXISTING SOLUTIONS

EXISTING MODEL

1. Limited dataset covering a few common crop diseases.
2. Offers general disease identification but lacks precision and comprehensive coverage.
3. Doesn't provide real-time solutions and the users need to consult additional resources for remedies.
4. Static model without continuous learning or adaptation to new diseases or data.

5. Might lack collaboration with domain experts, resulting in potential gaps in dataset relevance.
6. Dependent on cloud-based infrastructure, limiting deployment options.

PROPOSED SYSTEM

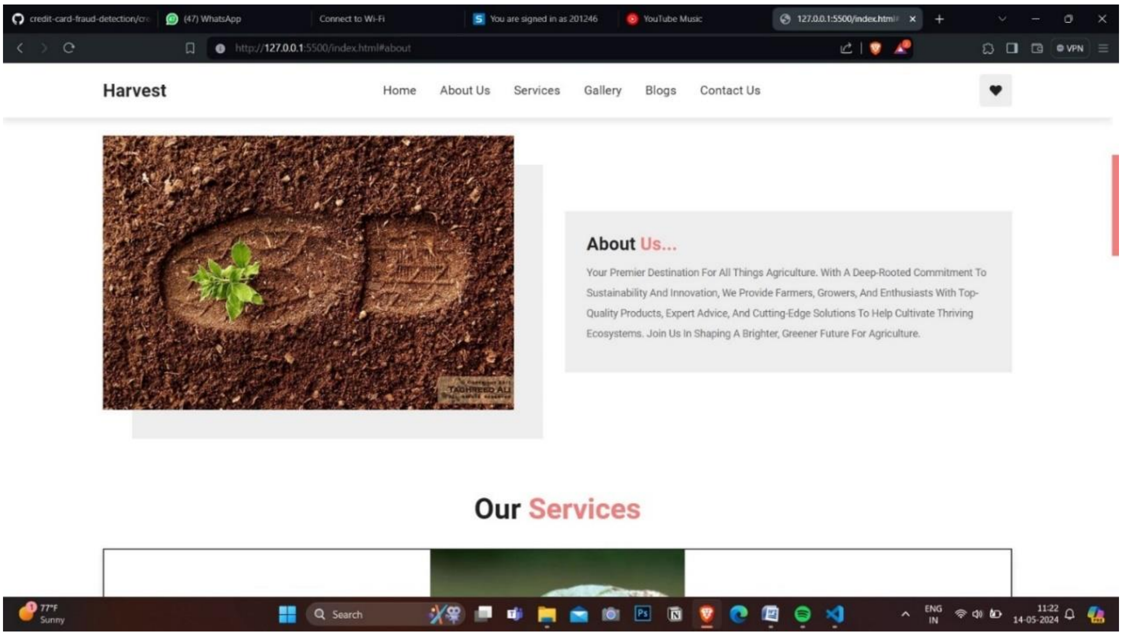


Figure 17: Home page of the website

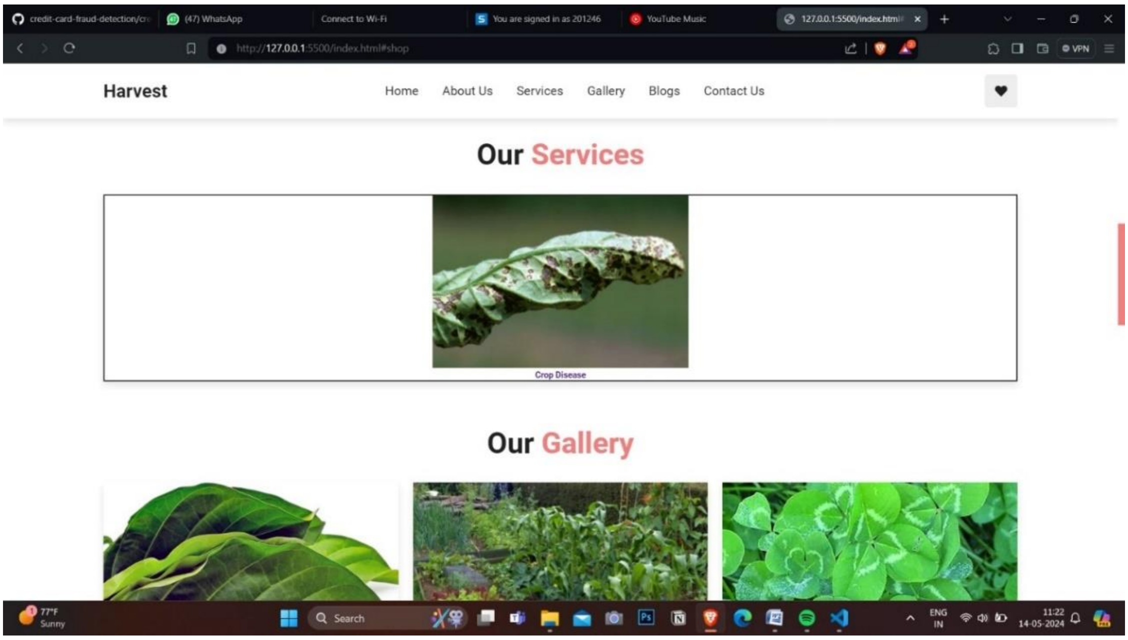


Figure 18: Snippet of the service page

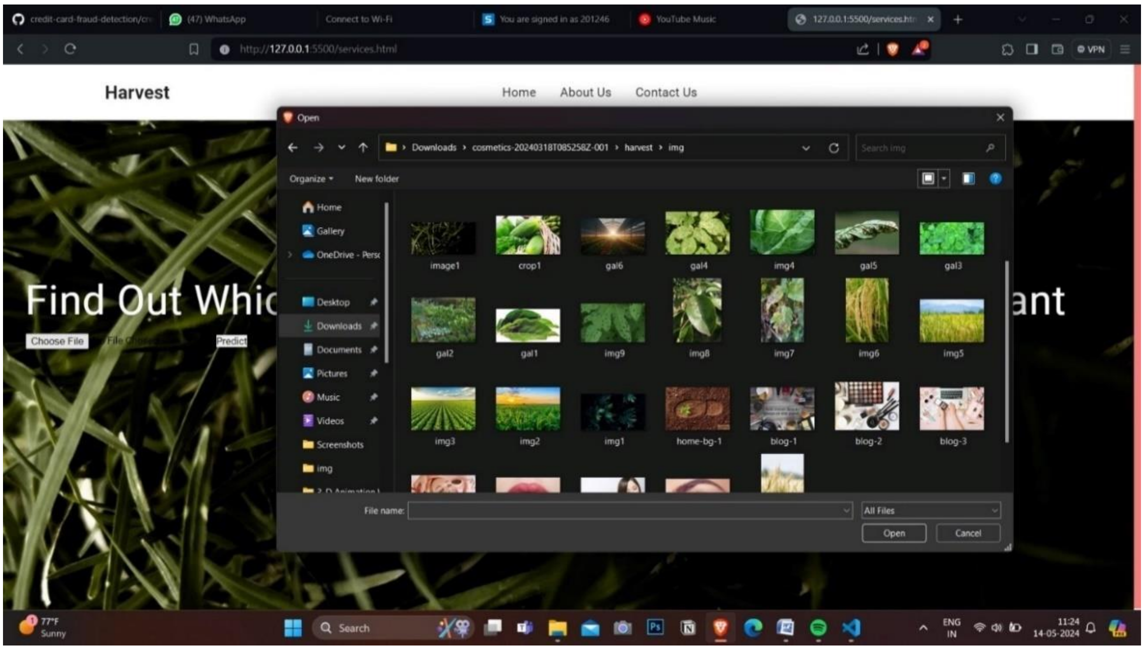


Figure 19: Page to upload the photo of the sample leaf

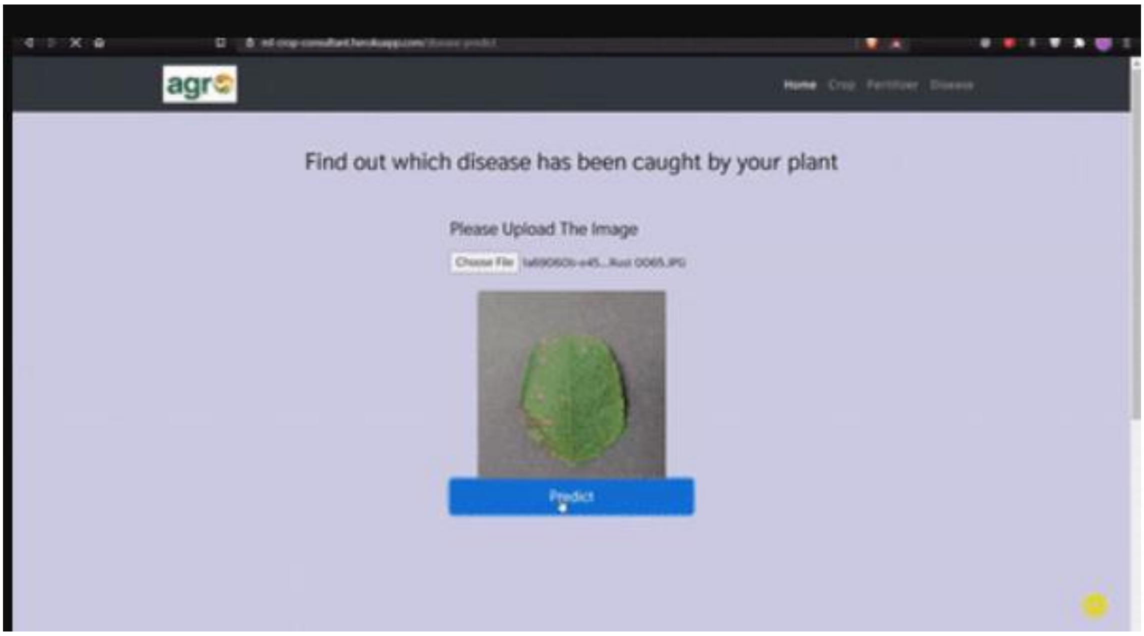


Figure 20: User Interface of the uploaded image

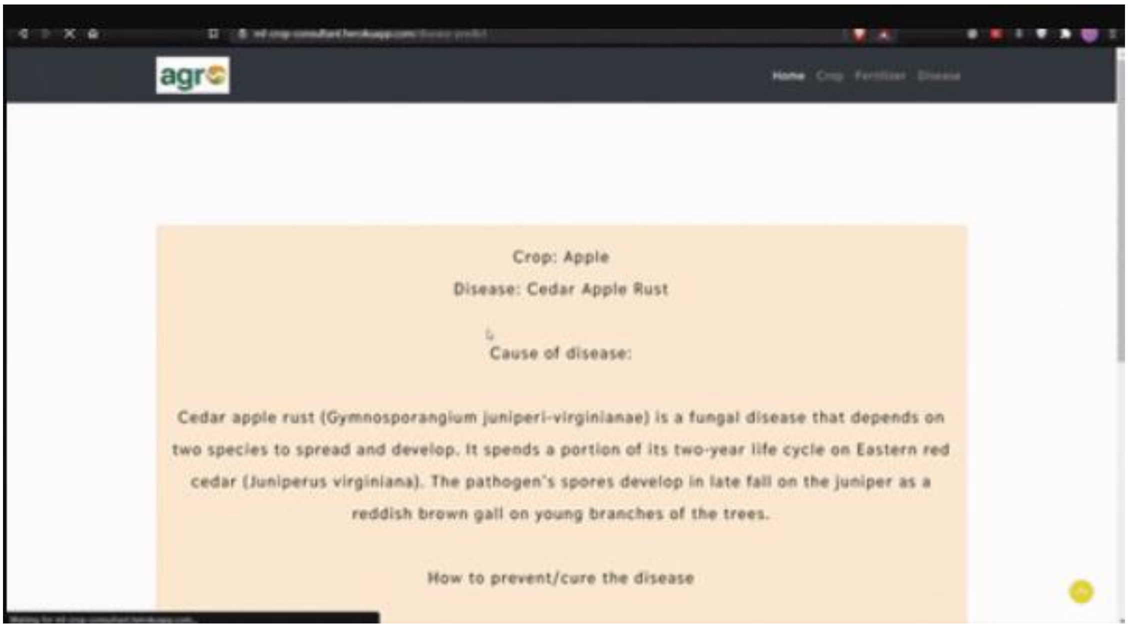


Figure 21: Output 1 (Showing the results based on the input photo)

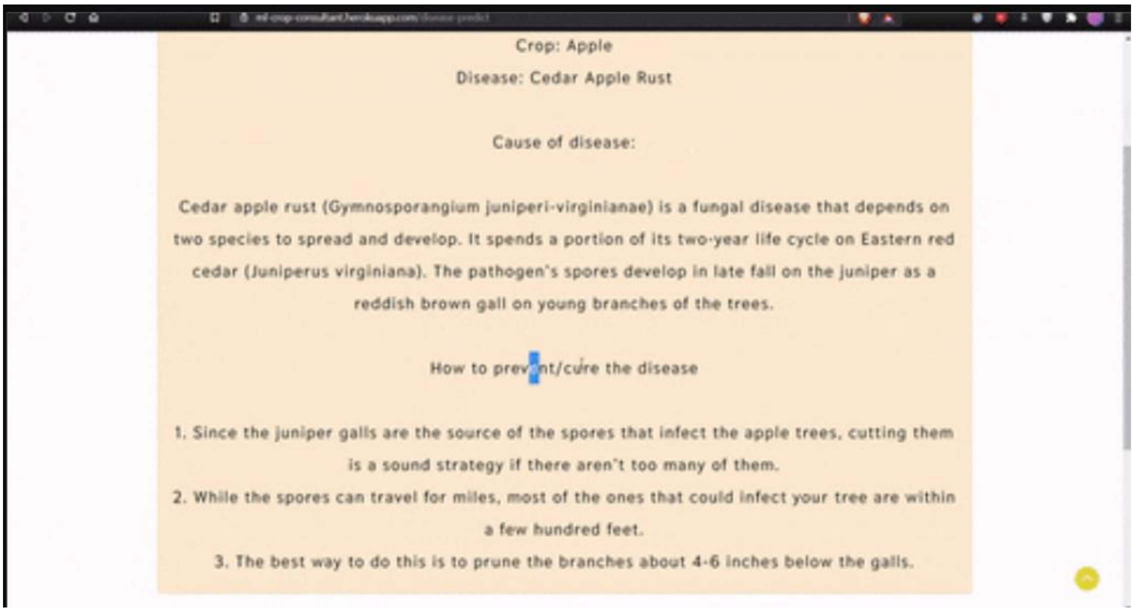


Figure 22: Snippet of output 2

In comparison to the aforementioned existing model, our project excels in accuracy, dataset diversity, real-time recommendations, continuous learning, expert collaboration, deployment flexibility, and security measures, positioning it as a more comprehensive and advanced solution for crop disease prediction and management.

CHAPTER-6 CONCLUSION AND FUTURE SCOPE

6.1) CONCLUSION

Finally, this comprehensive paper delves into the extensive field of crop disease prediction, utilizing cutting-edge technologies and innovative methods. The model's findings reflect a big step forward toward enhancing the precision of disease detection and prediction. While the model has demonstrated high precision and utility, it is critical to recognize its limitations. The dynamic nature of environmental conditions, developing disease strains, and the requirement for constant model development restrict its broad applicability. Nonetheless, these limitations highlight areas for further research and development. Moreover, the model's importance to agricultural and disease management is evident.It not only delivers timely and accurate insights to farmers, but it additionally establishes the framework for the integration of modern technology into the agricultural landscape. As we navigate the difficulties of crop disease prediction, this research demonstrates the potential of data science and machine learning to promote agricultural resilience and sustainability. The journey does not end here; rather, it invites more study and refinement, with a final objective of assuring food security and improving the livelihoods of those who rely on the richness of the land.

KEY FINDINGS

The fusion of machine learning algorithms, specifically Decision Trees, Random Forest, alongside sophisticated image processing techniques, has emerged as a groundbreaking approach in the domain of crop disease identification. This amalgamation has showcased remarkable potential in accurately discerning and categorizing various crop diseases by leveraging leaf images as primary inputs. The intricate interplay between these methodologies has resulted in a robust system capable of not only detecting diseases but also classifying them with a high degree of precision.

Similarly, it is crucial to note that through strategic collaboration with agricultural institutions, the effectiveness of the project has been greatly enhanced. Through these partnerships, it has been easy to access real-time data which ultimately strengthened the model and makes it more accurate and reliable. Broadness of the datasets that include various crop kinds, illnesses, and atmospheric conditions has played an important role. By doing this, therefore, the model has learned sufficiently so as to develop enhanced predictive capacity for multifaceted aspects underpinning diverse farming affairs.

Additionally, the incorporation of deep learning techniques such as CNNs represents a milestone in improving disease diagnosis to higher levels. These sophisticated models have proved great potential towards improving the accuracy and efficiency at which diseases are identified. These CNNs are characterized by extracting very subtle characteristics from images and this has greatly improved the precision in disease categorization.

In essence, this synthesis of machine learning paradigms, strategic data collaborations, and the incorporation of cutting-edge deep learning methodologies underscores the transformative impact on the project's outcomes. It heralds a new era in agricultural disease prediction, promising higher accuracy, adaptability to real-time scenarios, and a more comprehensive understanding of crop diseases, ultimately contributing to the enhancement of agricultural productivity and sustainability.

LIMITATIONS

While significant progress has been achieved in the domain of disease prediction within agriculture, certain limitations persist, casting shadows on the system's overall performance. One of the primary constraints revolves around the dataset—its size, diversity, and accessibility. Despite efforts to expand and diversify the dataset, there remains a persistent challenge in encompassing a truly comprehensive range of crop types, diseases, and environmental conditions. This limitation can potentially hinder the system's ability to accurately generalize across various agricultural scenarios, thereby impacting its predictive capabilities.

Another critical limitation pertains to the integration of real-time data. Despite collaborations with agricultural institutions, the seamless integration of up-to-the-minute data remains an ongoing challenge. The inability to access or integrate real-time data promptly may impede the system's ability to adapt swiftly to evolving agricultural conditions and emerging disease patterns. This limitation poses a significant hurdle in ensuring the system's responsiveness to dynamic scenarios.

Furthermore, the system faces inherent challenges in implementing robust security measures. While efforts are made to fortify the system's security, ensuring the integrity and confidentiality of data remains a concern. The implementation of stringent security measures, especially concerning user data and disease-related information, is crucial to foster user trust and maintain the system's credibility. Failure to address these security concerns adequately may hinder user adoption and trust in the system.

Additionally, scalability remains a challenge. The system's scalability is affected by the complexity of integrating diverse datasets, the computational demands of advanced machine learning models, and the need to accommodate a growing user base. Scaling the system while maintaining its performance efficiency presents a significant challenge that needs to be addressed for widespread adoption and usability.

In summary, limitations persist regarding dataset comprehensiveness, real-time data integration, security measures, and scalability. Addressing these challenges effectively will be pivotal in enhancing the system's reliability, responsiveness, and user acceptance, ultimately advancing its effectiveness in disease prediction and agricultural support.

CONTRIBUTION TO THE FIELD

This project marks a substantial contribution to the agricultural sector by pioneering a predictive model designed to revolutionize disease identification within crops. By harnessing the power of machine learning algorithms like Decision Trees, Random Forests, and advanced image processing techniques, this initiative facilitates the early detection of diseases in crops, playing a pivotal role in mitigating extensive crop damage.

At its heart, this project sets a benchmark within the agriculture sector industry. This, in turn, emphasizes the use of machine learning and image processing approaches for disease prognosis. This study therefore proves the effectiveness of these models in correctly discriminating and differentiating crop diseases from captured leaf images.

Furthermore, it highlights the importance of complete databases and collaboration with the agriculture institutions. This demonstrates the importance of building a strong basis for the increased dataset to represent different crop types, diseases and environmental conditions. This also emphasises the significance of working with agri institutions where real time data is accessible to ensure greater effectiveness of predictive models in an agric setting.

The fact that the project pays much attention to the security issue suggests that it cares about the protection of its client’s personal data and data integrity. This proposal addresses the challenges that are involved when securing IoT devices and it serves as an important basis on which other agri-technologies should be built to ensure they offer robust security options as well.

The contribution of this project goes beyond building a disease forecast model. It propels future research in agricultural sciences that calls for broadening of data sources, teamwork, and improved security. This foundation sets the stage for advanced and dependable agro-technology services targeting the improvement in crop health and yield.

In conclusion, while the system exhibits promising potential in revolutionizing disease prediction and solutions in agriculture, continuous efforts in dataset expansion, technological advancements, and security enhancement are imperative for its sustainable impact and widespread adoption in the agricultural sector.

6.2) FUTURE SCOPE

The project's scope for future advancement entails several critical facets aimed at enhancing its efficacy and accessibility within the agricultural domain. Firstly, the primary focus lies in broadening the dataset. This expansion aims to incorporate a diverse spectrum of crop types, encompassing various diseases, and accounting for a wide array of environmental conditions. By diversifying the dataset, the predictive model gains depth and versatility, enabling it to cater to a broader range of agricultural scenarios, thereby ensuring more accurate disease predictions.

Collaborating with esteemed agricultural institutions marks another pivotal step. Such partnerships facilitate access to real-time data sources, offering a wealth of updated information crucial for refining and fortifying the predictive model's accuracy. Real-time data infusion ensures that the system remains relevant and adaptive to the dynamic agricultural landscape.

The project's future trajectory also entails a deeper dive into advanced machine learning models, particularly exploring the potentials of deep learning techniques. Leveraging these advanced models promises heightened precision and sophistication in disease identification, potentially revolutionizing the accuracy of prediction and solution recommendation processes.

Simultaneously, the focus extends to the design aspect, specifically tailoring a user-friendly interface. This interface aims to cater directly to farmers and stakeholders, ensuring accessibility and ease of use. A well-crafted, intuitive interface fosters greater adoption and seamless interaction with the system's predictive capabilities.

Finally, there's a paramount need to develop a secure web portal fortified with user login features and comprehensive access to disease-related information. Prioritizing robust security measures, such as integrating OTP-based authentication, ensures the safeguarding of sensitive data while enabling seamless information access for users.

In essence, these future endeavors collectively contribute to the project's evolution, rendering it more robust, inclusive, and user-centric. This expansive vision aims to deliver a sophisticated agricultural disease prediction and solution system that caters comprehensively to the needs of farmers and stakeholders while upholding data security and technological innovation.

Expanding the dataset, collaborating for real-time data sources, designing a user-friendly interface, and fortifying security measures are key pathways to extend the project's impact, making it more comprehensive, accurate, and accessible for agricultural stakeholders.

REFERENCES

[1] Islam, Md Manowarul, et al. "DeepCrop: Deep learning-based crop disease prediction with web application." Journal of Agriculture and Food Research 14 (2023): 100764.

[2] Ashraf, Mahmood, et al. "A Convolutional Neural Network Model for Wheat Crop Disease Prediction." Computers, Materials & Continua 75.2 (2023).

[3] Nandhini, S., and K. Ashokkumar. "Machine learning technique for crop disease prediction through crop leaf image." Appl. Math 16.2 (2022): 149-158.

[4] Pokkuluri, Kiran Sree, SSSN Usha Devi Nedunuri, and Usha Devi. "Crop Disease Prediction with Convolution Neural Network (CNN) Augmented With Cellular Automata." INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY 19.5 (2022): 765-773.

[5] Sharma, Gargi, and Gourav Shrivastava. "Crop disease prediction using deep learning techniques—a review." *ICTACT J Data Sci Mach Learn* 3.02 (2022): 312-315.

[6] Yadav, Nikita, et al. "Crop disease prediction and solution." *International Research Journal of Engineering and Technology* 8.02 (2021).

[7] M. Ahmad, M. Abdullah, H. Moon and D. Han, "Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks with Stepwise Transfer Learning", IEEE Access, Vol. 9, pp. 140565-140580, (2021).

[8] C. Zhou, S. Zhou, J. Xing and J. Song, "Tomato Leaf Disease Identification by Restructured Deep Residual Dense Network", IEEE Access, Vol. 9, pp. 28822-28831, (2021).

[9] A. Luaibi and A. Miry, "Detection of Citrus Leaf Diseases using a Deep Learning Technique", International Journal of Electrical and Computer Engineering, Vol. 11, pp. 1719- 1727, (2021).

[10] Lee SH, Go¨eau H, Bonnet P, Joly A. Conditional multitask learning for plant disease identification, In2020 25th International Conference on Pattern Recognition (ICPR), IEEE, 3320-3327 (2021).

[11] Wagle SA. Comparison of Plant Leaf Classification Using Modified AlexNet and Support Vector Machine, Traitement du Signal, 38(1), (2021).

[12] Alaeddine H, Jihene M. Deep Batch-normalized eLU AlexNet For Plant Diseases Classification, In 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD), IEEE, 17-22 (2021).

[13] Manga NA. Plant Disease Classification using Residual Networks with MATLAB,Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(10), 3923-33 (2021).

[14] A. Khattak, "Automatic Detection of Citrus Fruit and Leaves Diseases Using Deep Neural Network Model", IEEE Access, Vol. 9, pp. 112942-112954, (2021).

[15] Miaomiao Ji, Lei Zhang and Qiufeng Wu, "Automatic Grape Leaf Diseases Identification via United Model based on Multiple Convolutional Neural Networks", Information Processing in Agriculture, Vol. 7, No. 3, pp. 1-13, (2020).

[16] Verma S, Chug A, Singh AP. Exploring capsule networks for disease classification in plants, Journal of Statistics and Management Systems, 23(2), 307-15 (2020).

[18] Arunnehru J, Vidhyasagar BS, Anwar Basha H. Plant Leaf Diseases Recognition Using Convolutional Neural Network and Transfer Learning, International Conference on Communication, Computing and Electronics Systems, Springer, Singapore, 221-229 (2020).

[17] Y. Zhang, C. Song and D. Zhang, "Deep Learning-Based Object Detection Improvement for Tomato Disease", IEEE Access, Vol. 8, pp. 56607-56614, (2020).

[18] B. Liu, C. Tan, S. Li, J. He and H. Wang, "A Data Augmentation Method Based on Generative Adversarial Networks for Grape Leaf Disease Identification", IEEE Access, Vol. 8, pp. 102188-102198, (2020).

[19] D. Jiang, F. Li, Y. Yang and S. Yu, "A Tomato Leaf Diseases Classification Method Based on Deep Learning", Proceedings of Chinese Conference on Control and Decision, pp. 1446-1450, (2020).

[20] Q. Wu, K. Zhang and J. Meng, "Identification of Soybean Leaf Diseases via Deep Learning", Journal of The Institution of Engineers (India): Series A, Vol. 100, pp. 659-666, (2019).

[21] Clarkson G, Dorward P, Osbahr H, Torgbor F, KankamBoadu I. An investigation of the effects of PICSA on smallholder farmers' decision-making and livelihoods when implemented at large scale–The case of Northern Ghana, Climate Services, 14 1-4 (2019).

[22] Aniah P, Kaunza-Nu-Dem MK, Ayembilla JA. Smallholder farmers' livelihood adaptation to climate variability and ecological changes in the savanna agro ecological zone of Ghana, Heliyon, 5(4), e01492 (2019).

[23] Phuong LT, Biesbroek GR, Sen LT, Wals AE. Understanding smallholder farmers' capacity to respond to climate change in a coastal community in Central Vietnam, Climate and Development, 10(8), 701-16 (2018).

[24] Tamilselvi, P., K.Ashok Kumar, Unsupervised machine learning for clustering the

infected leaves based on the leaf-colours (2017) ICONSTEM 2017 - Proceedings: 3rd IEEE International Conference on Science Technology, Engineering and Management, 106-110 (2018). DOI: 10.1109/ICONSTEM.2017.8261265

[25] Daba MH. Assessing local community perceptions on climate change and variability and its effects on crop production in selected districts of western Oromia, Ethiopia. J. Climatol. Weather Forecasting. 6, 216 (2018)

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

PLAGIARISM VERIFICATION REPORT

Date: ....

Type of Document (Tick): **PhD Thesis** | **M.Tech Dissertation/ Report** | **B.Tech Project Report** | **Paper**

Name: ______________________________**Department:** _________________**Enrolment No** _________

Contact No. ______________________________**E-mail.** _______________________________________

Name of the Supervisor: ____

Title of the Thesis/Dissertation/Project Report/Paper (In Capital letters): _________________________

____

____

UNDERTAKING

I undertake that I am aware of the plagiarism related norms/ regulations, if I found guilty of any plagiarism and copyright violations in the above thesis/report even after award of degree, the University reserves the rights to withdraw/revoke my degree/report. Kindly allow me to avail Plagiarism verification report for the document mentioned above.

Complete Thesis/Report Pages Detail:

- Total No. of Pages =
- Total No. of Preliminary pages =
- Total No. of pages accommodate bibliography/references =

(Signature of Student)

FOR DEPARTMENT USE

We have checked the thesis/report as per norms and found **Similarity Index** at................... (%). Therefore, we
are forwarding the complete thesis/report for final plagiarism check. The plagiarism verification report may be handed over to the candidate.

(Signature of Guide/Supervisor) **Signature of HOD**

FOR LRC USE

The above document was scanned for plagiarism check. The outcome of the same is reported below:

Copy Received on	Excluded	Similarity Index (%)	Generated Plagiarism Report Details (Title, Abstract & Chapters)	
	• All Preliminary Pages • Bibliography/Images/Quotes • 14 Words String		Word Counts	
Report Generated on			Character Counts	
		Submission ID	Total Pages Scanned	
			File Size	

Checked by
Name & Signature **Librarian**

....

Please send your complete thesis/report in (PDF) with Title Page, Abstract and Chapters in (Word File) through the supervisor at plagcheck.juit@gmail.com