



PREMIER UNIVERSITY  
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING  
*A Project Report On*  
**PlantGuard: Revolutionizing Agriculture with Image-Based Disease Detection**

**Course Title:** Artificial Intelligence Laboratory  
**Course Code:** CSE 318

Submitted To: Tashin Hossain  
Lecturer  
Department of Computer Science and Engineering

Submitted By:  
**Arnab Shikder** (0222210005101098)  
**Sayed Hossain** (0222210005101102)  
**Mohammad Ohidul Alam** (0222210005101123)  
**Mohammad Hafizur Rahman Sakib** (0222210005101118)  
**Mohammad Asmual Hoque Yousha** (0222210005101121)

# TABLE OF CONTENTS

<b>TITLE PAGE</b>	<b>i</b>
<b>TABLE OF CONTENTS</b>	<b>iii</b>
<b>LIST OF FIGURES</b>	<b>iv</b>
<b>LIST OF TABLES</b>	<b>1</b>
<b>1 Introduction</b>	<b>2</b>
<b>2 Objectives</b>	<b>2</b>
<b>3 Background and Motivation</b>	<b>3</b>
3.0.1 Literature Review . . . . .	3
<b>4 Problem Statement</b>	<b>4</b>
<b>5 Data Description</b>	<b>5</b>
5.1 Source of Data . . . . .	5
5.2 Type of Data . . . . .	5
5.3 Data Preprocessing Requirements . . . . .	5
<b>6 Methodology</b>	<b>6</b>
6.1 Proposed Approach . . . . .	6
6.2 Algorithms/Models Under Consideration . . . . .	6
6.3 Tools, Libraries, and Frameworks . . . . .	6
6.4 Workflow . . . . .	7
6.5 Cost-Benefit Analysis . . . . .	8
6.6 Gantt Chart . . . . .	8
<b>7 Evaluation Metrics</b>	<b>8</b>
7.1 Primary Metrics . . . . .	8
7.2 Additional Metrics . . . . .	9
7.3 Criteria for Success . . . . .	10
<b>8 Implementation Plan</b>	<b>10</b>
8.1 Phases of Implementation . . . . .	10
8.2 Estimated Timeline . . . . .	12
<b>9 Expected Outcome</b>	<b>12</b>

9.1	Tangible Deliverables . . . . .	12
9.2	Potential Business and Research Implications . . . . .	13
<b>10</b>	<b>Milestone</b>	<b>13</b>
10.1	Project Milestones and Timeline . . . . .	14
10.2	Visualization of Timeline . . . . .	14
	<b>REFERENCES</b>	<b>15</b>

## List of Figures

6.1	Workflow for Crop Disease Prediction System . . . . .	7
10.1	Sample Gantt Chart demonstrating schedule feasibility . . . . .	14

## List of Tables

6.1	Sample Cost-Benefit Analysis of the Proposed Project . . . . .	8
8.1	Timeline for Project Phases . . . . .	12

# **1. Introduction**

Crop diseases pose a significant challenge to global agriculture, impacting crop yield and food security. Timely and accurate identification of crop diseases is critical for effective management and prevention. With advancements in machine learning and computer vision, automated systems for disease detection have gained considerable attention.

This project focuses on utilizing a dataset of approximately 17,000 images categorized into five distinct classes of crop diseases. By leveraging these images, the project aims to develop a machine learning model capable of automatically classifying and predicting crop diseases. Such a system can assist farmers and agricultural stakeholders in early detection, enabling targeted interventions to mitigate crop loss.

The broader context of this project lies in enhancing agricultural productivity and sustainability. The use of technology in agriculture has the potential to address challenges such as disease outbreaks, reducing economic losses, and ensuring food security. This proposal outlines the methodology, objectives, and anticipated outcomes, contributing to the intersection of agriculture and artificial intelligence.

# **2. Objectives**

The primary objective of this project is to address the challenge of accurately identifying and classifying crop diseases through automated means. Crop diseases significantly impact agricultural productivity, leading to economic losses and threats to food security. This project aims to provide a technological solution to mitigate these issues.

The specific goals of this project include:

- Developing a machine learning model capable of classifying crop diseases using a dataset of approximately 17,000 images.
- Improving the speed and accuracy of crop disease detection compared to manual diagnosis.
- Reducing the dependency on expert intervention by providing farmers with an accessible and user-friendly diagnostic tool.

- Enhancing agricultural management practices by enabling early detection and targeted treatment of diseases.
- Contributing to sustainable agricultural development by minimizing crop losses and improving yield quality.

The intended outcome of this project is a robust, reliable, and scalable system that can assist agricultural stakeholders in efficiently managing crop health, ultimately supporting global food security.

### **3. Background and Motivation**

Agriculture plays a vital role in sustaining human life and economic development. However, crop diseases remain a persistent challenge, affecting global food production and supply chains. Early and accurate diagnosis of crop diseases is crucial for mitigating losses, but traditional methods rely heavily on human expertise, which can be time-consuming and error-prone.

The relevance of this problem lies in its potential impact on food security, economic stability, and environmental sustainability. By automating the identification of crop diseases using machine learning, this project seeks to address these challenges, empowering farmers with tools for proactive disease management.

Advancements in artificial intelligence and the availability of large-scale image datasets present a unique opportunity to tackle this issue effectively. The project aims to bridge the gap between traditional agricultural practices and modern technological solutions.

#### **3.0.1. Literature Review**

Several studies have explored the application of machine learning and computer vision in agriculture. For instance, convolutional neural networks (CNNs) have been widely adopted for image-based disease classification, achieving promising results. However, challenges such as dataset imbalance, variations in environmental conditions, and the need for real-time inference remain areas of concern [1, ?].

Opportunities lie in improving model generalization and integrating disease prediction systems with IoT-based agricultural tools for seamless field deployment. Recent research high-

lights the potential for transfer learning and data augmentation techniques to overcome dataset limitations [?]. These insights form the foundation for this project, which aims to build a robust system addressing these challenges while leveraging existing advancements.

## 4. Problem Statement

The agricultural sector faces significant challenges due to crop diseases, which lead to reduced yield, economic losses, and food insecurity. Traditional methods of disease diagnosis often rely on visual inspections by experts, which are time-consuming, labor-intensive, and prone to human error. Farmers in remote or resource-limited areas may not have access to expert knowledge, further exacerbating the issue.

This project seeks to address the problem of early and accurate identification of crop diseases using machine learning techniques. By leveraging a dataset of approximately 17,000 crop images across five disease classes, the project aims to develop an automated system capable of detecting and classifying crop diseases.

The solution will be evaluated based on:

- **Accuracy:** The classification performance of the model in correctly identifying diseases.
- **Speed:** The time taken by the system to process an image and provide a diagnosis.
- **Precision and Recall:** Metrics that ensure the system is not only accurate but also reliable in distinguishing between similar disease classes.
- **Scalability:** The ability of the system to handle new datasets and additional disease categories with minimal retraining.

By addressing these evaluation criteria, the project aims to provide a reliable, scalable, and accessible solution to assist farmers and agricultural stakeholders in managing crop health effectively.



## 5. Data Description

### 5.1. Source of Data

The dataset used for this project comprises approximately 17,000 labeled images of crops and their associated diseases. These images are sourced from publicly available datasets, including agricultural research repositories and academic datasets tailored for crop disease classification. Additionally, the dataset may be supplemented with synthetic data generated through data augmentation techniques to improve model performance and generalization.

### 5.2. Type of Data

The dataset primarily consists of unstructured image data. Each image belongs to one of five distinct classes, representing specific crop diseases. The data is organized in the following format:

- **Input Data:** High-resolution images of crops captured under various environmental conditions.
- **Labels:** Each image is annotated with its corresponding crop disease class, enabling supervised learning.

### 5.3. Data Preprocessing Requirements

To ensure optimal model performance and reliability, the dataset will undergo the following preprocessing steps:

- **Image Resizing:** All images will be resized to a uniform dimension to standardize input for the machine learning model.
- **Normalization:** Pixel values will be normalized to a common scale to enhance model convergence during training.
- **Data Augmentation:** Techniques such as rotation, flipping, and cropping will be applied to artificially increase dataset diversity and address class imbalance.
- **Noise Removal:** Images with poor quality, excessive noise, or irrelevant content will be filtered out.

- **Splitting:** The dataset will be divided into training, validation, and test sets to evaluate model performance effectively.

By ensuring proper preprocessing and leveraging diverse image data, the project aims to build a robust and accurate crop disease detection system.

## 6. Methodology

This section outlines the proposed approach for implementing the crop disease prediction system, including the machine learning techniques, tools, and resources used.

### 6.1. Proposed Approach

The project will utilize a deep learning-based supervised learning approach. The labeled dataset of crop images will serve as the input to train a convolutional neural network (CNN) for image classification tasks. The supervised learning approach ensures that the model learns to map input images to corresponding disease labels effectively.

### 6.2. Algorithms/Models Under Consideration

The following algorithms and models will be evaluated for their performance:

- **Convolutional Neural Networks (CNNs):** The primary model for image-based classification, leveraging architectures like VGG16, ResNet, and EfficientNet.
- **Transfer Learning:** Pre-trained models such as MobileNet and InceptionNet will be fine-tuned on the crop disease dataset to improve accuracy and reduce training time.
- **Ensemble Models:** Combining predictions from multiple models to enhance reliability and performance.

### 6.3. Tools, Libraries, and Frameworks

The implementation will utilize the following tools and libraries:

- **Programming Language:** Python
- **Libraries and Frameworks:** TensorFlow, Keras, PyTorch, OpenCV, and NumPy
- **Development Environment:** Jupyter Notebook, Google Colab
- **Visualization Tools:** Matplotlib, Seaborn, Plotly

## 6.4. Workflow

The proposed workflow for the project is depicted in Figure 6.1, which includes the phases of data preprocessing, model training, and deployment.

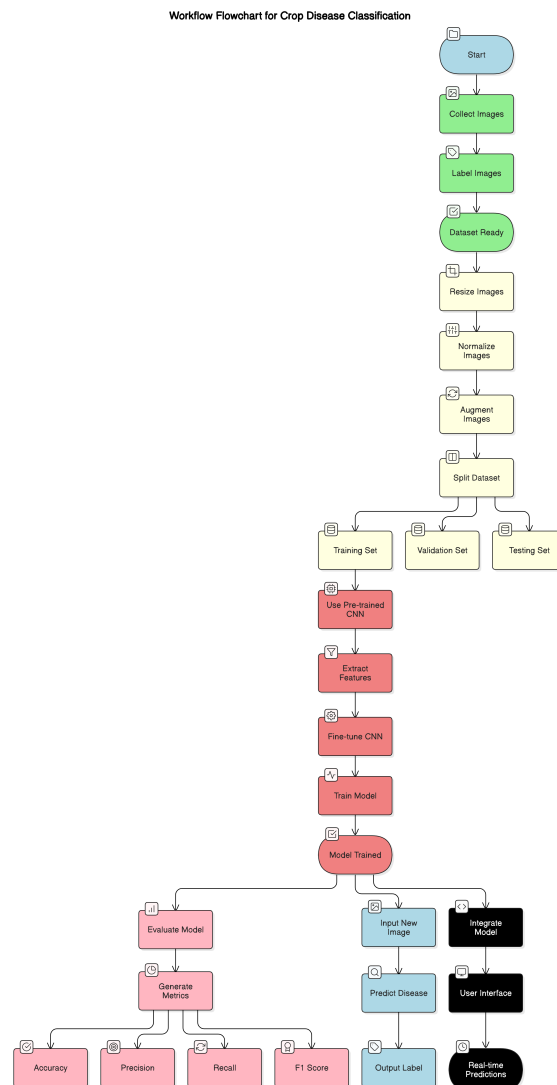


Figure 6.1: Workflow for Crop Disease Prediction System

## 6.5. Cost-Benefit Analysis

A sample cost-benefit analysis of the project is presented in Table 6.1.

Table 6.1: Sample Cost-Benefit Analysis of the Proposed Project

Item	Description	Cost (\$)	Benefit (\$)
Development Costs	Software Development	15,000	-
Hardware Costs	Servers and Equipment	5,000	-
Training Costs	User Training Sessions	2,000	-
Maintenance Costs	Annual Maintenance	1,000	-
<b>Total Costs</b>		<b>23,000</b>	-
Increased Efficiency	Time Savings	-	30,000
Improved User Satisfaction	User Feedback	-	10,000
Revenue Increase	New Customers	-	20,000
<b>Total Benefits</b>		-	<b>60,000</b>
<b>Net Benefit</b>		<b>23,000</b>	<b>37,000</b>

## 6.6. Gantt Chart

To plan and track the timeline for the project phases, a Gantt chart will be created using tools like <https://www.onlinegantt.com/#/gantt>. The Gantt chart will detail each phase's start and end dates, milestones, and dependencies.

# 7. Evaluation Metrics

To measure the success of the crop disease prediction model, various evaluation metrics will be used to assess its performance. These metrics are crucial for determining the effectiveness of the model in accurately identifying crop diseases and ensuring reliability for end-users.

## 7.1. Primary Metrics

The following metrics will be used to evaluate the model:

- **Accuracy:** Measures the overall correctness of predictions by calculating the proportion of correctly classified images out of the total dataset. It provides a general sense of performance but may not capture class imbalances.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Samples}}$$

- **Precision:** Focuses on the quality of positive predictions by measuring the proportion of correctly identified positive samples out of all predicted positives. Useful when false positives need to be minimized.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Recall (Sensitivity):** Indicates the ability to correctly identify positive samples out of all actual positives. It is crucial when false negatives have significant consequences, such as missing a disease diagnosis.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **F1 Score:** Provides a balanced measure of precision and recall by calculating their harmonic mean. It is particularly useful in scenarios with class imbalances.

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

## 7.2. Additional Metrics

The following additional metrics will be considered to gain deeper insights into the model's performance:

- **Confusion Matrix:** A tabular representation of actual versus predicted classifications that highlights true positives, true negatives, false positives, and false negatives.
- **Receiver Operating Characteristic (ROC) Curve and AUC:** Evaluates the model's ability to distinguish between classes by plotting true positive rates against false positive rates at various thresholds. The Area Under the Curve (AUC) quantifies this capability.
- **Cross-Entropy Loss:** Measures the difference between the predicted probability distribution and the actual labels during training, providing a sense of how well the model is learning.

- **Root Mean Squared Error (RMSE):** If applicable, RMSE will be used to evaluate models that involve regression components (e.g., severity prediction of a disease).

### 7.3. Criteria for Success

The success of the project will be determined based on the following criteria:

- Achieving an accuracy of at least 90% on the test dataset.
- Maintaining a precision and recall of at least 85% across all classes.
- Ensuring a high F1 Score, particularly for minority classes, to address any class imbalances.
- Demonstrating robustness and generalizability through consistent performance on unseen validation data.

By utilizing these metrics, the project ensures a comprehensive evaluation of the model's performance, guaranteeing its effectiveness and reliability in real-world applications.

## 8. Implementation Plan

This project will be executed in several phases, each focusing on a critical aspect of the crop disease prediction system. The steps and their respective timelines are as follows:

### 8.1. Phases of Implementation

#### 1. Data Collection and Understanding (Week 1–2):

- Collect publicly available datasets containing images of crop diseases.
- Analyze the dataset structure, distribution, and class balance.
- Supplement the dataset with synthetic data using data augmentation techniques if required.

#### 2. Data Preprocessing (Week 3–4):

- Standardize image dimensions by resizing.
- Normalize pixel values for consistent input across the model.
- Remove noisy or irrelevant data and handle class imbalance using augmentation.
- Split the dataset into training (70%), validation (15%), and test (15%) subsets.

### **3. Model Selection and Training (Week 5–8):**

- Evaluate various machine learning models, focusing on convolutional neural networks (CNNs).
- Train selected models on the preprocessed dataset using GPUs for faster computation.
- Fine-tune hyperparameters, such as learning rate, batch size, and number of epochs, to optimize performance.

### **4. Model Testing and Evaluation (Week 9–10):**

- Evaluate the trained model on the test dataset using metrics such as accuracy, precision, recall, and F1 score.
- Conduct error analysis to identify common misclassifications and refine the model.

### **5. System Deployment (Week 11–12):**

- Develop a user-friendly interface for farmers and agricultural stakeholders.
- Integrate the trained model into the system for real-time crop disease detection.
- Test the deployed system in a simulated or real-world environment to ensure reliability.

## 8.2. Estimated Timeline

Phase	Duration	Description
Data Collection	Week 1–2	Gather and analyze dataset(s).
Data Preprocess- ing	Week 3–4	Clean, augment, and split data.
Model Training	Week 5–8	Train and fine-tune machine learning models.
Model Testing	Week 9–10	Evaluate and optimize the trained model.
Deployment	Week 11–12	Deploy the system with a user-friendly interface.

Table 8.1: Timeline for Project Phases

By following this structured implementation plan, the project aims to achieve a robust and efficient solution for crop disease prediction within the specified timeline.

## 9. Expected Outcome

This project is expected to deliver tangible results that address the problem of crop disease identification and classification. The key outcomes and their potential implications are outlined below:

### 9.1. Tangible Deliverables

- **Predictive Model:** A robust machine learning model capable of accurately classifying crop diseases based on input images. The model will be trained on a dataset of approximately 17,000 images, achieving high accuracy, precision, and recall.
- **User-Friendly Interface:** A simple, intuitive application or web-based interface that allows users, such as farmers or agricultural advisors, to upload crop images and receive disease diagnoses in real time.
- **Visualization:** Clear and interpretable visualizations of the model’s predictions, including probability scores and explanations for the classification results, enhancing



user trust and understanding.

- **Insights:** Data-driven insights into common crop diseases, their frequency, and possible patterns, which could aid agricultural research and policy-making.

## 9.2. Potential Business and Research Implications

- **Business Implications:**

- The solution can be commercialized as a diagnostic tool for farmers, agribusinesses, and agricultural extension services.
- Reduced crop losses and improved yield quality can result in higher profits for farmers and agricultural stakeholders.
- Potential integration with agricultural IoT devices, enabling automated disease monitoring in smart farming ecosystems.

- **Research Implications:**

- The project contributes to advancements in applying machine learning and computer vision in agriculture.
- The generated dataset and model can serve as a baseline for future research on crop disease prediction and management.
- Insights from the project may inspire further studies on disease prevention strategies, including environmental and genetic factors affecting crop health.

By achieving these outcomes, the project aims to provide a practical, impactful solution that benefits both the agricultural industry and the research community, fostering sustainable agricultural practices and food security.

## 10. Milestone

This section outlines the key milestones for the successful completion of the project, along with their respective timelines. The milestones are divided into distinct phases to ensure clarity and manageability throughout the project lifecycle.

10.1. Project Milestones and Timeline

The major milestones for the project are as follows:

1. Data Collection and Preparation (Week 1-4):
- Collect the dataset of crop disease images.
  - Perform data cleaning, augmentation, and preprocessing to ensure quality and consistency.
2. Model Development and Training (Week 5-8):
- Explore and implement machine learning and deep learning models.
  - Train the models using the prepared dataset and tune hyperparameters.
3. Model Evaluation and Optimization (Week 9-10):
- Evaluate the models using defined metrics like accuracy, precision, and F1 score.
  - Optimize the model to improve performance and handle class imbalances.
4. Application Development (Week 11-12):
- Develop a user-friendly application or interface for disease prediction.
  - Integrate the trained model with the application for real-time predictions.
5. Testing and Deployment (Week 13-14):
- Test the application for accuracy, usability, and performance.
  - Deploy the final system for end-user accessibility.

10.2. Visualization of Timeline

The timeline and schedule feasibility for the project phases are depicted in the Gantt chart below (Figure 10.1).

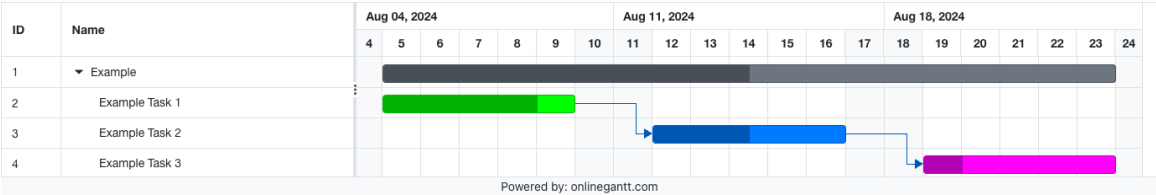


Figure 10.1: Sample Gantt Chart demonstrating schedule feasibility

## References

- [1] P. Neupane and M. Thakur, “Variational study of the impact of call graphs on precision of android taint analysis,” in *Proceedings of the 16th Innovations in Software Engineering Conference*, ser. ISEC '23. New York, NY, USA: Association for Computing Machinery, 2023. [Online]. Available: <https://doi.org/10.1145/3578527.3578545>