**POTATO DISEASE DETECTION USING DEEP LEARNING**

**Submitted by:** **SAMUEL KANYI**

**Registration Number: SCT221-0666/2021**

**Bachelor of Science in**

**JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY**

**Date: 27/11/2024**

### DECLARATION

This research project is my original work and has not been presented for a degree or any other award in any other University.

**Name:** Samuel Kanyi  
**Reg. No: SCT221-0666/2021**

**Signature:** ....................................

**Date:** …………………………….

This research project has been submitted for examination with my approval as the University Supervisor.

**Name of supervisor:**

**Signature:** ........................................

**Date:** ..............................................

### ****ACKNOWLEDGEMENTS****

Most importantly, I express my most heartfelt gratitude to the Almighty God, whose unwavering mercy, strength, and guidance have been my constant fount of hope and determination in this undertaking. It is by His hand that I have arrived thus far, and for this, I will forever be grateful.

I would like to sincerely thank Jomo Kenyatta University of Agriculture and Technology for giving me an academic setting that has nurtured my curiosity, honed my abilities, and motivated me to grow intellectually and as a person. The process of undertaking this degree and conducting this research has been a rewarding and life-altering experience.

To my supervisor, I owe a special debt. His continuous guidance, insightful criticisms, and unremitting confidence in my capacity have been instrumental in shaping the course and nature of this work. His tolerance and willingness to go with me through every hurdle and revision taught me not just about research, but also the values of endurance and academic integrity.

I also appreciate my friends and classmates, whose reflective feedback, support, and teamwork enriched this process. Their presence reminded me that this was not an individual effort but a joint process of knowledge gathering and building. To everyone who helped me directly or indirectly, thanks for being part of this moment in my life.

### ABSTRACT

Early disease detection for crops is important to improve yield and food security, especially for small-scale farmers. The project is focused on developing a deep learning system to detect potato leaf disease automatically by image classification. The system is focused on classifying three types: Early Blight, Late Blight, and Healthy.

The project follows the CRISP-DM methodology, beginning with an understanding of the need for affordable and fast disease diagnosis tools for farmers. A labeled dataset of potato leaf images is used for training and testing the model. Images are preprocessed through resizing, normalization, and augmentation to improve the model's generalization.

A Convolutional Neural Network (CNN) is designed and trained using TensorFlow on the basis of Python and other supporting libraries such as OpenCV and PIL. The model's performance is measured with parameters such as accuracy, precision, recall, and confusion matrix.

The trained model is found to have extremely high classification accuracy for leaf condition and can be deployed to the web or mobile for real-time disease identification. The result indicates that deep learning has potential as an efficient solution for helping agricultural decision-making and preventing crop loss.

**Table of Contents**

[DECLARATION 2](#_Toc196963446)

[ACKNOWLEDGEMENTS 3](#_Toc196963447)

[ABSTRACT 4](#_Toc196963448)

[LIST OF TABLES 7](#_Toc196963449)

[LIST OF FIGURES 7](#_Toc196963450)

[**CHAPTER 1 : INTRODUCTION** 8](#_Toc196963451)

[1.1 Background 8](#_Toc196963452)

[1.2 Project Overview 9](#_Toc196963453)

[1.3 Statement of the Problem 10](#_Toc196963454)

[1.4 Proposed Solution 10](#_Toc196963455)

[1.5 Objectives 11](#_Toc196963456)

[1.6 Research Questions 11](#_Toc196963457)

[1.7 Justification 12](#_Toc196963458)

[1.8 Proposed Research and System Methodologies 12](#_Toc196963459)

[1.9 Scope 13](#_Toc196963460)

[1.10 Requirements and Budgeting 14](#_Toc196963461)

[1.10.1Software requirements 14](#_Toc196963462)

[1.10.2 Hardware Requirements 14](#_Toc196963463)

[Budget justification 14](#_Toc196963464)

[1.10.3 Project Timeline 15](#_Toc196963465)

[TABLE 3: PROJECT TIME PLAN 15](#_Toc196963466)

[1.10.4 Project Cost Estimates 15](#_Toc196963467)

[**CHAPTER 2: LITERATURE REVIEW** 17](#_Toc196963468)

[2.1 Introduction 17](#_Toc196963469)

[2.2 Theoretical Review 17](#_Toc196963470)

[2.3 Case Study Review 19](#_Toc196963471)

[2.4 Integration and Architecture 20](#_Toc196963472)

[2.5 Summary 22](#_Toc196963473)

[2.6 Research Gaps 22](#_Toc196963474)

[**CHAPTER 3** **SYSTEM ANALYSIS AND DESIGN** 23](#_Toc196963475)

[3.1 Introduction 23](#_Toc196963476)

[3.2 Systems Development Methodology 23](#_Toc196963477)

[3.3 Feasibility Study 24](#_Toc196963478)

[3.4 Requirements Elicitation 25](#_Toc196963479)

[3.5 Data Analysis 25](#_Toc196963480)

[3.6 System Specification 27](#_Toc196963481)

[3.7 Requirements Analysis and Modeling 27](#_Toc196963482)

[**CHAPTER 4** 31](#_Toc196963483)

[**SYSTEM IMPLEMENTATION AND TESTING, CONCLUSIONS, AND RECOMMENDATIONS** 31](#_Toc196963484)

[4.1 Introduction 31](#_Toc196963485)

[4.2 Implementation Environment and Tools 31](#_Toc196963486)

[4.3 System Code Generation 31](#_Toc196963487)

[4.4 Testing 32](#_Toc196963488)

[4.5 User Guide 32](#_Toc196963489)

[4.6 Conclusions 35](#_Toc196963490)

[**REFERENCES** 36](#_Toc196963491)

[**APPENDICES** 38](#_Toc196963492)

[**Appendix 1: Interview questions** 38](#_Toc196963493)

[**Appendix 2: Questionnaire** 38](#_Toc196963494)

### LIST OF TABLES

Table 1: Software requirements………………………………………………………………. 13

Table 2: Hardware requirements……………………………………………………………… 13

Table 3: Project time plan ……………………………………………………………………...14

Table 4: Budget…………………………………………………………………………………15

### LIST OF FIGURES

Figure 1: CRISP-DM diagram…………………………………………………………………. 12

Figure 2: grant chart…………………………………………………………………………….15

Figure 3: system architecture…………………………………………………………………. 15

Figure 4: Visualizing sample images…………………………………………………………. 25

Figure 5: Use Case Diagram…………………………………………………………………... 27

Figure 6: Activity Diagram………………………………………………………………….... 28

Figure 7: Logical Design………………………………………………………………………. 29

Figure 8: Pipelineck…………………………………………………………………. ………... 29

Figure 9: Testing Graph…………………………………………………………………. ……. 31

Figure 10: Home Page…………………………………………………………………………. 32

Figure 11: Login…………………………………………………………………………….…. 32

Figure 12: Leaf Upload interface.……………………………………………………………. 33

Figure 13: Analyzing Image…………………………………………………………………. 33

Figure 14: Disease Detection………………………………………………………………….34

**CHAPTER 1 :  
INTRODUCTION**

## 1.1 Background

Agriculture continues to be a central pillar in global economies, contributing to food security, employment, and sustainable livelihoods. Among the world's most extensively used crop is the potato, the world's fourth most valuable food crop behind maize, rice, and wheat. Potatoes provide essential nutrients and are cultivated in over 150 nations and feed directly or indirectly over 1.3 billion people, according to the Food and Agriculture Organization (FAO). Nevertheless, the potato sector is confronted by serious plant diseases such as Early Blight and Late Blight leading to huge crop losses worldwide. The FAO estimates that plant diseases account for 20%–40% of global crop yield loss annually, adding to food insecurity.

Potatoes are the second most important food crop in Kenya, after maize, and are a regular source of food in most households. They also serve as a vital source of income for farmers in counties like Nyandarua, Nakuru, and Meru. However, potato farming in Kenya is heavily affected by fungal diseases, with Late Blight alone responsible for losses of up to 60% in some seasons, according to the Kenya Agricultural and Livestock Research Organization (KALRO). Early detection and diagnosis of these diseases are crucial for preventing large-scale infestations. Unfortunately, most local farmers depend on visual inspection or irregular visits by agricultural officers, which are often delayed or inaccurate. This not only increases the risk of spread but also leads to excessive pesticide use, harming both the environment and human health.

Technological innovation in artificial intelligence (AI), particularly deep learning, is changing agricultural practice across the globe. Fueled by Convolutional Neural Networks (CNNs), image recognition systems yield promising results in quick and precise identification of plant diseases. These models are being integrated into mobile applications and drone technology in countries such as India, China, and the Netherlands. Kenya has low levels of adoption due to low awareness, infrastructure, and localized solutions. There is an urgent need to develop locally adaptable, AI-based tools that can empower farmers to detect diseases early and accurately, reducing crop loss and improving yields.

## 1.2 Project Overview

This project falls under the research domain of artificial intelligence in agriculture, specifically leveraging deep learning for plant disease detection through image classification. It focuses on the design and implementation of a deep learning-based system to detect potato leaf diseases—namely Early Blight, Late Blight, and Healthy leaves—through analysis of leaf images. This research seeks to build a custom Convolutional Neural Network (CNN) model that learns from labeled datasets and classifies new images with high accuracy.

Globally, deep learning has emerged as a powerful tool in plant pathology. CNNs have demonstrated superior performance in image classification tasks, such as detecting crop diseases with accuracies exceeding 95%. Several international projects have explored this concept using plant datasets such as PlantVillage. These models are being integrated into mobile and web applications that farmers use in real-time.

In the Kenyan context, the integration of such technology in smallholder farming is still at a nascent stage. Farmers often lack access to affordable diagnostic tools or expert agronomists. This project, therefore, seeks to bridge the gap by designing a model tailored to the local environment and common diseases affecting Kenyan potato farmers.

From a computational standpoint, the key principles underpinning this project include image preprocessing (e.g., resizing, augmentation), supervised learning via CNNs, and performance evaluation using metrics such as accuracy, precision, and recall. Python and libraries like TensorFlow or PyTorch will be used to implement and train the model. The CRISP-DM methodology will guide the research lifecycle.

## 1.3 Statement of the Problem

Potato yields in Kenya are badly affected by diseases such as Early Blight and Late Blight, which, if not diagnosed early, cause enormous yield loss. The prominent method of disease diagnosis used by farmers currently is visual inspection, which is subjective, prone to errors, and typically too late to prevent the spread. Additionally, access to expert agronomists is limited in rural areas, leaving most farmers without reliable guidance. According to KALRO, fungal diseases account for an estimated 60%–70% of all potato crop losses in Kenya.

The unavailability of affordable and reliable diagnostic equipment is one of the reasons causing fungicide overuse, rising production costs, and environmental pollution. Elsewhere in the world, agricultural systems are increasingly adopting AI-based disease identification models, allowing for early diagnosis and cheap interventions. Kenya lags in this digital era, particularly in applying scalable and user-friendly disease identification systems.

This research problem centers on the absence of a robust, locally adaptable AI model that can assist Kenyan potato farmers in early disease detection using leaf image analysis. The proposed solution aims to fill this gap by applying deep learning techniques to automate disease detection with accuracy and efficiency. This goes beyond merely digitizing an existing manual process; it introduces intelligent, data-driven capabilities to support decision-making in farming.

## 1.4 Proposed Solution

This research seeks to develop a deep learning model that can accurately detect diseases in potato leaves, specifically focusing on Early Blight, Late Blight, and healthy conditions. The model will be trained using a labeled dataset of potato leaf images and will be capable of classifying new images based on features it has learned during training. The aim is not only to build the model but also to critically assess its performance and compare it with similar models developed around the world.

The key operations expected from the system include:

* Preprocessing images through resizing, normalization, and augmentation to improve training quality.
* Designing and training a convolutional neural network (CNN) tailored for classifying leaf images.
* Evaluating the model using metrics such as accuracy, precision, recall, and F1-score to determine its effectiveness.
* (Optionally) Developing a simple user interface that could be used in the field to test the model’s practical usability.

This solution builds on the latest advances in artificial intelligence and computer vision, particularly in the area of CNNs. Unlike older systems that focus merely on digitizing manual processes, this research emphasizes intelligent automation. The model will be benchmarked against global solutions to ensure it meets high performance standards while being adaptable to local agricultural contexts.

## 1.5 Objectives

**General Objective**  
To develop a deep learning model for detecting potato leaf diseases to support early diagnosis and improve crop management practices among Kenyan farmers.

**Specific Objectives**

1. To review and analyze current deep learning models used globally in plant disease detection.
2. To design and implement a Convolutional Neural Network (CNN) model for classifying potato leaf diseases.
3. To train and evaluate the model using labeled image datasets.
4. To assess the performance of the model using classification metrics such as accuracy and recall.

## 1.6 Research Questions

1. What deep learning models are most effective in detecting plant diseases through image classification?
2. How can a CNN be designed to detect and classify potato leaf diseases accurately?
3. What is the performance of the trained model in identifying Early Blight, Late Blight, and Healthy leaves?
4. How can image preprocessing techniques improve the accuracy and generalizability of the model?

## 1.7 Justification

This research is crucial as it addresses a real and pressing issue faced by thousands of potato farmers in Kenya—delayed and inaccurate disease detection. The proposed deep learning model offers a timely solution by leveraging emerging technology to solve agricultural challenges. By focusing on a crop that significantly contributes to food security and rural income, the research supports national goals such as the Big Four Agenda on food security.

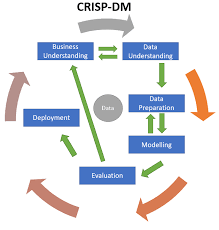
The project contributes to the growing field of agricultural informatics and expands on current AI applications in developing countries. It offers a scalable, affordable, and data-driven tool that, if adopted widely, can improve yield, reduce costs, and support sustainable farming. Furthermore, it advances research in AI by contextualizing models for low-resource environments.

## 1.8 Proposed Research and System Methodologies

**Research Methodology:**  
The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology will be used, encompassing the following phases:

* Business Understanding
* Data Understanding
* Data Preparation
* Modeling
* Evaluation
* Deployment (prototype-level)

**Figure 1: CRISP-DM**



**System Implementation Tools and Techniques:**

* **Programming Language:** Python
* **Libraries:** TensorFlow/Keras or PyTorch for model building, OpenCV for image handling
* **IDE:** Google Colab / Jupyter Notebook
* **Dataset:** PlantVillage or custom-collected images
* **Justification:** These tools are open-source, widely used, and suitable for deep learning applications. CRISP-DM offers a structured, iterative framework ideal for handling the data-centric nature of the project.

## 1.9 Scope

This project will focus on the classification of potato leaf images into Early Blight, Late Blight, and Healthy categories using a deep learning model. The primary geographical focus is Kenya, with the model tailored for use by smallholder farmers and agricultural officers.

The scope excludes other disease types not present in the dataset and does not address soil or environmental factors contributing to disease. Limitations may include dataset imbalance, limited computational resources, and lack of real-time deployment infrastructure. The project will focus on offline model development and testing, with any deployment considerations being secondary.

## 1.10 Requirements and Budgeting

Creating a budget involves considering various factors such as development costs, software licenses, hardware requirements, ongoing maintenance, and potential additional resources.

### 1.10.1Software requirements

Table 1: SOFTWARE requirements

|  |  |
| --- | --- |
| Software | Pricing |
| Smart contract | Free |
| Programming Languages (solidity ,JavaScript) | Free |
| UML Design Tools | Free |
| Antivirus Software – Kaspersky | free |
| Operating systems(windows) | Free |

### 1.10.2 Hardware Requirements

Table 2: HARDWARE requirements

|  |  |
| --- | --- |
| Hardware | Pricing |
| Laptop  Core i3 and above  Internal Hard disk drive – 500GB  Processor Speed – 2.5GHz  RAM – 4GB and above | 40000(Provided) |
| Total |  |

## Budget justification

1. Computer Device (Laptop): A computer device is necessary for the development and testing of the system. It should meet the minimum specifications required by the development tools and software used in the project, such as a processor with sufficient processing power, an adequate amount of memory, and ample storage capacity.
2. Operating System: The system will be developed on windows operating system.
3. Mobile phone: A mobile phone can be used to take pictures of the infected leaf and upload directly to the website.

## ****1.10.3 Project Timeline****

The project is structured to follow a systematic schedule over a period of 14 weeks. Each phase focuses on a key deliverable aligned with the CRISP-DM methodology and system development process.

## ****TABLE 3: PROJECT TIME PLAN****

|  |  |  |  |
| --- | --- | --- | --- |
| **Task** | **Start Date** | **End Date** | **Duration In Days** |
| Define Research Problem | **29/9/2024** | **30/9/2024** | **1** |
| **Project Overview** | **10/10/2024** | **11/10/2024** | **1** |
| **Statement of the problem** | **12/10/2024** | **17/10/2024** | **5** |
| **Proposed Solution** | **18/10/2024** | **25/10/2024** | **6** |
| **Objective** | **25/10/2024** | **5/11/2024** | **9** |
| **Research Question** | **6/11/2024** | **6/12/2024** | **1** |
| **Justification** | **7/11/2024** | **22/11/2024** | **21** |
| **Proposed Research System Methodology** | **22/11/2024** | **27/11/2024** | **5** |
| **Scope** | **27/11/2024** | **29/11/2024** | **2** |
| **Theoretical and Case Study Review** | **29/11/2024** | **7/12/2024** | **8** |
| **Intergration and architecture** | **8/12/2024** | **11/12/2024**   |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  | | **3** |
| **Summary and Research Gaps** | **12/12/2024** | **17/12/2024** | **5** |

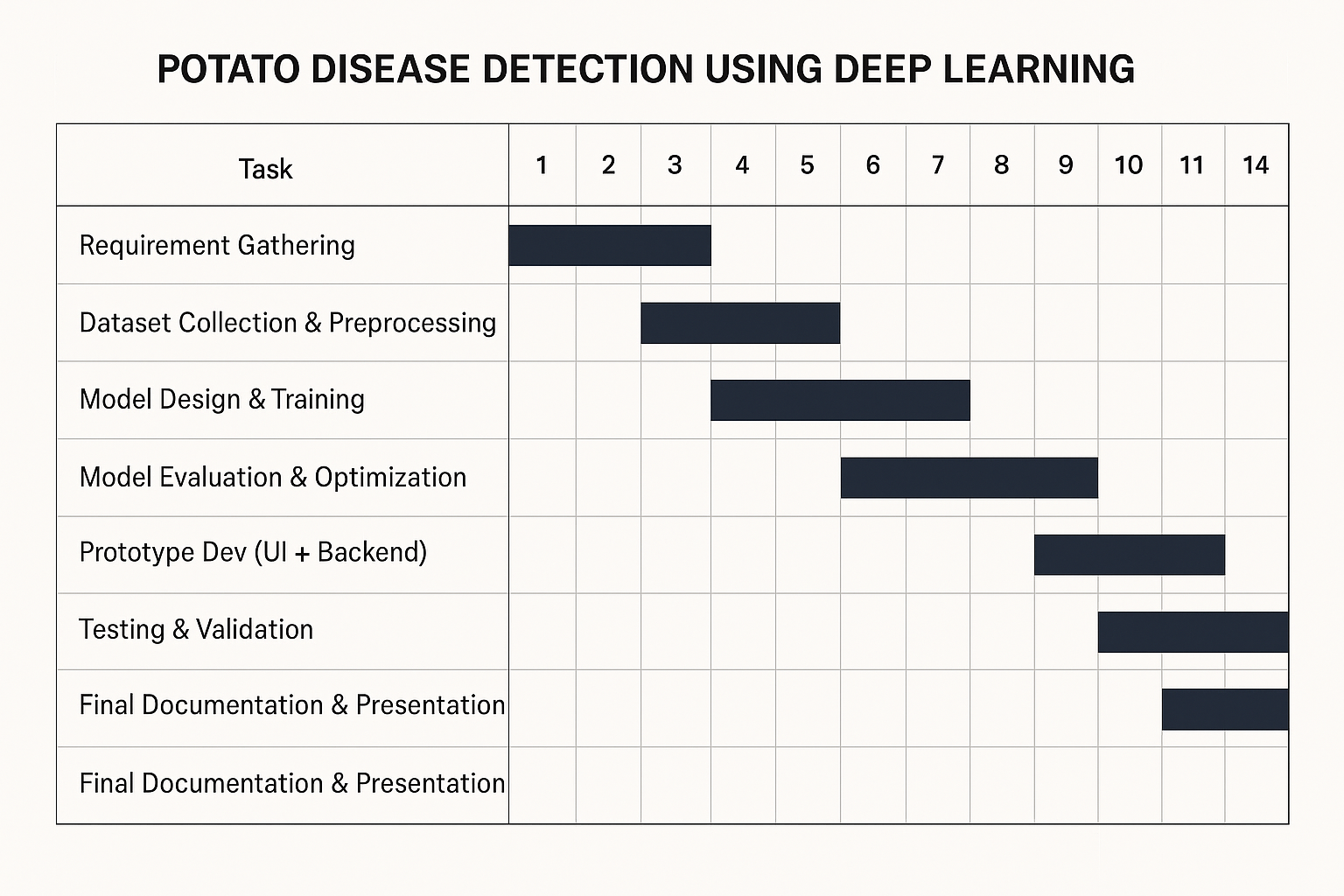
### ****1.10.4 Project Cost Estimates****

The estimated cost of the project remains minimal due to the reliance on open-source tools and publicly available datasets. However, minor costs are projected for deployment and miscellaneous requirements.

Table 4: BUDGET

| **Item** | **Cost (Estimated)** |
| --- | --- |
| Dataset (PlantVillage) | $0 |
| Software & Libraries (TensorFlow, OpenCV, etc.) | $0 |
| Hosting (for prototype deployment) | $30 |
| Miscellaneous (e.g., internet, storage) | $30 |
| **Total Estimated Cost** | $60 |

**Gantt Chart**



**Figure 2: Gantt Chart**

# **CHAPTER 2: LITERATURE REVIEW**

## 2.1 Introduction

This chapter reviews the literature surrounding the use of deep learning techniques for detecting plant diseases, particularly focusing on the use of Convolutional Neural Networks (CNNs) for the identification of potato leaf diseases. The application of artificial intelligence (AI) and machine learning (ML) in agriculture has gained significant attention in recent years, particularly for disease detection. This relevance is heightened by the expanding impact of plant diseases on crop yields and food security. With technology continuing to advance, deep learning has become a viable way of automating disease diagnosis, offering solutions that can scale with the increasing need for precision agriculture.

The objective of this literature review is to provide a comprehensive overview of the current research on deep learning-based plant disease detection, namely CNNs, and to explore how these technologies can be applied to potato disease detection. The review will cover key concepts, theoretical frameworks, case studies, and integration strategies, ultimately highlighting the research gaps that this study seeks to address.

## 2.2 Theoretical Review

**Key Concepts and Variables**

A number of key concepts form the basis of deep learning research in plant disease detection. These are:

•Data: Deep models require a lot of data to learn effectively. In detecting plant diseases, these data are labeled images of plant leaves by disease type (e.g., Early Blight, Late Blight, Healthy). The quantity and diversity of such data play a very important role in the ability of the model to generalize across environments.

• Features and Labels: Features refer to the physical attributes in an image that the model uses to classify, such as leaf texture, color, and shape. Labels refer to the class that each image is assigned based on the disease it possesses, such as "Early Blight" or "Healthy." The model learns to identify these features with their respective labels.

• Algorithms: Convolutional Neural Networks (CNNs) are the primary deep learning algorithms used in plant disease detection. CNNs are particularly effective in image classification tasks because they can automatically extract relevant features from raw pixel data, reducing the need for manual feature extraction.

• Training: Deep learning models are trained using an algorithm called backpropagation, in which the model's weights are optimized based on the gap between actual and predicted labels. Through this back-and-forth exchange, the model learns to make accurate predictions.

**Theoretical Divisions in Plant Disease Detection**

Within plant disease detection using deep learning, there are different schools of thought that guide the application of these techniques:

• Supervised Learning: This is the most common approach in plant disease detection. In supervised learning, the model is trained on a labeled dataset, where each input image is associated with a known disease category. Supervised learning algorithms are said to provide high accuracy if sufficient data are provided, but they require a large amount of labeled data, which can be resource-intensive to procure.

Pros: Highly accurate, suitable for applications where labeled data is present.

Cons: Relying on extensive labeled data, which may be expensive and time-consuming to source.

• Unsupervised Learning: Contrary to supervised learning, unsupervised learning does not need to be based on labeled data. Instead, the model identifies patterns in the data autonomously. While promising, this approach is used less often in plant disease identification due to the need for specific classification.

Pros: No labeled data is needed.

Cons: It is challenging to achieve high classification accuracy since there are no labels, thus less suitable for disease classification.

• Transfer Learning: Transfer learning refers to the practice of taking a pre-trained model and applying it to a big dataset (e.g., ImageNet) and then fine-tuning it to a specific task (e.g., plant disease detection). This is a useful method when data available in the domain is limited. This method allows the model to leverage information from the general dataset to improve performance.

Advantages: Reduces the quantity of large datasets needed, saves time, and can potentially deploy the model sooner.

Limitations: Can sometimes not be best-fitted for domain-related tasks and may need to undergo significant fine-tuning.

**Advantages and Limitations**

* Supervised Learning: It primarily generates good models but is highly reliant on large labeled datasets. If the dataset is small or imbalanced, the model's performance could worsen.
* Unsupervised Learning: The advantage of this approach is that it does not require labeled data, but its lack of labeled examples makes it harder to apply in plant disease diagnosis, where accurate diagnoses are necessary.
* Transfer Learning: Transfer learning has the advantage of employing existing models, leading to quicker training and better performance on small datasets. However, the transferability of features learned from general datasets to specific plant disease tasks may not always be ideal.

## ****2.3 Case Study Review****

Several studies have applied deep learning techniques to detect plant diseases, with varying degrees of success. Key case studies include:

* **PlantVillage Dataset**: A significant study using the PlantVillage dataset, which contains over 50,000 images of different plant diseases, applied CNNs to identify disease types in various crops. The CNN model achieved high classification accuracy but faced challenges such as imbalanced datasets, which impacted the model's performance on underrepresented diseases like Late Blight.

Key Success: High accuracy in detecting common diseases such as Early Blight, providing valuable insights for large-scale agricultural use.

Challenges: The model struggled with overfitting on the less represented classes, highlighting the need for better data balancing techniques.

* **Tomato Disease Detection**: A study on tomato disease detection used deep learning models to classify images of tomato leaves affected by diseases such as Early Blight and Late Blight. The model performed well in a controlled environment but faced difficulties when applied to real-world data due to variations in lighting conditions and background noise.

Key Success: Effective detection of tomato diseases under controlled conditions. Challenges: The model's performance decreased in outdoor settings, suggesting the need for data preprocessing to handle environmental variations.

* **Maize Disease Detection**: In India, deep learning models were applied to detect maize diseases such as Leaf Blight and Rust. While the models showed promise, their performance deteriorated when exposed to varying environmental conditions like changes in soil type or plant variety.

Key Success: The deep learning model demonstrated high accuracy in controlled conditions.

Challenges: The need for a more adaptable model that can handle diverse environmental conditions.

## ****2.4 Integration and Architecture****

The incorporation of deep learning models into agriculture systems requires data acquisition, processing, and deployment methods in mind. The integration options are:

• Data Acquisition: Images can be captured through mobile phones, drones, or fixed cameras to detect plant diseases. Drones are employed for capturing high-resolution images from different directions, which can detect slight symptoms of disease.

• Model Deployment: After training, the model can be deployed on mobile apps or cloud platforms, where farmers are able to upload images for real-time disease diagnosis. Such integration enables farmers to receive actionable feedback and suggestions on disease management practices.

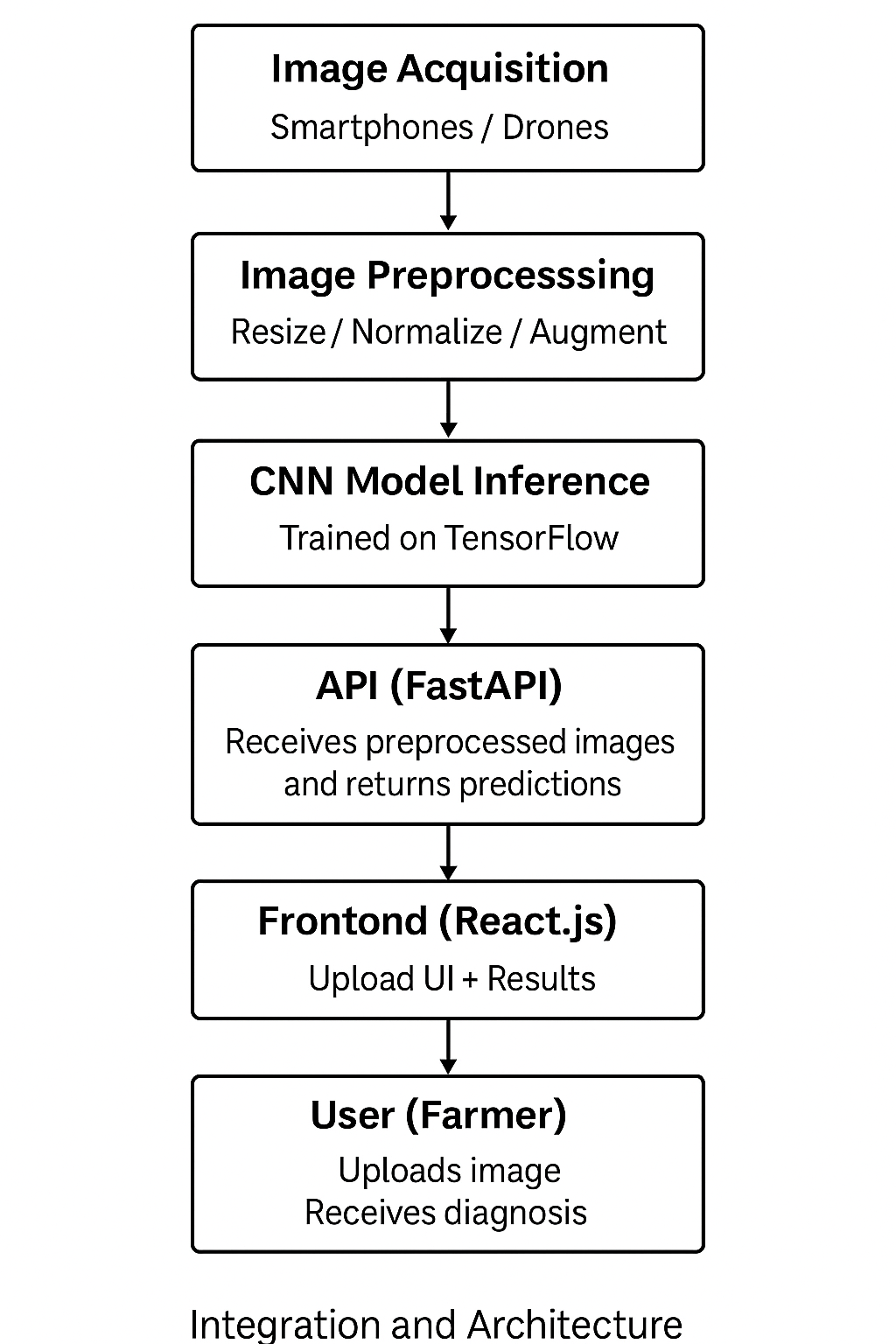
System Structure: A typical deep learning system for plant disease detection comprises:

1. Image Acquisition: Capture images using smartphones or drones.

2. Preprocessing: Images are preprocessed through resizing, normalization, and data augmentation.

3. Model Inference: Trained CNN model infers the class of the images and identifies the disease.

4. User Interface: Simple-to-use interface provides feedback and suggestions for disease management.



**Figure 3: 2.4 System Integration and Architecture**

## ****2.5 Summary****

The literature reviewed demonstrates greater potential in plant disease identification using deep learning, particularly in agriculture. Convolutional Neural Networks (CNNs) have been successful in plant disease classification, with significant improvement compared to traditional methods. Nevertheless, the research also identifies some challenges that involve high data requirements, generalization to practical environments, and data imbalance.

## ****2.6 Research Gaps****

Despite the promising advances, several research gaps remain:

1. Environmental Variability: The deep learning models developed in controlled settings tend to perform poorly under real-world conditions. Model robustness to environmental conditions like lighting and background noise needs to be enhanced through research.

2. Data Imbalance: The majority of plant disease datasets are imbalanced, and some diseases are underrepresented. Future work must explore methods for data imbalance handling to improve model accuracy across all classes of diseases.

3. User Accessibility: The majority of solutions today are not merely accessible to the farmer, particularly in areas with limited resources. More affordable, user-friendly solutions that are ready to roll-out on universally accessed mobile phones need to be developed.

This research seeks to bridge these gaps using a heterogeneous dataset, implementing advanced data augmentation techniques, and developing a mobile application that provides farmers with real-time disease diagnosis and management advice.

# **CHAPTER 3** **SYSTEM ANALYSIS AND DESIGN**

## ****3.1 Introduction****

This chapter provides the general overview of the system analysis and design process for the proposed research project. It includes the methodologies, feasibility studies, requirements elicitation, data analysis, and detailed system specifications. It also provides the design and modeling of the system with emphasis on both logical and physical designs. The chapter outlines how the system will be built and configured to meet the specific needs of the project and how it will integrate into the existing environment.

The aim of this chapter is to establish a clear and practical approach to system design, beginning from the planning and requirement gathering phase to the final stages of physical implementation. By breaking down the project into manageable sections, this chapter ensures that each component of the system will work together to address the identified problem effectively.

## ****3.2 Systems Development Methodology****

The **Cross-Industry Standard Process for Data Mining (CRISP-DM)** methodology is adopted for this project due to its structured approach in machine learning development. The key phases include:

* **Business Understanding**: Potato leaf diseases, such as Early Blight and Late Blight, cause significant agricultural losses, particularly in regions where potatoes are a major food crop. Early and accurate disease detection is critical for timely intervention and minimizing yield loss. Traditional diagnostic methods rely on expert inspection, which can be time-consuming, expensive, and sometimes inaccurate.
* **Data Understanding** The dataset used for this project was sourced from the PlantVillage dataset available on Kaggle. It contains approximately 2176 labeled images of potato leaves categorized as "Healthy," "Early Blight," or "Late Blight." The dataset is relatively balanced and was organized for ease of use with TensorFlow’s Image Libraries.
* **Data Preparation**: The images undergo preprocessing steps, including resizing, normalization, augmentation, and splitting into training, validation, and test sets to enhance model performance.
* **Modeling**: Different deep learning architectures, including Convolutional Neural Networks (CNNs), are trained and evaluated to determine the most effective model for potato leaf diseases classification.
* **Evaluation**: The trained model is assessed using standard performance metrics.
* **Deployment**: The system's backend was developed using FastAPI, and the trained model was deployed using Hugging Face's. The frontend, developed using React, enables users to upload potato leaf images and instantly receive classification results via a user-friendly web application.

## ****3.3 Feasibility Study****

A detailed feasibility study was conducted to ensure that the proposed system is viable across four critical dimensions: **economic**, **technical**, **operational**, and **legal**.

* **Economic Feasibility**:  
  The project is economically feasible, primarily because it leverages open-source technologies that reduce licensing and operational costs. The use of cloud-based infrastructure also eliminates the need for expensive physical hardware, making it cost-effective. A cost-benefit analysis suggests that the anticipated benefits, such as improved efficiency and user satisfaction, will outweigh the costs incurred during development. The project also has the potential for future scalability, which can provide an extended return on investment (ROI).
* **Technical Feasibility**:  
  On the technical front, the system is feasible due to the wide availability and support for the tools and technologies being used. The chosen technologies—such as cloud computing platforms, relational databases, and open-source frameworks—are reliable and well-suited to handle the system's requirements. Furthermore, the development team is equipped with the necessary skills and experience to implement these technologies successfully.
* **Operational Feasibility**:  
  Operationally, the system is highly feasible because it is designed to integrate with existing workflows and processes. The user interface is intuitive, and the functionality is closely aligned with the needs of the stakeholders. Furthermore, comprehensive training will be provided to end-users to ensure smooth adoption. Operational feasibility is supported by the inclusion of continuous monitoring tools and user support services to handle potential issues post-implementation.
* **Legal and Compliance Feasibility**:  
  The dataset is publicly available, ensuring compliance with data privacy regulations.

## ****3.4 Requirements Elicitation****

**Data Collection**

The dataset used for training and evaluation in this project is publicly available on **Kaggle**. It contains approximately **2,176** labeled images of potato leaves categorized into three classes: *Healthy*, *Early Blight*, and *Late Blight*. The images are high-quality and curated to support machine learning research in potato disease detection.

The dataset was collected from controlled environments to ensure consistency in image quality and labeling. It provides a sufficient number of samples for effective model training, validation, and evaluation.

To optimize the model’s learning process and ensure reliable performance, the dataset was split into three subsets:

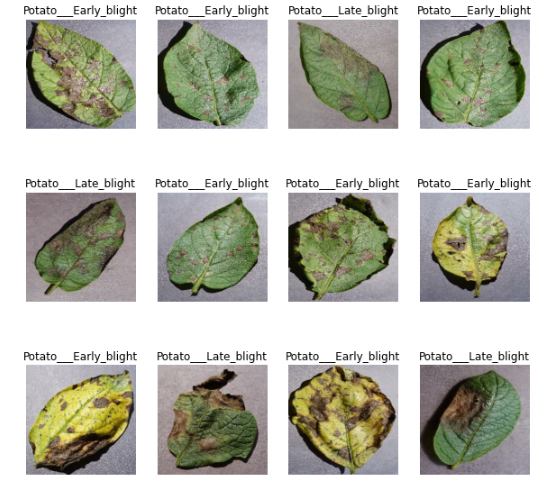
* **Training Set:** 80% of the data
* **Validation Set:** 10% of the data
* **Testing Set:** 10% of the data

This structured split allows for efficient model optimization and accurate evaluation on unseen data.

## ****3.5 Data Analysis****

The dataset is analyzed to understand its distribution and characteristics. The steps include:

* **Data Cleaning**: Removing corrupted images and ensuring proper labeling.
* **Exploratory Data Analysis (EDA)**: Visualizing sample images;



**Figure 4: Visualizing sample images**

* **Feature Extraction**: Extracting important patterns using CNN architectures.
* **Splitting the Data**: Dividing the dataset into training (80%), validation (10%), and testing (10%) sets.

.

## ****3.6 System Specification****

The **system specifications** are divided into **functional** and **non-functional** requirements, which provide a comprehensive outline of what the system needs to accomplish.

* **Functional Requirements**:
  1. The system must support **user registration and authentication**, with a secure login mechanism.
  2. It must allow **data input and retrieval** from the database, with real-time processing capabilities.
  3. The system should generate **dynamic reports** based on the data processed.
  4. It must allow users to access the system from different devices, ensuring **cross-platform compatibility**.
* **Non-functional Requirements**:
  1. **Scalability**: The system should be able to handle increasing data loads and user numbers over time.
  2. **Performance**: Response time for user queries should not exceed 5 seconds under normal usage.
  3. **Security**: Data should be encrypted both during transmission and storage to prevent unauthorized access.
  4. **Usability**: The system interface must be user-friendly, with clear navigation and minimal learning curve.
  5. **Availability**: The system should be highly available, with minimal downtime for maintenance.

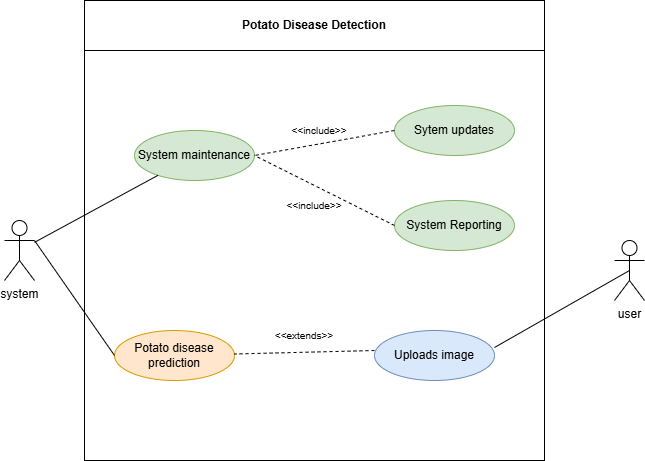
## 3.7 Requirements Analysis and Modeling

**Use Case Diagram**

System (Admin/System actor): Represents backend operations or administrators who manage and maintain the system.

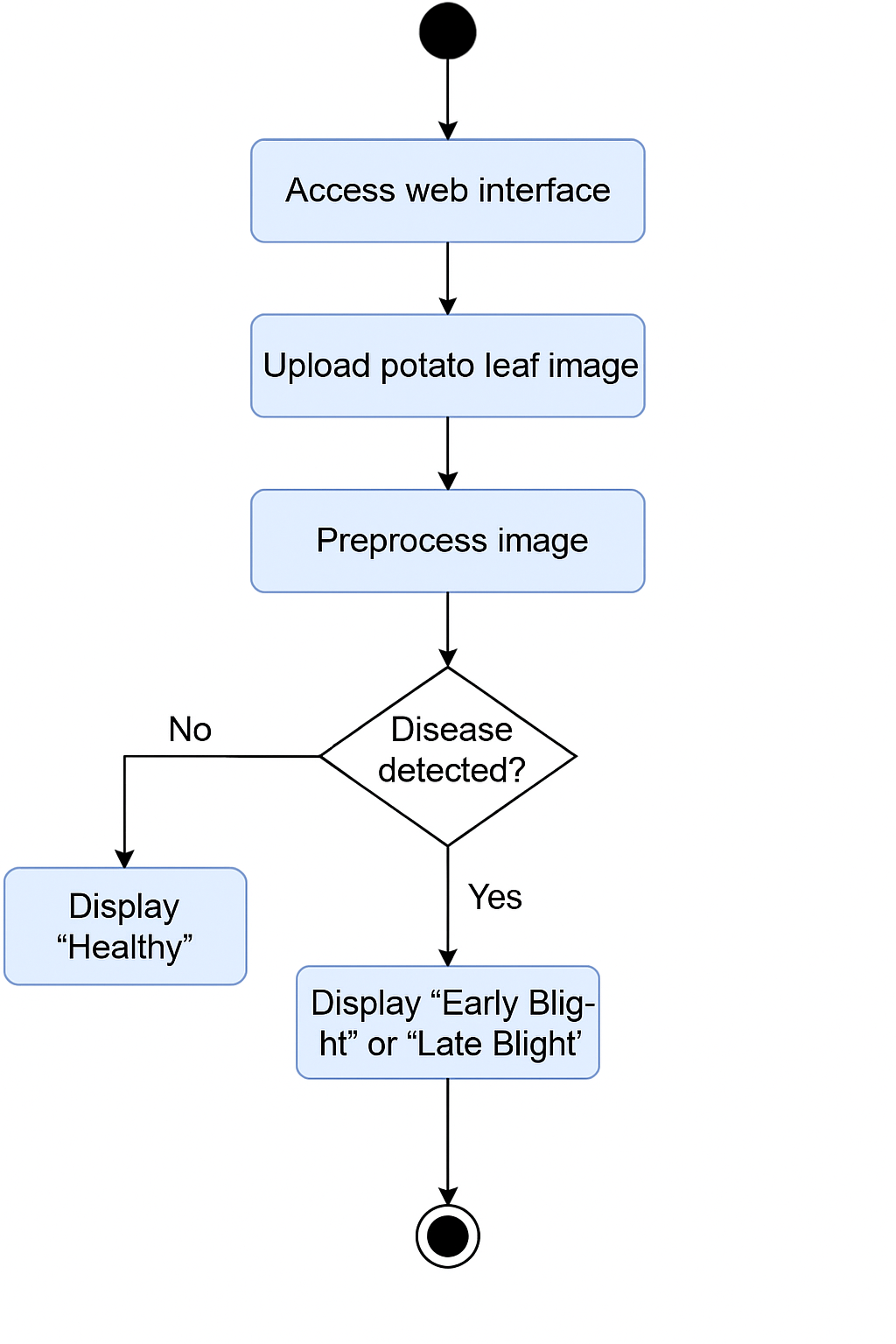
User: Represents end-users (e.g., farmers or agricultural officers) who interact with the system to detect potato diseases.

1. **System Maintenance**:
   * A primary administrative function, responsible for ensuring the system runs correctly.
   * **Includes**:
     + **System Updates**: Regular updates to software, model improvements, or security patches.
     + **System Reporting**: Generation of logs, performance metrics, or user activity reports.
2. **Potato Disease Prediction**:
   * Core system functionality triggered by user input (specifically image upload).
   * **Extends**:
     + **Uploads Image**: This is a prerequisite action — the user must upload an image before disease prediction can occur.
3. **Uploads Image***:*
   * The main interaction from the **user** side. This action initiates the disease detection process.
   * This use case is **extended** by Potato Disease Prediction because image upload is essential for prediction to happen.



**Figure 5: Use Case Diagram**

**Activity diagram**

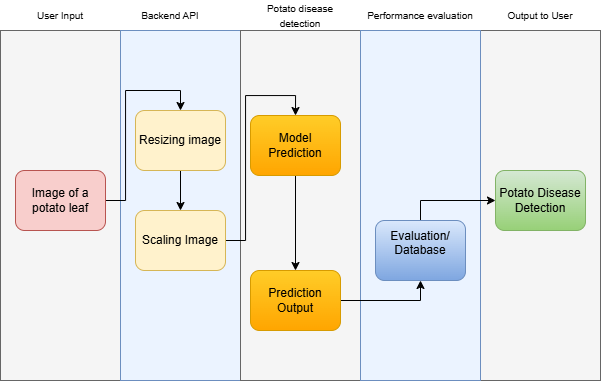


**Figure 6: Activity Diagram**

1. Start: The user initiates the process by accessing the web interface.
2. Upload Image: The user uploads a potato leaf image to the system.
3. Preprocessing: The image is resized, normalized, and augmented.
4. Decision Point – Disease Detected?
5. If No: The system classifies the leaf as “Healthy”.
6. If Yes: The system classifies it as “Early Blight” or “Late Blight”.
7. Display Result: The result is shown to the user.
8. End: The process concludes.

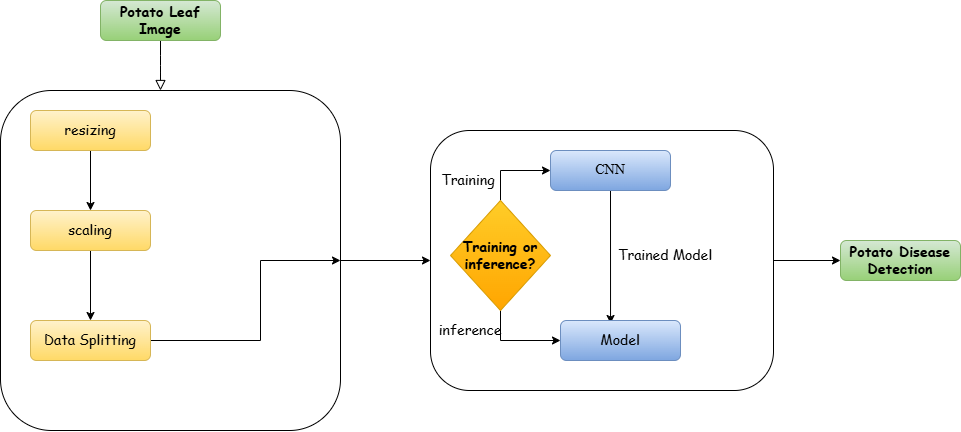
**3.8 Logical Design**

**Overview of the proposed Machine Learning Model**



**Figure 7: Logical Design**

**Pipeline**



**Figure 8: Pipeline**

# **CHAPTER 4**

# **SYSTEM IMPLEMENTATION AND TESTING, CONCLUSIONS, AND RECOMMENDATIONS**

## 4.1 Introduction

This chapter presents the implementation and testing processes of the system developed as part of this research project. It covers the environment and tools used in the development, the process of system code generation, the testing strategies employed, the user guide, and concludes with an evaluation of the system's success in addressing the client’s problem. Furthermore, the chapter discusses the limitations of the system, challenges faced during the study, and provides recommendations for future improvements.

## 4.2 Implementation Environment and Tools

* **Programming Languages**: Python (backend), JavaScript (frontend)
* **Frameworks**: FastAPI for the backend API, React for UI
* **Libraries**: TensorFlow, Pandas, Matplotlib for deep learning and image processing
* **Cloud Deployment**: Hugging Face for model hosting, Github for web application deployment

## 4.3 System Code Generation

Key implementation components include:

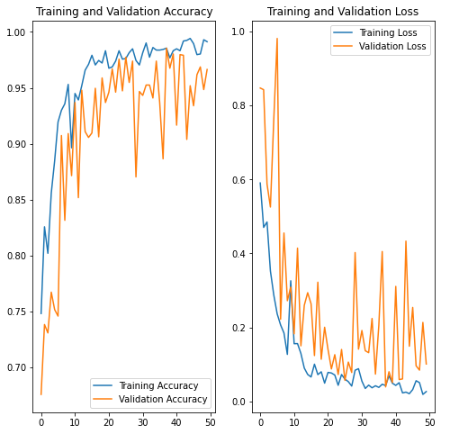
* **Data Preprocessing Module**: Cleans and processes potato leaf images, applying resizing, normalization, and augmentation techniques.
* **ML Model Module**: Loads the trained CNN model
* **API Module**: Handles communication between the frontend and backend using RESTful APIs.
* **User Interface Module**: Provides web-based user interaction, allowing users to upload images and receive classification results.

Link to the website: <https://peter-kanyi-project.github.io/potato_classification/>

## 4.4 Testing

Following the training phase, the system underwent extensive testing to evaluate its ability to accurately detect potato leaf diseases. The model was tested using an unseen dataset and assessed using standard performance metrics, including accuracy, precision, recall, and F1-score.

The model achieved an impressive accuracy of 96%, demonstrating its high effectiveness in distinguishing between “Healthy,” “Early Blight,” and “Late Blight” categories.

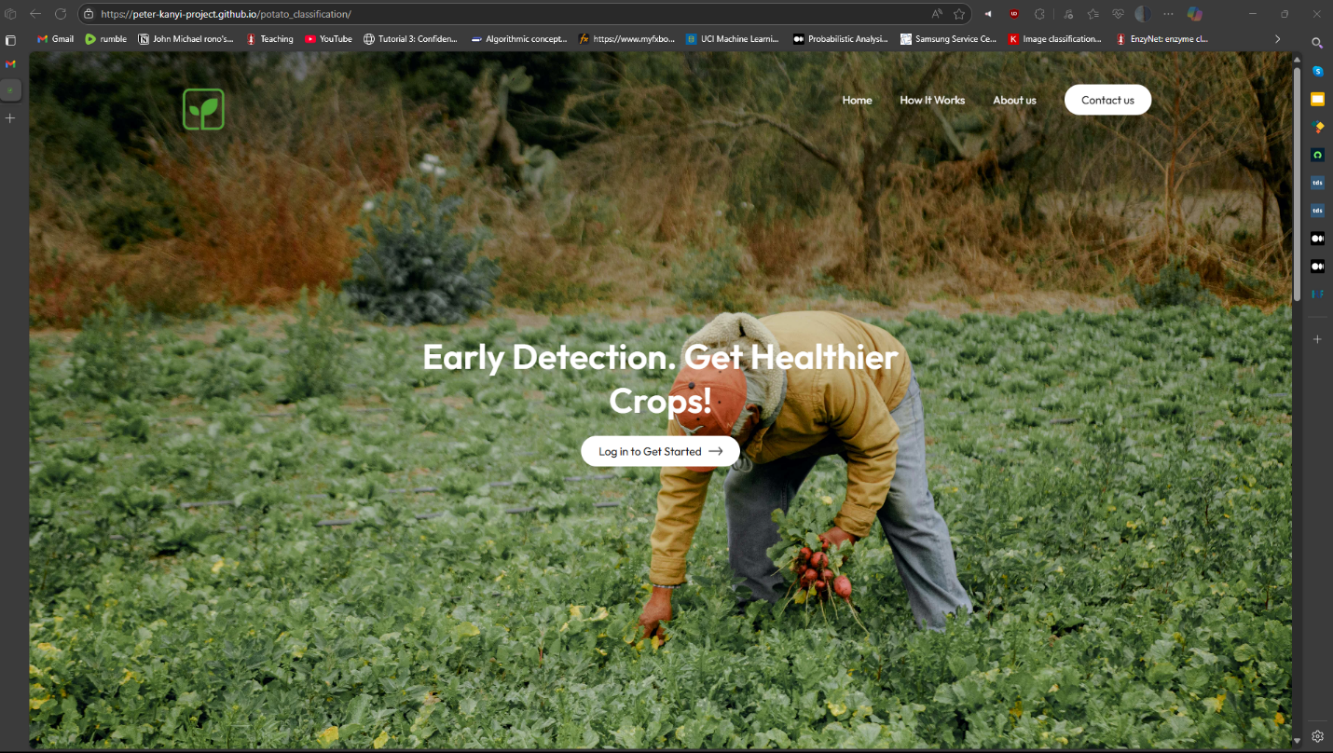


**Figure 9: Testing Graph**

## 4.5 User Guide

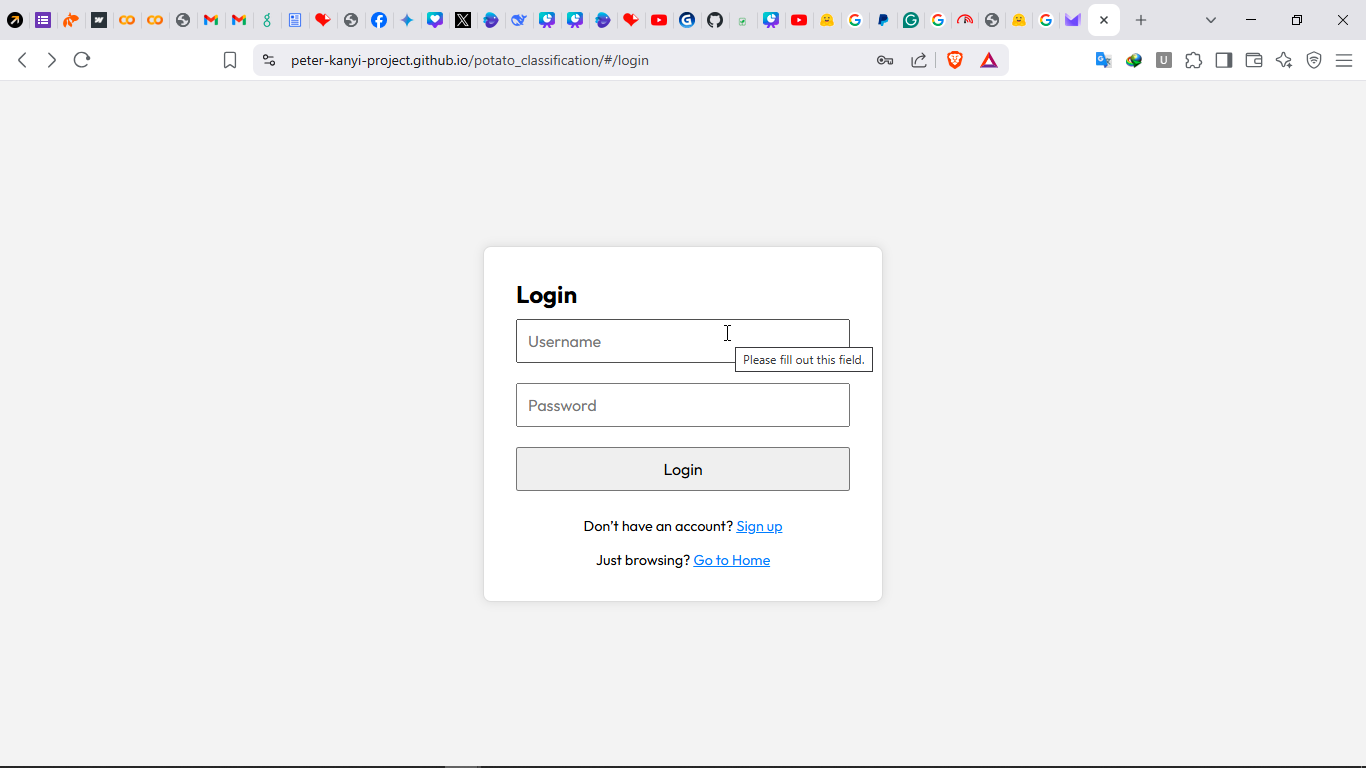
The system is accessible via a web-based platform. Users can follow these steps to utilize the potato leaf disease detection system:

1. Go to the [**website**](https://peter-kanyi-project.github.io/potato_classification/)



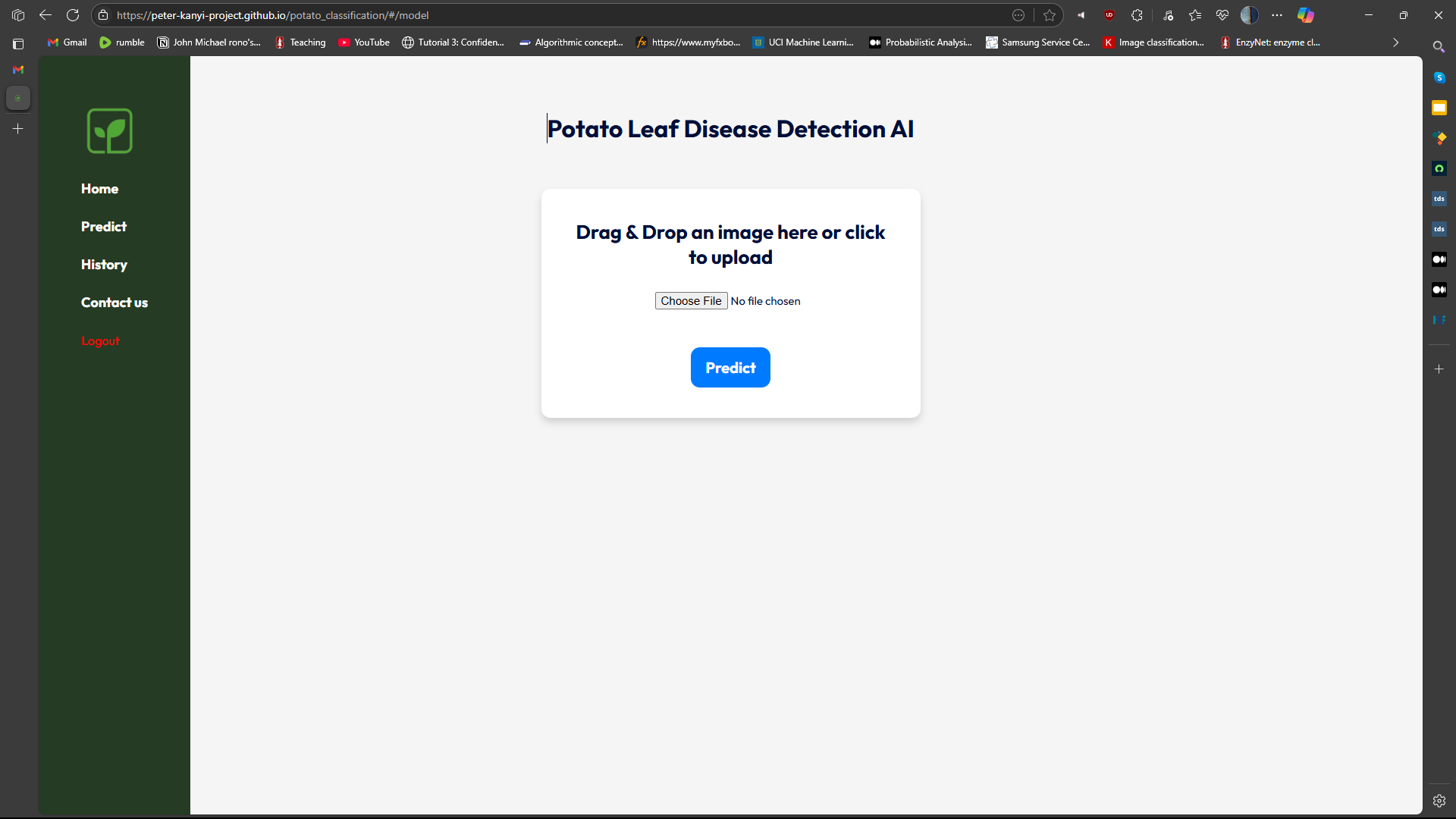
**Figure 10: Home page**

1. Login into the account or create a new profile



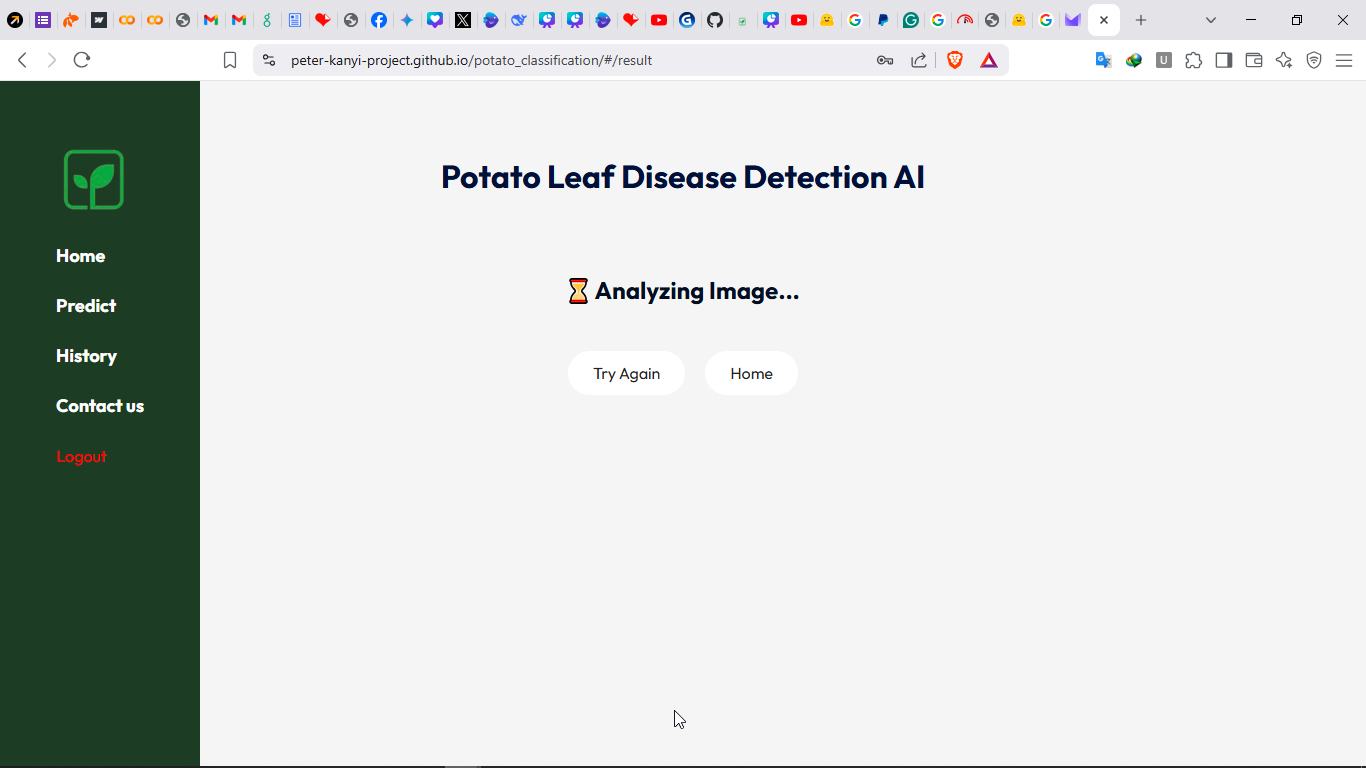
**Figure 11: Login**

1. **Click the "Upload Image" button** to select a potato leaf image.



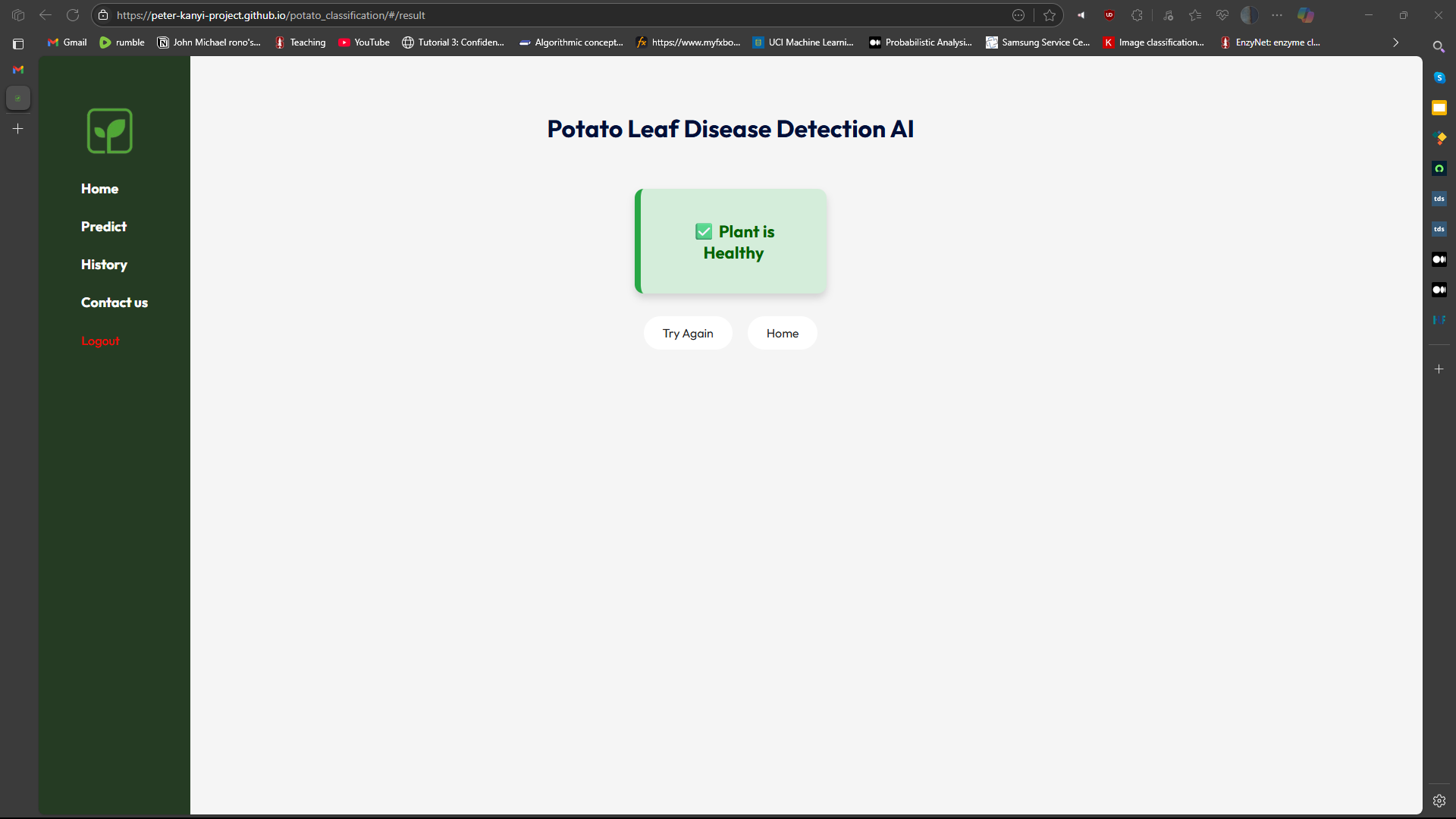
**Figure 12: Leaf Upload interface**

1. **Wait for processing** as the image is analyzed using the deep learning model.



**Figure 13: Analyzing Image**

1. **Receive results** displaying whether the sample is “Healthy”, has “Early Blight” or has “Late Blight”



**Figure 14: Disease Detection**

## 4.6 Conclusions

The developed system successfully classifies potato leaf images into the categories of "Healthy," "Early Blight," and "Late Blight" with high accuracy and user-friendliness. The use of deep learning, combined with modern web technologies, provides an effective solution for early detection of potato leaf diseases, which can significantly aid farmers and agricultural experts.

However, to further enhance the system’s applicability, it is recommended that the model be tested with larger and more diverse datasets, including real-world field images under varying environmental conditions. Additionally, integration with agricultural advisory platforms could broaden the system’s impact and usefulness in practical farming scenario

# **REFERENCES**

1. FAO. (2023). *Home*. PRD-FAO Home. <https://www.fao.org/home/en/>
2. Agricultural, K. (2021). *Kenya Agricultural and Livestock Research Organization*. Kalro.org. <https://www.kalro.org/>
3. *PlantVillage*. (n.d.). Plantvillage.psu.edu. <https://plantvillage.psu.edu/>
4. Roselyne Akoth. (2024, September 18). *Potato Farming in Kenya: Comprehensive Guide*. Farming in Kenya | Farming in Kenya. <https://farminginkenya.co.ke/potato-farming-in-kenya-a-comprehensive-guide/>
5. Botero-Valencia, J., García-Pineda, V., Valencia-Arias, A., Valencia, J., Reyes-Vera, E., Mateo Mejia-Herrera, & Ruber Hernández-García. (2025). Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives. *Agriculture*, *15*(4), 377–377. <https://doi.org/10.3390/agriculture15040377>
6. Hotz, N. (2024, December 9). *What Is CRISP DM?* Data Science Project Management. <https://www.datascience-pm.com/crisp-dm-2/>
7. Agong, S., Mwangi, M., Kahuthia-Gathu, R., & Waceke, W. (2021). POTATO PRODUCTION PRACTICES AND LATE BLIGHT MANAGEMENT IN NYANDARUA COUNTY, KENYA. *Journal of Agricultural, Food and Environmental Sciences*, *75*(2), 28–36. <https://doi.org/10.55302/jafes21752028a>
8. *Using Artificial Intelligence (AI) for Potato Disease Detection in Kenya - Centre for Intellectual Property and Information Technology law*. (2024, September 13). Centre for Intellectual Property and Information Technology Law - Centre for Intellectual Property and Information Technology Law. <https://cipit.strathmore.edu/using-artificial-intelligence-ai-for-potato-disease-detection-in-kenya/>
9. Alhammad, S. M., Doaa Sami Khafaga, El-hady, W. M., Samy, F. M., & Hosny, K. M. (2025). Deep learning and explainable AI for classification of potato leaf diseases. *Frontiers in Artificial Intelligence*, *7*. <https://doi.org/10.3389/frai.2024.1449329>
10. Keshika Jangde. (2024). POTATO LEAF DISEASE DETECTION AND CLASSIFICATION USING CNN. *Journal of Nonlinear Analysis and Optimization*, *Vol. 15*(Issue. 01, No. 16), 57–62. <https://www.researchgate.net/publication/389600786_POTATO_LEAF_DISEASE_DETECTION_AND_CLASSIFICATION_USING_CNN>
11. Catal Reis, H., & Turk, V. (2024). Potato leaf disease detection with a novel deep learning model based on depthwise separable convolution and transformer networks. *Engineering Applications of Artificial Intelligence*, *133*, 108307. <https://doi.org/10.1016/j.engappai.2024.108307>
12. Gülmez, B. (2024). A Comprehensive Review of Convolutional Neural Networks based Disease Detection Strategies in Potato Agriculture. *Potato Research*. <https://doi.org/10.1007/s11540-024-09786-1>
13. Li, J., Wu, J., Liu, R., Shu, G., Liu, X., Zhu, K., Wang, C., & Zhu, T. (2024). Potato late blight leaf detection in complex environments. *Scientific Reports*, *14*(1). <https://doi.org/10.1038/s41598-024-82272-3>
14. Naeem, M. A., Saleem, M. A., Sharif, M. I., Akber, S., Saleem, S., Akhtar, Z., & Siddique, K. (2025). *Deep Learning-Based Approach for Identification of Potato Leaf Diseases Using Wrapper Feature Selection and Feature Concatenation*. ArXiv.org. <https://arxiv.org/abs/2502.03370>
15. Rashid, J., Khan, I., Ali, G., Almotiri, S. H., AlGhamdi, M. A., & Masood, K. (2021). Multi-Level Deep Learning Model for Potato Leaf Disease Recognition. *Electronics*, *10*(17), 2064. <https://doi.org/10.3390/electronics10172064>
16. Sharma, J., Al-Huqail, A. A., Almogren, A., Doshi, H., B Jayaprakash, B Bharathi, Rehman, A. U., & Hussen, S. (2025). Deep learning based ensemble model for accurate tomato leaf disease classification by leveraging ResNet50 and MobileNetV2 architectures. *Scientific Reports*, *15*(1). <https://doi.org/10.1038/s41598-025-98015-x>
17. Wang, X., & Liu, J. (2024). An efficient deep learning model for tomato disease detection. *Plant Methods*, *20*(1). <https://doi.org/10.1186/s13007-024-01188-1>
18. Zhang, Y., Wa, S., Liu, Y., Zhou, X., Sun, P., & Ma, Q. (2021). High-Accuracy Detection of Maize Leaf Diseases CNN Based on Multi-Pathway Activation Function Module. *Remote Sensing*, *13*(21), 4218. <https://doi.org/10.3390/rs13214218>
19. Haque, Md. A., Marwaha, S., Deb, C. K., Nigam, S., Arora, A., Hooda, K. S., Soujanya, P. L., Aggarwal, S. K., Lall, B., Kumar, M., Islam, S., Panwar, M., Kumar, P., & Agrawal, R. C. (2022). Deep learning-based approach for identification of diseases of maize crop. *Scientific Reports*, *12*(1). <https://doi.org/10.1038/s41598-022-10140-z>

‌

**APPENDICES**

**Appendix 1: Interview questions**

1. What is your level of familiarity with deep learning and CNNs?
2. Have you used any agricultural technology tools before?
3. How often do you experience potato crop disease on your farm?
4. What method do you currently use to detect plant diseases?
5. How confident are you in your ability to identify Early Blight or Late Blight manually?
6. Would you be willing to use a mobile app for disease detection?
7. Do you have regular internet access on your farm?
8. What device do you usually use to access the internet (e.g., smartphone, computer)?
9. How important is disease detection to your overall farming success?
10. Would you be interested in receiving automated advice based on leaf image analysis?

**Appendix 2: Questionnaire**

* + 1. How easy was it to use the potato disease detection system?

A. Very Easy

B. Easy

C. Difficult

D. Very Difficult

ii. Did the system correctly identify the disease in your potato leaf image?

A. Yes

B. No

C. Not Sure

D. Did Not Try

iii. How would you rate the clarity of the results provided?

A. Very Clear

B. Clear

C. Unclear

D. Very Confusing

iv. Do you believe the system's results were accurate?

A. Very Accurate

B. Somewhat Accurate

C. Not Accurate

D. I Don't Know

v. How likely are you to use this system during your farming activities?

A. Very Likely

B. Likely

C. Unlikely

D. Very Unlikely

vi. Do you have access to the internet or smartphone to use this tool regularly?

A. Always

B. Sometimes

C. Rarely

D. Never

vii. How often do you face potato disease problems on your farm?

A. Very Often

B. Sometimes

C. Rarely

D. Never

viii. Would you recommend this system to other farmers?

A. Definitely

B. Maybe

C. Probably Not

D. Definitely Not

ix. What feature would you like added to this system?

x. What challenges did you experience when using the system?