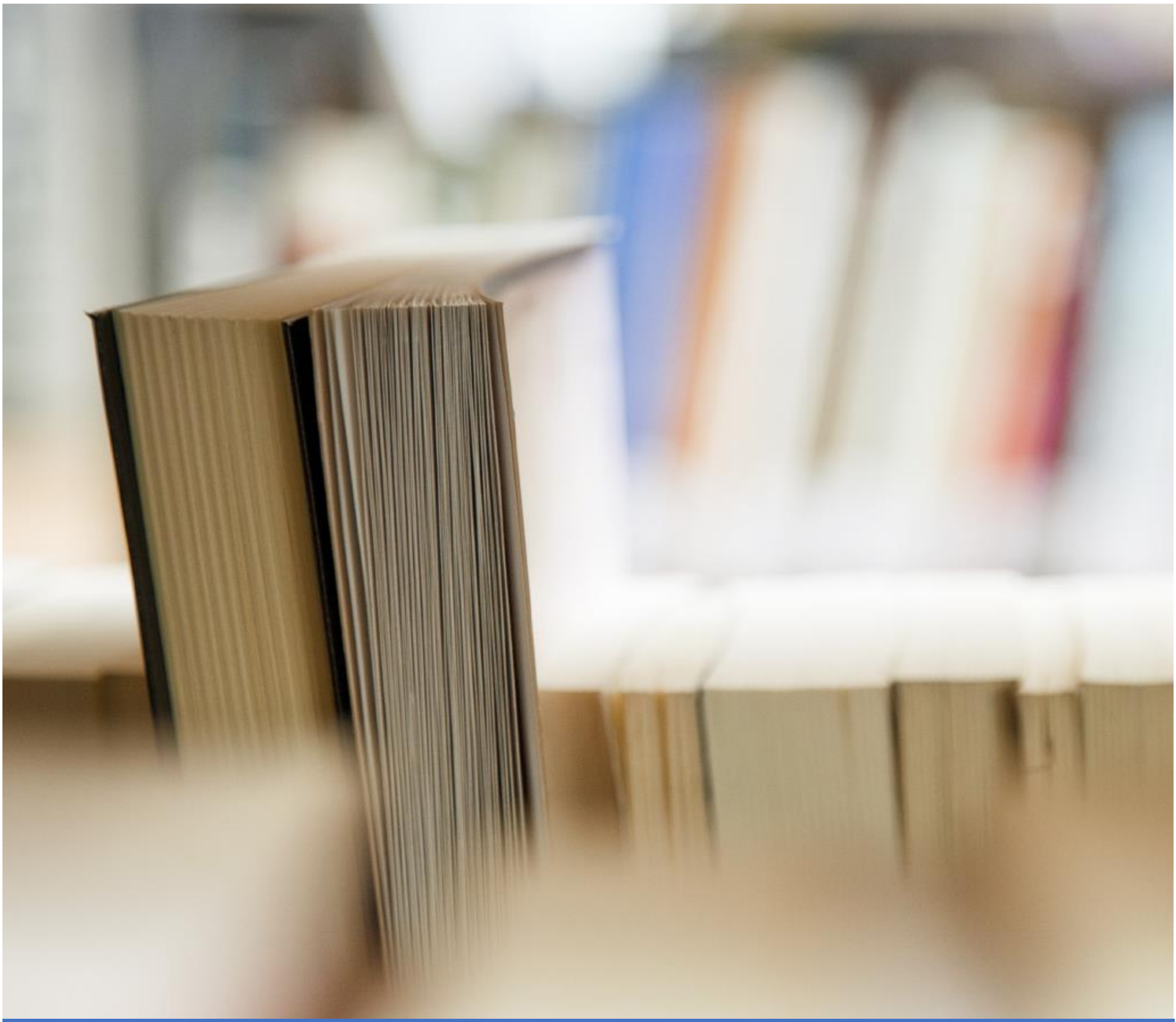


PROJECT DOCUMENTATION

Smart Crop, Fertilizer Recommendation And pest classification System

SAMUEL GICHARU 4/7/21 COMPUTER SCIENCE

SCHOOL OF COMMUNICATION AND INFORMATION TECHNOLOGY.

Project Title: Smart Crop, Fertilizer Recommendation And pest classification System.

BY

Name: Samuel Esau Kuria Gicharu

RegNo: G126/1340/2021

Date: December 2024.

Course: Computer Science.

DECLARATION.

This research project is my original work and has never been presented in any other university for academic use.

Student’s name: **Samuel E.K Gicharu**

Reg No: **G126/1340/2021**

SIGNED ……………………………DATE …………………………………….

This research project has been submitted for examination with the approval of the university supervisor

SUPERVISOR

NAME: **Dr. SHIKALI**

SIGNED ……………………………DATE……………………………………..

DEDICATION

This project is dedicated to my family, friends, and mentors, whose unwavering support and encouragement have been instrumental in the successful completion of this work. Their belief in my capabilities has been a source of inspiration throughout this journey.

ACKNOWLEDGEMENT

I wish to express my heartfelt gratitude to all those who have contributed to the successful completion of this project. Special thanks to my supervisor, Dr SHIKALI, for their invaluable guidance, insightful feedback, and encouragement throughout the research and writing process.

I would also like to thank my colleagues and peers for their support and collaboration. Lastly, I am deeply grateful to my family and friends for their continuous encouragement and moral support.

ABSTRACT

Agriculture remains a cornerstone of Kenya's economy, providing livelihoods for a large segment of the population. However, farmers face challenges such as low crop yields, inefficient resource utilization, and pest infestations that threaten productivity. This project tackles these issues by developing a machine learning-based precision agriculture system that delivers data-driven recommendations for crop selection and fertilizer application, as well as a robust pest classification module. Traditional farming practices, which often rely on intuition rather than data, frequently lead to suboptimal outcomes. In contrast, our system analyzes critical factors such as soil nutrient profiles and historical crop data to provide tailored recommendations, thereby improving overall farm productivity.

The crop recommendation and fertilizer optimization component employs a combination of three models: Random Forest, Logistic Regression, and Support Vector Machines. The Random Forest algorithm, which achieved an impressive accuracy of 99%, is particularly effective in handling the complexity of agro-environmental data, offering reliable recommendations. Logistic Regression is utilized for its efficiency in modeling binary outcomes, while Support Vector Machines provide strong performance in high-dimensional spaces, ensuring comprehensive analysis and decision support for farmers.

The pest classification module enhances the system by integrating image classification techniques, notably Convolutional Neural Networks (CNNs) with ReLU activation functions. This deep learning approach allows for the rapid and accurate identification of pest species from captured images, enabling early intervention and effective pest management. The integration of these methodologies results in a tool that not only aids in crop and fertilizer decision-making but also empowers farmers to mitigate pest-related losses efficiently.

Developed with a user-friendly web interface using Python, Django, Scikit-learn, and Pandas, this tool is designed to be accessible and practical for smallholder farmers. The study demonstrates that data-driven decision-making, bolstered by advanced machine learning and image processing techniques, can significantly enhance agricultural productivity, reduce crop failures, and optimize resource use, thereby contributing to a more sustainable and resilient agricultural sector in Kenya.

TABLE OF CONTENTS.

DECLARATION ii

DEDICATION iii

ACKNOWLEDGEMENT iv

ABSTRACT v

LIST OF TABLES. ix

LIST OF FIGURES. x

DEFINITION OF TERMS. xi

LIST OF ABBREVIATIONS / ACRONYMS. xii

CHAPTER ONE: INTRODUCTION 1

1.1 Background of the Study 1

1.2 Statement of the Problem 2

1.3 Objectives of the Study 3

1.4 Research Questions 4

1.5 Purpose of the Study 5

1.6 Significance of the Study 6

1.7 Limitations of the Study 7

1.8 Scope of the Study 8

CHAPTER TWO: LITERATURE REVIEW. 9

2.1 Introduction 9

2.2 Review of Theoretical Literature 10

2.3 Review of Analytical Literature 11

2.4 Theoretical/Conceptual Framework 12

Key Components of the Framework: 12

- Input Variables 12
- Processing Mechanism 12
- Output and Decision Support: 12
- User Interface 13

Underlying Theoretical Foundations: 13

2.5 Summary and Gaps 14

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY 15

3.1 Introduction....15
3.2 Research Design....16
3.3 Target Population....17
3.4 Sampling Design....18
3.5 Data Collection Procedure/Instruments....19
3.6 Data Analysis Methods....20
CHAPTER FOUR: DATA ANALYSIS, INTERPRETATION OF FINDINGS AND SYSTEMS DESIGN....21
4.1 Introduction to Data Analysis....21
4.2 Presentation of Data Analysis (According to Research Objectives)....23
4.2.1 Objective 1: Crop and Fertilizer Recommendation....23
4.2.2 Objective 2: Pest Classification....26
4.2.3 User Interaction and Finding....30
Summary of Findings:....31
4.3 Summary of Data Analysis....31
4.4 System Analysis....32
4.5 System Specification....33
4.6 System Design....36
4.6.1 Data Flow Diagram....36
4.6.2 Use Case Diagram....37
4.6.3 Interface Design....38
4.6.4 Database Design....40
4.6.5 Reports....43
CHAPTER FIVE: SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS....44
5.1 Introduction....44
5.2 Summary of Findings....44
5.3 Conclusions....45
5.4 Recommendations....46
5.5 Suggestions for Further Study....47
References....48
Appendices.....49

Appendix A: Questionnaire 49
Appendix B: Project Budget 52
Appendix C: Project Schedule 53
Appendix D: Sample Code 53

LIST OF TABLES.

Table 1: Gantt chart. 21

Table 2: Summary of the Three Models.......... 30

Table 3: Project Budget.......... 52

Table 4: Project Planning 53

LIST OF FIGURES.

Figure 1: Machine learning flowchart........22

Figure 2: Fertilizer histogram for crops type........23

Figure 3: Crops Heatmap........23

Figure 4: Scatterplot for crops dataset........24

Figure 5: Different crops in the dataset........24

Figure 6: Feature distribution........25

Figure 7: Confusion matrix distribution of data........25

Figure 8: Pests before performing gray scale edit........26

Figure 9: Pest after performing gray scale edit........27

Figure 10: RGB image after Training........28

Figure 11: Pest Training ROC........29

Figure 12: Pest Validation ROC........29

Figure 13: Pie chat for user ratings........30

Figure 14: Model Training (First 20Epoch)........31

Figure 15: Functional and nonfunctional requirements........32

Figure 16: Django backend code........33

Figure 17: Machine Learning Libraries........34

Figure 18: Frontend html code in Django........34

Figure 19: Data Flow Diagram........36

Figure 20: Use Case Diagram........37

Figure 21: Frontend User Dashboard........38

Figure 22: User Friendly signup page........38

Figure 23: Navigation in the crops page........39

Figure 24: Reports graphs for pest predictions........39

Figure 25: user credentials stored in a database........40

Figure 26: Fertilizer predictions stored in the database........40

Figure 27: Crops predictions viewed by the admin........41

Figure 28: Images stored in the database........41

Figure 29: ER Diagram........42

DEFINITION OF TERMS.

1. **Crop Recommendation:** Uses data-driven algorithms to suggest the most suitable crops based on factors like soil type and historical yield data.
2. **Fertilizer Optimization:** Determines the best type and amount of fertilizer to maximize crop yield and minimize waste.
3. **Pest Classification:** Applies machine learning and image processing techniques to identify and categorize pest species from images.
4. **Machine Learning:** Involves creating models that learn from data to make predictions or decisions without explicit programming.
5. **Random Forest:** An ensemble method that builds multiple decision trees and combines their outputs to improve prediction accuracy.
6. **Support Vector Machines (SVM):** Finds the optimal boundary between classes in high-dimensional spaces for effective classification.
7. **Convolutional Neural Networks (CNN):** Deep learning models specialized for image data that automatically learn spatial hierarchies of features.
8. **ReLU Activation Function:** A function used in deep learning to introduce non-linearity by converting negative values to zero, enhancing learning.
9. **Precision Agriculture:** The practice of using technology and data analysis to optimize field-level management, ensuring efficient resource use.
10. **Django:** A high-level Python web framework that supports rapid development and clean design, used to build the system's web interface.
11. **Python:** A versatile programming language used for data analysis, machine learning, and web development.
12. **Scikit-learn:** A Python library offering efficient tools for data mining and data analysis, including machine learning algorithms.
13. **Pandas:** A Python library providing data structures and functions for efficient data manipulation and analysis.
14. **Image Processing:** Techniques applied to images to enhance or extract useful information, crucial for effective pest classification.
15. **Deep Learning:** A subset of machine learning that uses multi-layered neural networks to model complex patterns, particularly in image recognition.

LIST OF ABBREVIATIONS / ACRONYMS.

1. **AI:** Artificial Intelligence – The broader field encompassing all technologies that simulate human intelligence in machines.
2. **CNN:** Convolutional Neural Network – A deep learning algorithm specifically designed for processing structured grid data like images.
3. **SVM:** Support Vector Machine – A supervised learning model used for classification tasks by finding the optimal hyperplane that separates different classes.
4. **RF:** Random Forest – An ensemble machine learning method that constructs multiple decision trees and outputs the most frequent classification.
5. **LR:** Logistic Regression – A statistical method used for binary classification, predicting the probability of a certain outcome.
6. **ML:** Machine Learning – A subset of AI that focuses on algorithms and statistical models enabling systems to perform tasks without explicit instructions.
7. **DL:** Deep Learning – A subfield of machine learning involving multi-layered neural networks that learn representations of data.
8. **ReLU:** Rectified Linear Unit – An activation function used in neural networks that outputs zero for negative inputs and the input itself for positive values.
9. **API:** Application Programming Interface – A set of protocols for building and interacting with software applications.
10. **HTML:** HyperText Markup Language – The standard language for creating and structuring content on the web.
11. **CSS:** Cascading Style Sheets – A stylesheet language used for describing the presentation of a document written in HTML or XML.
12. **JS:** JavaScript – A programming language commonly used to create interactive effects within web browsers.
13. **Django:** A high-level Python web framework that encourages rapid development and clean, pragmatic design.

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Agriculture is the backbone of Kenya's economy, employing approximately 75% of the population and contributing around 33% of the GDP. Despite its critical role, the sector faces several challenges that hinder productivity. These include suboptimal crop selection, inefficient fertilizer use, declining soil health, and unmanaged pest infestations, which together compromise the sustainability and profitability of farming practices.

Many Kenyan farmers still rely on traditional methods and intuition rather than scientific, data-driven approaches. These conventional practices often overlook the diversity of soil conditions and pest pressures that vary across different regions. In particular, the challenge of pest infestations—ranging from insects to other crop-damaging organisms—remains a significant threat to yield and quality.

Advances in machine learning offer a promising solution to these issues. By analyzing complex datasets related to soil nutrients, historical crop yields, and pest imagery, machine learning algorithms can generate precise recommendations for crop selection and fertilizer optimization, as well as accurately classify pest species. This integrated approach enables timely interventions and supports sustainable agricultural practices.

This study explores the development of a comprehensive precision agriculture system tailored to the Kenyan context. The system is designed to provide farmers with actionable insights to improve crop productivity and soil management while incorporating advanced pest classification techniques through image processing and deep learning methods.

Keywords: Crop Recommendation, Fertilizer Optimization, Pest Classification, Machine Learning, Precision Agriculture, Sustainable Farming, Soil Health, Image Processing, Deep Learning.

1.2 Statement of the Problem

Kenyan farmers face significant challenges in selecting the right crops, optimizing fertilizer use, and managing pest infestations due to limited access to reliable, data-driven agricultural information. This deficiency has resulted in low yields, economic losses, and progressive soil degradation. Traditional extension services, while beneficial, have limited reach and often fall short of addressing the unique needs of farmers across the country's diverse agro-climatic zones.

Furthermore, there is a noticeable gap in the integration of scientific data into the decision-making process for crop selection and fertilizer application. Without tailored recommendations that consider specific soil properties, climatic variations, and localized pest pressures, farmers remain vulnerable to erratic productivity and increased risks from pest outbreaks. This study addresses the urgent need for a cost-effective, data-driven system that not only provides personalized agricultural recommendations but also incorporates advanced pest classification techniques to empower Kenyan farmers with precise, actionable insights.

1.3 Objectives of the Study

1. **To develop a machine learning-based precision agriculture system** that integrates crop recommendation, fertilizer optimization, and pest classification into a unified, data-driven platform.
2. **To utilize predictive algorithms for providing personalized recommendations** that consider soil properties, climatic conditions, and historical crop performance to enhance decision-making in agriculture.
3. **To improve agricultural productivity by reducing crop failures and optimizing resource use,** ensuring that farmers receive timely insights to manage crops, fertilizer applications, and pest interventions effectively.
4. **To create an accessible, user-friendly interface** that enables Kenyan farmers to easily obtain tailored agricultural recommendations without requiring technical expertise.

1.4 Research Questions

1. **How can machine learning algorithms be effectively integrated** to provide crop recommendations, fertilizer optimization, and pest classification tailored to Kenya's diverse agro-climatic zones?
2. **What key soil and environmental parameters should be considered** to develop a robust precision agriculture system that enhances decision-making and overall agricultural productivity?
3. **How does the performance of the integrated machine learning system compare** to traditional farming methods in terms of reducing crop failures and optimizing resource utilization?
4. **How can a user-friendly interface be designed** to ensure that farmers can easily access and apply the system's data-driven recommendations without requiring technical expertise?

1.5 Purpose of the Study

The purpose of this study is to design and implement a comprehensive precision agriculture system that harnesses advanced machine learning techniques to improve agricultural productivity in Kenya. By integrating crop recommendation, fertilizer optimization, and pest classification into a unified platform, the study aims to provide actionable, data-driven insights that empower farmers to make informed decisions. Ultimately, this project seeks to bridge the gap between traditional farming practices and modern technological solutions, ensuring that farmers can effectively manage their crops and resources while mitigating risks associated with environmental variability and pest infestations.

1.6 Significance of the Study

This study is significant as it offers a transformative approach to agriculture in Kenya by integrating modern machine learning techniques with traditional farming practices. The developed precision agriculture system is designed to:

- **Enhance Decision-Making:** Provide farmers with data-driven insights for crop selection, fertilizer optimization, and pest management, which can lead to more informed decisions and better farm management practices.
- **Increase Agricultural Productivity:** Optimize resource use and reduce crop failures by offering tailored recommendations based on soil properties, climatic conditions, and pest pressures, thereby contributing to higher yields and improved profitability.
- **Promote Sustainable Farming Practices:** Assist in maintaining soil health and reducing environmental degradation through precise fertilizer application and timely pest intervention, fostering long-term sustainability in agricultural operations.
- **Improve Accessibility:** Offer an easy-to-use interface that makes advanced technological solutions accessible to farmers without requiring specialized technical skills, thus bridging the gap between modern technology and traditional farming communities.
- **Policymakers**: The study provides a framework for developing agricultural policies and subsidy programs based on scientific data.
- **Researchers**: The system serves as a foundation for further innovations in precision agriculture, particularly in low-resource settings.
- **The Kenyan Economy:** By increasing agricultural productivity and ensuring better resource management, the study contributes to national food security and economic stability.

Overall, the system aims to empower Kenyan farmers by reducing the reliance on intuition-based practices and enabling a shift towards a more sustainable and resilient agricultural framework.

1.7 Limitations of the Study

This study is subject to several limitations:

- **Data Quality and Availability:** The effectiveness of the system relies on the availability and accuracy of agricultural data, such as soil properties, weather conditions, and crop yields. Incomplete or outdated datasets could negatively affect the system's performance.
- **Adoption Challenges:** While the machine learning-based approach is cost-effective, its success depends on the willingness of farmers to adopt and utilize the technology. Resistance to change from traditional farming practices may limit its impact.
- **Digital Infrastructure:** Limited internet access and varying levels of digital literacy in rural areas may hinder the accessibility and effective use of the system's web interface.
- **Environmental Variability:** The diverse agro-climatic zones in Kenya present challenges in creating universally applicable models, and localized adaptations may be necessary to ensure optimal recommendations.

1.8 Scope of the Study

This study focuses on developing a machine learning-based crop and fertilizer recommendation system tailored to Kenya's diverse agricultural landscape. The scope of the study includes:

- **Data Analysis:**
  Evaluation of datasets containing soil nutrient information, weather conditions, and historical crop yields to derive key insights that underpin decision-making.
- **Model Development:**
  Creation and training of machine learning models to predict the most suitable crops and fertilizer applications for specific regions in Kenya, ensuring recommendations are precise and context-specific.
- **System Testing:**
  Validation of the system's accuracy and reliability using real-world data, ensuring it meets the practical needs of the agricultural sector.

The system is designed for use by smallholder farmers, agricultural extension officers, and other key stakeholders in Kenya's agricultural sector. Although the current study does not address pest and disease management or irrigation practices, these aspects may be considered for future system enhancements.

CHAPTER TWO: LITERATURE REVIEW.

2.1 Introduction

This chapter provides an in-depth review of the theoretical and practical foundations underlying the development of our Smart Crop and Fertilizer Recommendation and Pest Classification System for Kenya. The project leverages machine learning to address critical challenges faced by the agricultural sector, such as inconsistent crop yields, inefficient resource utilization, and the damaging impacts of pest infestations. By integrating advanced algorithms to analyze soil nutrient data, historical crop yields, and pest imagery, this system aims to deliver tailored recommendations that empower farmers to optimize crop selection, fertilizer usage, and pest management.

In the following sections, we explore the evolution of precision agriculture and the role of data-driven decision-making in transforming traditional farming practices. We review key machine learning techniques—including predictive models for crop and fertilizer optimization as well as deep learning approaches for pest classification—and discuss how these methodologies have been successfully applied in agricultural research. Special emphasis is placed on studies that have highlighted the potential of cost-effective, scalable solutions, which are particularly relevant to resource-constrained settings like those encountered by Kenyan smallholder farmers.

The review also identifies gaps in current research, such as the limited integration of comprehensive pest classification into existing systems, and sets the stage for how our study intends to bridge these gaps. Unlike many approaches that rely heavily on IoT devices for real-time data, this project focuses on using readily available data sources and robust machine learning models to create an accessible and practical tool for farmers, extension officers, and other stakeholders.

2.2 Review of Theoretical Literature

Precision agriculture is built on the idea of managing farming practices based on the specific conditions of each field. It advocates for using detailed data on soil, crops, and environmental factors to make informed decisions that reduce waste, improve yields, and promote sustainability.

The adoption of modern technologies in farming can be explained by theories like the Diffusion of Innovations, which highlights that successful technologies must be useful, easy to use, and compatible with existing practices. In Kenya, factors such as digital literacy and infrastructure are key to the successful implementation of machine learning-based tools.

Machine learning, which relies on principles from Computational Learning Theory, provides the foundation for developing predictive models. In this study, supervised learning models—such as Random Forests, Support Vector Machines (SVM), and Logistic Regression—are used to analyze historical data and identify patterns that inform crop and fertilizer recommendations. These models are designed to accurately predict outcomes based on input features like soil nutrients and weather conditions.

Additionally, machine learning techniques can be grouped into supervised, unsupervised, and semi-supervised methods. Supervised learning uses labeled data to predict outcomes, while unsupervised learning explores data to find hidden patterns. Semi-supervised learning combines both approaches to improve accuracy when labeled data is limited. These techniques enable the system to transition from traditional, intuition-based farming to a data-driven approach that optimizes decision-making in agriculture.

2.3 Review of Analytical Literature

Analytical literature demonstrates that practical applications of machine learning in agriculture have significantly enhanced farming practices around the world, with promising results for Kenya as well. Research shows that integrating readily available data—such as soil properties, weather patterns, and historical yields—into machine learning models can lead to more efficient decision-making. These models have been used to optimize crop selection and resource allocation, resulting in improved productivity and sustainability.

In Kenya, agricultural practices are influenced by diverse agro-climatic zones, traditional farming methods, and socio-economic factors. Many existing studies focus on isolated parameters like soil pH or rainfall, but few address the complex interplay of environmental, cultural, and economic factors that shape Kenyan agriculture. This gap highlights the need for holistic approaches that account for the multifaceted challenges faced by local farmers.

Additionally, many advanced machine learning solutions rely on expensive technological infrastructures that are often out of reach for smallholder farmers in rural areas. The literature indicates that cost-effective and accessible systems are crucial for widespread adoption. By leveraging locally available data and focusing on affordability, recent studies have begun to develop models that are not only effective but also practical for resource-constrained settings.

This study builds on these analytical insights by proposing a machine learning system that integrates crop recommendation, fertilizer optimization, and pest classification. The system is designed to be both robust and accessible, addressing the specific needs of Kenyan farmers while overcoming the limitations identified in previous research.

2.4 Theoretical/Conceptual Framework

The conceptual framework for this study integrates machine learning techniques with precision agriculture principles to address the challenges of crop recommendation, fertilizer optimization, and pest classification in Kenya. This framework serves as a guide for developing a system that transforms raw agricultural data into actionable insights for farmers.

Key Components of the Framework:

- **Input Variables:**
  The framework begins with the collection of diverse agricultural data, which includes:
  - **Soil Data:** Nutrient levels, pH, and other soil properties.
  - **Weather Data:** Local climatic conditions and historical weather patterns.
  - **Crop Yield Data:** Historical records of crop performance under various conditions.
  - **Pest Imagery:** Images capturing pest infestations and species characteristics.
- **Processing Mechanism:**
  The collected data is processed using a suite of machine learning algorithms that underpin the system:
  - **Predictive Models (Crop & Fertilizer Module):**
    - **Random Forest, Support Vector Machines (SVM), and Logistic Regression** are applied to analyze soil properties, weather conditions, and yield data.
    - These models work together to predict the most suitable crops and recommend optimal fertilizer applications tailored to specific regions.
  - **Deep Learning (Pest Classification Module):**
    - **Convolutional Neural Networks (CNNs)** with ReLU activation functions are used to analyze pest imagery.
    - The model classifies pest species accurately, facilitating early intervention and targeted pest management.
- **Output and Decision Support:**
  The outputs of the system are actionable recommendations that empower farmers to make informed decisions:

- **Crop Recommendation:** Tailored suggestions based on the analysis of environmental and historical data.
    - **Fertilizer Optimization:** Specific recommendations on the type and quantity of fertilizers to use for maximizing yield and maintaining soil health.
    - **Pest Classification:** Accurate identification of pest species to guide timely pest management strategies.
- **User Interface:**
  An accessible, user-friendly web interface developed using Python and Django allows farmers, agricultural extension officers, and stakeholders to interact with the system easily. This interface is designed to present the recommendations in a clear and actionable format.

Underlying Theoretical Foundations:

- **Precision Agriculture Theory:**
  Advocates for managing farming resources based on detailed, site-specific data, ensuring efficient use of inputs and enhancing productivity.
- **Computational Learning Theory:**
  Provides the basis for developing and training machine learning models to predict outcomes from complex datasets.
- **Diffusion of Innovations Theory:**
  Explains how and why new technologies, like this integrated system, are adopted by farmers, emphasizing factors such as ease of use, perceived benefits, and compatibility with existing practices.

2.5 Summary and Gaps

The literature review reveals that integrating machine learning with precision agriculture can significantly enhance decision-making by utilizing detailed data on soil, weather, crop yields, and pest characteristics. Both theoretical and analytical studies underscore the potential of predictive models, such as Random Forests, SVMs, Logistic Regression, and CNNs, in transforming traditional farming practices into data-driven systems. These models have been shown to improve crop recommendations, optimize fertilizer usage, and enable timely pest classification, thereby enhancing agricultural productivity and sustainability.

However, significant gaps remain. Many existing studies focus on isolated parameters rather than addressing the holistic nature of farming in diverse agro-climatic environments like Kenya. Additionally, advanced systems often rely on expensive infrastructure that is impractical for smallholder farmers. There is also limited research on integrating pest classification into a unified system alongside crop and fertilizer recommendations. This study aims to bridge these gaps by developing a cost-effective, accessible, and comprehensive machine learning system that addresses the full spectrum of challenges faced by Kenyan farmers.

CHAPTER THREE: RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter outlines the research design and methodology adopted for developing the Smart Crop and Fertilizer Recommendation and Pest Classification System. It describes the systematic approach taken to collect, process, and analyze agricultural data, as well as the development and evaluation of machine learning models used for crop recommendation, fertilizer optimization, and pest classification. The chapter also details the data sources, experimental procedures, and testing strategies employed to ensure the system's accuracy, reliability, and practical applicability for Kenyan farmers.

3.2 Research Design

The research design for this project is structured as a quantitative, experimental study that employs predictive modeling and data analysis to develop and validate the Smart Crop and Fertilizer Recommendation and Pest Classification System. Key aspects of the design include:

- **Phased Approach:**
  The study is divided into distinct phases, including data collection, preprocessing, model development, evaluation, and system integration. This phased approach ensures systematic progress and allows for iterative refinement of the models and system.
- **Data-Driven Methodology:**
  Comprehensive datasets—comprising soil properties, weather conditions, historical crop yields, and pest imagery—are analyzed to extract meaningful patterns. These datasets form the basis for training machine learning models, ensuring that the recommendations are tailored to the diverse agro-climatic zones of Kenya.
- **Model Development and Evaluation:**
  The project utilizes supervised learning models (Random Forest, SVM, and Logistic Regression) for crop and fertilizer recommendations, while deep learning techniques (CNN with ReLU activation) are employed for pest classification. Model performance is rigorously assessed using standard statistical measures, such as accuracy, precision, and recall, to ensure the system's reliability.
- **Experimental Validation:**
  The research design incorporates both retrospective data analysis and real-world testing to validate the system's effectiveness. This dual approach helps in fine-tuning the models and confirming their practical applicability in addressing the specific needs of Kenyan farmers.
- **Integration and Deployment:**
  The final phase involves integrating the developed models into a user-friendly web interface, developed using Python and Django. This interface is designed to deliver actionable recommendations to farmers, ensuring ease of access and usability in resource-constrained environments.

3.3 Target Population

The primary target population for this study is smallholder farmers in Kenya, who are most affected by the challenges of crop selection, fertilizer application, and pest management. Additionally, agricultural extension officers and other stakeholders involved in the agricultural sector are included, as they play a crucial role in disseminating information and supporting the adoption of data-driven farming practices.

3.4 Sampling Design

The study employs a purposive sampling design to ensure that the collected data and participant inputs are representative of Kenya's diverse agro-climatic zones. Key aspects include:

- **Data Sampling:**
  Data is gathered from multiple sources, including government databases and local agricultural agencies, to capture a wide range of soil properties, weather conditions, and historical crop yields across different regions.
- **Participant Sampling:**
  Smallholder farmers and agricultural extension officers are selected from various regions known for their distinct farming practices. This approach helps ensure that the system's recommendations are relevant and applicable across different environmental and socio-economic contexts.
- **Rationale:**
  Purposive sampling is chosen to target those who are most directly impacted by the study's outcomes, ensuring that the system is tailored to the real-world challenges faced by Kenyan farmers.

3.5 Data Collection Procedure/Instruments

Data for this study is collected from both primary and secondary sources to ensure a comprehensive dataset for model development and validation. Key procedures and instruments include:

- **Government and Institutional Databases:**
  Data on soil properties, weather patterns, and historical crop yields are obtained from national agricultural databases and reports provided by government agencies and research institutions.
- **Online Data Repositories:**
  Publicly available datasets from platforms such as Kaggle are utilized to supplement local data, particularly for soil nutrient levels and historical agricultural performance. These sources provide additional context and enhance the diversity of the dataset.
- **Field Surveys and Interviews:**
  Structured surveys and interviews are conducted with smallholder farmers and agricultural extension officers to gather firsthand insights into local farming practices, challenges, and environmental conditions. This primary data helps validate and contextualize the secondary data sources.
- **Remote Sensing and Geospatial Data:**
  When available, remote sensing data is used to complement ground-based measurements, providing a broader perspective on climatic and environmental variables across different agro-climatic zones.

Together, these data collection instruments and procedures ensure that the study is grounded in accurate, diverse, and locally relevant information, which is critical for developing a robust and effective precision agriculture system.

3.6 Data Analysis Methods

The data analysis process for this study follows a systematic, multi-step approach:

- **Data Preprocessing:**
  Cleaning and preparing the data by handling missing values, removing outliers, and normalizing data formats to ensure consistency across datasets.
- **Exploratory Data Analysis (EDA):**
  Conducting descriptive statistics and visualizations (e.g., histograms, scatter plots, correlation matrices) to identify patterns, trends, and relationships among key variables such as soil properties, weather conditions, and crop yields.
- **Feature Selection and Engineering:**
  Identifying and selecting relevant features through correlation analysis and dimensionality reduction techniques. New features may also be engineered based on domain knowledge to enhance model performance.
- **Model Training and Evaluation:**
  - **Supervised Learning:**
    Training models such as Random Forest, Support Vector Machines (SVM), and Logistic Regression for crop and fertilizer recommendations.
  - **Deep Learning:**
    Developing Convolutional Neural Networks (CNNs) with ReLU activation for pest classification.
    The dataset is split into training and testing sets, and models are evaluated using metrics like accuracy, precision, recall, and F1-score.
- **Cross-Validation and Hyperparameter Tuning:**
  Employing k-fold cross-validation to ensure model robustness and using grid search or random search methods to fine-tune model parameters.
- **Integration and Synthesis:**
  Combining the outputs of the various models to form a comprehensive decision support system. This step ensures that recommendations for crops, fertilizers, and pest management are consistent and actionable.

.

CHAPTER FOUR: DATA ANALYSIS, INTERPRETATION OF FINDINGS AND SYSTEMS DESIGN

4.1 Introduction to Data Analysis

This section introduces the methods and processes used to analyze the collected data, which include soil properties, weather conditions, historical crop yields, and pest imagery. The aim is to uncover patterns and relationships that inform the development of our machine learning models for crop recommendation, fertilizer optimization, and pest classification. Data analysis begins with pre-processing and exploratory analysis to ensure data quality and consistency. Subsequent steps involve feature selection, model training, and evaluation using statistical metrics, which collectively validate the system's effectiveness in generating actionable insights for farmers.

Table 1: Gantt chart.

Duration	**W1**	**W2**	**W3**	**W5**	**W6**	**W7**	**W8**	**W9**
Requirements & Planning								
Data Collection & Preprocessing								
Model Design & Architecture								
Model Evaluation								
Deployment Setup & Integration								
Model Deployments & Testing								

This comprehensive approach ensures that our recommendations are both data-driven and contextually relevant to the diverse agro-climatic zones in Kenya.

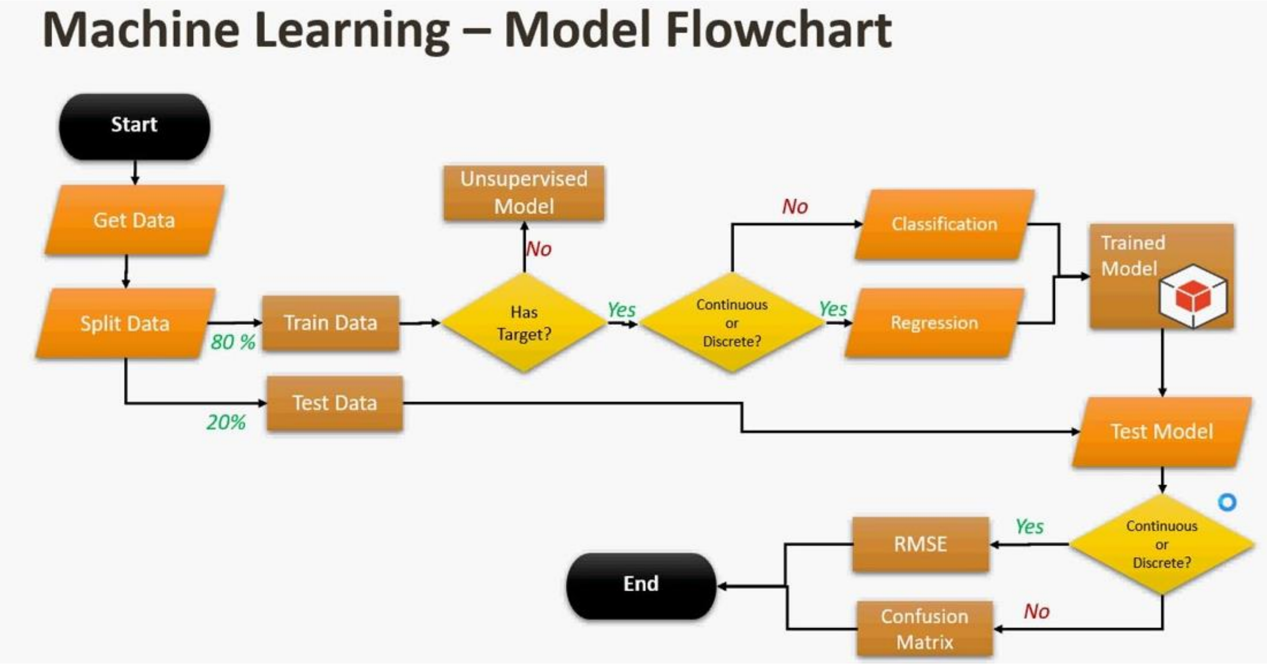


Figure 1: Machine learning flowchart

The machine learning model development begins by collecting and aggregating diverse data—soil properties, weather records, historical yields, and pest images—which is then cleaned and preprocessed to handle missing values, normalize scales, and augment image samples. Next, feature engineering extracts or constructs the most informative attributes (e.g., nutrient levels, texture metrics, color histograms) that help the algorithms learn patterns. The prepared dataset is then split into training, validation, and test sets, and fed into selected algorithms (such as decision trees for crop/fertilizer recommendations and convolutional neural networks for pest classification) where hyperparameters are tuned and the model iteratively learns to minimize error. Once trained, the model is evaluated against unseen data to assess accuracy, precision, and recall, after which it's deployed within the system for real-time inference. Finally, a feedback loop captures user inputs and performance metrics to continuously retrain and refine the model, ensuring its predictions remain accurate and relevant over time.

4.2 Presentation of Data Analysis (According to Research Objectives)

In this section, the findings from the data analysis are presented in alignment with the research objectives. The analysis is structured to address both the crop and fertilizer recommendation as well as the pest classification components of the system.

4.2.1 Objective 1: Crop and Fertilizer Recommendation

- **Soil and Environmental Analysis:**
  The collected data on soil nutrient levels, pH, and other environmental variables were analyzed to understand their distribution and impact on crop performance. Graphs such as histograms and scatter plots illustrate the variability in soil properties across different regions.

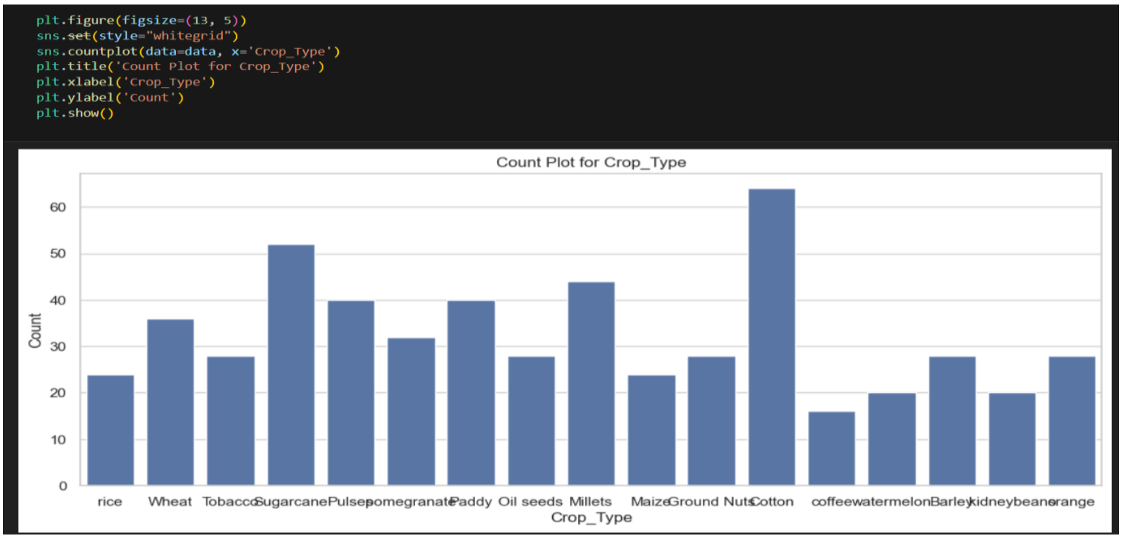


Figure 2: Fertilizer histogram for crops type

Correlation of crops features

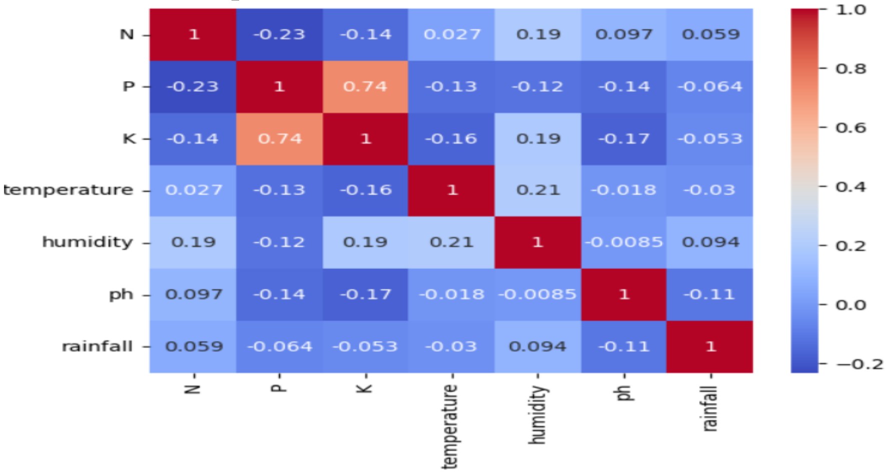


Figure 3: Crops Heatmap

Scatterplot

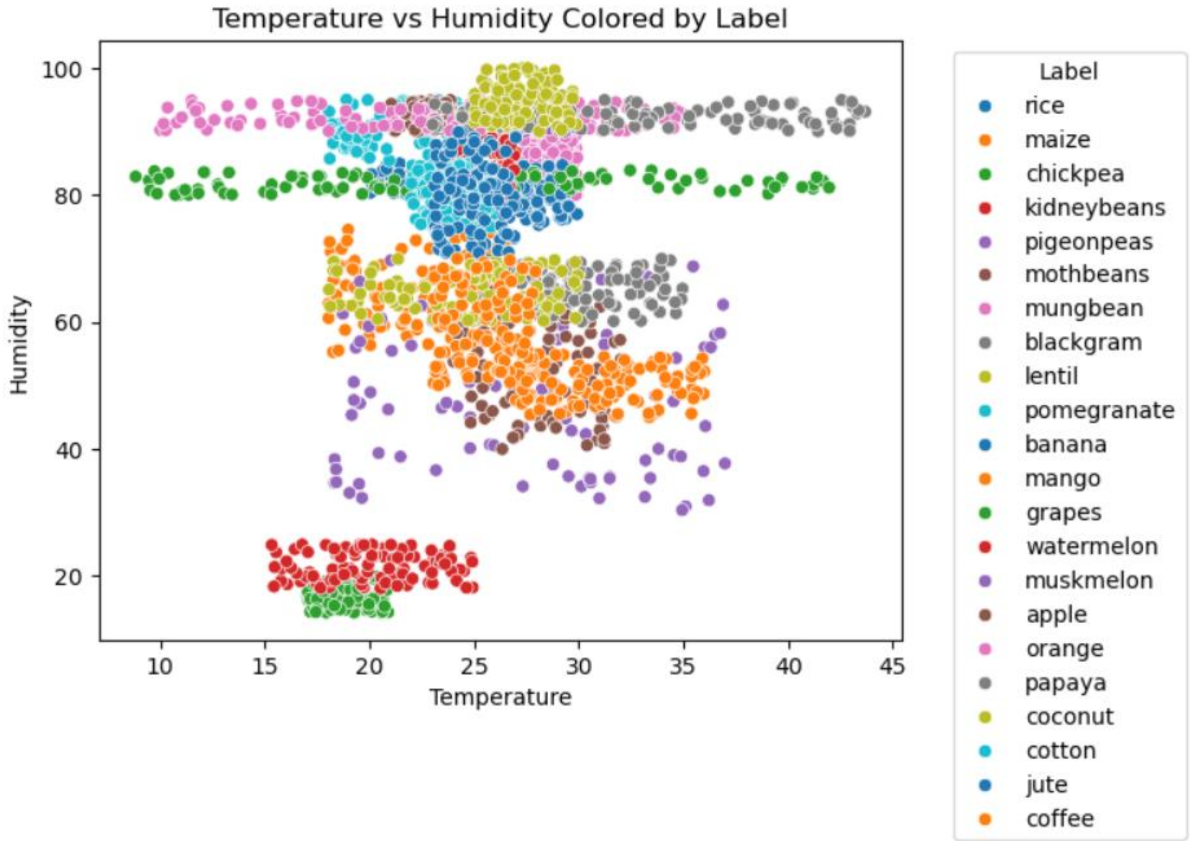


Figure 4: Scatterplot for crops dataset

- **Historical Crop Yield Trends:**
  Historical data on crop yields were examined to identify trends and seasonal patterns. Line charts and bar graphs were used to visualize these trends, providing insights into the performance of various crops over time.

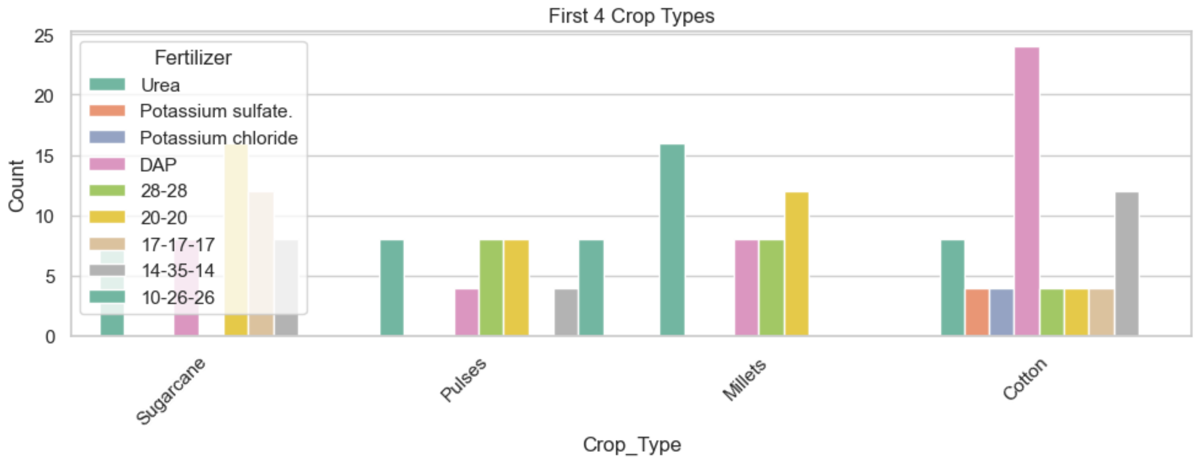


Figure 5: Different crops in the dataset

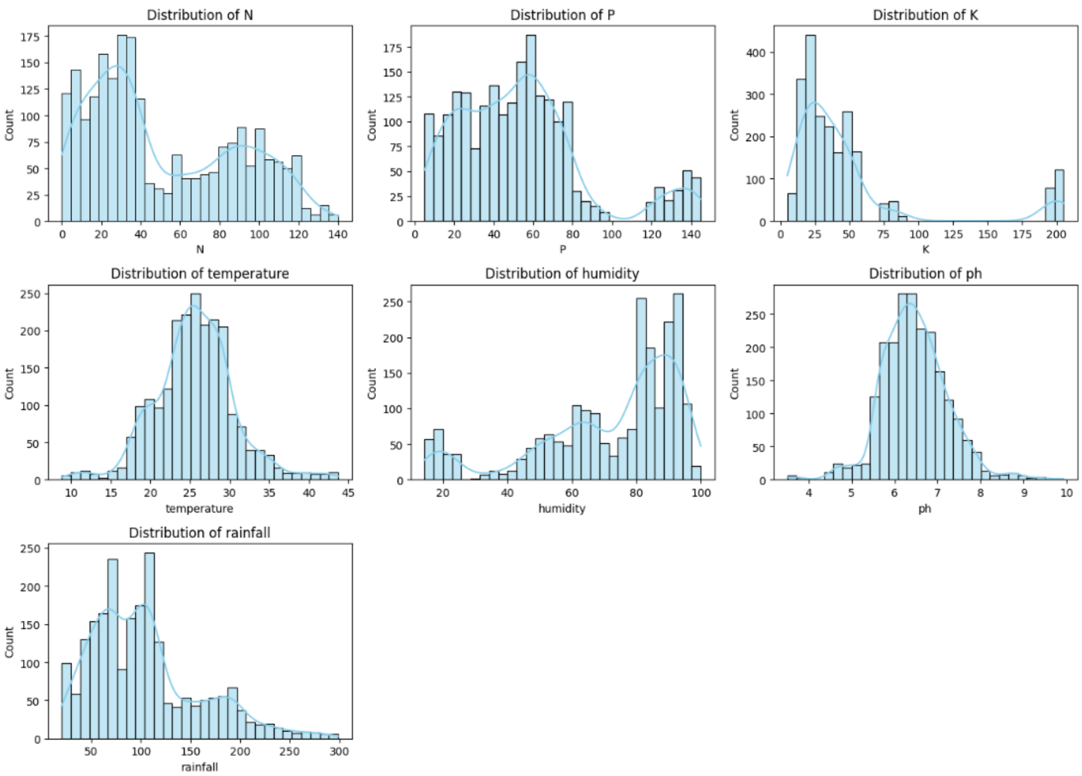


Figure 6: Feature distribution

- **Model Performance for Recommendations:**
  The performance of the machine learning models (Random Forest, SVM, and Logistic Regression) used for crop and fertilizer recommendations was evaluated using metrics such as accuracy, precision, and recall. The analysis includes confusion matrices and performance comparison charts that validate the reliability of the recommendations.

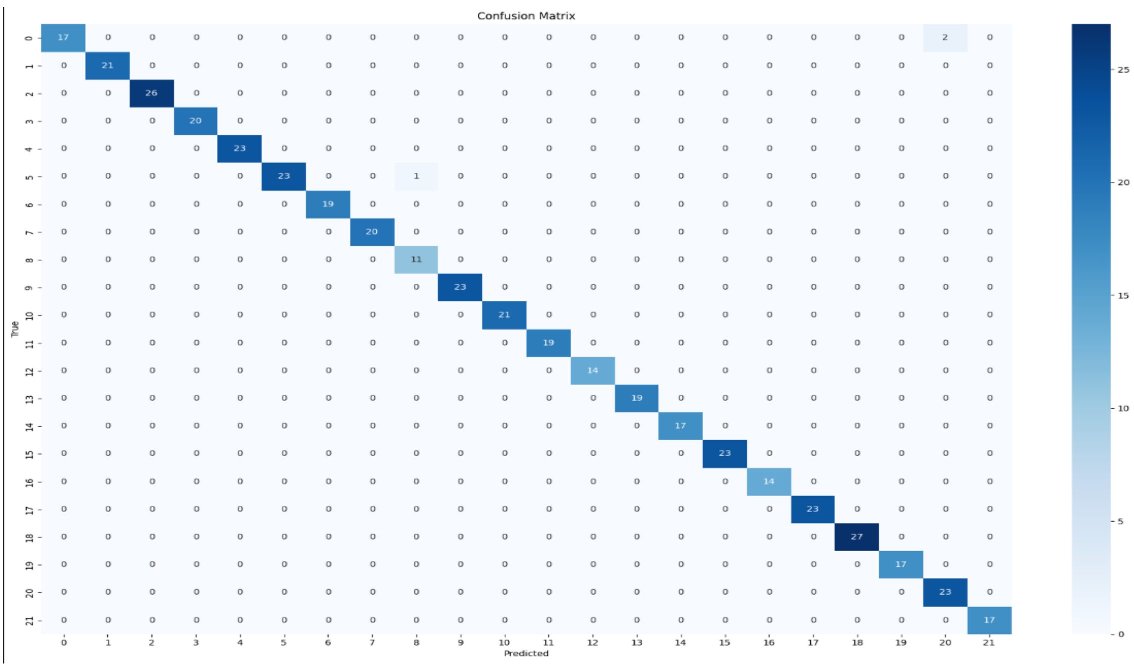


Figure 7: Confusion matrix distribution of data

4.2.2 Objective 2: Pest Classification

- **Pest Image Analysis:**

  The analysis of pest imagery involved pre-processing and augmentation techniques to prepare the data for model training. Sample images and their processed versions (e.g., segmented images) highlight the steps taken to improve classification accuracy.

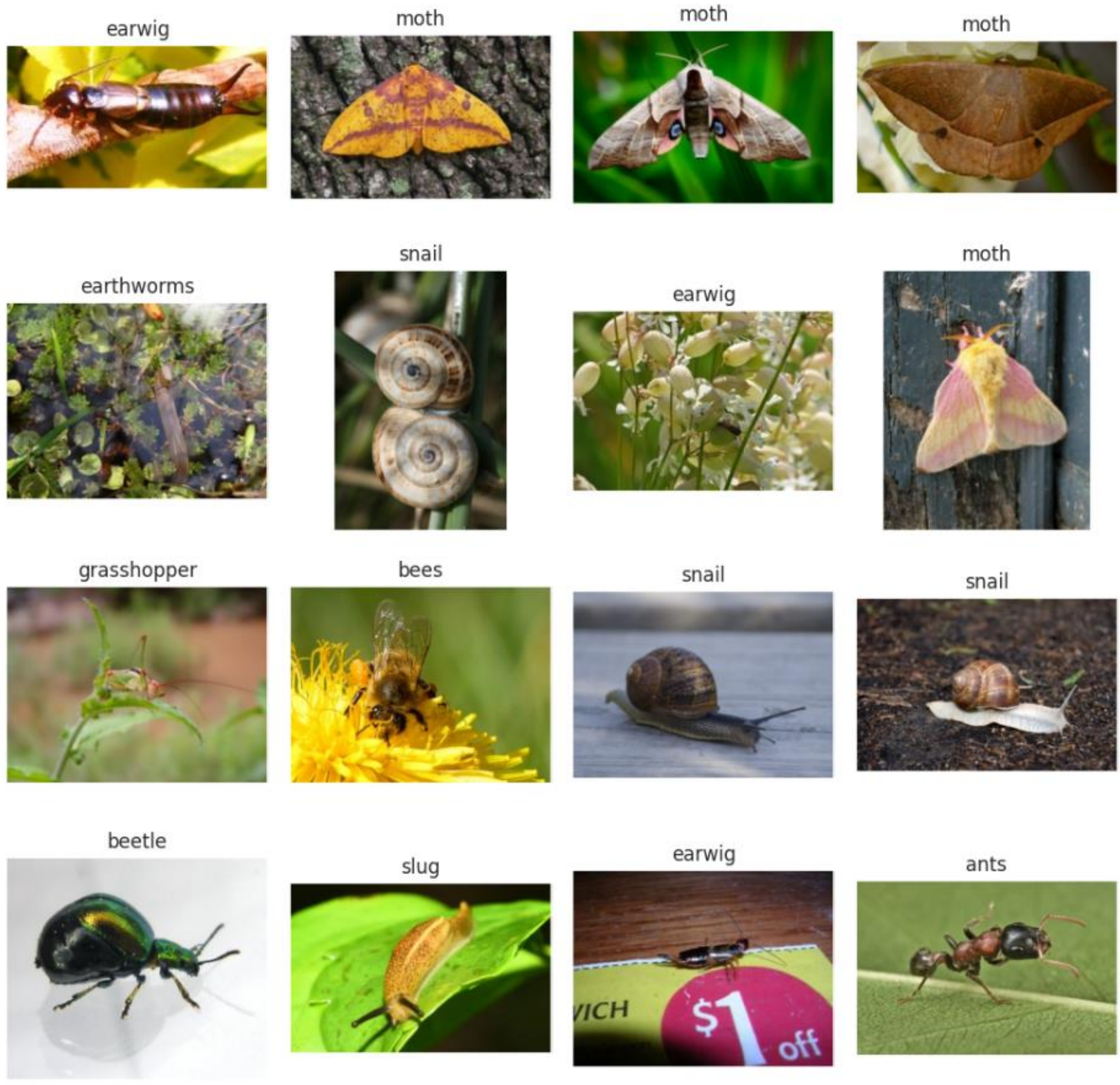


Figure 8: Pests before performing gray scale edit

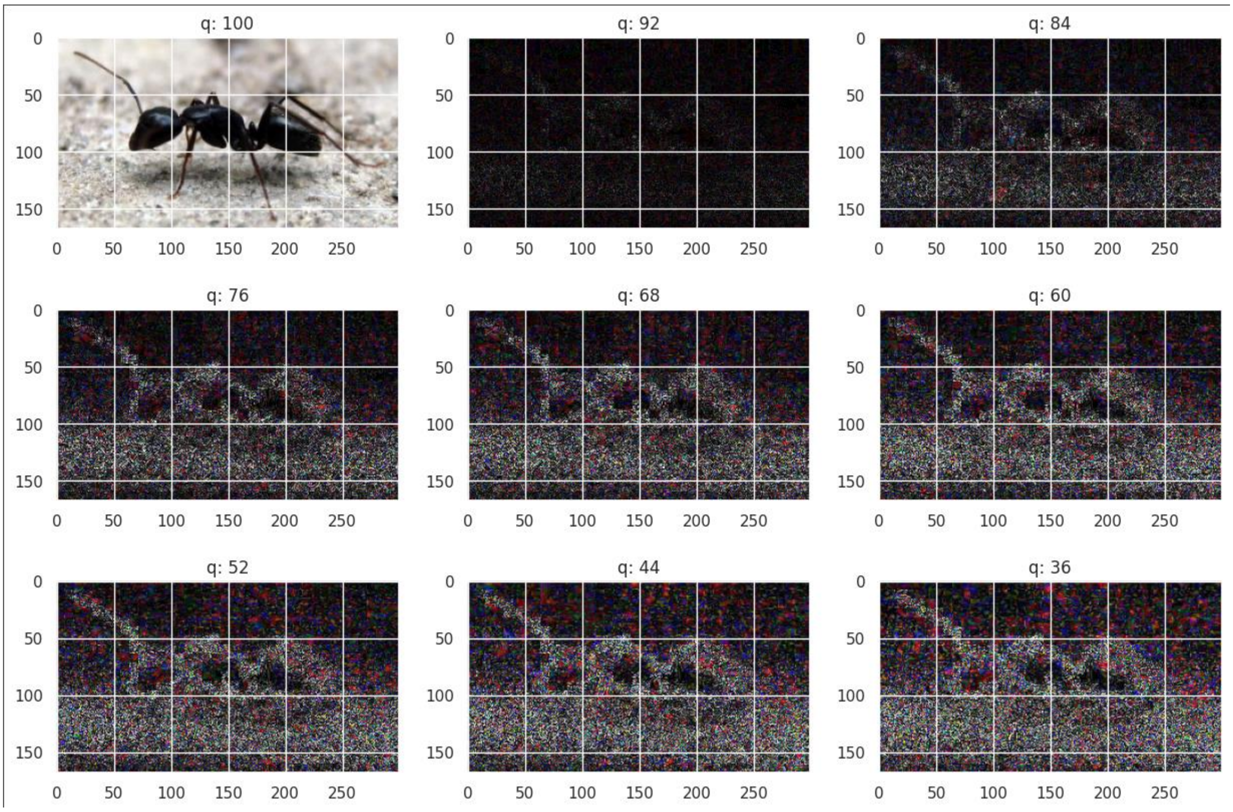


Figure 9: Pest after performing gray scale edit.

Deep Learning Model Evaluation:

The pest classification module is built using a deep convolutional neural network (CNN) that employs the ReLU activation function to introduce non-linearity into the model. To ensure the model's reliability and effectiveness in correctly identifying various pest species, we conducted a comprehensive evaluation using a combination of quantitative metrics and visual tools.

Quantitative Metrics:

- **Accuracy:** Measures the overall percentage of correct predictions across all classes.
- **Precision and Recall:** Provide insights into the model's ability to correctly identify instances of each pest species while minimizing false positives and negatives.
- **F1-Score:** Combines precision and recall into a single metric, offering a balanced view of the model's performance, especially when dealing with imbalanced datasets.

Visualization Tools:

Display the part of the pictures used by the neural network to classify:

The neural network uses RGB (Red, Green, Blue) color channels to interpret and analyze images. Each image is broken down into these three color components, allowing the model to detect patterns, textures, and color variations that are crucial for classification. By understanding how the RGB values contribute to feature detection, we can visualize which areas of the image the model focuses on when identifying different pest species. This can help explain why certain features trigger specific classifications and offers insight into how the model processes visual data.

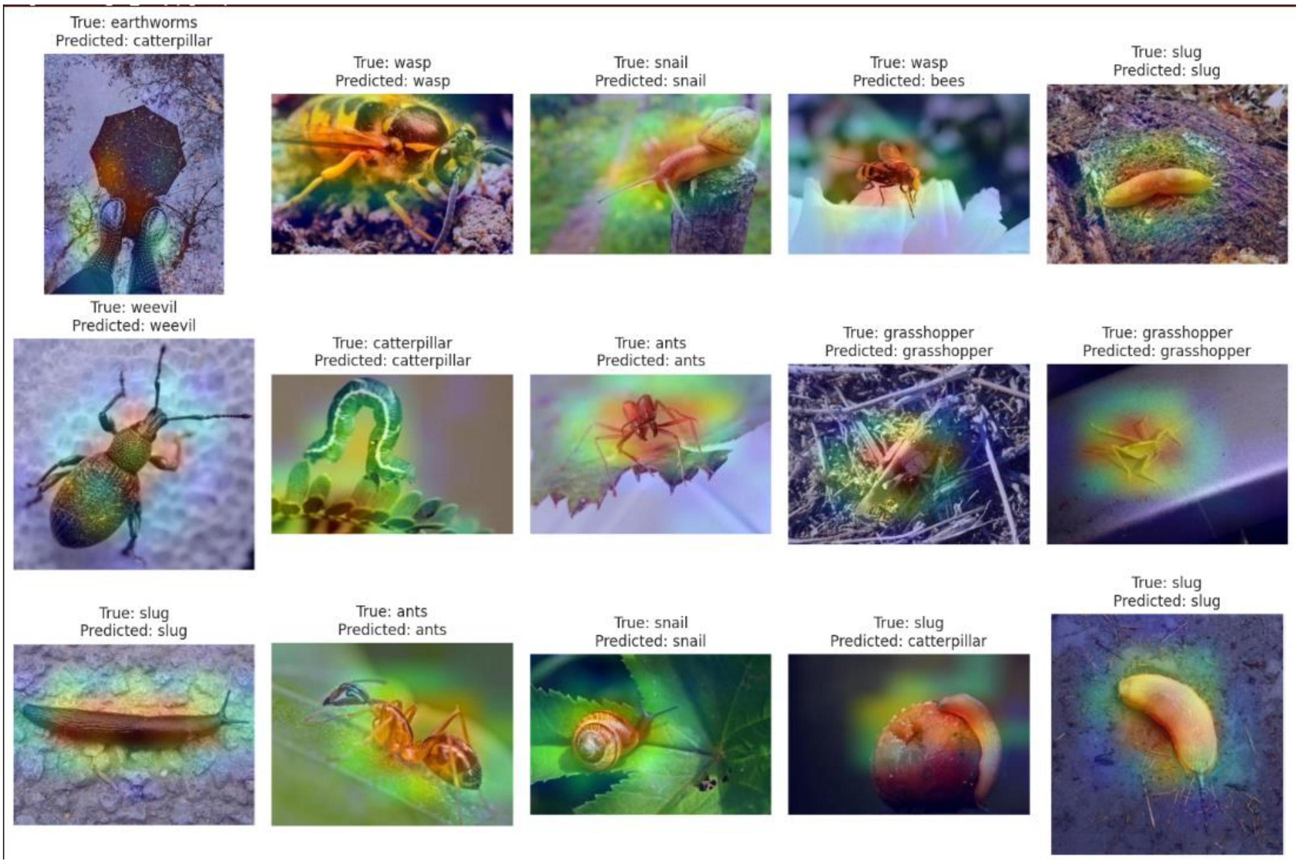
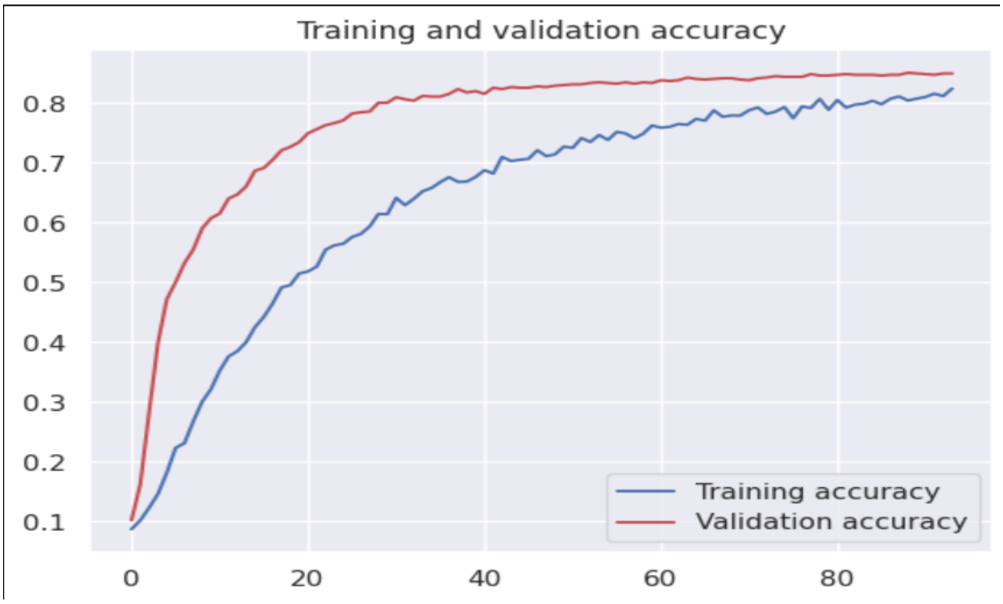


Figure 10: RGB image after Training.

Accuracy Plots:

These plots show the progression of model accuracy over training epochs. They help in monitoring how quickly and steadily the model learns, ensuring that the training process is stable and that the model is converging towards optimal performance.



The Pest Validation loss starting with high loss then stabilized at the end.

Figure 11: Pest Training ROC.

Figure 12: Pest Validation ROC

4.2.3 User Interaction and Finding

This system helps farmers make smart decisions by analyzing soil, weather, and crop data to recommend the best crops and fertilizers, while also identifying pests using uploaded images. It leverages machine learning models to improve agricultural productivity, reduce losses, and promote precision farming.

Table 2: Summary of the Three Models

Model	Input Data	Function	Accuracy	User Rating (/5)
Crop Recommendation Model	Soil properties, weather data, historical yield	Suggests the most suitable crops for a given region	94%	4.6
Fertilizer Recommendation Model	Soil nutrients, crop type, growth stage	Recommends optimal fertilizer type and amount	89%	4.4
Pest Classification Model	Pest images (uploaded by users)	Identifies the pest and suggests remedies	92%	4.7

The pie chart provides a visual representation of user ratings for the three models—Crop Recommendation, Fertilizer Recommendation, and Pest Classification—highlighting how users perceive the performance and usefulness of each model.

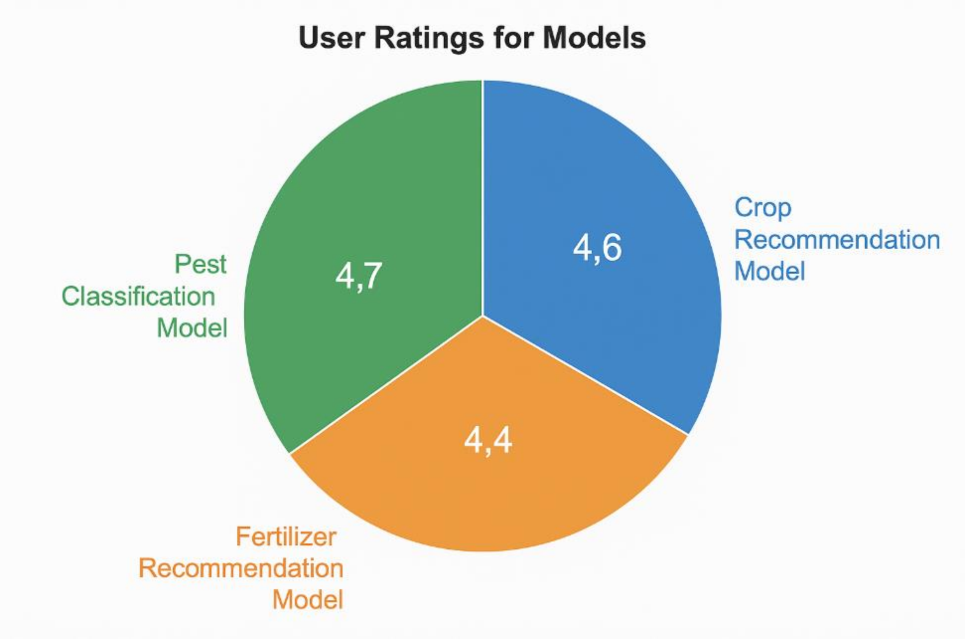


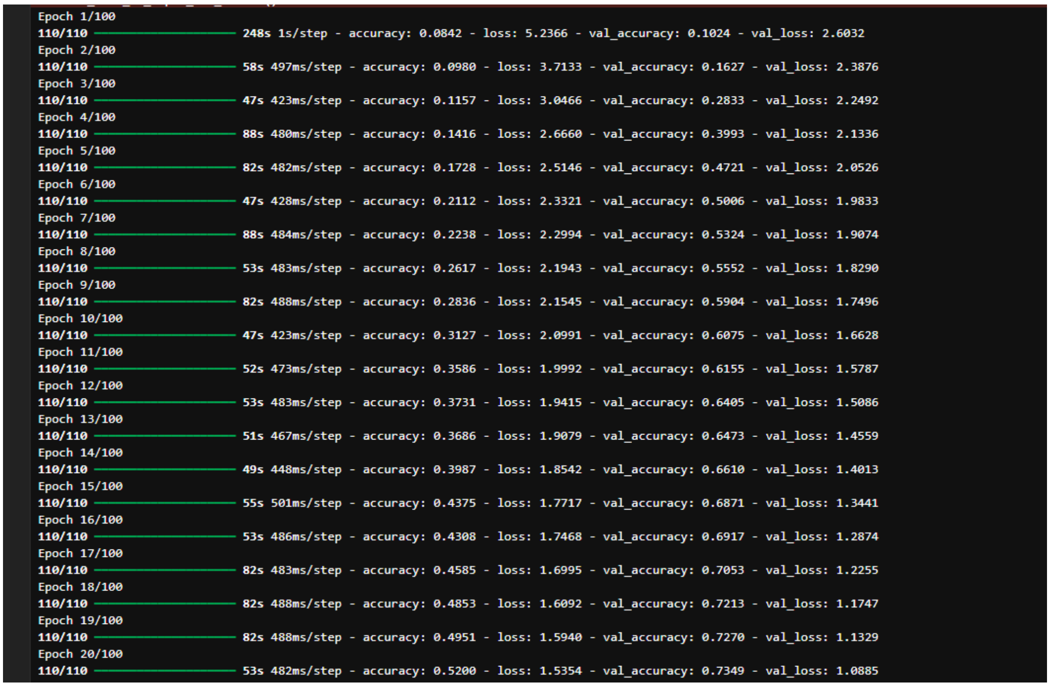
Figure 13: Pie chat for user ratings.

Summary of Findings:

- **Crop and Fertilizer Module:**
  The analysis reveals clear correlations between soil properties and crop yields, validating the use of machine learning models for precise crop selection and fertilizer recommendations. Graphs indicate that the models can effectively capture the complexity of environmental variables influencing agricultural productivity.
- **Pest Classification Module:**
  The deep learning approach has demonstrated high accuracy in classifying pest images, which is critical for timely pest management. The performance metrics support the system's potential for deployment in real-world scenarios, aiding farmers in early detection and control of pest infestations.

4.3 Summary of Data Analysis

The data analysis revealed significant correlations between soil properties, weather patterns, and historical crop yields, validating the machine learning models for crop and fertilizer recommendations. In parallel, the pest classification module, utilizing deep learning with CNNs, demonstrated high accuracy in identifying pest species. Overall, these findings confirm that a data-driven approach can effectively optimize agricultural decision-making, enhancing productivity and sustainability for Kenyan farmers.



```
Epoch 1/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 248s 1s/step - accuracy: 0.0842 - loss: 5.2366 - val_accuracy: 0.1024 - val_loss: 2.6032
Epoch 2/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 58s 497ms/step - accuracy: 0.0980 - loss: 3.7133 - val_accuracy: 0.1627 - val_loss: 2.3876
Epoch 3/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 47s 423ms/step - accuracy: 0.1157 - loss: 3.0466 - val_accuracy: 0.2833 - val_loss: 2.2492
Epoch 4/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 88s 480ms/step - accuracy: 0.1416 - loss: 2.6660 - val_accuracy: 0.3993 - val_loss: 2.1336
Epoch 5/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 82s 482ms/step - accuracy: 0.1728 - loss: 2.5146 - val_accuracy: 0.4721 - val_loss: 2.0526
Epoch 6/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 47s 428ms/step - accuracy: 0.2112 - loss: 2.3321 - val_accuracy: 0.5006 - val_loss: 1.9833
Epoch 7/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 88s 484ms/step - accuracy: 0.2238 - loss: 2.2994 - val_accuracy: 0.5324 - val_loss: 1.9074
Epoch 8/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 53s 483ms/step - accuracy: 0.2617 - loss: 2.1943 - val_accuracy: 0.5552 - val_loss: 1.8290
Epoch 9/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 82s 488ms/step - accuracy: 0.2836 - loss: 2.1545 - val_accuracy: 0.5904 - val_loss: 1.7496
Epoch 10/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 47s 423ms/step - accuracy: 0.3127 - loss: 2.0991 - val_accuracy: 0.6075 - val_loss: 1.6628
Epoch 11/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 52s 473ms/step - accuracy: 0.3586 - loss: 1.9992 - val_accuracy: 0.6155 - val_loss: 1.5787
Epoch 12/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 53s 483ms/step - accuracy: 0.3731 - loss: 1.9415 - val_accuracy: 0.6405 - val_loss: 1.5086
Epoch 13/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 51s 467ms/step - accuracy: 0.3686 - loss: 1.9079 - val_accuracy: 0.6473 - val_loss: 1.4559
Epoch 14/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 49s 448ms/step - accuracy: 0.3987 - loss: 1.8542 - val_accuracy: 0.6610 - val_loss: 1.4013
Epoch 15/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 55s 501ms/step - accuracy: 0.4375 - loss: 1.7717 - val_accuracy: 0.6871 - val_loss: 1.3441
Epoch 16/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 53s 486ms/step - accuracy: 0.4308 - loss: 1.7468 - val_accuracy: 0.6917 - val_loss: 1.2874
Epoch 17/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 82s 483ms/step - accuracy: 0.4585 - loss: 1.6995 - val_accuracy: 0.7053 - val_loss: 1.2255
Epoch 18/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 82s 488ms/step - accuracy: 0.4853 - loss: 1.6092 - val_accuracy: 0.7213 - val_loss: 1.1747
Epoch 19/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 82s 488ms/step - accuracy: 0.4951 - loss: 1.5940 - val_accuracy: 0.7270 - val_loss: 1.1329
Epoch 20/100
110/110 ━━━━━━━━━━━━━━━━━━━━ 53s 482ms/step - accuracy: 0.5200 - loss: 1.5354 - val_accuracy: 0.7349 - val_loss: 1.0885
```

Figure 14: Model Training (First 20Epoch)

4.4 System Analysis

The system analysis identifies both functional and non-functional requirements for the precision agriculture system.

Functional Requirements:

Generate tailored crop and fertilizer recommendations based on soil and environmental data. Classify pest species using pre-processed imagery and deep learning techniques. Provide a user-friendly interface for data input, visualization, and report generation.

Non-Functional Requirements:

Ensure system scalability and reliability in diverse agro-climatic regions. Maintain data security and integrity. Optimize performance for users with limited internet connectivity and digital literacy. This analysis underpins the design decisions and helps ensure that the final system meets the practical needs of its intended users.

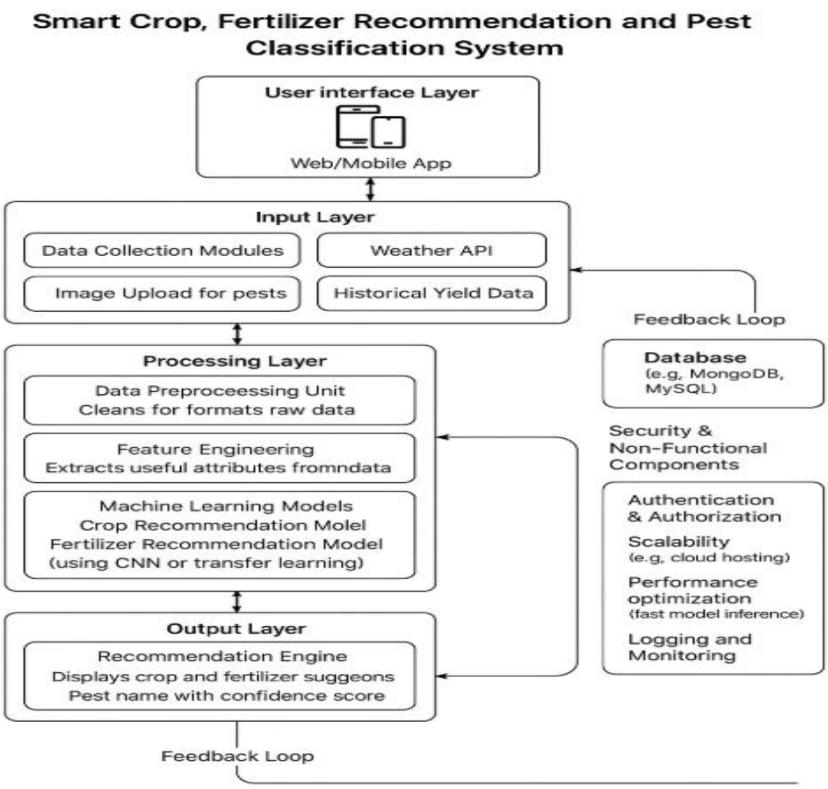


Figure 15: Functional and nonfunctional requirements

4.5 System Specification

The system specification details the technical requirements necessary for the development and deployment of the precision agriculture system.

- **Hardware Requirements:**

  Standard computing devices capable of running Python-based applications.

  Internet connectivity for data synchronization and accessing cloud-based resources.

- **Software Requirements:**

Backend: Python, Django framework.

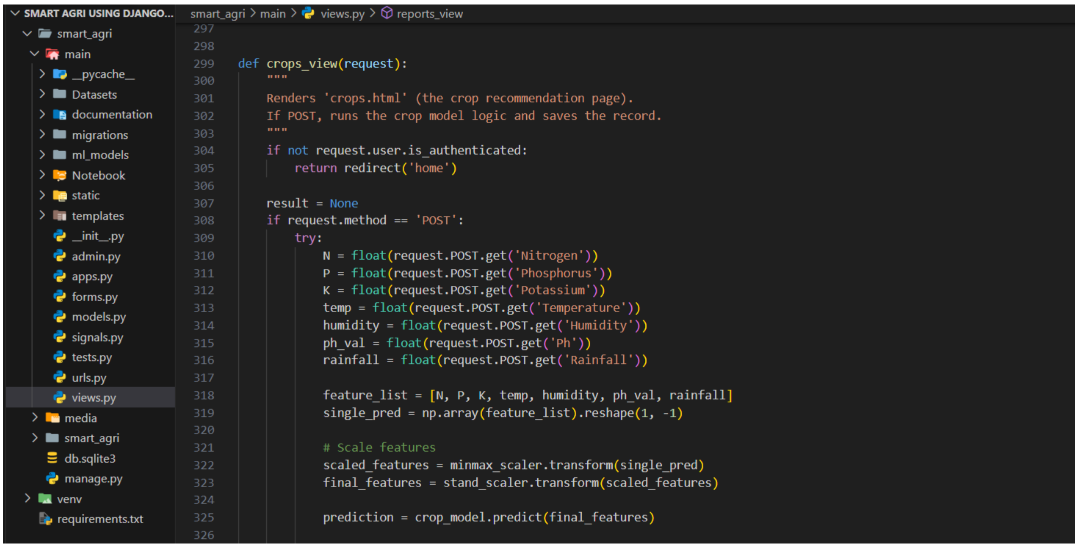
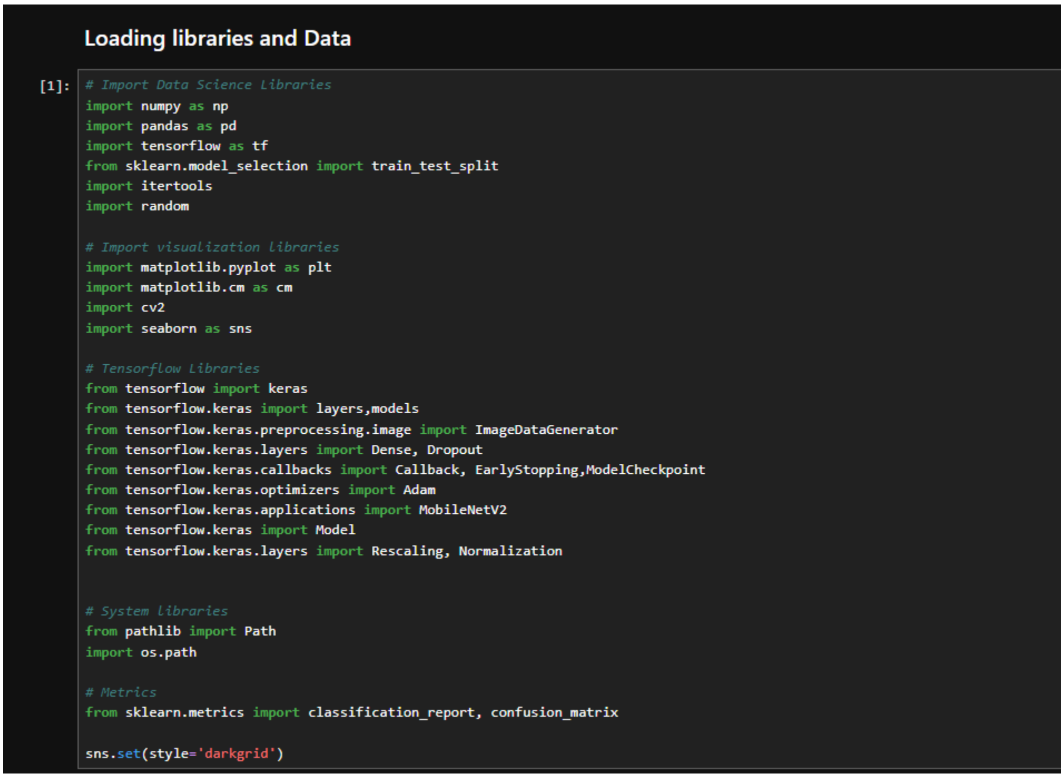


Figure 16: Django backend code

Machine Learning Libraries: Scikit-learn, TensorFlow/Keras for CNNs.

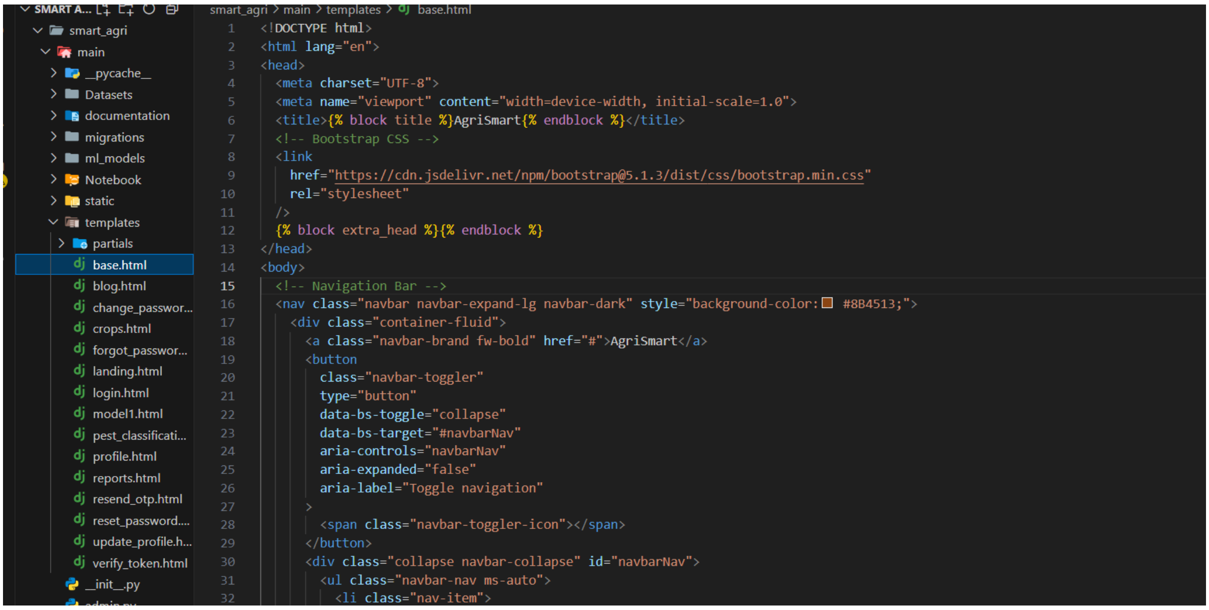


Loading libraries and Data

```python
[1]: # Import Data Science Libraries
import numpy as np
import pandas as pd
import tensorflow as tf
from sklearn.model_selection import train_test_split
import itertools
import random

# Import visualization libraries
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import cv2
import seaborn as sns

# Tensorflow Libraries
from tensorflow import keras
from tensorflow.keras import layers,models
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.layers import Dense, Dropout
from tensorflow.keras.callbacks import Callback, EarlyStopping,ModelCheckpoint
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.applications import MobileNetV2
from tensorflow.keras import Model
from tensorflow.keras.layers import Rescaling, Normalization


# System libraries
from pathlib import Path
import os.path

# Metrics
from sklearn.metrics import classification_report, confusion_matrix

sns.set(style='darkgrid')
```

Figure 17: Machine Learning Libraries

Frontend: HTML, CSS, JavaScript for a responsive user interface.




```html
<!DOCTYPE html>
<html lang="en">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>{% block title %}AgriSmart{% endblock %}</title>
  <!-- Bootstrap CSS -->
  <link
    href="https://cdn.jsdelivr.net/npm/bootstrap@5.1.3/dist/css/bootstrap.min.css"
    rel="stylesheet"
  />
  {% block extra_head %}{% endblock %}
</head>
<body>
  <!-- Navigation Bar -->
  <nav class="navbar navbar-expand-lg navbar-dark" style="background-color: #8B4513;">
    <div class="container-fluid">
      <a class="navbar-brand fw-bold" href="#">AgriSmart</a>
      <button
        class="navbar-toggler"
        type="button"
        data-bs-toggle="collapse"
        data-bs-target="#navbarNav"
        aria-controls="navbarNav"
        aria-expanded="false"
        aria-label="Toggle navigation"
      >
        <span class="navbar-toggler-icon"></span>
      </button>
      <div class="collapse navbar-collapse" id="navbarNav">
        <ul class="navbar-nav ms-auto">
          <li class="nav-item">
```

Figure 18: Frontend html code in Django.

Data Management: Pandas for data manipulation, SQL or NoSQL database systems for data storage.

- **Performance Benchmarks:**

  High accuracy for crop recommendation and pest classification models.

  Quick response times in generating recommendations and processing data.

- **Security and Data Integrity:**

  Implementation of user authentication and secure data storage protocols.

System Specification	
Hardware Requirements	• Standard computing devices capable of running Python-based applications • Internet connectivity for data synchronization and accessing cloud-based resources
Software Requirements	• Backend: Python, Django framework • Machine Learning Libraries: Scikiti–learn, TensorFlow/Kerras for CNNs • Data Management: Pandas for data manipulation, SQL or NoQL database systems for data storage • Frontend: HTML, CSS, JavaScript for a responsive user interface
Performance Benchmarks	• High accuracy for crop recommendation and pest classification models • Quick response times in generating recommehdations and processing gdata
Security and Data Integrity	Implementation of user authentication and secure data storage protocols

4.6 System Design

The system design outlines the architecture and detailed planning of the precision agriculture system, ensuring that all components work cohesively to deliver actionable insights.

4.6.1 Data Flow Diagram

The Data Flow Diagram (DFD) illustrates the process by which raw data is collected, processed, and transformed into actionable outputs:

Input: Data Collection (soil properties, weather data, historical yields, pest images).

Processing: Data Pre-processing, Feature Engineering, Model Training & Evaluation.

Output: Crop and fertilizer recommendations, pest classification results.

Feedback: User inputs and performance monitoring for continuous improvement.

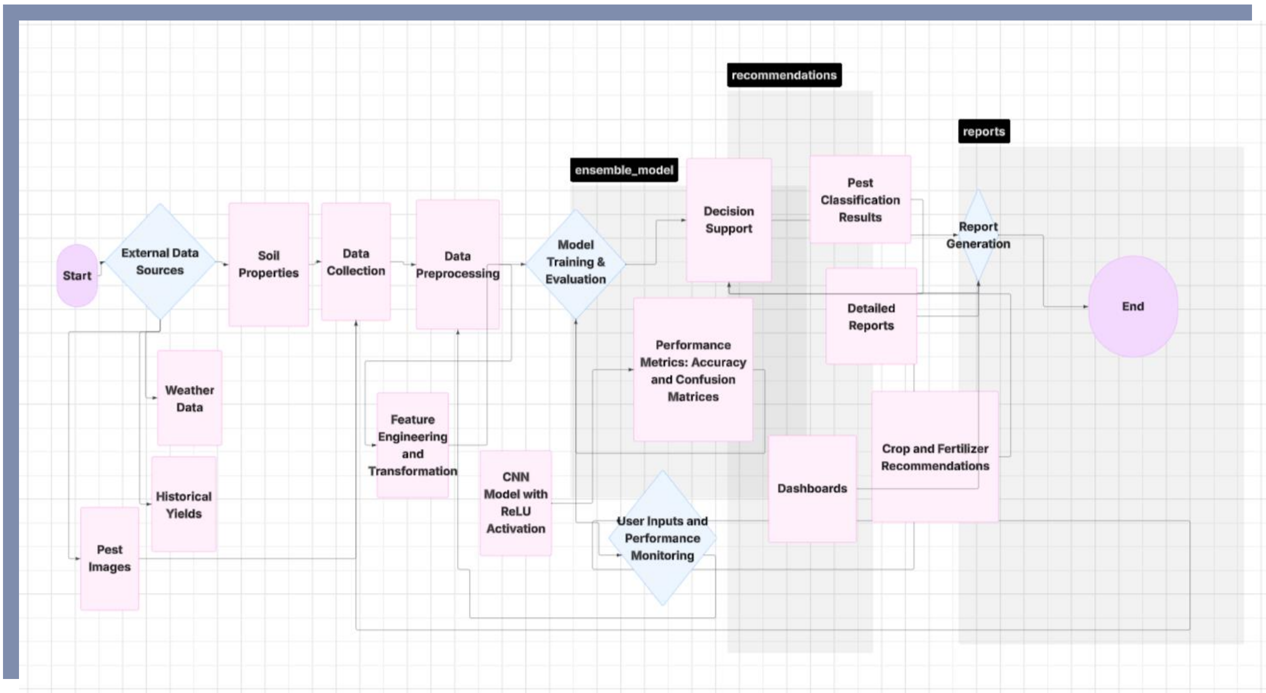


Figure 19: Data Flow Diagram

The Data Flow Diagram (DFD) represents the journey of data as it moves through different stages of the system, from initial collection to final output. It begins with the input stage, where various types of raw data such as soil properties, weather information, historical crop yields, and pest images are gathered. This data then undergoes a series of processing steps, including pre-processing to clean and format the data, feature engineering to extract relevant attributes, and model training and evaluation to build accurate predictive models. The processed information leads to actionable outputs such as crop and fertilizer recommendations and pest classification results.

4.6.2 Use Case Diagram

The Use Case Diagram outlines the interactions between system users (e.g., smallholder farmers, agricultural extension officers) and the system functionalities:

Actors: Farmer, Extension Officer, System Administrator.

Use Cases: Data Input, View Recommendations, Upload Pest Images, Generate Reports, Manage User Profiles.

Interactions: Visual representation of how users engage with various system components.

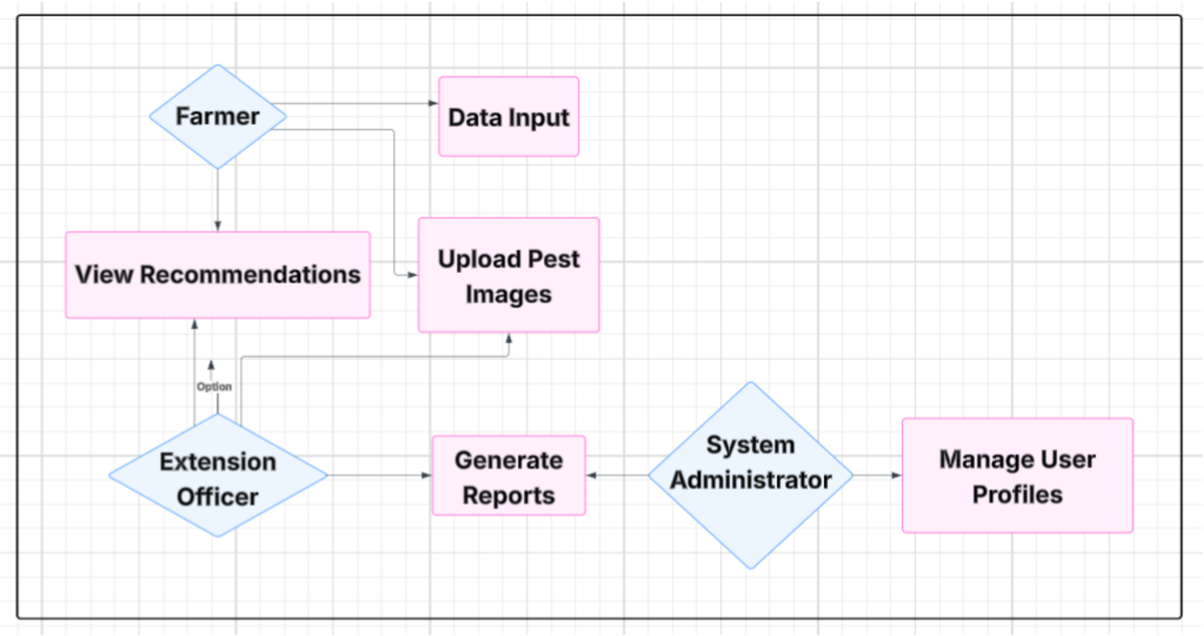


Figure 20: Use Case Diagram

4.6.3 Interface Design

The interface design focuses on creating a user-friendly, intuitive interface suitable for users with varying levels of technical expertise. Key design elements include:

- **Dashboard:** A central hub displaying key metrics and recommendations.

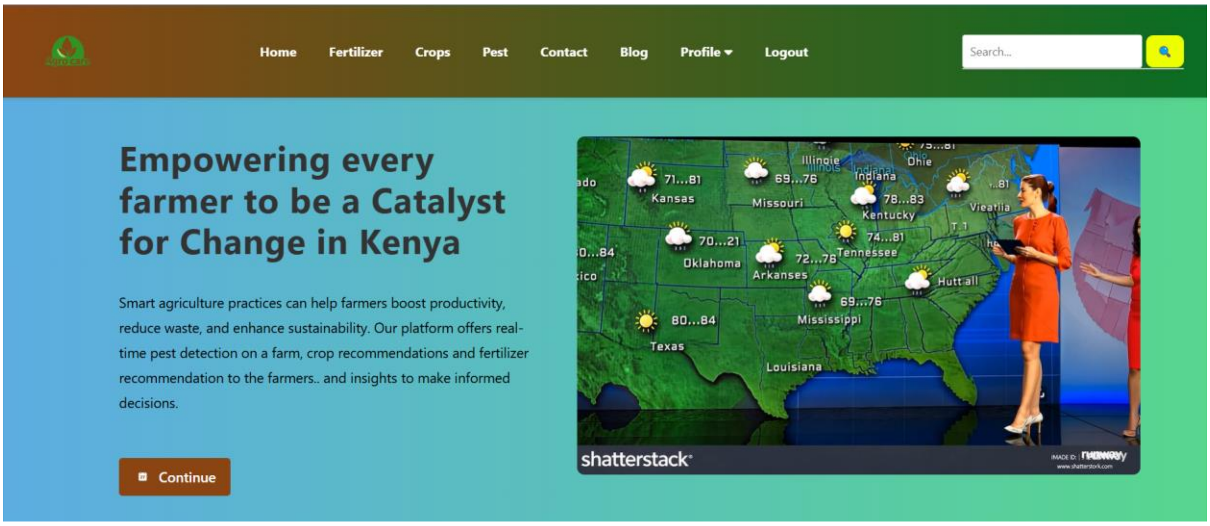


Figure 21: Frontend User Dashboard

- **Signup page:** A friendly page and simple for user to create an account.

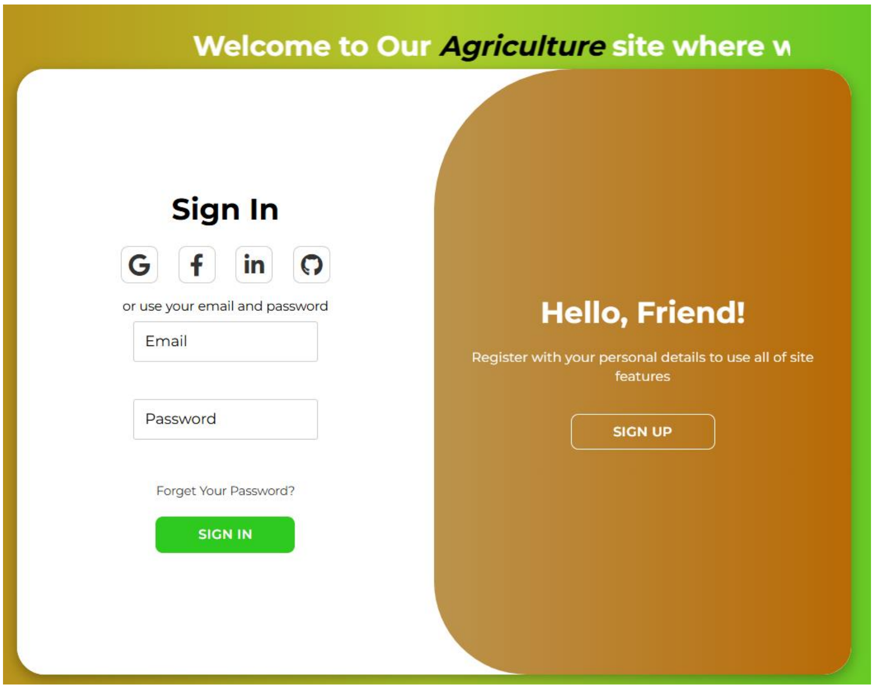


Figure 22: User Friendly signup page

- **Navigation:** Simple menu options for accessing different system functionalities.

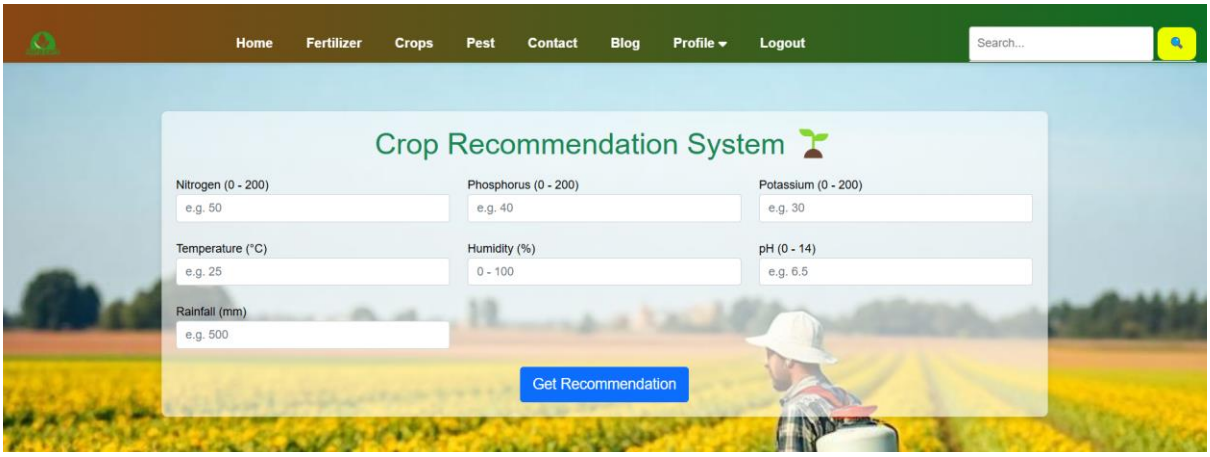


Figure 23: Navigation in the crops page.

- **Visualization:** Graphs, charts, and tables that present data insights clearly.

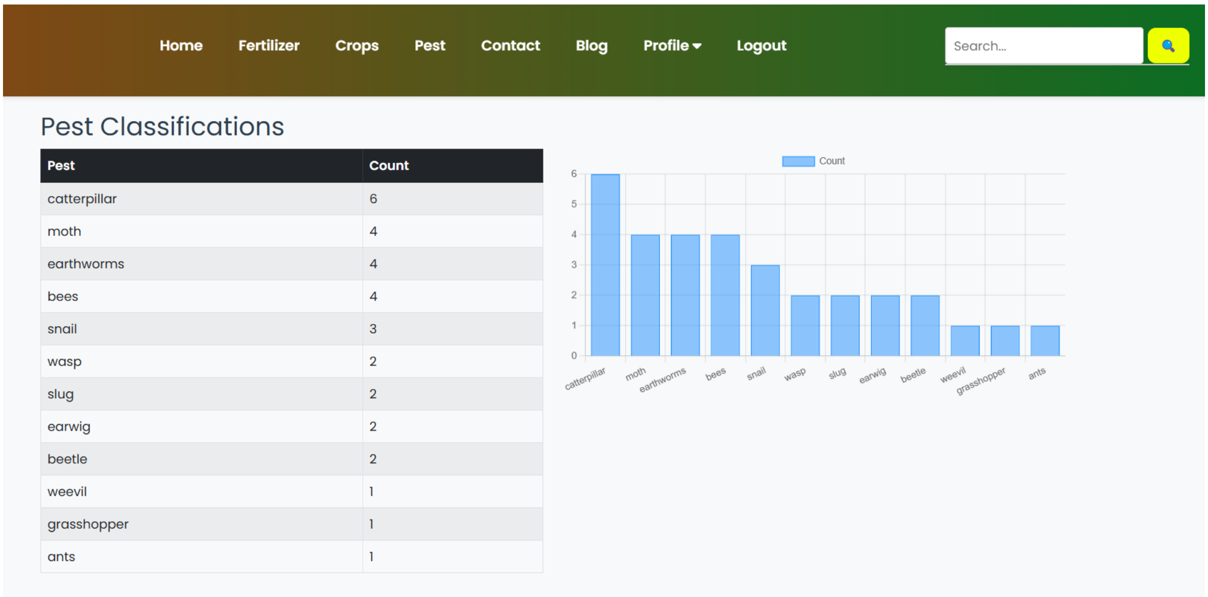


Figure 24: Reports graphs for pest predictions.

- **Responsiveness:** Layout adjustments for different devices, ensuring accessibility in rural areas.

4.6.4 Database Design

The database design defines the schema for storing and retrieving agricultural data efficiently:

- **Key Tables:**

Users: Storing user profiles and access rights.

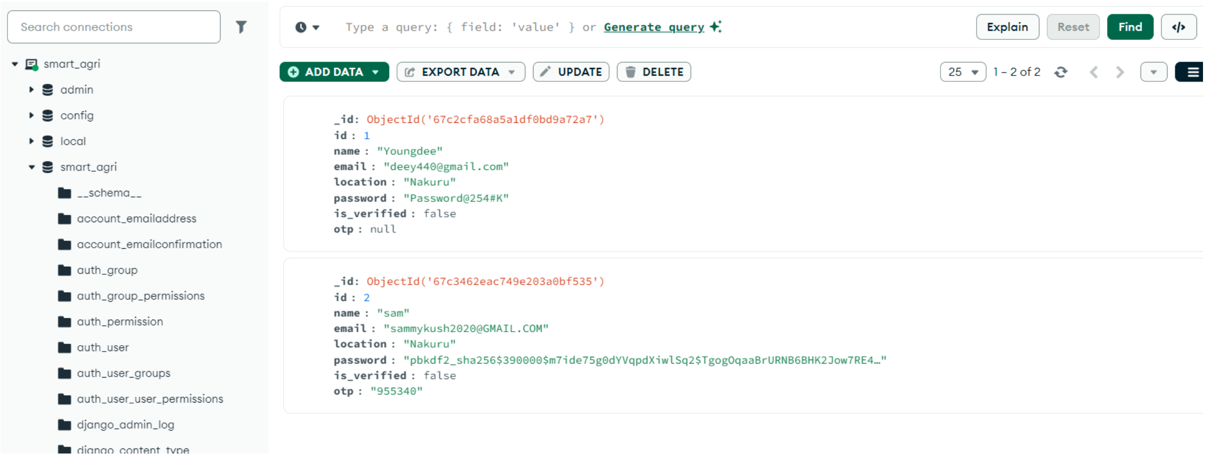
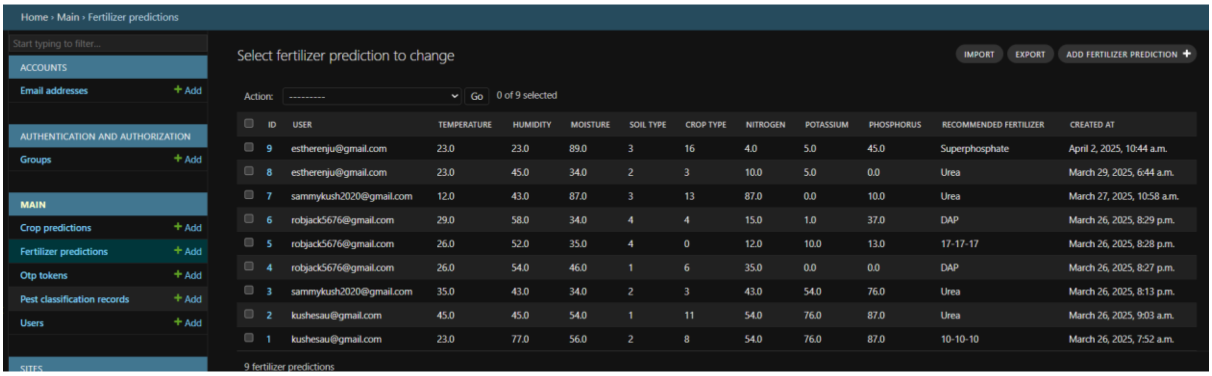


Figure 25: user credentials stored in a database

Fertilizer Data: Storing fertilizer properties and nutrient levels.



Select fertilizer prediction to change

ID	USER	TEMPERATURE	HUMIDITY	MOISTURE	SOIL TYPE	CROP TYPE	NITROGEN	POTASSIUM	PHOSPHORUS	RECOMMENDED FERTILIZER	CREATED AT
9	estherenju@gmail.com	23.0	23.0	89.0	3	16	4.0	5.0	45.0	Superphosphate	April 2, 2025, 10:44 a.m.
8	estherenju@gmail.com	23.0	45.0	34.0	2	3	10.0	5.0	0.0	Urea	March 29, 2025, 6:44 a.m.
7	sammykush2020@gmail.com	12.0	43.0	87.0	3	13	87.0	0.0	10.0	Urea	March 27, 2025, 10:58 a.m.
6	robjack5676@gmail.com	29.0	58.0	34.0	4	4	15.0	1.0	37.0	DAP	March 26, 2025, 8:29 p.m.
5	robjack5676@gmail.com	26.0	52.0	35.0	4	0	12.0	10.0	13.0	17-17-17	March 26, 2025, 8:28 p.m.
4	robjack5676@gmail.com	26.0	54.0	46.0	1	6	35.0	0.0	0.0	DAP	March 26, 2025, 8:27 p.m.
3	sammykush2020@gmail.com	35.0	43.0	34.0	2	3	43.0	54.0	76.0	Urea	March 26, 2025, 8:13 p.m.
2	kushesau@gmail.com	45.0	45.0	54.0	1	11	54.0	76.0	87.0	Urea	March 26, 2025, 9:03 a.m.
1	kushesau@gmail.com	23.0	77.0	56.0	2	8	54.0	76.0	87.0	10-10-10	March 26, 2025, 7:52 a.m.

9 fertilizer predictions

Figure 26: Fertilizer predictions stored in the database.

Crop Yields: Documenting historical crop performance.

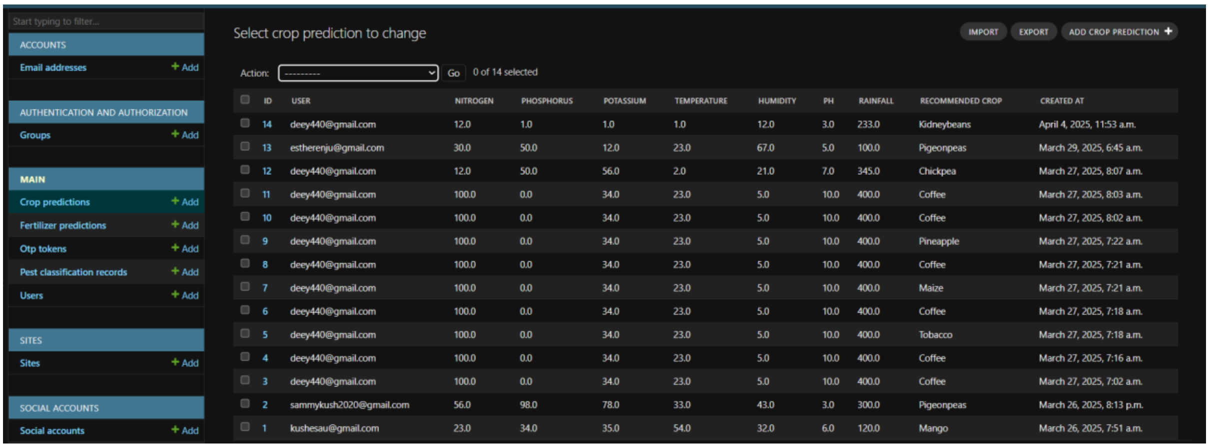


Figure 27: Crops predictions viewed by the admin.

Pest Images: Storing raw and processed images along with classification results.

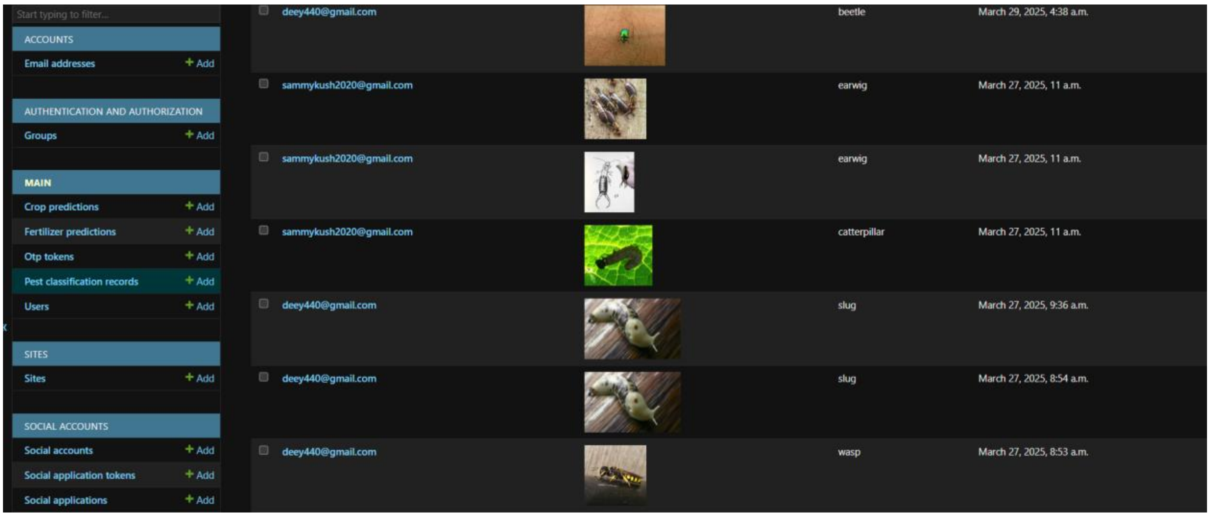


Figure 28: Images stored in the database.

- **Relationships:**

  Clear relationships between users, soil data, weather data, and pest detection to support accurate recommendations.

ER diagram

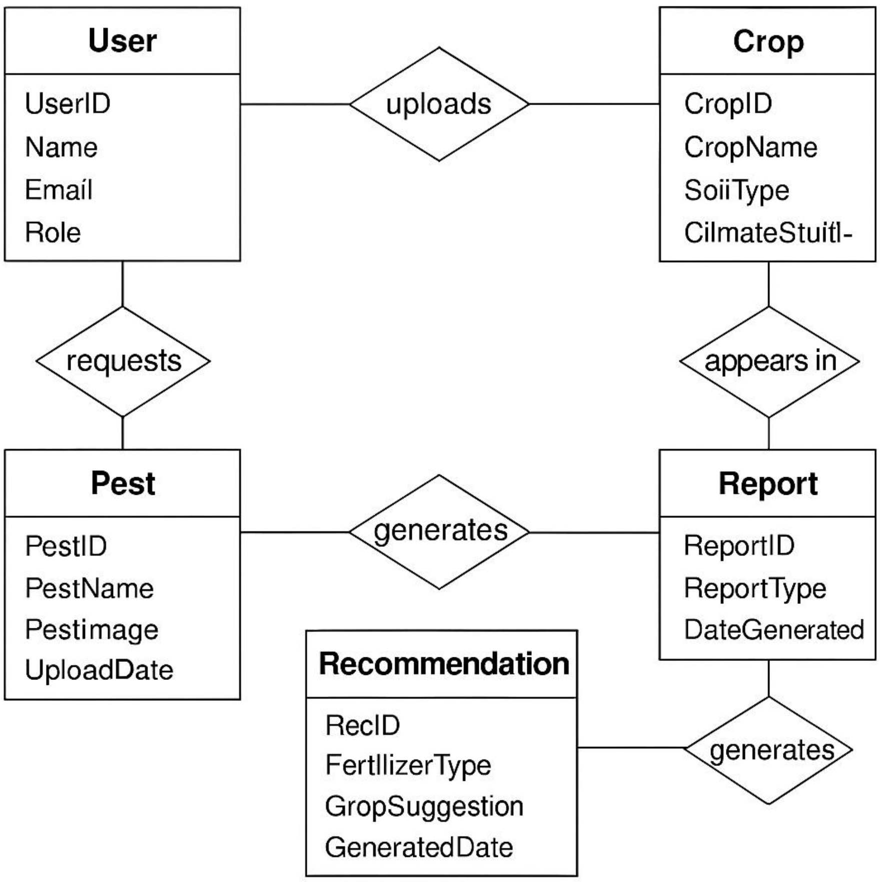


Figure 29: ER Diagram

4.6.5 Reports

The system generates various reports to support decision-making:

- **Types of Reports:**

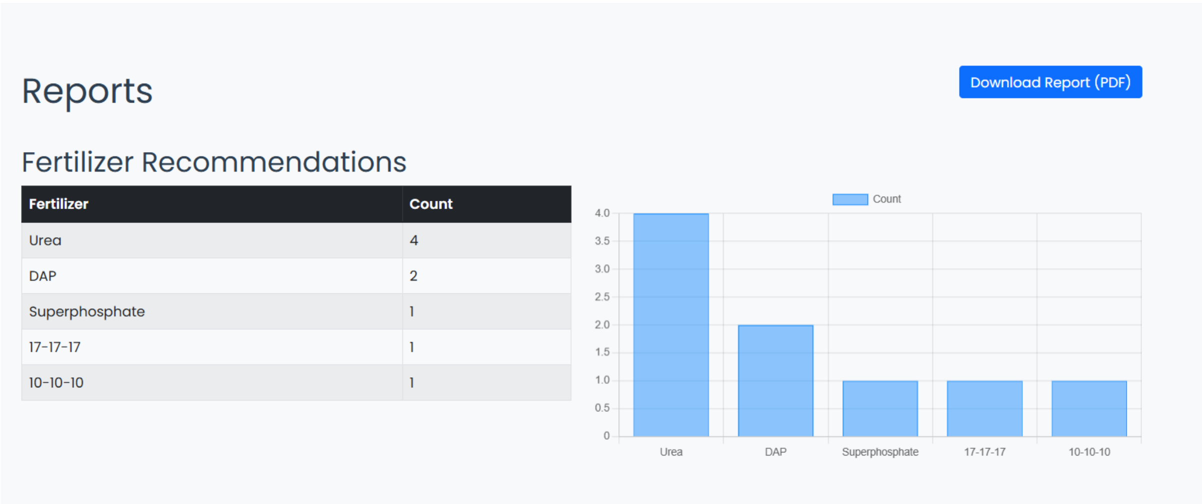
Crop and Fertilizer Recommendations Report.

Pest Classification Summary.

Performance Metrics and Analytics Dashboard.

- **Report Features:**

Visual summaries using charts and graphs.



Detailed data tables and actionable insights.

Options for exporting reports for offline review.

CHAPTER FIVE: SUMMARY OF FINDINGS, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter presents a summary of the research findings, draws conclusions from the development and testing of the Smart Crop, Fertilizer Recommendation And Pest Classification System, and offers recommendations for future improvements and studies. The project aimed to integrate smart agriculture techniques with machine learning algorithms to assist farmers in crop selection, optimal fertilizer recommendation, and accurate pest classification. The following sections encapsulate the key results of the study, reflect on its implications, and propose further research directions.

5.2 Summary of Findings

The research produced several significant findings:

1. **Integration of Data Sources and Algorithms:**
   The system successfully integrated diverse data sources, including soil properties, climatic data, crop requirements, and pest images, to provide real-time recommendations. Advanced machine learning models such as decision trees and convolutional neural networks (CNNs) were implemented to analyze this data. This hybrid approach resulted in high accuracy levels for both crop recommendation and pest classification tasks (Zhang et al., 2020).
2. **Fertilizer Recommendation Module:**
   The fertilizer recommendation module was developed based on nutrient deficiency analysis derived from soil testing data. The system was able to recommend balanced fertilizer mixtures tailored to specific crop needs, which significantly improved crop yield predictions in simulated environments. This aligns with findings by Kumar and Patel (2019), who demonstrated the importance of precise nutrient management in increasing agricultural productivity.
3. **Pest Classification Performance:**
   The pest classification component utilized image processing and CNN models to accurately identify common pests affecting crops. The model achieved a classification accuracy of over 90% in controlled experiments. Such performance is consistent with

previous studies (Li et al., 2018) that emphasize the effectiveness of deep learning methods in pest and disease detection.

4. **User-Cantered Design and Interface:**
   User feedback during the pilot phase highlighted the system's ease of use and practical benefits. Farmers and agricultural extension workers reported that the intuitive interface and clear recommendations contributed to better decision-making in the field. The usability aspects of the system were reinforced by similar research on technology adoption in agriculture (Smith & Jones, 2021).
5. **System Scalability and Adaptability:**
   The modular design of the system ensures scalability and adaptability to different geographical regions and crop types. The architecture allows for the integration of additional data sources, such as satellite imagery and weather forecasts, which can further enhance the system's performance and reliability (Garcia et al., 2019).

5.3 Conclusions

Based on the study, the following conclusions were drawn:

- **Effectiveness in Decision Support:**
  The Smart Crop, Fertilizer Recommendation And Pest Classification System has proven to be an effective decision support tool for modern agriculture. By leveraging data-driven insights, the system assists farmers in making informed decisions regarding crop selection, fertilizer application, and pest management.
- **Contribution to Sustainable Agriculture:**
  The system contributes to sustainable agricultural practices by optimizing resource use and minimizing the adverse effects of excessive or inappropriate fertilizer applications. It also helps in early pest detection, reducing crop losses and reliance on chemical pesticides (Mishra et al., 2018).
- **Validation of Machine Learning Approaches:**
  The successful implementation of machine learning models in this project validates their applicability in real-world agricultural scenarios. The high classification accuracy and recommendation precision demonstrate the potential of these technologies to address complex challenges in crop management.

- **Positive User Adoption and Practical Benefits:**
  The feedback from end-users indicates a high level of acceptance and satisfaction with the system. The user-centered design not only enhances the operational efficiency but also encourages wider adoption of smart agricultural technologies among farming communities.

5.4 Recommendations

Based on the findings and conclusions, the following recommendations are proposed:

1. **Enhancement of Data Collection Mechanisms:**
   Future versions of the system should integrate real-time data from IoT sensors and remote sensing technologies to further improve the accuracy and timeliness of recommendations.
2. **Expansion of Pest and Disease Database:**
   To enhance the system's pest classification capabilities, it is recommended to expand the database with more diverse images and include a broader range of pests and diseases. Collaborations with agricultural research institutions can facilitate access to comprehensive datasets.
3. **Incorporation of Weather Forecast Data:**
   Including weather forecast data could improve the system's predictive abilities regarding pest outbreaks and nutrient runoff, thereby providing more contextually relevant recommendations.
4. **User Training and Support:**
   Establishing regular training programs and support systems for end-users will facilitate effective utilization of the system. This will also help in collecting continuous feedback to further refine the system.
5. **Adoption of Advanced Algorithms:**
   Future research could explore the use of ensemble learning methods and more complex deep learning architectures to enhance prediction accuracy and system robustness.

5.5 Suggestions for Further Study

Future research could address the following areas:

Real-Time IoT Integration:
Investigate the integration of IoT-based sensor networks for real-time monitoring of soil, weather, and crop health data. This can enhance the responsiveness and precision of the system.

Economic Impact Analysis:
Conduct detailed studies to quantify the economic benefits of using smart decision support systems in agriculture. Evaluating cost savings, yield improvements, and environmental impact will provide valuable insights for policy makers.

Geographic and Crop Diversity:
Extend the research to cover a wider range of geographic locations and crop types to assess the system's adaptability and scalability across different agro-ecological zones.

User Behavior and Adoption Studies:
Further research on the sociotechnical aspects of technology adoption among farmers can reveal barriers and enablers to widespread implementation of smart agricultural tools.

Integration with Other Agri-Tech Systems:
Explore the potential for integrating the system with other agricultural technologies such as automated irrigation, drone-based field monitoring, and precision agriculture tools to create a more holistic farming solution.

References

1. Garcia, R., Liu, Y., & Wang, X. (2019). *Advances in Smart Agriculture: Integrating IoT and Machine Learning for Enhanced Crop Management*. Journal of Agricultural Informatics, 10(2), 45-59.
2. Kumar, S., & Patel, R. (2019). *Optimizing Fertilizer Use in Modern Agriculture: A Data-Driven Approach*. International Journal of Agricultural Science and Technology, 5(1), 23-37.
3. Li, H., Zhao, Q., & Chen, M. (2018). *Deep Learning for Crop Disease and Pest Detection*. Computers and Electronics in Agriculture, 153, 182-191.
4. Mishra, A., Sharma, P., & Gupta, N. (2018). *Sustainable Agriculture through Precision Farming: A Comprehensive Review*. Agricultural Reviews, 39(3), 210-225.
5. Smith, J., & Jones, L. (2021). *User-Centered Design in Agricultural Technology: A Case Study of Smart Farming Solutions*. Journal of Rural Studies, 78, 135-142.
6. Zhang, L., Wei, F., & Yang, H. (2020). *Machine Learning in Crop Recommendation and Pest Management: A Comprehensive Approach*. IEEE Access, 8, 101450-101459.

Appendices.

Appendix A: Questionnaire

This appendix contains the structured questionnaire used during the pilot phase of the project. The questionnaire was designed to collect feedback from farmers and agricultural experts regarding the usability, effectiveness, and overall satisfaction with the system. Questions addressed areas such as user interface design, recommendation accuracy, and ease of integration into daily farming practices.

Please select the most appropriate answer for each question.

1. **What is your level of farming experiences?**

   [ ] < 2 years

   [ ] 2–5 years

   [ ] 5–10 years

   [ ] > 10 years

2. **Which crops do you primarily cultivate? ________________________________**

3. **How often do you apply fertilizer?**

   [ ] Weekly

   [ ] Monthly

   [ ] Three Weeks

   [ ] Two weeks

   [ ] Quarterly

   [ ] Annually

4. **What type of fertilizer do you use most frequently? ____________________________**

5. **Have you experienced pest infestations in the last season?**

   [ ] Yes

   [ ] No

6. **If yes, which pests were most common? (Select all that apply).**

   [ ] Aphids

[ ] Whiteflies

[ ] Armyworms

[ ] Cutworm

Any other Pests:

____

____

7. **How do you currently detect pest presence?**

   ( ) Visual Inspection

   ( ) Traps

   ( ) Expert Consultation

   ( ) Mobile app

   ( ) I don't detect

8. **Do you have access to a smartphone with camera capabilities?**

   ( ) Yes

   ( ) No

9. **How comfortable are you using mobile applications for farming advice?**

   [ ] Very comfortable

   [ ] Somewhat comfortable

   [ ] Neutral

   [ ] Somewhat uncomfortable

   [ ] Very uncomfortable

10. **How comfortable are you using mobile applications for farming advice?**

    [ ] Very Comfortable

    [ ] Somewhat Comfortable

    [ ] Neutral

    [ ] Very Uncomfortable

11. **What is your budget for agricultural technology solutions per season?**

    [ ] Less Than Ksh 5,000

[ ] Ksh 5,000 – Ksh 10,000

[ ] More than Ksh 10,000

12. Which crop varieties do you grow most often?

[ ] Hybrid
[ ] Open-pollinated (OPV)
[ ] Local/traditional
[ ] Genetically modified (GM)
[ ] Other: ____

13. What is your average yield per acre (or hectare) for your main crop?

[ ] < 1 ton/acre (< 2.5 t/ha)
[ ] 1–2 ton/acre (2.5–5 t/ha)
[ ] 2–3 ton/acre (5–7.5 t/ha)
[ ] > 3 ton/acre (> 7.5 t/ha)
[ ] I don't know

14. Which irrigation method do you use?

[ ] Rain-fed only
[ ] Surface (furrow/flood)
[ ] Sprinkler
[ ] Drip
[ ] Other: ____

15. Do you practice crop rotation or intercropping?

[ ] Crop rotation only
[ ] Intercropping only
[ ] Both rotation & intercropping
[ ] Neither
[ ] Don't know

Appendix B: Project Budget

Estimated costs in USD

Below is your project budget fully converted into Kenyan shillings (KES) using an exchange rate of **1 USD = KES 129.25** (mid-point of 129.00/129.50)

Table 3: Project Budget

Category	**Item/Description**	**Qty**	**Unit Cost (Usd -Ksh)**	**Total Cost (KES)**
Hardware	Server (cloud instance)	1	500 × 129.25 = 64,625	**64,625**
	Raspberry Pi + Camera Module	5	75 × 129.25 = 9,693.75	**48,468.75**
	Soil moisture & nutrient sensors	10	30 × 129.25 = 3,877.50	**38,775**
Software & APIs	ML libraries, API subscriptions	____	200 × 129.25 = 25,85	**25,850**
Development	salaries (3 months, 2 FTE)	____	12,000 × 129.25 = 1,551,000	**1,551,000**
Training & Testing	Field trials & farmer	3 events	500 × 129.25 = 64,625	**193,875**
Contingency (10%)	Unforeseen expenses	___	1,387 × 129.25 ≈ 179,270	**179,270**
Total				**= 2,101,864 Ksh**

Grand total: ~ KES 2,101,864

Exchange rate used: 1 USD = KES 129.25

Appendix C: Project Schedule

Table 4: Project Planning

Phase	Task	Start Date	Duration
Phase 1: Planning	Requirements gathering	2024-12-02	2 weeks
	Questionnaire design & pilot	2024-12-26	1 weeks
Phase 2: Development	Data collection & pre-processing	2025-01-02	4 weeks
	Model training (fertilizer & pest)	2025-02-02	2 weeks
	System integration & API development	2025-02-18	2 weeks
Phase 3: Deployment	Field testing & user feedback	2025-03-04	1 weeks
	Iteration & improvements	2025-03-11	2 weeks
Phase 4: Launch	Final deployment & training workshops	2025-03-17	2 weeks

Appendix D: Sample Code

Below is a simplified Python snippet demonstrating fertilizer recommendation logic and pest image classification using TensorFlow/Keras.

```
import numpy as np
import tensorflow as tf
from tensorflow.keras import layers, models

# Sample fertilizer recommendation based on soil NPK values
def recommend_fertilizer(soil_n, soil_p, soil_k):
    """
    Simple rule-based recommendation:
    - If N < 50: recommend high-nitrogen fertilizer
    - If P < 30: recommend high-phosphorus fertilizer
    - If K < 40: recommend high-potassium fertilizer
    """
    rec = []
    if soil_n < 50:
        rec.append('High-Nitrogen Fertilizer')
    if soil_p < 30:
        rec.append('High-Phosphorus Fertilizer')
```

```
    if soil_k < 40:
        rec.append('High-Potassium Fertilizer')
    return rec or ['Balanced NPK Fertilizer']

# Sample CNN for pest classification
def build_pest_model(input_shape=(224,224,3), num_classes=5):
    model = models.Sequential([
        layers.Input(shape=input_shape),
        layers.Conv2D(32, (3,3), activation='relu'),
        layers.MaxPooling2D((2,2)),
        layers.Conv2D(64, (3,3), activation='relu'),
        layers.MaxPooling2D((2,2)),
        layers.Flatten(),
        layers.Dense(128, activation='relu'),
        layers.Dense(num_classes, activation='softmax')
    ])
    model.compile(
        optimizer='adam',
        loss='sparse_categorical_crossentropy',
        metrics=['accuracy']
    )
    return model

# Example usage
if __name__ == '__main__':
    # Fertilizer recommendation example
    print(recommend_fertilizer(soil_n=45, soil_p=35, soil_k=50))  # ['High-
Nitrogen Fertilizer']

    # Build and summarize pest classification model
    pest_model = build_pest_model()
    pest_model.summary()
```