MACHINE LEARNING IN PRECISION AGRICULTURE

BY

OTOGHILE ETINOSA SAMUEL

ENG1905129

DEPARTMENT OF COMPUTER ENGINEERING

FACULTY OF ENGINEERING

UNIVERSITY OF BENIN

BENIN CITY

FEBRUARY 2025

MACHINE LEARNING IN PRECISION AGRICULTURE

BY

OTOGHILE ETINOSA SAMUEL

ENG1905129

A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER ENGINEERING, FACULTY OF ENGINEERING, UNIVERSITY OF BENIN, BENIN CITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF BACHELOR OF ENGINEERING (B.ENG) DEGREE IN COMPUTER ENGINEERING.

FEBRUARY, 2025

CERTIFICATION

This project was carried out by **OTOGHILE ETINOSA SAMUEL, ENG1905129,** in the Department of Computer Engineering, Faculty of Engineering, University of Benin, Benin City, and is hereby certified in accordance with rules and regulations of the university of Benin.

______________________________ ______________________________

ENGR. DR. O. I. OMOIFO **DATE**

(PROJECT SUPERVISOR)

______________________________ ______________________________

ENGR. DR. MRS. O. OKOSUN **DATE**

(HEAD OF DEPARTMENT)

DEDICATION

This project is humbly dedicated to God Almighty for his faithfulness, insight and direction. May this project serve as a testament to Your boundless love, favor and grace and providence, inspiring others to seek Your guidance in all their endeavors. All glory and honor belong to You, now and forevermore. Amen.

It is also dedicated to our parents and siblings for their love and support.

ACKNOWLEDGEMENT

We attribute this work to GOD ALMIGHTY, the ultimate source of wisdom and guidance, who has been a constant companion and source of inspiration all through this undertaking. He has given us the strength and knowledge to complete this project, so we appreciate him for his exceeding grace.

We extend our profound appreciation to our esteemed project supervisor, Dr. Omoifo who provided invaluable supervision and mentorship during the course of this project. We are really thankful for your support, cooperation, guidance and suggestions.

Also, our sincere appreciation goes to our parents, for their unending love, encouragement, support, advice and their sacrifices that ensured our comfort and success throughout our academic journey.

This acknowledge will not be complete without expressing our appreciation to our beloved siblings.

ABSTRACT

Precision agriculture has emerged as a vital approach for improving agricultural efficiency and sustainability by leveraging advanced technologies. This project focuses on developing a machine learning model that utilizes historical data for precision agriculture applications. The system aims to assist farmers in making data-driven decisions by predicting suitable crops, forecasting crop yields, and detecting potential crop diseases.

The proposed solution integrates wireless sensor networks to collect environmental parameters such as rainfall, temperature, humidity, Potassium, Nitrogen and soil pH, which are combined with historical agricultural data. Machine learning models, including Voting Classifier (Support Vector Machines, Gaussian Naïve Bayes, Random Forest) and Linear regression model, are trained and deployed to analyze the data for actionable insights. The Voting classifier model achieved a 99.3% accuracy in predicting suitable crops, while the Linear regression model provided yield forecasts with an R^2 score of 0.91. The Convolutional Neural Network (CNN) detected crop diseases with an accuracy of 93.8%.

The system's implementation demonstrates the effectiveness of combining historical data with machine learning techniques to enhance precision agriculture practices. By providing accurate and timely information, this solution helps optimize crop selection, improve resource allocation, and mitigate the risk of crop diseases. Recommendations for integration with mobile applications, weather data integration and farmer education and training are proposed to further enhance the system's usability and impact.

This project offers a promising step toward sustainable and smart farming practices, contributing to food security and agricultural productivity.

TABLE OF CONTENTS

CHAPTER ONE

1.1 BACKGROUND OF THE STUDY

The agricultural sector has undergone a significant revolution with the advent of cutting-edge technologies, transitioning from traditional practices to data-driven approaches. A novel concept, precision agriculture, seeks to maximize crop yields while minimizing waste by harnessing innovative tools such as wireless sensor networks (WSNs) and machine learning (ML) algorithms. These networks comprise distributed sensors that monitor environmental parameters, providing farmers with real-time data to inform their decisions. However, the mere collection of data is insufficient; efficient processing and analysis are crucial. Machine learning algorithms play a vital role in analyzing large volumes of sensor data to predict crop yields, optimize irrigation schedules, and detect early signs of disease. The integration of machine learning-based WSNs in precision agriculture has the potential to reduce water consumption, optimize fertilizer application, and enhance overall productivity. This convergence of technologies can transform conventional agricultural practices into a more sustainable, resource-efficient, and environmentally friendly model.

Precision agriculture is an innovative farming management concept that leverages information technology to ensure crops and soils receive exactly what they need for optimal health and productivity. This concept relies on technology to gather, process, and analyze data from various sources, including sensors, drones, satellites, and other tools. The rapid advancement of technology has led to significant changes in agriculture, enabling real-time monitoring of environmental and soil parameters. However, raw data collected from sensors requires significant manual interpretation, which can be inefficient and prone to human error. Machine learning offers the ability to process vast amounts of data and identify patterns and relationships that would be difficult for humans to detect.

This study aims to develop a system that leverages both WSNs and machine learning to automate decision-making in precision agriculture. Such a system can improve the accuracy of irrigation and fertilizer applications, reduce waste, increase crop yield, and ultimately support more sustainable farming practices. Agriculture plays a vital role in global food security and economic development, especially in regions where the majority of the population depends on farming. However, climate change, resource limitations, and inefficient farming practices hinder productivity and sustainability. Precision agriculture has emerged as a solution to these challenges, leveraging advanced technologies such as WSNs and machine learning to optimize resource usage and enhance crop yield through data-driven decision-making.

Historical data, such as past crop yields, rainfall patterns, soil moisture levels, and temperature records, plays a critical role in enhancing the accuracy of predictive models. By combining real-time environmental data with historical trends, farmers can predict crop performance, optimize irrigation, and even forecast potential diseases, leading to increased efficiency in agricultural processes. The shift to precision agriculture involves using data to enhance farming processes, optimize resource use, and increase crop yields. Traditional farming practices largely relied on farmers' experience and seasonal observations, but this method lacks precision, especially in today's context of climate change and unpredictable weather patterns. Precision agriculture, when combined with WSNs and machine learning, offers a powerful solution to these challenges.

1.2 PROBLEM STATEMENT

Farmers face numerous challenges related to environmental unpredictability, improper irrigation schedules, and unoptimized fertilizer use, which contribute to reduced crop yields. The inability to anticipate environmental changes and plant health issues based on historical patterns results in suboptimal farming practices. Although wireless sensor networks provide real-time data on current farm conditions, they lack the capacity to leverage historical data for predicting future outcomes. Thus, there is a need for an integrated solution that combines real-time sensor data with historical agricultural data to improve decision-making and boost farm productivity.

Modern farming is confronted with increasing challenges, such as unpredictable weather, water scarcity, and soil degradation. Many farmers lack access to tools that provide timely, accurate information on which they can base decisions. While wireless sensor networks provide real-time information, they lack the predictive power that can be achieved by analysing historical data. For example, without historical weather and crop yield data, farmers are unable to anticipate droughts or disease outbreaks effectively. There is a pressing need for a system that integrates historical data with real-time information to offer actionable insights that can help farmers optimize resource usage and improve crop performance.

1.3 AIM AND OBJECTIVES

Aim:

To develop a machine learning-based system integrated with wireless sensor network for precision agriculture, utilizing historical and real time data to optimize resources usage, enhance crop monitoring, and support data-driven decision-making for improved agricultural productivity and stability.

Objectives:

i. Data collection: search for suitable historical environmental and crop data to support the machine learning model development.
ii. Data analysis: analyse historical data to identify patterns and trends influencing crop growth and farm conditions.

iii. Machine learning model development: develop and train machine learning models for predicting suitable crops, crop yield and pest/disease outbreaks.

iv. System integration: incorporate the machine learning models with real time data to enable dynamic decision-making.

v. Automation and decision support: provide automated recommendations for suitable crops and pest control based on predictive analytics.

vi. User interface development: design a user-friendly interface to present actionable insights to farmers for informed decision-making.

1.4 SCOPE OF THE STUDY

This research endeavors to elucidate the pivotal factors influencing crop growth, specifically examining the interplay between soil moisture, temperature, and humidity. By leveraging advanced Wireless Sensor Network technology, the study will gather precise data on a targeted crop type, such as wheat or rice, within a controlled and manageable field setting.

1.5 SIGNIFICANCE OF THE STUDY

This research presents a pioneering solution for the agricultural sector, offering a sophisticated system that monitors and optimizes resource allocation, thereby boosting productivity and minimizing waste. Moreover, it advances the field of precision agriculture by exploring the synergy between machine learning and Wireless Sensor Networks (WSNs), ultimately promoting data-informed farming practices. The significance of this study lies in its potential to address the inefficiencies of traditional farming methods by harnessing technology to make agriculture more precise and data-driven. Furthermore, this research contributes to the development of a framework that integrates WSNs with machine learning, which can be applied to a broad range of agricultural contexts.

The fusion of machine learning and WSNs in agriculture is poised to transform farming, making it more data-driven, efficient, and sustainable. This study will contribute to:

i. Optimizing water resource management through precision irrigation scheduling, ensuring timely and efficient water allocation.
ii. Enhancing crop yields by enabling data-driven interventions and timely decision-making.
iii. Reducing labour and input costs by automating crop management decision-making processes.

By leveraging cutting-edge technology to optimize crop production, this research promotes environmentally friendly and sustainable farming practices. The system's advanced data analysis capabilities enable efficient use of water and fertilizers, reduce crop diseases, and increase productivity. The insights generated by this system can empower farmers, particularly in developing countries, to adopt precision agriculture techniques. This study seeks to contribute to sustainable farming practices by incorporating historical data into precision agriculture systems. By improving the accuracy of predictions and recommendations, this system will help farmers minimize waste, optimize water usage, and enhance crop yields. The project's outcomes are particularly valuable for regions facing water scarcity and unpredictable weather conditions, where historical data can help mitigate risks. The integration of historical data with real-time WSN data provides a comprehensive understanding of the farming environment. This hybrid approach increases the accuracy of predictions, such as crop yields and irrigation needs, thereby enabling more efficient use of resources. In regions experiencing water scarcity or inconsistent weather, the system's predictive capabilities can help mitigate risks. The results of this study can lead to more informed farming practices, reduced waste, and improved crop productivity.

1.6 DEFINITION OF TERMS

i. Precision Agriculture (PA): A cutting-edge farming management approach that leverages data-driven insights to optimize resource allocation and maximize crop yields.

ii. Wireless Sensor Network (WSN): A decentralized network of sensors strategically deployed across a farm to gather real-time data on environmental conditions, including temperature, humidity, and soil moisture levels.

iii. Machine Learning (ML): A sophisticated subset of artificial intelligence that utilizes complex algorithms to analyze data, identify patterns, and make informed predictions or decisions.

iv. Soil Moisture: The critical amount of water present in the soil, essential for facilitating healthy plant growth and development.

v. Optimization: The systematic process of refining a system to achieve peak efficiency and effectiveness, thereby minimizing waste and maximizing output.

CHAPTER TWO

2.1 OVERVIEW OF PRECISION AGRICULTURE

Modern farming has witnessed a significant breakthrough with the advent of precision agriculture, a cutting-edge approach that focuses on providing crops and soil with tailored resources to optimize growth. At the heart of precision agriculture lies the strategic integration of advanced technologies, including geographic information systems (GIS), global positioning systems (GPS), and sensor networks. This fusion enables site-specific management practices, allowing farmers to tailor their approach to the unique conditions of each section of the field, including soil type and nutrient content.

In contrast to traditional farming methods, precision agriculture facilitates targeted applications of resources such as water, fertilizers, and pesticides, thereby minimizing waste and maximizing yields. The underlying theoretical framework involves the integration of wireless sensor networks with machine learning algorithms to optimize decision-making in precision agriculture. By collecting and processing real-time data from sensor nodes deployed across agricultural fields, farmers can predict irrigation needs, fertilizer application, and overall crop health. This approach ensures that resources are applied at the right time and in the right amount, thereby minimizing waste and maximizing crop yield.

Historical data analysis also provides valuable insights into recurring patterns and trends, informing future agricultural decisions. By examining past records of climate conditions, soil properties, irrigation practices, and crop performance, farmers can gain a long-term perspective, crucial for understanding recurring agricultural trends. This information enables farmers to make informed decisions about which crops are most suitable for specific environmental conditions, and how to optimize their farming practices accordingly.

2.2 MACHINE LEARNING IN AGRICULTURE

The significance of machine learning in agricultural data analysis cannot be overstated. Studies have consistently demonstrated the efficacy of supervised and unsupervised learning techniques in addressing issues such as crop yield prediction, disease detection, and resource optimization. Supervised Learning has been widely used for yield prediction and disease classification, while Unsupervised Learning models have aided in grouping similar farm zones based on soil properties. Deep Learning techniques, particularly Convolutional Neural Networks, have been applied to analyze satellite and drone imagery to assess crop health.

Various algorithms, including Support Vector Machines, Random Forest, and Neural Networks, have been successfully employed for tasks such as yield prediction, disease detection, and crop classification. For example, Support Vector Machines have proven highly effective in classifying crops, while Neural Networks have demonstrated remarkable accuracy in yield prediction. The ability of machine learning to process vast datasets and recognize patterns makes it an indispensable tool in precision agriculture.

Furthermore, time series models, regression analysis, and deep learning techniques have been widely used to predict agricultural outputs based on historical data and real-time environmental conditions, enabling farmers to forecast crop yields, irrigation needs, and disease outbreaks with remarkable accuracy. By applying machine learning algorithms to agricultural data, it is possible to predict future trends, and detect diseases earlier, thereby enhancing agricultural productivity and sustainability.

2.3 WIRELESS SENSOR NETWORKS IN AGRICULTURE

In agricultural settings, wireless sensor networks (WSNs) have emerged as a vital tool for monitoring and managing environmental conditions. These networks consist of dispersed sensors that track parameters such as temperature, soil moisture and humidity, transmitting the collected data wirelessly to a central hub for analysis. By leveraging WSNs, farmers can make informed, real-time decisions regarding crop choice, irrigation and fertilization, thereby optimizing crop yields and resource allocation. Spatially distributed sensor nodes in WSNs continuously monitor agricultural fields, capturing data on variables like soil moisture, temperature, humidity, and light intensity. This data is then transmitted wirelessly to a central station for storage and analysis, enabling farmers to make data-driven decisions.

The widespread adoption of WSNs in agriculture has enabled the monitoring of micro-environments within farms, providing real-time insights into environmental conditions. By combining these insights with historical data, farmers can gain a comprehensive understanding of their farm environment, enabling them to make informed decisions that balance current conditions with long-term predictability. informing decisions that optimize yields and resource usage.

The applications of WSNs in agriculture are diverse, including soil condition monitoring for irrigation management, crop disease detection, and weather tracking.

2.4 INTEGRATION OF WSN AND MACHINE LEARNING

The fusion of Wireless Sensor Networks (WSNs) and machine learning technology has the potential to revolutionize agricultural practices by enabling data-driven decision-making. By harnessing the power of advanced machine learning models, such as Random Forest, Decision Trees, and Deep Learning, vast amounts of sensor data can be analyzed to inform strategic decisions. This integration has the potential to elevate WSNs from mere data collection tools to sophisticated systems capable of providing actionable recommendations.

The synergy between WSNs and machine learning enables precision agriculture practices. WSNs provide real-time environmental and soil condition data, which is then processed by machine learning algorithms to generate predictive and prescriptive insights. For instance, a

WSN can track soil moisture levels and transmit the data to a machine learning model, which predicts the optimal irrigation schedule based on historical trends and current weather conditions.

This integration facilitates more automated decision-making in agriculture, reducing the need for manual data analysis and human judgment. Research has demonstrated that machine learning can significantly enhance the accuracy of predictions and improve the overall performance of WSNs in agricultural applications.

2.5 TOOLS AND TECHNOLOGIES

i. IoT and Wireless Sensors: Various sensors like soil moisture sensors, temperature sensors, and humidity sensors provide real-time data from the field. These sensors communicate wirelessly, using technologies such as ZigBee, Wi-Fi, or LoRaWAN, to transmit data to a central server.

ii. Machine Learning Libraries: Tools like Scikit-learn, TensorFlow, and Keras are used for building machine learning models that analyse both historical and real-time data. These libraries provide pre-built algorithms for regression, classification, and deep learning, making them ideal for predicting crop yields or detecting diseases.

iii. Time Series Forecasting Tools: Time series models like ARIMA (Autoregressive Integrated Moving Average) or LSTM (Long Short-Term Memory) are used to predict future agricultural outcomes based on historical data. These tools can capture patterns like seasonality and long-term trends, which are vital for precision agriculture.

2.6 RELATED WORKS

1. "Machine Learning Techniques in Wireless Sensor Network-Based Precision Agriculture" (2020)

- Authors: Raj et al.
- Summary: This study reviews various machine learning techniques, such as decision trees and SVM, applied to WSN data for soil monitoring and weather prediction in precision agriculture.
- Relation to the Topic: It highlights the potential of WSN and machine learning integration for efficient crop management but lacks a robust system for analysing historical data comprehensively.
- Gap Identified: Limited focus on integrating historical data for more accurate long-term predictions and decision-making.
- Improvement: This project aims to bridge this gap by utilizing historical data alongside real-time WSN data to improve predictive accuracy and support informed decision-making.

2. "IoT-Based Precision Agriculture Using Machine Learning Algorithms" (2021)

- Authors: Patel and Singh
- Summary: This research investigates IoT-enabled precision agriculture, using ML models to optimize irrigation and resource usage. The system focuses mainly on immediate decision-making.
- Relation to the Topic: The use of machine learning aligns with This project; however, the emphasis was primarily on IoT rather than long-term prediction.
- Gap Identified: Lack of historical data analysis for predictive analytics and decision-making over longer periods.

- Improvement: This project will enhance their work by focusing on predictive insights using historical and real-time data, leading to smarter resource management.

3. "A Wireless Sensor Network Approach for Smart Farming" (2022)

- Authors: Kumar et al.
- Summary: The study proposes a WSN framework for monitoring temperature, humidity, and soil moisture in real-time to improve agricultural productivity.
- Relation to the Topic: This work provides a robust WSN infrastructure but lacks machine learning-based data analysis for optimized decision-making.
- Gap Identified: No integration of advanced ML techniques for predictive insights.
- Improvement: By incorporating machine learning, this project will provide predictive analytics for pest control, irrigation, and resource management, enhancing efficiency.

4. "Deep Learning in Agriculture: A Survey" (2023)

- Authors: Gupta and Sharma
- Summary: This research focuses on deep learning applications in precision agriculture, particularly image-based crop disease detection.
- Relation to the Topic: Although related to machine learning in agriculture, it focuses mainly on image-based applications rather than environmental sensor data.
- Gap Identified: Limited use of sensor-based data analytics for resource management in agriculture.
- Improvement: This project integrates sensor data and machine learning for a more holistic approach to precision agriculture, extending beyond disease detection to include environmental monitoring and resource optimization.

2.7 META-ANALYSIS TABLE FOR "MACHINE LEARNING-BASED PRECISION AGRICULTURE USING HISTORICAL DATA"

S/N	Author(s)	Year	Title	Objective	Methodology	Key findings	Limitations	Relevance to current study
1	Raj et al	2020	machine learning techniques in wireless sensor network-based precision agriculture	To explore machine learning techniques for analyzing WSN data in precision agriculture	Used decision trees and SVM for soil monitoring and weather prediction	Improved real-time monitoring	No focus on historical data	Forms the foundation for integrating WSN with ML in this project
2	Patel and Singh	2021	IoT-based precision agriculture using machine learning algorithms	Optimize irrigation through IoT and ML	ML algorithms for resource optimization	Immediate resource efficiency improvement	Lack of long-term data analysis	Highlights the need for predictive insight using historical data
3	Kumar et al.	2022	A wireless sensor network approach	Develop a WSN framework for real time	WSN for temperature, humidity and soil moisture	Enhanced real-time monitoring	No predictive analysis or ML	Provides a WSN framework for integration

			for smart farming	monitoring	monitoring		integration	with ML
4	Gupta and Sharma	2023	Deep learning in agriculture : A survey	Survey of deep learning applications in agriculture	Image-based crop disease detection	High accuracy in disease detection	Focused on image data only	Demonstrates potential for advanced ML techniques
5	Chandra et al.	2023	Precision on agriculture using historical data and machine learning	Analyze historical weather and soil data for decision support	Used Random Forest and ANN for crop yield prediction	Accurate long-term yield predictions	Limited real-time sensor integration	Provides insights into historical data analysis methods

2.8 GAP ANALYSIS AND HOW THIS STUDY FILLS THE GAPS

1. Limited use of historical data utilization: Previous works primarily focused on real-time data without leveraging historical data for long-term predictions.

- Improvement: Incorporating historical data for training ML models to enhance prediction accuracy.

2. Limited integration of WSN and ML: Most studies focused on either WSN or ML, but not their seamless integration for data-driven decision-making.

- Improvement: Combining WSN and ML for dynamic decision-making based on both real-time and historical data.

3. Narrow focus on specific applications: Previous works often targeted specific issues like irrigation or disease detection.

- Improvement: Providing a comprehensive solution for precision agriculture, including irrigation management, crop yield prediction, and pest control.

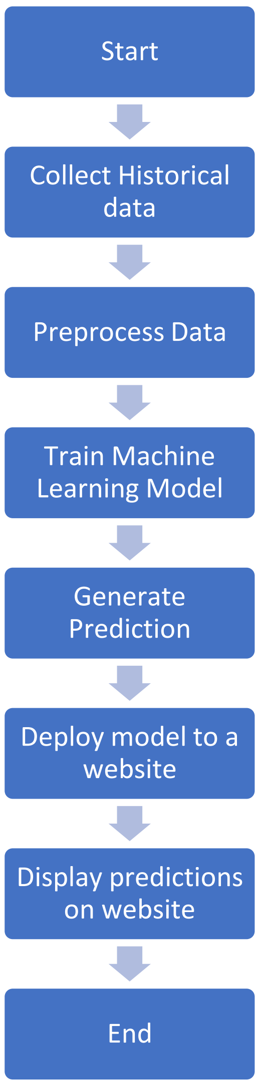
CHAPTER THREE

3.1 SYSTEM DESIGN

This system employs a data-driven design approach, where agricultural data is used to develop machine learning models for predicting best suitable crops, crop yield and disease detection. The project involves data collection, preprocessing, feature engineering, model training and evaluation.

3.2 FLOWCHART

The process flowchart for the system is as follows:



3.3 CROP CLASSIFICATION

3.3.1 DATA COLLECTION

Data collection focused on gathering agricultural data relevant to crop classification.

1. Data Features:
    - Nitrogen
    - Phosphorus
    - Potassium
    - Temperature
    - Humidity
    - Ph value
    - Rainfall
    - Crop
2. Data Sources:
    - Government agricultural databases (USDA)
    - Open-source datasets (Kaggle, FAO, google dataset)

3.3.2 DATA PREPROCESSING

To ensure the quality of data for machine learning, the following preprocessing steps were applied:

1. Exploratory Data Analysis: Plotting the values of the data features to understand patterns and identify errors in the data.

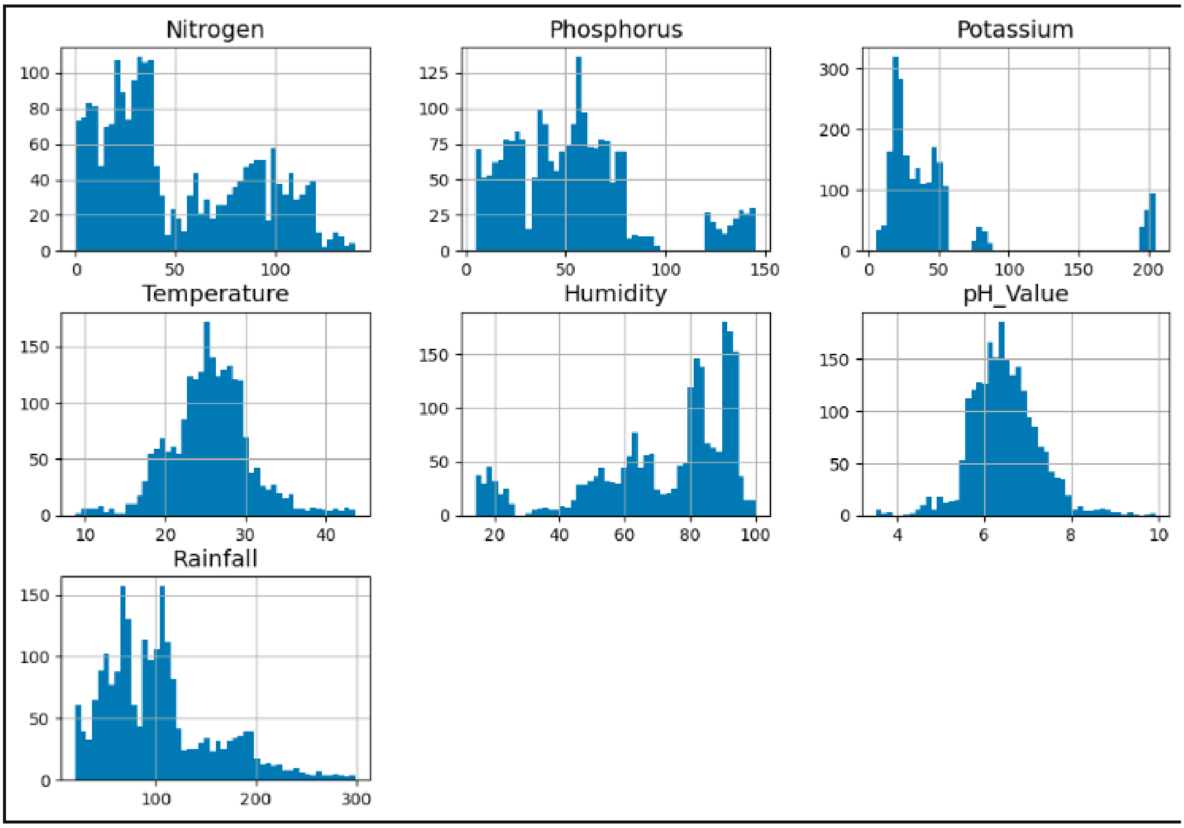


FIG 3.1: Exploratory Data Analysis Chart

2. Data Encoding: Conversion of categorical data to numerical values using the label encoder.

3. Data Splitting: The dataset was divided into 80% for training and 20% for testing to assess the models' generalization capabilities.

3.3.3 MODEL SELECTION

The success of a machine learning project largely depends on selecting suitable models for specific tasks. Selecting the right machine learning model is crucial for achieving high prediction accuracy and reliable decision-making in precision agriculture. The following models were considered and selected based on their suitability for the required tasks:

1. Support Vector Machine (SVM): A classification algorithm that finds the optimal hyperplane to separate classes in the feature space. SVM performs well with small datasets and complex decision boundaries, making it ideal for classifying crops.
2. Gaussian Naïve Bayes: This is a probabilistic machine learning model commonly used in classification tasks, including crop classification. It is based on Bayes' Theorem and assumes that features are conditionally independent given the class label, which is why is called "Naïve".
3. Random Forest Classifier: Random Forest is robust in handling high-dimensional and noisy data, making it suitable for classifying crops based on multiple environmental features. It is a robust ensemble-based model that combines the outputs of multiple decision trees.
4. Voting Classifier: This is an ensemble learning technique in machine learning that combines the prediction of multiple individual models to produce a final prediction.

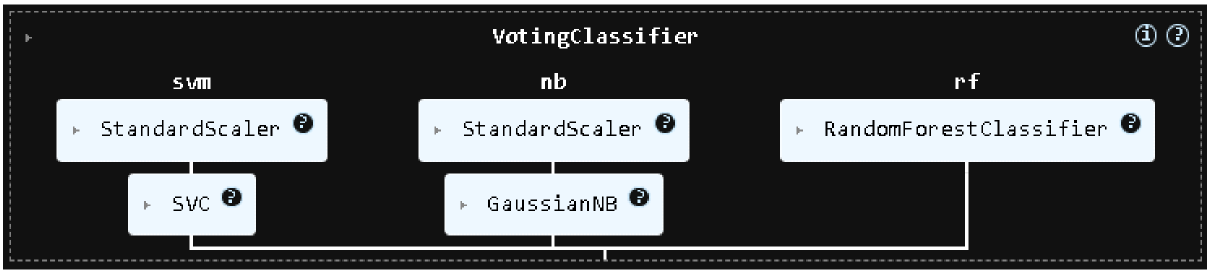


FIG 3.2: Models used

3.3.4 MODEL TRAINING

Training the models involved feeding preprocessed data into each algorithm. A pipeline is created for each of the algorithms containing a standard scaler and each of the algorithms.

3.3.5 MODEL EVALUATION

The model is evaluated using the Accuracy Metric. The accuracy metric is one of the most commonly used evaluation metrics in machine learning for classification tasks. It measures the proportion of correctly predicted instances (both true positives and true negatives) out of the total number of instances in the dataset.

3.4 YIELD PREDICTION

3.4.1 DATA COLLECTION

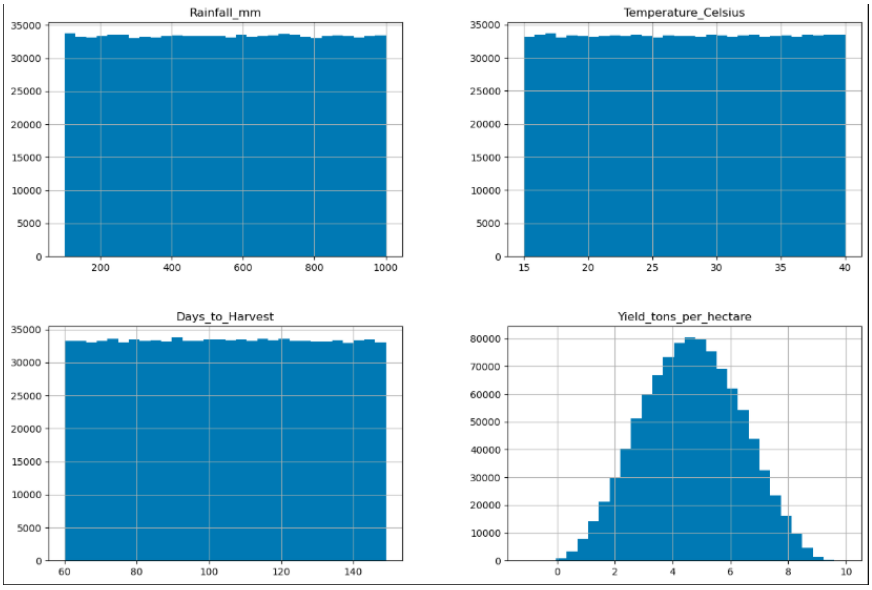
Data collection focused on gathering data relevant to yield prediction

1. Data features:
    - Region
    - Soil type
    - Crop
    - Rainfall
    - Temperature
    - Fertilizer
    - Irrigation
    - Weather
    - Days to Harvest
    - Yield (tons per hectare)
2. Data Sources:
    - Government agricultural databases (USDA)
    - Open-source datasets (Kaggle, FAO, google dataset)

3.4.2 DATA PREPROCESSING

To ensure the quality of data for machine learning, the following preprocessing steps were applied:

1. Exploratory Data Analysis: Plotting the values of the data features to understand patterns



and identify errors in the data.

FIG 3.3: Exploratory Data Analysis

2. Data Encoding: Conversion of categorical data to numerical values using the label encoder.

3. Correlation Analysis: This is used to identify the relationships between different variables in a dataset, which allows to understand how features are connected to each other and to the target variable.

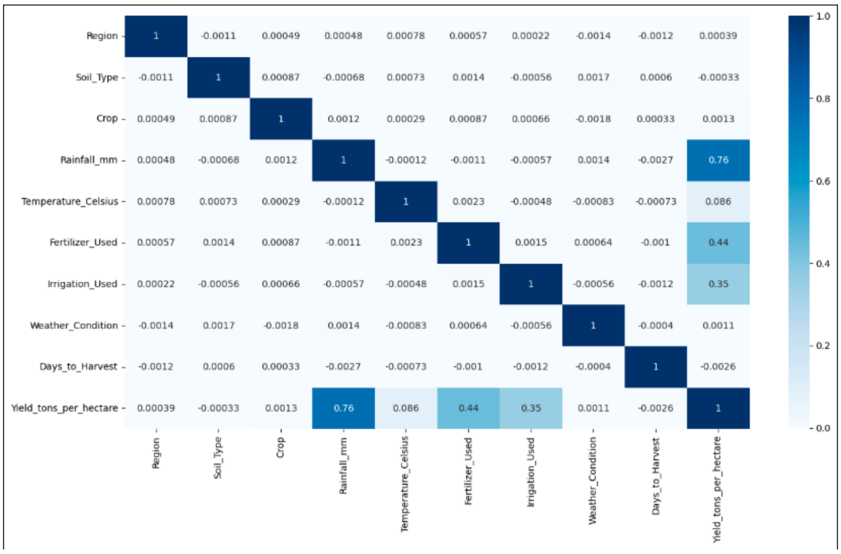


FIG 3.4: Correlation heatmap

4. Data Splitting: The dataset was divided into 70% for training and 30% for testing to assess the models' generalization capabilities.

3.4.3 MODEL SELECTION

The algorithm selected for this task is Linear Regression:

Linear Regression: This is a statistical method used to model the relationship between a dependent variable and one or more independent variables. It quantifies the relationship between predictors and the yield.

3.4.4 MODEL EVALUATION

The evaluation metrics used in this model are; Mean Squared Error (MSE) and r^2 score:

1. Mean Squared Error: This a widely used metric in machine learning to evaluate the performance of a regression model. It measures the average squared difference between the predicted values and the actual value.

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

MSE = mean squared error
n = number of data points
Y_i = observed values
$\hat{Y}_i$ = predicted values

2. R^2 score: This measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It ranges from 0 to 1 and is calculated as:

$$R^2 = 1 - \frac{RSS}{TSS}$$

R^2 = coefficient of determination

RSS = sum of squares of residuals

TSS = total sum of squares

3. Predicted vs actual plot: This helps visualize how well the model's predictions align with the true observed values. If the predicted values closely match the actual values, the points on the plot will fall close to the 45 degrees line (y=x). This indicates that the model is accurate.

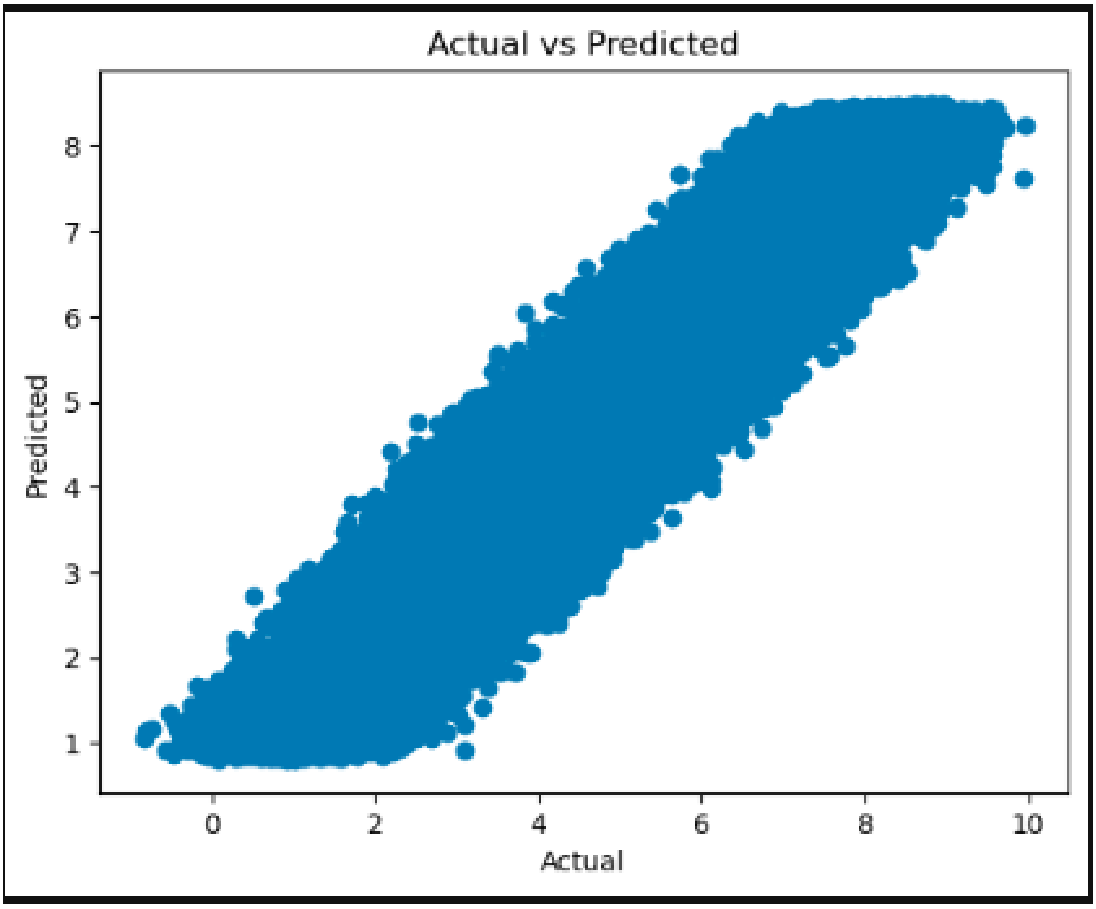


FIG 3.4: Actual values vs Predicted values

3.5 DISEASE CLASSIFICATION

3.5.1 DATA COLLECTION

The dataset consists of multiple folders containing images of both diseased and healthy crops.

3.5.2 DATA PREPROCESSING

1. An Image data generator normalizes the pixel values from the original range of [0, 255] to [0, 1] and splits the dataset into 80% training and 20% validation
2. The Images are resized for uniform input size, batch processing is set to load images in batches of 32. One Hot encoding is applied which converts labels for multi class classification
3. The same preprocessing steps are then applied to the validation set.

3.5.3 MODEL TRAINING

A Convolutional Neural Network model is designed for the plant disease classification. It processes 150×150 RGB images and consists of three convolutional blocks, each with Conv2D, Max Pooling, and Dropout layers to extract features while preventing overfitting. The extracted features are flattened and passed through fully connected layers, ending with a SoftMax output layer for multi-class classification.

The model is compiled using the Adam optimizer and categorical cross-entropy loss, making it suitable for multi-class problems. The Dropout layers reduce overfitting, and batch processing ensures efficient training.

Model: "sequential"

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 146, 146, 32)	2,432
max_pooling2d (MaxPooling2D)	(None, 73, 73, 32)	0
dropout (Dropout)	(None, 73, 73, 32)	0
conv2d_1 (Conv2D)	(None, 69, 69, 64)	51,264
max_pooling2d_1 (MaxPooling2D)	(None, 34, 34, 64)	0
dropout_1 (Dropout)	(None, 34, 34, 64)	0
conv2d_2 (Conv2D)	(None, 30, 30, 128)	204,928
max_pooling2d_2 (MaxPooling2D)	(None, 15, 15, 128)	0
dropout_2 (Dropout)	(None, 15, 15, 128)	0
flatten (Flatten)	(None, 28800)	0
dense (Dense)	(None, 128)	3,686,528
dropout_3 (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 19)	2,451

Total params: 3,947,603 (15.06 MB)
Trainable params: 3,947,603 (15.06 MB)
Non-trainable params: 0 (0.00 B)

FIG 3.5: Summary of the CNN model

3.5.4 MODEL EVALUATION

The model evaluation uses Classification Report and Confusion Matrix to assess performance. The classification report provides precision, recall, and F1-score for each class, helping measure accuracy and class balance. The confusion matrix shows how well the model distinguishes between different classes by comparing actual vs. predicted labels. These metrics help identify misclassifications, detect class imbalances, and improve model performance beyond just accuracy.

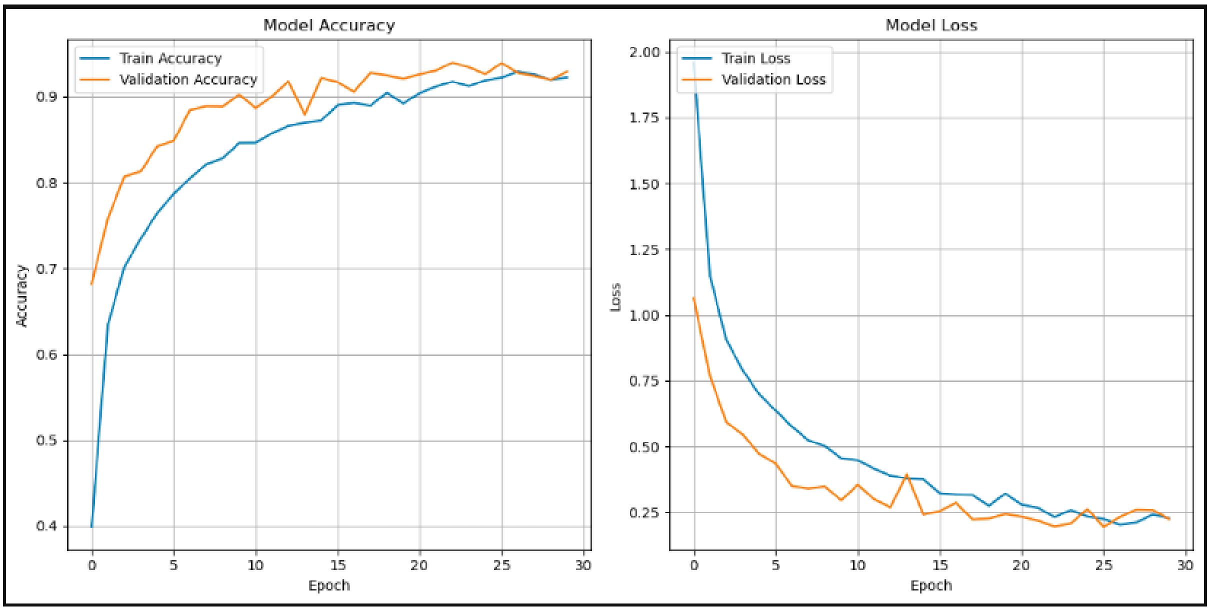


FIG 3.6: Model accuracy and Model loss chart

3.6 SYSTEM IMPLEMENTATION

The final machine learning models were deployed using a backend API developed with Flask. Key components include:

- Data Processing Pipeline: Handles data preprocessing.
- Model Inference Engine: Provides real-time predictions based on new input data.
- User Interface: A simple dashboard allowing users to visualize predictions and analyze agricultural data trends.

3.7 SYSTEM TESTING

- Functional Testing: Ensured the correct operation of data preprocessing, model prediction, and API endpoints.
- Performance Testing: Assessed model inference speed and response time.
- Accuracy Testing: Verified that model predictions aligned with test data outcomes.

3.8 SOFTWARE TOOLS AND FRAMEWORKS

- Backend: Python, Flask
- Frontend: HTML, CSS, JavaScript
- Machine learning libraries: Numpy, Pandas, Scikit Learn, TensorFlow
- Visualization: Matplotlib, Seaborn

CHAPTER FOUR

4.1 INTRODUCTION

This chapter presents the results obtained from training, building and evaluating machine learning models for precision agriculture using historical data. The system focuses on three key tasks: predicting suitable crops based on environmental conditions, estimating crop yield, and detecting potential crop diseases. The performance of various models is analyzed, and the decision-making system generated actionable insights that were communicated to the farmers via a user-friendly dashboard.

4.2 MODEL TRAINING AND TESTING RESULTS

The models were trained and tested using preprocessed historical data, including environmental factors, soil properties, and crop health indicators. The dataset was split into training and testing sets to evaluate model generalization.

4.2.1 PREDICTING SUITABLE CROPS

The task was modelled as a multi-class classification problem, where the system predicts the most suitable crop for a given set of environmental conditions (Nitrogen, Phosphorus, Potassium, temperature, humidity, soil pH, and rainfall).

Model Selected: Voting classifier (SVM, GaussianNB, Random Forest Classifier)

Evaluation Metrics:

Metric	**Value**
Accuracy	99.3%

Insights:

- The model successfully classified the optimal crop for most environmental scenarios.
- Slight misclassifications occurred for crops with overlapping environmental requirements.

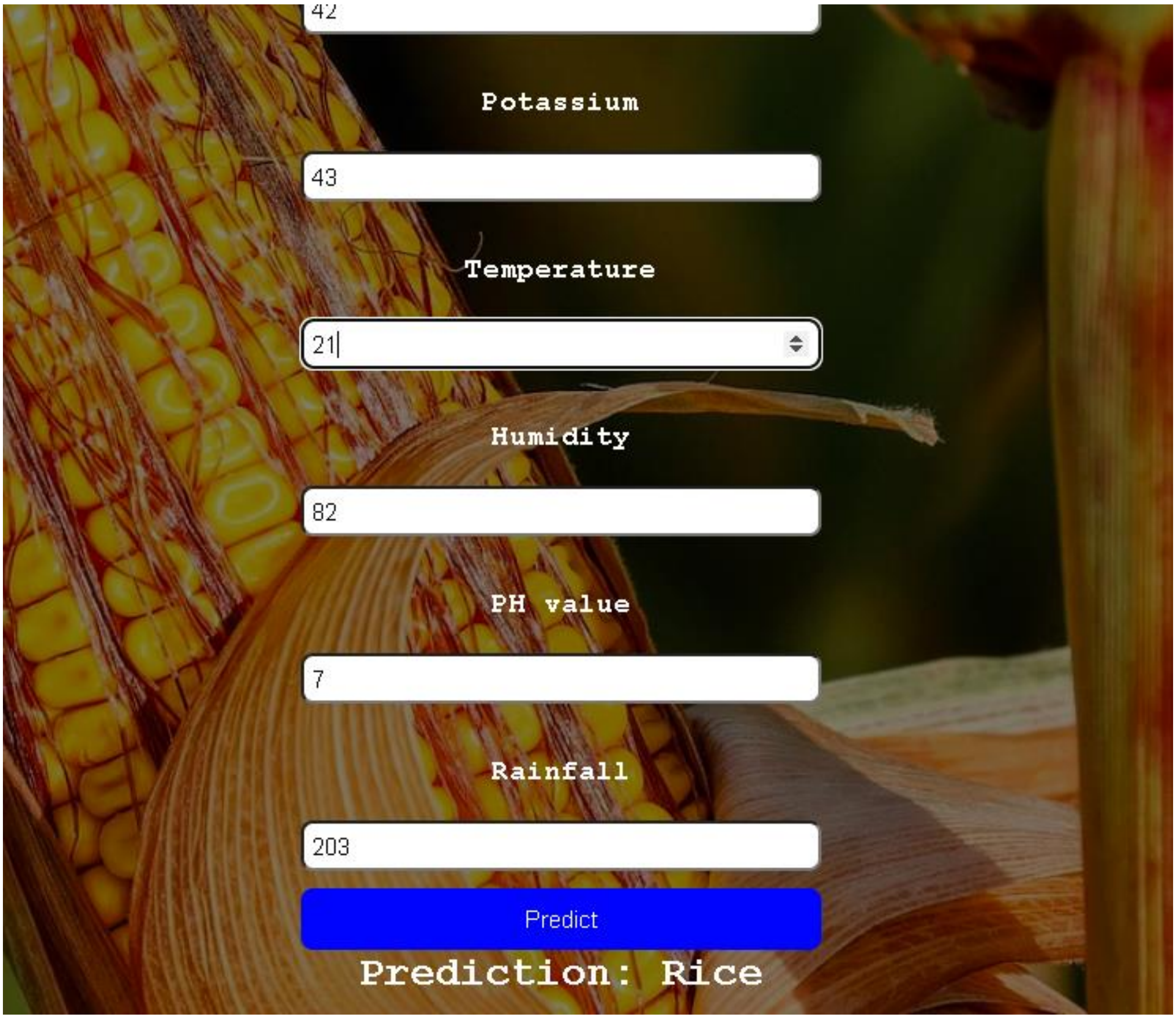


FIG 4.1: Crop prediction

4.2.2 PREDICTING CROP YIELD

The crop yield prediction was modelled as a regression task. The system forecasts the yield per hectare based on environmental and management factors.

Model Selected: Linear regression

Evaluation Metrics:

Metric	Value

MSE	0.25
R score	0.91

Insights:

- The model provided accurate yield forecasts, with minimal errors between predicted and actual values.
- High R^2 value (0.91) indicated that the model explained 91% of the variance in the data.

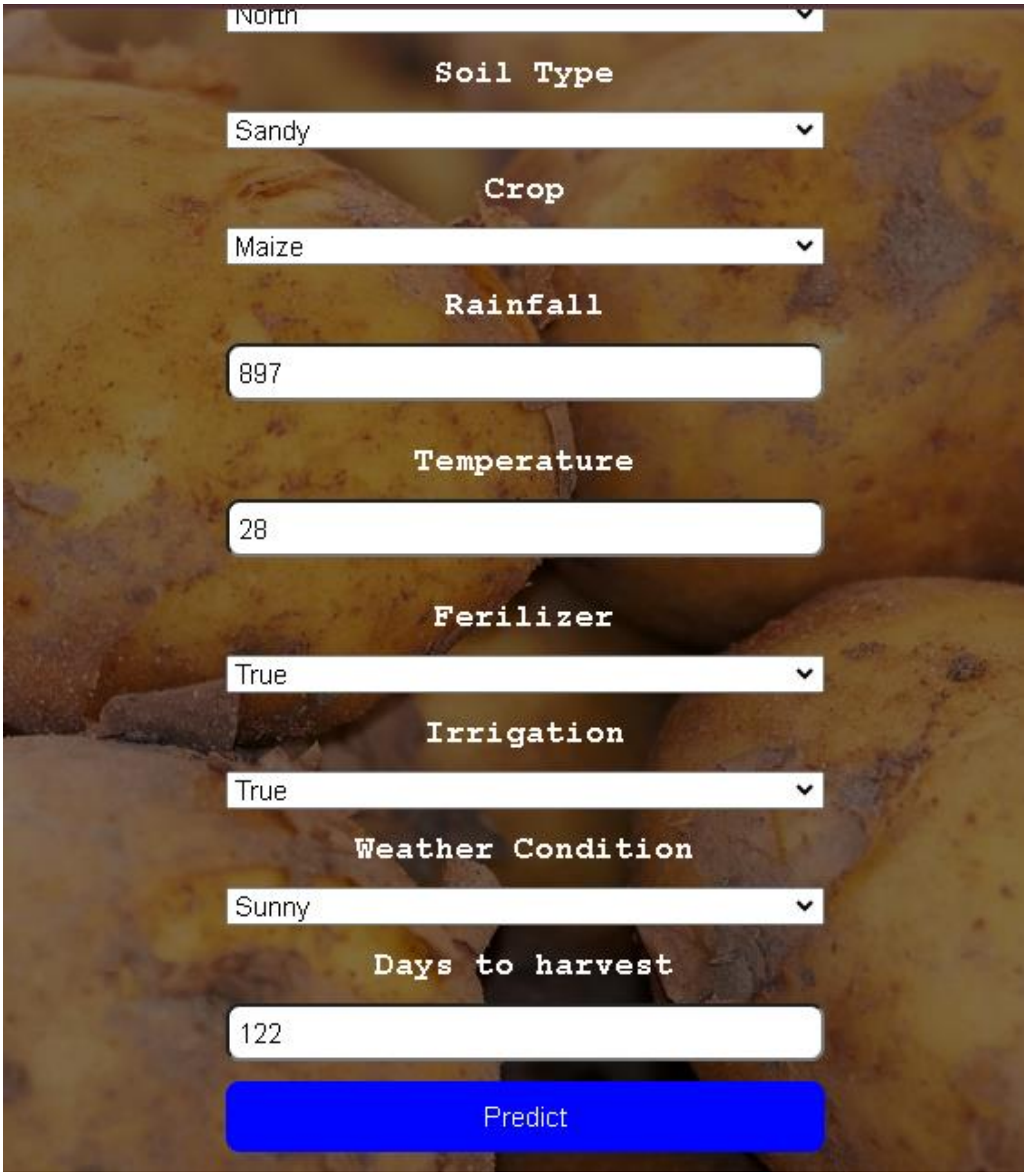


FIG 4.2: Yield prediction

4.2.3 DISEASE DETECTION

The disease detection task was modelled as a multi classification problem, where the system predicts whether a crop is infected with a disease based on images of the leaf.

Model Selected: Convolutional Neural Network

Evaluation Metrics: Accuracy

Metric	Value
Accuracy	93.8%

Insights:

- The model effectively detected crop diseases, helping to ensure timely intervention.
- Slight false positives occurred due to overlapping symptoms between diseases.

FIG 4.3: Detecting Disease

4.3 MODEL EVALUATION AND PERFORMANCE ANALYSIS

1. Suitable Crop Prediction

The Voting Classifier demonstrated high accuracy in predicting suitable crops, making it reliable for recommending crops suited to specific environmental conditions. Having multiple algorithms in the voting classifier increased the accuracy across all the algorithms.

2. Crop Yield Prediction (Linear Regression)

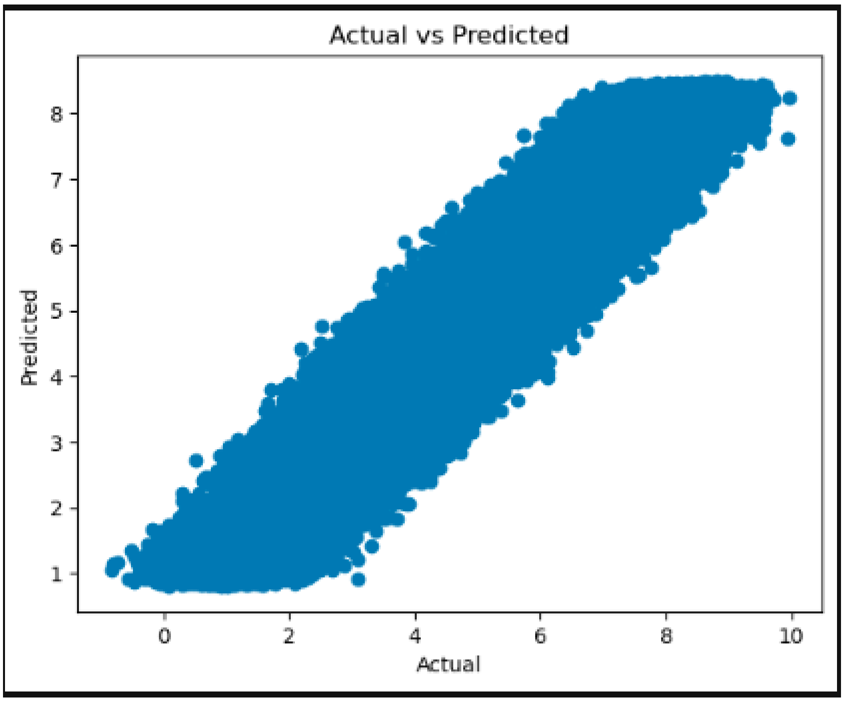
Linear Regression provided accurate predictions with an R^2 score of 0.91, indicating that the model explained 91% of the variance in crop yield data. The prediction errors were minimal and symmetrically distributed, indicating model stability. This makes it a highly effective tool for yield forecasting.

3. Disease Detection

The CNN effectively identified the presence of crop diseases in the images, achieving an accuracy score of 0.93. This performance ensures accurate detection and prevention of potential crop health issues.

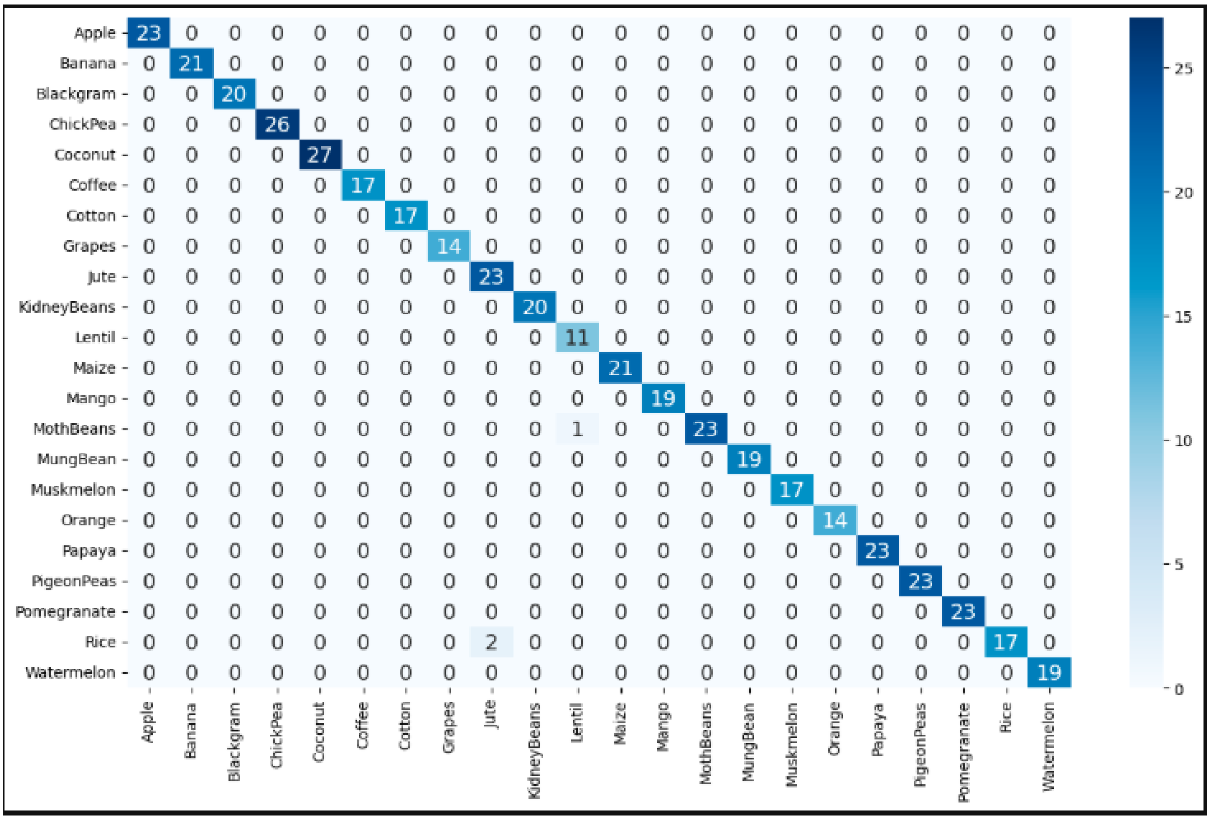
4.4 VISUALIZATIONS

1. Crop Yield Prediction (Actual vs. Predicted)



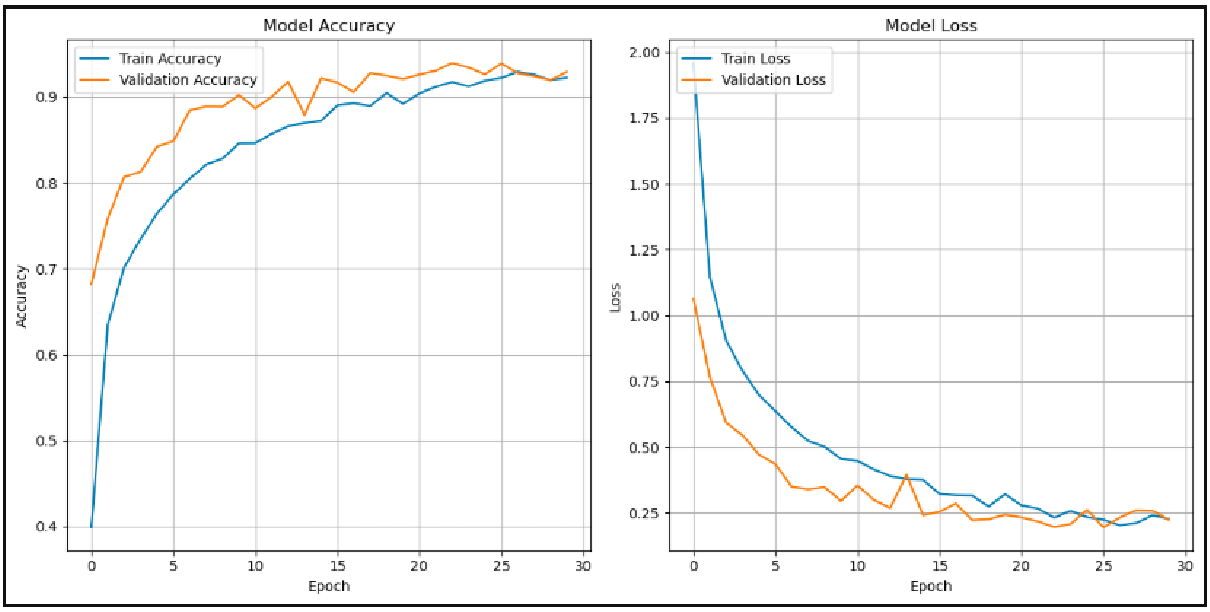
A plot showed that the predicted yield closely aligned with the actual yield, indicating the model's accuracy.

2. Confusion Matrix for Suitable Crop Prediction (SVM Model)



The matrix visualized the correct and incorrect classifications, highlighting the model's effectiveness.

3. Disease Detection Results (CNN)



A graph visualizing the train and validation accuracy and the train and validation loss for disease detection.

4.5 DISCUSSION OF RESULTS

1. Crop Suitability Prediction:

- The Voting Classifier model demonstrated reliable performance, enabling farmers to make informed decisions on crop selection.
- High accuracy values indicated the model's robustness.

2. Crop Yield Forecasting:

- The Linear Regression model's accurate predictions provided valuable insights for resource allocation and market planning.
- The low MAE and high R^2 score emphasized the model's reliability.

3. Disease Detection:

- The Convolutional Neural network successfully detected diseases, ensuring timely intervention to prevent crop losses.
- Despite the high accuracy, false positives highlighted the need for more diverse disease data.

CHAPTER FIVE

5.1 INTRODUCTION

This chapter provides a summary of the project's achievements, limitations and recommendations for future improvements. It highlights how the developed machine learning models contribute to precision agriculture. The study aimed to build a machine learning model for precision agriculture using historical data to predict suitable crops, forecast crop yields, and detect crop diseases.

5.2 SUMMARY OF FINDINGS

The study successfully demonstrated that the integration of machine learning with WSN could lead to significant improvements in precision agriculture. The system was able to provide actionable insights to farmers and improve crop yield.

The primary objective of this project was to build a machine learning-based system using historical data to enhance decision-making in precision agriculture. The developed system achieved the following:

1. Suitable Crop Prediction:

The Voting Classifier model achieved an accuracy of 99.3%, successfully recommending suitable crops based on environmental conditions such as Nitrogen, Phosphorus, Potassium, Temperature, Humidity, Ph value, Rainfall.

This model helps farmers choose the most suitable crops, increasing efficiency and reducing losses.

2. Crop Yield Prediction:

The Linear Regression model used in the prediction of yield performed accurately, with a low Mean Square Error and a High Variance.

Significance: Accurate yield forecasting assists in planning resource allocation, optimizing fertilizer usage, and managing market strategies.

3. Disease Detection:

The CNN detected crop diseases with an accuracy of 93.8%, offering timely responses for potential crop health issues, thereby reducing crop losses and maintaining farm health.

5.3 CONTRIBUTIONS OF THE PROJECT

1. Data-Driven Agricultural Insights: The system empowers farmers with reliable information for crop selection, yield optimization, and disease prevention. The use of historical data ensures that decisions are based on reliable and empirical evidence.
2. Precision Agriculture Adoption: The use of machine learning models encourages precision farming techniques, making agricultural practices more efficient and sustainable.
3. Support for Resource Optimization: By predicting yields and detecting diseases, farmers can reduce waste and manage resources effectively.
4. Environmental Sustainability: The system promotes better land use and informed decision-making, contributing to sustainable agriculture.
5. Decision Support for Farmers: The system provides actionable insights to improve crop management practices.
6. Increased Agricultural Efficiency: Accurate yield predictions and early disease detection help optimize agricultural operations.

5.4 LIMITATIONS OF THE STUDY

Despite its success, the project has some limitations:

1. Limited Disease Dataset:

The disease detection model's accuracy could be improved with a more diverse and comprehensive disease dataset.

2. Data Quality and Availability:

Historical data inconsistencies, such as missing values and measurement errors, affected model training and predictions.

3. Dynamic Weather Factors:

The current model relies on historical weather data, which may not capture sudden weather changes or anomalies.

4. Computational Requirements:

Training large datasets for machine learning models required significant computational resources.

5.5 RECOMMENDATIONS

To further enhance the system's effectiveness and practical applicability, the following recommendations are proposed:

1. Model Optimization: Experiment with other advanced machine learning models such as deep learning techniques for better yield and disease detection accuracy.

2. Integration with Mobile Applications: Develop a user-friendly mobile application to make the system accessible to farmers in the field.

3. Weather Data Integration: Include dynamic weather forecasting data to improve crop yield predictions by accounting for future weather patterns.

4. Farmer Education and training: Organize workshops and training sessions for farmers on the effective use of the system for decision-making and interpretation of the results.

5.6 CONCLUSION

The project successfully demonstrated that machine learning models could support precision agriculture by predicting suitable crops, forecasting crop yields, and detecting diseases. By leveraging historical data, the system offers valuable insights to farmers, enabling them to make informed decisions that enhance productivity and sustainability.

By addressing the identified limitations and implementing the recommended improvements, this system can become an invaluable tool for modern agricultural practices. The integration of real-time data, advanced models, and user-friendly interfaces will further enhance its adoption and effectiveness.

REFERENCES

1. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

2. Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.

3. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674.

4. Patil, M. D., Kolhe, S. R., & Dey, N. (2017). Precision agriculture using machine learning models: A comprehensive survey. Computers and Electronics in Agriculture, 142, 69-90.

5. Zhang, X., Liu, L., & He, J. (2020). IoT-based precision agriculture for crop monitoring and prediction: A case study. Journal of Agricultural Informatics, 11(2), 23-32.

6. Mohana Priya, N., Amudha, G., Dhurgadevi, M., Malathi, N., Balakrishnan, K., & Preetha, M. (2024). IoT and Machine Learning based Precision Agriculture through the Integration of Wireless Sensor Networks. Journal of Electrical Systems, 20(4s), 2292-2299.

7. Kocian, A., & Incrocci, L. (2020). Learning from Data to Optimize Control in Precision Farming. arXiv preprint arXiv:2007.05493.

8. Akhter, R., & Sofi, S. A. (2022). Precision Agriculture Using IoT Data Analytics and Machine Learning. Journal of King Saud University – Computer and Information Sciences, 34, 5602–5618.

9. Yang, C. (2020). Remote Sensing and Precision Agriculture Technologies for Crop Disease Detection and Management with a Practical Application Example. Engineering, 6(5), 528–532.

10. Google Scholar (https://scholar.google.com/Crop Yield Prediction Models)

11. IEEE Xplore (https://ieeexplore.ieee.org/Disease Detection Using Machine Learning)

12. ResearchGate (https://www.researchgate.net/Machine Learning in Precision Agriculture)

13. Google Scholar (https://scholar.google.com/Machine Learning in Precision Agriculture)