A

Project Report

on

Agri-Go

submitted as partial fulfilment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

Computer Science & Engineering

by

Navya Rajvanshi (2100290100108)

Mudita Shakya (2100290100101)

Abhishree Bisht (2100290130008)

Akanksha Tiwari (2100290130015)

Under the supervision of

Ms. Ayushi and Mr. Himan Kalita

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May, 2025

DECLARATION

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgement has been made in the text.

Signature

Name: Navya Rajvanshi

Roll No.: 2100290100108

Signature

Name: Mudita Shakya

Roll No.: 2100290100101

Signature

Name: Abhishree Bisht

Roll No.: 2100290130008

Signature

Name: Akanksha Tiwari

Roll No.: 2100290130015

Date:07/05/2025

CERTIFICATE

This is to certify that the Project Report entitled "**Agri-Go**" which is submitted by **Akanksha, Abhishree , Navya , Mudita** in partial fulfillment of the requirement for the award of degree B. Tech. in the department of Computer Science and Engineering of KIET Group of Institutions, Delhi NCR affiliated to Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

Supervisor Name	**Dr. Vineet Sharma**
(Designation)	**Dean(CSE)**

Date: 07/05/25

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech project undertaken during B. Tech. Final Year. We owe a special debt of gratitude to the supervisor name, Department of Computer Science and Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavours have seen the light of day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, dean of Computer Science and Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

Last but not least, we acknowledge our friends for their contribution to the completion of the project.

Date: 07/05/25
Signature:
Name : Navya Rajvanshi , Mudita Shakya
Roll No.: 2100290100108, 2100290100101

ABSTRACT

The agriculture industry stands as an economic backbone of India by offering employment to numerous workers while simultaneously making major contributions to national GDP statistics. Various persistent agricultural obstacles including aquatic weather, depleted soil quality and regular plant disease occurrences reduce both crop productivity and farmer profits. The successful resolution of interconnected sustainable food safety issues depends on modern technological solutions which this research examines through an advanced machine learning (ML) operation system focused on crop suggestion and plant health monitoring and fertilizer management. The platform reaches its analysis results through Convolutional Neural Networks (CNNs) together with Random Forest and XGBoost models and soil information and weather data and crop health monitoring to establish actionable conclusions. XGBoost scored the highest among these algorithms at 99.3% accuracy in its prediction results and ensemble models showed high ability to recommend both crops and fertilizers. CNNs excel as image-based plant disease detectors.

Keywords: Machine Learning, crop recommendation, plant disease diagnosis, fertilizer optimization, CNNs, Random Forest, XGBoost

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
AI	Artificial Intelligence
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
EDA	Exploratory Data Analysis
eNAM	National Agriculture Market
ICAR	Indian Council of Agricultural Research
IMD	India Meteorological Department
IoT	Internet of Things
ML	Machine Learning

KNN	K-Nearest Neighbors
SMOTE	Synthetic Minority Oversampling Technique
NPK	Nitrogen, Phosphorus, Potassium
VGG	Visual Geometry Group (used in CNN architecture)
SVM	Support Vector Machine
RGB	Red, Green, Blue (color model for images)
TP	True Positives
F1 Score	Harmonic Mean of Precision and Recall
TN	True Negatives
FP	False Positives
FN	False Negatives

TABLE OF CONTENTS

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Indian agricultural economics rests upon two big pillars because it maintains half of the workforce while contributing 16% to the gross domestic product.[1]Regardless of its societal value the industry continues to deal with issues which include weather volatility and deteriorating soil nutrients and cyclical diseases in plants. The current environment conditions and crop diseases create severe obstacles for agricultural farming and farmer income thus requiring immediate deployment of eco-friendly cultivating approaches. The rising forecast of global hunger up to three times higher by 2050 demands precision agriculture because it simultaneously ensures efficient natural resource use and prevents resource overexploitation.

The modernized farming sector has received considerable disruption through machine learning (ML) and artificial intelligence (AI) which emerged as disruptive technologies during recent years. Through these tools farmers alongside stakeholders gain accessibility to data-based choices when determining their crop selection and disease identification and soil measurement and resource allocation.

These challenges are effectively managed through Decision Trees, Random Forest, XGBoost and Convolutional Neural Networks (CNNs) and [2][3]. Plenty of research work shows how CNNs utilize image-based analysis to detect plant illnesses accurately while Random Forest alongside XGBoost demonstrate outstanding behavior in predictive analytics to recommend crops.

The proposed research initiative works to solve current limitations through its delivery of a complete platform implemented with state-of-the-art machine learning technologies. Multiple elements including environmental conditions and plant health markers and soil elements integrate within the platform which offers appropriate recommendations to farmers. The system helps farmers through three key functions that include selecting crops and identifying plant diseases through image processing and delivering appropriate fertilizer recommendations according to soil and plant conditions.[4]The platform depends on Decision Trees, Naive Bayes and XGBoost and CNNs to deliver precise results that farmers can trust. The plant health diagnostic ability of CNNs works best with image-based assessment and XGBoost and Random Forest serve for predictive analytics to deliver farmers practical solutions for their difficult agricultural issues.

The successful operation of the platform results from its models delivering high precision results. The accuracy levels for XGBoost at 99.3% exceed those of Logistic Regression at 94.7% and Naive Bayes at 98.8% while surpassing Decision Tree at 91.5% but CNNs achieve better performance in disease identification. The high accuracy of ML-powered solutions indicates their ability to advance agricultural production alongside environmental hazard reduction initiatives for widespread sustainable practices adoption.

This technique lines up with worldwide initiatives to build farming systems which stand up to environmental change and feed increasing food needs simultaneously. The platform provides farmers with state-of-the-art resources through which it tackles major Indian agricultural problems while simultaneously ensuring worldwide food security and sustainable development.

Computational intelligence represents an emerging set of important systems which transform conventional agricultural methodologies. The implementation of these technologies delivers data-based insights which enable knowledgeable choices in several sectors such as:

- Crop Recommendation identifies suitable plant selections by using environmental factors like natural climate along with geographic results of previous harvests.

- The use of image-based diagnosis helps plants maintain healthy status at the earliest detection point before harm spreads to extensive crop areas.

- Soil and crop requirements drive the process of precise fertilizer application recommendation during fertilizer optimization.

- The evaluation of soil pH together with moisture levels and nutrient content through soil analysis allows for maintaining soil fertility.

1.2 PROJECT DESCRIPTION

Crop prediction, plant disease diagnosis, and fertilizer recommendation are three critical issues facing the agriculture industry that this project aims to address by utilizing cutting-edge machine learning (ML) and artificial intelligence (AI) approaches. Giving farmers—particularly those in rural and resource-constrained areas—accurate, data-driven insights that can boost sustainability and productivity is the main goal. The project intends to increase the accessibility and effect of contemporary agricultural technology at the local level by combining these elements into a single, user-friendly platform.

1. Crop Prediction System

The algorithm suggests the best crops for a certain season and region based on environmental data, such as soil type, temperature, and rainfall, to help farmers make better agricultural decisions. This component analyzes past trends and environmental circumstances using machine learning models including Support Vector Machines (SVM), Logistic Regression, and Decision Trees. Better yields and more effective land use result from farmers being able to make more informed decisions based on scientific facts rather than solely depending on intuition or traditional knowledge.

2. Plant detection module

The method for detecting plant diseases is the second essential component. Here, possible plant illnesses are detected using a Convolutional Neural Network (CNN) trained on the popular PlantVillage dataset, which has more than 87,000 annotated photos of crop leaves in good and bad condition. The portal allows farmers to upload photos of their leaves, which are then analyzed by the algorithm to identify early disease symptoms. This reduces crop loss and the requirement for excessive pesticide use by enabling prompt intervention. The system maintains excellent accuracy while remaining sufficiently efficient to operate in low-resource contexts by utilizing transfer learning through models such as ResNet50 and VGG16.

3. Fertilizer Recommendation System

The third element tackles the problem of using fertilizer correctly, which is frequently disregarded. The system uses a combination of machine learning (ML) techniques, such as K-Nearest Neighbors (KNN), Gaussian Naive Bayes, and Random Forests, to create fertilizer recommendations based on the farmer's unique requirements and soil nutrient and crop requirements data. By preventing excessive or insufficient fertilizer application, this improves soil health and lowers expenses.

4. Data Processing and Interface

To guarantee the precision and effectiveness of the models, a significant amount of work is put into preprocessing. The input data is standardized with the aid of methods like encoding, augmentation, and data normalization. Strategies like SMOTE (Synthetic Minority Over-sampling Technique) are used to address issues like missing values and class imbalance. By concentrating on the most pertinent elements of the data, feature engineering enhances model performance even further.

A web-based platform created with simplicity in mind unifies all of this. A user-friendly interface that is available in various languages and designed to function in places with restricted internet bandwidth allows farmers to interact with the system. Because of its versatility, the system can be used in a variety of agricultural areas without requiring highly skilled technical personnel.

5. Broader Impact and Future Scope

This project's link with the worldwide push towards sustainable agriculture makes it particularly pertinent. The platform seeks to encourage ethical farming methods that benefit the farmer and the environment by fusing AI with real-world farming requirements. To provide recommendations that are even more accurate and location-aware, the platform may eventually be extended to include satellite imagery, IoT sensor data, and climate forecasting technologies. In a world with uncertain climate concerns, such linkages would improve food security by opening the door for next-generation precision agriculture.

This research is topical and has a significant impact because it is highly relevant to the current worldwide focus on sustainable agriculture. There is a pressing need for farming methods that are both economically and environmentally sustainable due to the growing worries about soil degradation, climate change, and food security. This platform acts as a link between contemporary technology and conventional agriculture by fusing artificial intelligence with useful farming applications. Instead of depending only on experience or regional heuristics, it empowers farmers to make well-informed decisions based on data and predictive analytics. This leads to increased output, decreased waste, and more intelligent use of resources—all of which are essential components of sustainable farming.

Additionally, the project encourages moral and inclusive farming methods, especially for rural or disadvantaged populations who frequently do not have access to these kinds of technology resources. It promotes fair access to innovation by providing affordable, user-friendly, and low-infrastructure-compatible solutions.

The platform has a lot of room to grow in the future. Future iterations might incorporate climate forecasting to predict weather patterns, IoT devices to gather real-time field data like pH and soil moisture, and satellite imaging for land and crop analysis. When combined, these technologies would produce an intelligent, flexible precision agriculture system that increases agricultural output and promotes world food security. Such innovation is essential for sustainable agriculture in a changing environment.

CHAPTER 2

LITERATURE REVIEW

A technological and analytical perspective in agriculture initiates a transformative revolution that leads to increased productivity along with enhanced sustainability and profitability outcomes. The digital transformation has a significant effect on India and other nations, given that a large portion of their population relies on agricultural activities, which constitute 16% of the national GDP. Agriculture is undergoing a new technological transition, as Industry 4.0 technologies combined with Artificial Intelligence (AI), Machine Learning (ML), the Internet of Things (IoT), and big data applications have given rise to the concept of digital agriculture [5].

These advanced technologies are transforming traditional farming practices by incorporating data-driven decision-making into everyday farm operations. They facilitate smarter monitoring, precise planning, and quicker response times, allowing farmers to manage scarce resources like water, fertilizers, and labor more effectively. Furthermore, digital platforms are closing the information gap between farmers and specialists by providing real-time insights on soil conditions, weather patterns, pest invasions, and market needs.

These innovations apply their capabilities to agricultural enhancements concentrated on three critical operational domains: crop selection suggestions, plant health detection, and optimal fertilizer application. The use cases that have been implemented play vital roles in bolstering decision-making and increasing agricultural yields while reducing negative environmental impacts. By embracing such intelligent systems, agriculture becomes more resilient to climate changes, economically sustainable, and accessible to even small and marginal farmers.

Plant Disease Detection

Plant disease diagnosis progressed rapidly because of machine learning and deep learning algorithms that delivered faster and more precise diagnostic capabilities. These techniques provide a non-invasive, scalable, and economical way to control agricultural health, particularly in areas with limited resources or in rural areas. Automating the identification process lessens the need for manual inspection, which is frequently laborious and prone to errors, particularly in situations where skilled pathologists are not easily accessible.

The researchers at Simona E. Grigorescu et al used Gabor filter-based texture analysis to enhance feature extraction for Convolutional Neural Networks (CNNs) which improved precise disease identification in images through analysis. Higher classification accuracy resulted from the model's ability to detect subtle textural variations between photos of healthy and sick leaves thanks to the addition of Gabor filters. This method highlights how crucial preprocessing and improved feature engineering are to the diagnosis of agricultural and medical images.

Dheeb Al Bashish et al. implemented a neural network classifier based on statistical pattern recognition methods to recognize and sort various plant diseases with precise results. Multiple illness types, including complicated ones with overlapping visual traits, might be effectively distinguished using their system. In order to increase the general resilience of these models, researchers are experimenting with several designs, including DenseNet, InceptionNet, and attention-based mechanisms, as well as larger datasets.

By using advanced image processing technology deep learning models can identify diseases in their pre-symptomatic stages through analysis of leaf pattern modifications. For prompt intervention, lowering the possibility of significant yield losses, and decreasing the usage of pesticides, early-stage identification is essential. By using high-resolution photos and hyperspectral photography, detection accuracy is further increased, increasing the models' sensitivity to minute changes that are not detectable to the human eye.

The new models obtain assessment accuracy levels which surpass human-operated traditional methods reaching greater than 95% success rates.When trained on high-quality annotated datasets, especially in controlled testing situations, these algorithms can routinely beat even seasoned agronomists. This demonstrates how combining AI with traditional farming methods can improve decision-making and have revolutionary effects.

The technique faces obstacles preventing its expansion to various plant species. The system only functions under specific climatic situations thus preventing mass application. Research must concentrate on developing standardized prediction models that function across various geographic locations and crop

kinds in order to achieve broad adoption. This necessitates creating balanced and varied datasets that represent several disease types in a range of field settings. Research labs, IT developers, and agricultural institutes working together could result in the development of unified platforms. Furthermore, putting mobile-friendly versions of these models into use could help with real-time, on-field diagnosis and rural distribution.

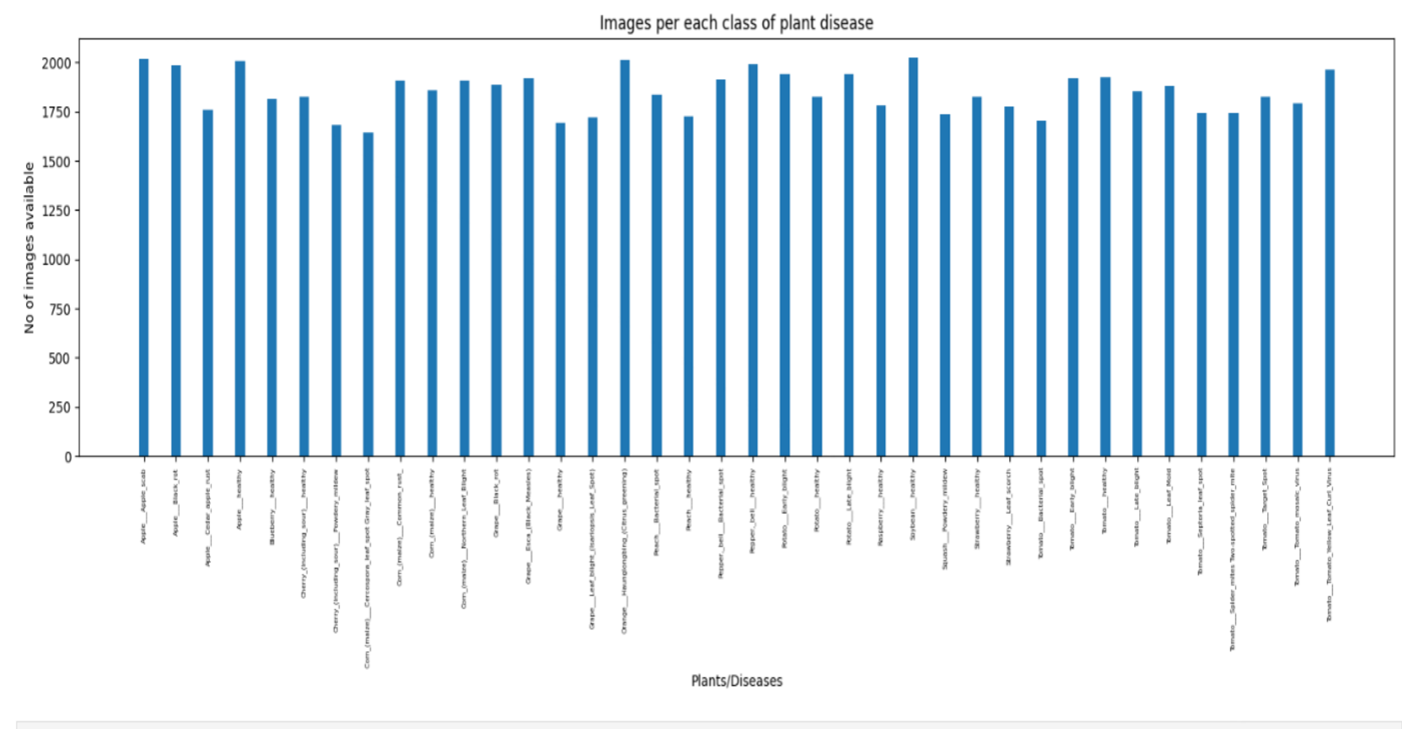

Fig. 1: Graph of Plant Disease Detection

Crop Recommendation Systems

Crop recommendation systems provide farmers with assistance for selecting optimal crops while relying on environmental data combined with soil characteristics alongside climatic conditions of their land. These systems provide customized advice based on scientific discoveries rather than customs or gut feeling, with the goal of increasing agricultural productivity and decreasing guesswork. Farmers can make more educated decisions that are in line with soil-specific and seasonal factors by utilizing technology.

The authors of Taj et al. constructed a hybrid model by merging Artificial Neural Networks (ANNs) which performed regression analysis together with K-Nearest Neighbors (KNN) for generating classified region-specific crop recommendations. In addition to learning intricate correlations between environmental factors, this combination of supervised learning techniques enables the system to accurately classify historical data from the region. Particularly in areas with changing climatic circumstances, the hybrid approach greatly increases the system's accuracy and adaptability.

The researchers of Banavlikar and colleagues upgraded the crop advice system with live monitoring of temperature and soil components as input for neural network modeling to optimize personalized guidance. By incorporating real-time data, the system may take into consideration daily variations in soil or weather conditions, producing recommendations that are more immediate and pertinent. At the farm level, this real-time decision support aids in resource optimization and increased crop yield.

Through utilization of these models farmers achieve better efficiency in their agricultural operations while making their decisions based on data analysis. The main challenge emerges from system implementation into diverse agricultural areas which experience variable weather patterns or variable soil environments. By incorporating machine learning, models may be continuously improved as new data becomes available, gradually increasing the systems' intelligence and adaptability. Additionally, by suggesting appropriate crops that need less irrigation or fertilizer depending on the availability of local resources, such systems can save input costs.

Real-time sensor data together with predictive weather analytics needs mandatory integration to achieve operational scalability of these systems.

Furthermore, the integration of IoT sensor networks and satellite data may help future implementations by offering a more thorough picture of circumstances on the ground. By enabling remote access via user-friendly mobile or web interfaces, cloud computing can also facilitate centralized model training. Such developments could greatly speed up the adoption of precision agriculture in developing nations if they are in line with local government regulations and farmer education initiatives.

Fertilizer Optimization

Efficient fertilizer application presents two key benefits: it promotes optimal crop production while also reducing the environmental harm caused by excessive chemical use, such as water contamination and topsoil depletion. Achieving precision in nutrient application is essential for sustainable agriculture, particularly given the decline in soil fertility and the impact of changing climate conditions. Approaches based on machine learning, like those suggested by Hussain et al., evaluate soil characteristics along with the requirements of crops and climatic factors to recommend customized fertilizer application strategies.

These AI-enhanced systems assist farmers in supplying nutrients according to the precise demands of their crops, which minimizes waste and fosters healthier plant growth. However, many traditional models concentrate solely on a limited set of soil variables, such as pH and moisture levels. They often overlook vital factors including micronutrient concentrations, historical cropping patterns, and microbial activity—each playing a significant role in soil health and fertilizer efficacy. For comprehensive optimization of fertilizers, advanced models need to combine these various interrelated factors to create more effective and contextually relevant advisory systems.During recent times we witness both government organizations and private companies unite to develop AI and data-driven agricultural solutions at an increasing pace.

ICAR (Indian Council of Agricultural Research) formed a partnership with private firms during 2021 to build AI-based disease prediction systems and precision farming solutions through IoT and remote sensing technology for optimized resource allocation especially for pesticide and water efficiency.[6]

Labour and government agency NITI Aayog together with IBM introduced AI-powered weather prediction systems allowing farmers to make better decisions regarding planting and irrigation in 2019.[7]

The World Bank-supported "Sustainable Agriculture in a Changing Climate" initiative applied AI to help farmers select specific farming practices which adapted to climate fluctuations and protected natural resources during 2020.

The eNAM platform integrates machine learning algorithms to assess market supply-demand patterns in order to provide equitable market access in National Agriculture Market since its launch in 2016.

The advisory system IBM Watson Decision Platform for Agriculture uses satellite imagery as well as sensor data along with weather forecasts for generating customized recommendations. AI platforms achieve predictions regarding soil conditions and farm disease threats and yield expectations with a precision rate of up to 90% which enables farmers to base their decisions on data-based information.[9].

These technology-driven fertilizer advisory platforms could significantly transform India's agricultural sector. They enhance input efficiency and environmental sustainability, while also equipping farmers—especially those in resource-limited or climate-sensitive areas—with the ability to make smarter, more adaptive choices.

CHAPTER 3

PROPOSED METHODOLOGY

1. Importance of Data in Agriculture

In modern agriculture, data serves as the backbone of informed decision-making. Systematically collected and well-curated agricultural data plays a transformative role in addressing some of the most critical challenges faced by farmers and the agricultural ecosystem. Issues such as the outbreak of crop diseases, inefficient farming techniques, unpredictable weather patterns, and declining crop yields can be proactively tackled by harnessing accurate and timely data.

Agriculture is uniquely vulnerable to a wide array of external factors, including climate change, soil degradation, pest infestations, and resource scarcity. In such a scenario, real-time access to data becomes essential for building adaptive and resilient farming strategies. It enables researchers and policymakers to detect early warning signs, understand emerging trends in crop health, and recognize regional disparities in farming practices. Moreover, a data-centric approach opens avenues for precise, localized interventions that can scale up to support national-level agricultural reforms.

The ability to transform raw data into actionable insights empowers stakeholders at all levels—from smallholder farmers to agricultural scientists and government agencies. By continuously monitoring and analyzing data on environmental factors, farming behavior, and soil health, the system can support proactive strategies for sustainable growth, improved crop management, and increased productivity.

2. Data Collection and Sources

The approach begins with gathering a wide range of agricultural datasets from trustworthy sources to assure credibility, representativeness, and functionality. Due to the intricate and variable nature of agriculture in India, information is sourced from three respected entities:

- **India Meteorological Department (IMD)**: Provides an extensive collection of historical and current weather information, including rainfall, temperature, humidity, and wind speed.
- **Indian Council of Agricultural Research (ICAR)**: Delivers scientifically verified research data related to farming techniques, types of crops, regional agroclimatic disparities, and soil properties.
- **Accredited Commercial Soil-Testing Laboratories**: Accredited Commercial Soil-Testing Laboratories: These offer localized and officially sanctioned soil test reports that detail pH levels, organic matter content, macronutrient and micronutrient concentrations, and contamination indicators.

These sources are selected based on their expertise in the field, reliability as institutions, and their thorough coverage of India's agricultural landscape. The combination of this diverse data promotes a comprehensive understanding of actual farming conditions, which facilitates the development of models that reflect genuine agricultural dynamics.

Historical weather information helps to recognize long-term climatic trends, identify irregularities such as droughts and floods, and formulate strategies for risk reduction. In a similar way, ongoing soil health data reveals shifts in fertility, erosion patterns, and contamination threats, informing choices related to crop rotation, fertilizer usage, and land management. By capturing both large-scale climatic changes and detailed soil attributes, the data forms a solid base for precision agriculture.

3. Data Preprocessing

Before developing any machine learning model, it is essential to clean and preprocess the raw data to guarantee high-quality and consistent input. Raw agricultural datasets frequently contain missing values, noisy entries, inconsistent formats, and irrelevant features that can adversely affect the performance of predictive models. Therefore, a comprehensive preprocessing pipeline is implemented to address these issues effectively.

The initial step involves **addressing missing values**, which are prevalent in weather and soil reports due to equipment failures, delayed updates, or data transmission problems. Techniques like mean/mode imputation or K-Nearest Neighbors imputation are utilized based on the data type and context. For categorical data, missing entries are either replaced with the most common category or labeled as "unknown" when suitable.

Subsequently, **normalization and scaling** are performed on numerical features such as temperature, rainfall, nitrogen levels, or pH. Since these data points often exist on varying scales, algorithms that depend on distance metrics (like KNN or SVM) may exhibit bias. Standardization (z-score normalization) or Min-Max scaling is applied to align all features within a comparable range.

Feature encoding is then executed to transform categorical variables (like crop names or soil types) into numeric formats that machines can interpret. Depending on the model and feature type, either one-hot encoding or label encoding is utilized**. Noise reduction and outlier detection** are carried out through statistical methods or visual tools like box plots to identify anomalies that might distort predictions.

Finally, the dataset is **split into training and testing subsets**, often employing stratified sampling to maintain class distribution. This practice ensures that the model can be trained effectively and assessed impartially, which supports better generalization on unseen agricultural data.

4. Machine Learning Models

To capture the intricate and varied agricultural situations in India, various machine learning models are utilized and assessed across different applications such as crop selection, fertilizer forecasting, and disease identification. The choice of algorithms strikes a balance between interpretability.

- **Decision Tree**Provides a high level of interpretability by visualizing decision-making rules, making it suitable for comprehending how particular factors like soil nutrients or weather conditions influence crop choices. It performs effectively with non-linear datasets and can tackle both classification and regression challenges.

- **Naive Bayes Classifier**: This model relies on Bayes' Theorem and presumes that features are independent, making it efficient for high-dimensional data and effective for tasks such as fertilizer classification where the features may have weak correlations.

- **Random Forest**By functioning as an ensemble of decision trees, it enhances accuracy by mitigating overfitting and improving generalization. It is particularly advantageous in crop recommendation due to the complex interactions among numerous input factors like rainfall, pH, and temperature.

- **XGBoost**: This robust gradient boosting algorithm is recognized for its speed and efficacy. It adeptly manages large datasets and intricate relationships, making it ideal for precise forecasts in crop or fertilizer applications.

- **Convolutional Neural Networks (CNNs)**: These are utilized for identifying plant diseases through the analysis of leaf images. CNNs autonomously extract spatial features from images, such as blemishes, discoloration, and patterns, allowing for accurate disease classification, including issues like blight, rust, or mildew.

Each model undergoes training with cross-validation methods and is assessed using metrics such as accuracy, precision, recall, and F1-score. Hyperparameter optimization (using grid search or random search) is also conducted to enhance the performance of each algorithm.

5. Model Evaluation

After training machine learning models, it is crucial to evaluate their performance to ensure they are dependable, precise, and applicable to real-world agricultural situations. The evaluation process entails utilizing a distinct test dataset that was not used during training, facilitating an impartial assessment.

Various performance metrics are employed depending on the nature of the task:

- **In classification tasks** (such as crop recommendation, predicting fertilizer types, or diagnosing plant diseases), the primary metrics include:
    - **Accuracy** The ratio of correct predictions to the total number of predictions made.
    - **Precision**: Evaluates how many predicted positive cases are genuinely positive, which is important when false positives carry significant costs.
    - **Recall (Sensitivity)**: Assesses how many actual positive cases were accurately identified, crucial when overlooking positives is critical.
    - **F1-score**: The harmonic mean of precision and recall, beneficial for dealing with imbalanced datasets.
    - **Confusion Matrix**: Offers a detailed account of prediction outcomes for each category.
- For **regression tasks** (if relevant, such as predicting fertilizer amounts), the metrics include:
    - **Mean Absolute Error (MAE)**: The average of absolute differences between predicted and actual values.
    - **Root Mean Squared Error (RMSE)**: Highlights significant errors by squaring the differences.
    - **R-squared (R^2)**: Indicates the proportion of variance in the dependent variable that the model explains.

Additionally, **k-fold cross-validation** is applied to validate models across different subsets of data, ensuring consistent performance. This prevents overfitting and increases model reliability across diverse conditions. **ROC-AUC curves** may also be used for binary classification models to visualize the trade-off between true positive and false positive rates.

Ultimately, the models are compared based on these metrics, and the top-performing ones are chosen not solely on accuracy but also considering interpretability, robustness, and appropriateness for field deployment.

6. System Design

The agricultural support system's overall architecture is crafted to unify various modules—crop suggestions, fertilizer forecasting, and disease identification—into a seamless, user-friendly platform. The system is designed for scalability, efficiency, and accessibility for farmers and agricultural specialists.

At its foundation, the backend operates on trained machine learning models, each managed through APIs utilizing a Python-based framework such as Flask or FastAPI. These APIs manage incoming data, perform necessary preprocessing, execute predictions, and return the outcomes. The backend's modular structure allows for independent updates or enhancements to particular models without impacting the overall system.

The frontend features a web-based or mobile interface that enables users to input soil parameters (such as pH and NPK levels), environmental conditions (including temperature, humidity, and rainfall), or upload images of leaves for disease identification. The interface is designed to be user-friendly, offering multilingual support and visual aids like icons or example images to ensure usability in rural settings.

For predicting plant diseases, the system includes an image upload feature. The uploaded images are analyzed and processed through a pre-trained CNN model, providing information on the disease classification along with possible treatments or preventive strategies.

All input data and results can be stored, if desired, in a **centralized database** (for instance, MongoDB or PostgreSQL) for monitoring, analytics, and future enhancements of models. Additionally, a logging and feedback system is included, allowing users to indicate whether predictions are helpful or not, facilitating ongoing learning.

To safeguard against misuse, security measures such as HTTPS and input validation are implemented. The design is prepared for cloud deployment, making it compatible with platforms like Heroku, AWS, or GCP to ensure accessibility across various devices and locations.

7. Deployment

Once the models have been trained and evaluated, the subsequent phase is to launch the system so that users—particularly farmers—can access the recommendations in practical settings. Deployment marks the shift from a research prototype to a usable, accessible application.

The trained models are saved (utilizing tools like joblib or pickle for machine learning, and TorchScript or ONNX for deep learning models) and incorporated within a backend server using frameworks such **as Flask, FastAPI, or Django.** These frameworks offer **RESTful APIs** that accept user inputs (like soil data or images), process the information, execute predictions, and return results in a structured format (JSON)

For wider accessibility, the backend is deployed on cloud services such **as Heroku, AWS EC2, or Google Cloud,** enabling scalable and continuous service. A frontend interface—typically developed with **React.js, HTML/CSS**, or mobile frameworks like **Flutter**—interacts with the backend APIs and displays results to users in a user-friendly and straightforward manner.

To facilitate real-time usage, the system is optimized for **low-latency** predictions, particularly for disease detection where image processing can be resource-heavy. Strategies such as image resizing, batching, and GPU utilization (on platforms like Google Colab or cloud-based GPUs) are employed for performance enhancement.

Basic authentication and **rate limiting** are put in place to ensure security and guard against misuse. The system can also be containerized using Docker for smoother deployment across various environments, making the application portable and sustainable over time.

Finally, deployment testing is performed to verify that all components operate properly in the live setting, and feedback from early users is gathered to inform future enhancements.

8. User Interface

The user interface (UI) is the most critical part for end-users, especially farmers who may have limited experience with digital systems. The goal is to make the UI **simple, responsive, and intuitive**, while maintaining clarity and functionality for different features—crop recommendation, fertilizer suggestion, and disease detection.

The UI provides **clear input fields** for entering soil parameters like nitrogen (N), phosphorus (P), potassium (K), pH level, temperature, humidity, and rainfall. It also includes **dropdowns or auto-suggestions** for selecting crops or fertilizers, and **file upload functionality** for plant disease detection through leaf images.

To ensure **accessibility**, the UI is designed with:

- **Large buttons and readable fonts** for mobile devices.
- **Icons and visual aids** to help users understand input types.
- **Multilingual support**, allowing users to operate in their regional language.
- **Color coding** for different types of alerts or recommendations (e.g., red for disease detected, green for healthy plants).

The interface is also **responsive**, meaning it adapts seamlessly to desktops, tablets, and smartphones. On submission, the UI provides real-time results like the most suitable crop, required fertilizer amounts, or identified plant disease along with its treatment suggestions.

Additionally, the system includes a **feedback form** where users can rate the usefulness of predictions or report errors. This user feedback helps in refining the model and UI over time. Optional features like **saving prediction history**, **chatbot assistance**, or **voice input** may be added later for an enhanced experience.

9. Evaluation Metrics

To assess the effectiveness and reliability of the models used in this system, several **quantitative evaluation metrics** are employed. These metrics vary based on the task—classification (for crop, fertilizer, and disease prediction) or image-based detection (for plant diseases).

For **classification models** (e.g., crop and fertilizer prediction), the following metrics are used:

- **Accuracy**: Measures the proportion of correct predictions out of total predictions. It is useful when class distribution is balanced.
- **Precision**: Measures the ratio of correctly predicted positive observations to the total predicted positives. High precision means fewer false positives.
- **Recall (Sensitivity)**: Measures the ratio of correctly predicted positives to all actual positives. High recall ensures fewer false negatives.
- **F1-score**: Harmonic mean of precision and recall. This metric is especially useful when class distribution is imbalanced, as it balances both false positives and false negatives.
- **Confusion Matrix**: Provides a detailed breakdown of model predictions versus actual values, helping to analyze where the model is making mistakes.

For **plant disease detection** using CNN models, in addition to the above metrics, we consider:

- **Validation Loss**: Measures how well the model performs on unseen data during training.
- **Training vs. Validation Accuracy**: Helps detect overfitting or underfitting.

- **Receiver Operating Characteristic (ROC) Curve and AUC (Area Under Curve)**: For multi-class disease classification, this provides insight into the trade-off between sensitivity and specificity.

These metrics are calculated during both training and testing phases. They help determine not only how accurate the models are, but also how **generalizable and robust** they are when exposed to new or unseen data.

10. Integration and Testing

After training and validating the models, the next essential step is to integrate all individual components—crop recommendation, fertilizer prediction, and disease detection—into a **unified system**. This phase ensures smooth communication between the frontend, backend, and model layers.

Integration includes:

- Connecting UI input fields to the appropriate backend API endpoints.
- Ensuring input data is correctly formatted and sanitized before model consumption.
- Structuring output responses from models into user-friendly result cards or messages.

Once integrated, **system testing** is conducted across various dimensions:

- **Unit Testing**: Individual components (such as each API endpoint or model function) are tested in isolation to ensure they work correctly.
- **Integration Testing**: Checks how well different modules (frontend, backend, models) interact with each other.
- **Functional Testing**: Ensures that all user-facing features—like submitting soil data, uploading images, receiving recommendations—work as expected.
- **Usability Testing**: A small group of potential end-users (farmers or students) test the platform to identify any user experience issues.

- **Performance Testing**: Evaluates how the system handles multiple requests, large image uploads, or prolonged usage.

Any bugs or issues identified during testing are documented and resolved through an iterative process. Logs are maintained for debugging, and **version control systems like Git** are used to manage changes and ensure deployment stability.

11. Deployment

After successful integration and testing, the entire system is prepared for deployment, making it accessible to real users such as farmers, students, and agricultural researchers. Deployment ensures that the application is hosted on a stable and secure platform and can handle real-time requests efficiently.

The deployment process involves several steps:

- **Model Serialization**: All trained machine learning models (e.g., crop prediction, fertilizer recommendation, and CNN-based disease detection) are saved using formats like .pkl (pickle), .joblib, or .h5 (for deep learning models). This allows them to be loaded and used without retraining every time.

- **Backend Hosting**: The backend, developed using frameworks like Flask or Node.js, is hosted on cloud platforms such as **Heroku**, **Render**, or **AWS EC2**. These services allow the model APIs to stay live and handle incoming requests from the frontend.

- **Frontend Deployment**: The frontend, typically built using HTML/CSS/JavaScript or React, is deployed on platforms like **Netlify** or **Vercel** for quick, scalable delivery to users.

- **Database Integration**: If user interaction or data storage is required (e.g., saving user queries or historical predictions), a database like **MongoDB Atlas** or **Firebase** is connected and deployed alongside the backend.

- **API Configuration**: APIs are exposed securely using proper **CORS** settings and rate limiting to avoid abuse. All APIs are tested post-deployment to confirm they are working with real-time data.
- **Monitoring and Logging**: Tools like **LogRocket**, **Sentry**, or built-in cloud logs are used to track performance, errors, and uptime. This helps in quick issue resolution and better system health monitoring.
- **Responsive Design & Mobile Access**: Since many farmers may access the system through mobile devices, the UI is made responsive and tested on multiple screen sizes to ensure usability.

The system is now available online and can start delivering smart, ML-powered agricultural assistance to users, helping them make informed decisions in real-time.

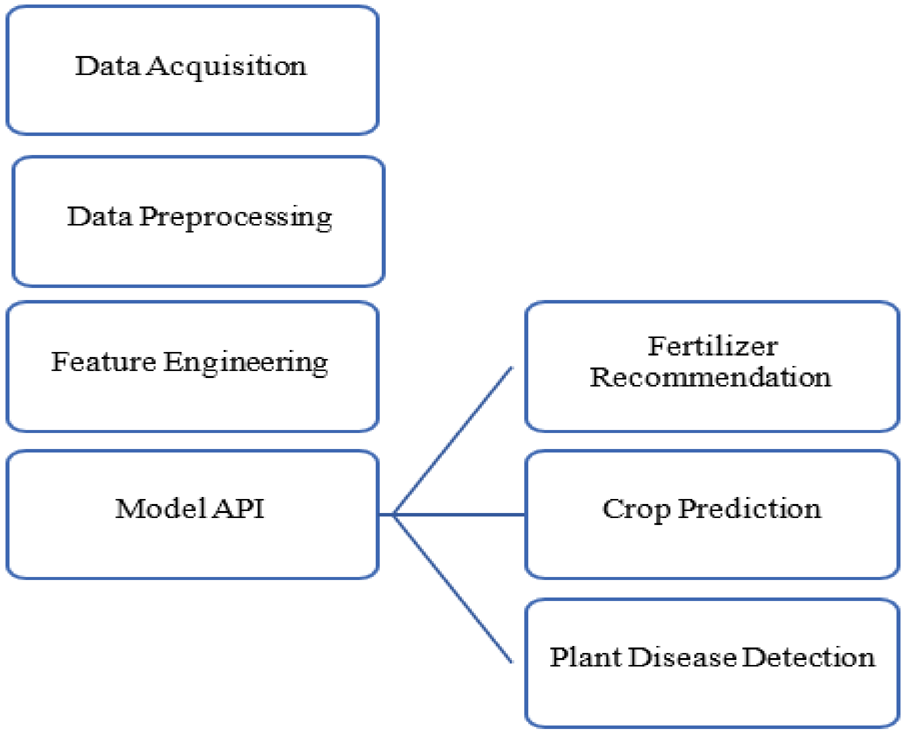

Fig. 2:

Flowchart for the proposed model

CHAPTER 4

RESULTS AND DISCUSSION

A. Classifiers Used:

The following classifiers and methods were used for crop prediction, fertilizer recommendation, and plant disease detection:

1. *Logistic Regression*

A popular classification approach that works particularly well for binary classification issues is logistic regression. Based on one or more input features, it calculates the likelihood that a specific class or event—such as "Yes" or "No"—will occur.

The main purposes of this study's use of logistic regression were crop prediction and fertilizer suggestion. It simulates how different input factors, including as temperature, rainfall, soil pH, nitrogen, phosphorus, and potassium levels, affect desired results. We can better comprehend how each environmental component influences the final decision thanks to the algorithm's visible and interpretable findings, which are produced by its linear nature.

The capacity of logistic regression to use its coefficients to describe the impact of specific variables is one of its advantages. In agricultural decision systems, where stakeholders favor comprehensible models over opaque solutions, this feature is especially advantageous.

Despite its binary nature, the One-vs-Rest (OvR) method was used to expand Logistic Regression to multi-class crop forecast scenarios. Using this method, many binary classifiers are constructed, one for each class, and the class with the highest confidence score is chosen as the prediction outcome.

When input features and outcomes have linear connections, the model performed reliably despite its simplicity. But when nonlinear interactions became more noticeable, its accuracy declined, highlighting the necessity for more sophisticated classifiers in these situations.

2. *Support Vector Machine (SVM)*

A potent supervised learning model called the Support Vector Machine (SVM) creates an ideal decision boundary, or hyperplane, between classes in a high-dimensional field. It is very helpful in complex classification settings since it may use kernel methods to convert non-linear data into a separable form.

SVM was used in this study for a variety of tasks, such as crop categorization, fertilizer recommendations, and disease detection of leaf-based plants. When appropriately configured, it can handle high-dimensional data with little overfitting, which is its strength.

A number of kernel functions were tested, such as the RBF (Radial Basis Function) kernel, which performed better with data that had non-linear correlations, and the linear kernel for data that was linearly separable. Optimizing model performance was mostly dependent on fine-tuning hyperparameters such the kernel type, gamma, and C-value (penalty parameter).

SVM was used to categorize photos of healthy and diseased leaves in order to detect plant diseases. When paired with pre-processed picture datasets, the model successfully captured pixel-level characteristics and subtle patterns.

Furthermore, by adding soil nutrient content, historical crop data, and temperature/humidity patterns, SVM's efficacy in recommending fertilizer was enhanced. It provided extremely detailed instructions by taking advantage of slight changes in the input.

All things considered, SVM was a reliable option in situations where dimensionality and data complexity presented difficulties for more straightforward models such as logistic regression.

3. *Decision Tree*

Decision trees are a popular option for applications involving rule-based decisions since they are among the most user-friendly and interpretable machine learning models. Because decision trees are transparent and efficient at handling both numerical and categorical information, they were used in this study for a number of purposes, including crop prediction and fertilizer suggestion.

The dataset is recursively divided into subsets by the model according to feature values. A decision rule based on a single attribute (such as soil pH, temperature, or rainfall) is represented by each internal node in the tree; the rule's result is shown by each branch; and the final prediction classes (such as crop type or fertilizer category) are represented by the leaf nodes. Explainability is crucial for establishing user trust in agricultural contexts, and this framework offers a clear visual mapping from input variables to output options.

To enhance model performance and avoid overfitting, several hyperparameters were fine-tuned. These included:

- **Max depth**: Controls how deep the tree can grow to limit overly specific decisions.
- **Min samples split**: Sets the minimum number of data points required to perform a node split.
- **Splitting criterion**: Evaluated using Gini impurity and entropy to measure the purity of partitions.

Though they might not always provide the highest accuracy on their own, Decision Trees served as a solid foundation for comparison analysis and were an essential part of more complex ensemble models like Random Forest. During real-time user interactions, they were helpful in producing timely and comprehensible recommendations because to their organized, rule-based logic.

4. *Gaussian Naive Bayes*

Based on Bayes' Theorem, Gaussian Naive Bayes is a probabilistic classification model that assumes each input characteristic has a normal (Gaussian) distribution and is independent of the others. Because of its simplicity, it can process high-dimensional data remarkably quickly and at a comparatively low computing cost.

Both crop prediction and fertilizer recommendation tasks in the current study were handled by Gaussian Naive Bayes. Because of its solid performance with smaller datasets and lightweight design, it was particularly helpful in the early stages of model comparison. It did remarkably well while handling multivariate environmental data, including temperature, humidity, and soil nutrient concentrations, despite its high assumption of feature independence.

In this instance, the model's main advantages were:

- Efficiency in computation, which is beneficial in real-time decision-making systems where rapid responses are critical.
- Noise tolerance, making it suitable for noisy agricultural data collected from varied sources.
- Scalability, enabling fast predictions across diverse crop types and regions.

Its incorporation into the ensemble approach improved the overall robustness of the system, especially when dealing with high-dimensional input vectors or sparse training data. Additionally, Naive Bayes offered a helpful alternative to more intricate models like as SVM and Random Forest, particularly when interpretability and performance-speed trade-offs were taken into account.

5. *Random Forest*

Based on a group of decision trees, Random Forest is a potent ensemble learning technique. Because of its scalability, resilience, and high reliability in handling both structured and unstructured datasets, it was used in all three of the main modules of this study: crop prediction, fertilizer recommendation, and plant disease detection.

Using bagging (Bootstrap Aggregation), in which several Decision Trees are trained on distinct subsets of the training data, each chosen with replacement, is the basic idea underlying Random Forest. Majority voting (for classification tasks) or averaging (for regression tasks) determines the final output during prediction, greatly lowering variance and lowering the possibility of overfitting, a common problem with single Decision Trees.

In this project, Random Forest achieved superior classification performance due to its ability to:

- **Adapt well to a variety of agricultural datasets**, such as plant traits, soil properties, and meteorological conditions.

- **Manage hierarchical and non-linear data relationships**, making it particularly appropriate for jobs like predicting the risk of pests and diseases and classifying crop types.

- **Provide accurate metrics for feature significance,** enabling the system to determine and prioritize the factors that have the greatest influence on decisions. Interestingly, in all three modules, soil pH, rainfall, and macronutrient values (such as the amount of potassium, phosphorus, and nitrogen) were often listed as the best predictors.

The model was fine-tuned using parameters such as:

- **n_estimators:** The number of trees in the forest that are optimized for balanced performance.
- **max_depth:** Controlled tree complexity and minimized overfitting.
- **max_features and min_samples_split:** Managed node splitting and feature selection.

6. ***XGBoost (Extreme Gradient Boosting)***

The three main objectives implemented XGBoost as this model featured both excellent predication abilities and scalability features which many users appreciated. The gradient boosting development process runs sequentially by constructing new models from previous stages to solve errors made in the past. The algorithm uses regularization features as a built-in mechanism to stop overfitting and supports fast parallel computations and processing. Organizations adopted the system because of its strong performance with extensive structured data and its efficient boundary condition management which resulted in its use for crop recommendation systems and disease detection services together with fertilizer advisory services.

Such algorithms served as a base for creating smart agricultural support systems that brought effectiveness through data-driven techniques. Each job task received its optimal model selection through assessments based on accuracy rates alongside F1-score and precision, recall, and confusion matrix data collection.

The following are some of the key benefits and technical features of XGBoost that contributed to its selection:

- **Learns from past mistakes:** XGBoost builds its model step by step, where each new tree corrects the errors made by the previous ones. This helps the model continuously improve and reduce prediction errors.
- **Built-in regularization:** The model includes techniques to prevent overfitting by penalizing complexity. This helps ensure that predictions stay reliable even when the data has noise or variability.
- **Highly scalable:** XGBoost can handle both small and large datasets. It works well on a single machine but can also be scaled to run on distributed systems, making it ideal for large-scale agricultural use across multiple regions.
- **Fine-tuning flexibility:** Key parameters like learning rate, tree depth, and sampling ratios can be adjusted and optimized using cross-validation. This allows for better control over model performance and accuracy.
- **Handles missing data effectively**: XGBoost can process datasets even when some values are missing, which is common in real-world agricultural data. This reduces the need for time-consuming data cleaning.

XGBoost was particularly effective in:

- Crop recommendation: It correctly identified the best crops for a certain input environment by capturing the intricate linkages between soil composition, weather, and past crop data.

- Fertilizer recommendation: To produce accurate fertilizer input recommendations, the model took into consideration the complex relationships between soil nutrient levels and historical yield data.

- Plant disease detection: XGBoost was a high-performance classifier for disease categories, providing dependable distinction between visually similar leaf infections when combined with image-based feature extraction models like CNNs.

The algorithm was crucial in providing precise, timely, and location-specific agricultural insights because of its ability to function well with both tabular and derived characteristics. A wide range of metrics, such as accuracy, precision, recall, F1-score, and confusion matrix analysis, were used to assess performance to make sure each model's output satisfied strict quality requirements for every module.

To sum up, XGBoost was a key element of the AI-driven smart farming system. Because of its integration, agricultural decision-making might be intelligently automated, increasing sustainability and efficiency. It is a favored model for practical implementation in precision agricultural frameworks because to its interpretability, rapidity, and strong prediction strength.

A. *Model Comparison and Performance Evaluation:*

1. **XGBoost:**

   XGBoost surpassed all other models, achieving an impressive accuracy rate of 99.3%. This outstanding performance is due to its gradient boosting framework, which emphasizes reducing the errors of prior iterations. The built-in regularization feature of XGBoost helps to avoid overfitting, making it particularly effective for managing large agricultural datasets. By accurately capturing intricate relationships between factors such as soil conditions, weather patterns, and crop growth cycles, XGBoost excelled in tasks related to crop prediction, fertilizer suggestions, and disease identification. Its capability to effectively handle both structured and unstructured data further enhanced its exceptional performance.

2. **SVM (Support vector machine)**

   SVM demonstrated excellent performance in classification tasks, especially when handling non-linear data using the RBF kernel. This kernel was particularly useful in challenging situations such as detecting plant diseases, where the distinction between healthy and infected plants is not always linearly separable. Nonetheless, SVM's effectiveness lagged slightly behind that of XGBoost and required careful hyperparameter tuning to reach its best performance. While it performs well with smaller datasets and particular classification issues, it faced challenges with scalability when confronted with larger, more varied datasets, which limited its applicability in certain agricultural contexts.

3. **Decision Trees:**

   Decision Trees offered clear and interpretable decision-making rules, making them a valuable tool for understanding the significance of different features. While they were not as precise as ensemble methods such as XGBoost, their straightforwardness and clarity facilitated better visualization of how input features affected the output. They performed adequately in tasks related to crop prediction and fertilizer recommendations but faced challenges with more complex data relationships. The model's difficulty in managing intricate, non-linear patterns often resulted in underfitting in certain situations, underscoring the necessity for more advanced models for intricate tasks.

B. *Feature Importance and Model Interpretability:*

1. **Feature Importance:**

   The Random Forest and XGBoost algorithms played a crucial role in pinpointing essential features such as soil pH, rainfall, and nutrient levels, which had a significant effect on the precision of crop selection and fertilizer advice. Through ensemble learning, these models delivered a more detailed comprehension of how various environmental factors interacted to affect agricultural results. The capacity to prioritize features according to their significance enabled stakeholders to focus on the most impactful variables, ensuring the models' predictions were both accurate and applicable in real-world farming situations. This assessment of feature importance was vital for enhancing the system's recommendation capabilities and optimizing interventions effectively.

2. **Model Interpretability:**

   Regarding interpretability, Decision Trees and Logistic Regression emerged as the most comprehensible options. Decision Trees offered clear decision pathways that connected input features to the output, allowing for easy tracing of how specific conditions led to certain recommendations or predictions. This level of clarity is highly appreciated in agricultural contexts, where stakeholders favor models that are not only precise but also easy to understand. Likewise, Logistic Regression, despite being less complex, similarly provided a direct correlation between features and outcomes. These models were particularly beneficial when engaging with users lacking extensive technical expertise but who still required confidence in the recommendations made by the system.

C. *Model Limitations and Areas for Improvement:*

1. **Logistic Regression:**

   Although Logistic Regression is a useful method for modeling straightforward, linear relationships, it faced challenges when addressing more intricate, non-linear patterns in agricultural data. Its linearity assumptions between features and outcomes rendered it less effective in situations where multiple factors, such as soil conditions, weather patterns, and crop health, interact in non-linear manners. In cases where there are complex interactions, the model's performance was not optimal. To improve its usefulness, future enhancements might involve incorporating non-linear methods or integrating it with other models to achieve better accuracy.

2. **SVM:**
   Even with its advantages, SVM encountered limitations because of its dependence on the proper selection of kernel and hyperparameters. Without meticulous tuning, the model's efficacy could decline significantly, particularly when handling high-dimensional agricultural datasets. Furthermore, SVM proved less scalable than models like XGBoost, resulting in difficulties when trying to manage larger datasets effectively. This restricted its use in practical agricultural applications, where data can be extensive and changing. Further refinements and advancements in kernel selection and scaling methods would enhance its performance on large-scale projects.

3. **Gaussian Naive Bayes:**

   Naive Bayes demonstrated to be a rapid and efficient model; however, its fundamental assumption of feature independence was frequently unrealistic for agricultural data. Many agricultural features, such as soil nutrients and weather patterns, exhibit high correlation, and this assumption resulted in inaccuracies within the model's predictions. While Naive Bayes functioned reasonably well for less complicated tasks, it struggled with more complex relationships, such as the interactions between environmental factors and crop outcomes. Improving this model could involve addressing the dependencies among features or combining it with more sophisticated ensemble methods to strengthen its predictive capability in agricultural contexts.

D. *Results:*

The research findings highlighted the effectiveness of **ResNet-9**, which leveraged its deep learning architecture to extract intricate patterns from plant leaf images. Its use of residual connections helped overcome typical training issues found in deep neural networks, enabling smooth learning even with complex image data. This allowed ResNet-9 to deliver highly accurate predictions in identifying plant diseases. Its ability to automatically extract deep features from images gave it a strong edge over traditional models like Decision Trees and SVMs. As a result, ResNet-9 emerged as a superior model for plant disease diagnosis, offering both high performance and practical value for optimizing real-world agricultural practices.

In the crop recommendation analysis, various machine learning models were assessed, including **Naïve Bayes, XGBoost, Random Forest, SVM, Decision Trees,** and **Logistic Regression.** Their performances were compared using essential metrics like F1-Score, Precision, and Recall. Among them, XGBoost recorded the highest accuracy at **99.3%,** demonstrating its robust performance across diverse agricultural prediction tasks. The application of SMOTE to address class imbalance, together with meticulous feature engineering, significantly contributed to enhancing the accuracy of the models. These enhancements not only improved prediction quality but also ensured more dependable and balanced results, which are vital when working with real-world agricultural data sets.

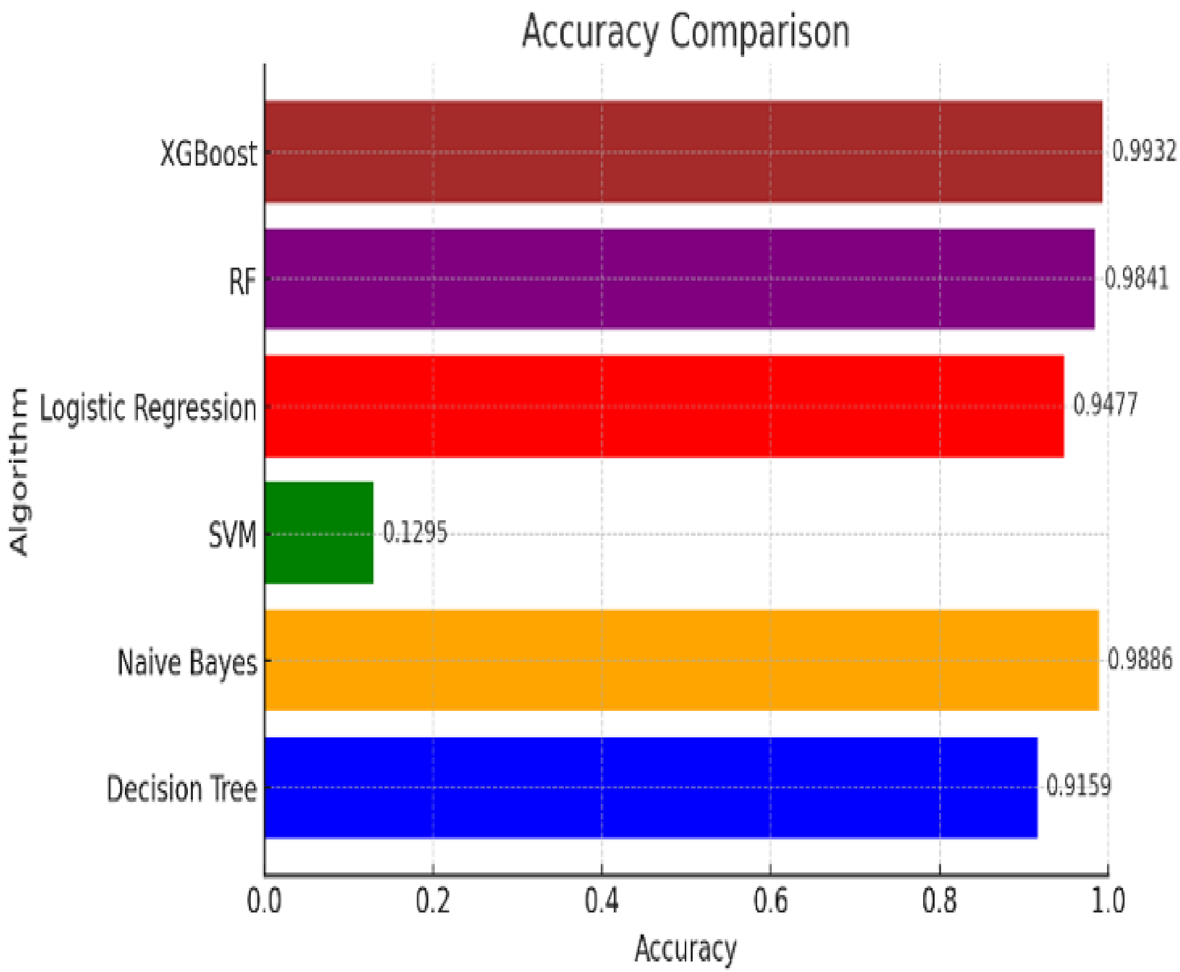

Fig. 3: Model Accuracy Comparison of Different Classifiers

Tabular representation of the accuracy comparision:

S.No	Algorithm	Accuracy
1.	Decision Tree	91.59%
2.	Naïve Bayes	98.86%
3.	SVM	12.95%
4.	Logistic Regression	94.77%
5.	RF	98.41%
6.	XG Boost	99.32%

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

This study introduces a comprehensive agricultural support platform powered by artificial intelligence, designed to tackle several critical issues in contemporary farming, namely crop forecasting, disease detection, and fertilizer suggestions. By merging various machine learning models and deep learning methodologies, the platform provides smart and actionable information that can greatly improve productivity, especially for farmers operating in areas with limited data or resources.

The module for detecting plant diseases utilizes Convolutional Neural Networks (CNNs) to precisely recognize diseases from leaf images. The deployment of advanced deep learning structures guarantees accurate classification, enabling timely and targeted actions that reduce crop loss and enhance the quality of yields. These findings highlight the importance of the system in precision agriculture and sustainable farming approaches.

For forecasting crops and assessing fertilizers, the platform utilizes a mix of traditional and ensemble machine learning methodologies, such as Logistic Regression, Support Vector Machines (SVMs), Decision Trees, and Random Forests. Random Forests, in particular, have shown outstanding results thanks to their capacity to manage feature interactions and decrease overfitting, thus improving the quality of recommendations.

Farmers can utilize a user-friendly interface to input various agricultural parameters like soil type, pH levels, temperature, and rainfall. In response, they receive tailored advice that helps them choose the most appropriate crops and fertilizers for their specific environmental circumstances. These functionalities make the system practical and accessible for real-world use.

The effectiveness of the system has been assessed using standard evaluation metrics, including accuracy, precision, recall, and F1-score, which confirms its dependability across various agricultural contexts [15]. Nevertheless, to further enhance accessibility and performance—particularly in rural and low-bandwidth regions—certain optimizations are required. These modifications include refining the lightweight design of the interface and incorporating multilingual support to promote inclusivity across different language speakers.

Partnering with reputable organizations like the Indian Council of Agricultural Research (ICAR) and other national agricultural agencies would be crucial for broadening the system's reach and practical testing. Feedback from real-world usage would aid in further refining the model and improving the quality of training data, resulting in better generalization across different areas and farming practices.

Moreover, incorporating farmer-specific data, such as landholding patterns and crop histories from government or cooperative farm registries, would lead to even more personalized and context-sensitive recommendations. This integration would upgrade the system from a generalized advisory instrument to a precision farming assistant capable of managing resources, enhancing pest control strategies, and optimizing crop planning [16].

REFERENCES

[1] Domingues, T.; Brandão, T.; Ferreira, J.C. Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey. *Agriculture* 2022, *12*, 1350.

[2] Anand & Eswari, *Precision Agriculture Using Machine Learning: Applications and Challenges*, IEEE Xplore (2020)

[3] Prajapati et al., *Crop Recommendation System using ML Algorithms*, IEEE Xplore (2021)

[4] Chauhan et al., *Performance Evaluation of ML Models for Agricultural Applications*, IEEE Xplore (2023)

[5] R. Chand, J. Singh, From Green Revolution to Amrit Kaal Lessons and Way Forward for Indian Agriculture, NITI Working Paper (2023).

[6] Delfani, P., Thuraga, V., Banerjee, B. *et al.* Integrative approaches in modern agriculture: IoT, ML and AI for disease forecasting amidst climate change. *Precision Agric* 25, 2589–2613 (2024)

[7] ABHISHEK PATEL, SHILPA S SELVAN, ASEEYA WAHID, D K YADAV, RAJESHWAR SANODIYA: TRANSFORMING AGRICULTURE WITH ARTIFICIAL INTELLIGENCE (AI): A PROMISING FUTURE.

[8] Delfani, P., Thuraga, V., Banerjee, B. *et al.* Integrative approaches in modern agriculture: IoT, ML and AI for disease forecasting amidst climate change.

[9] Delfani, P., Thuraga, V., Banerjee, B. *et al.* Integrative approaches in modern agriculture: IoT, ML and AI for disease forecasting amidst climate change.

[10] Ravi, P., et al. (2020). "Agriculture crop prediction using machine learning techniques." Materials Today: Proceedings, 29, 924-928.

[11] Singh, A., et al. (2020). "Soil quality prediction using machine learning models: A review." Environmental Monitoring and Assessment, 192(10), 616.

[12] Iglewicz, B., & Hoaglin, D. (2021). How to handle missing data in datasets. Wiley-Interscience.

[13] Dong, Z., et al. (2020). "Data preprocessing and feature engineering in agriculture for machine learning models." Computers and Electronics in Agriculture.

[14] Bisht, D., et al. (2020). "Feature engineering and selection for crop prediction: A systematic review." Artificial Intelligence in Agriculture, 3.

[15] Shloka Gupta, Nishit Jain, Farmer's Assistant: A Machine Learning Based Application for Agricultural Solutions.

APPENDIX 1

A comprehensive description of all datasets together with the preprocessing techniques follows within this appendix. This appendix provides additional information to Methodology by describing specific details about the data origins and their properties throughout data preparation and the methods employed to prepare data suitable for model development and assessment.

1. Datasets

1.1 PlantVillage Dataset (for Plant Disease Detection)

- Source: PlantVillage Dataset Repository
- Data_Description:
  A total of approximately 87,000 high-resolution RGB images of crop leaves serve as the foundation of this dataset which comprises different disease conditions from three major agricultural crops including tomato, potato and grape and corn.
- Split Ratio:
  - Training/Validation: 80%
  - Testing: 20%
- Characteristics:
  - The dataset includes RGB format images with different original pixel dimensions which received standardized processing to work with convolutional neural networks (CNNs) through an input shape of 224×224 pixels.
  - Data_Balance:
    A balanced class distribution was implemented after augmentation because it reduced model preference for more frequent classes.

- Classes:

  Includes common diseases such as Early Blight, Late Blight, Leaf Curl Virus, Mosaic Virus, among others.

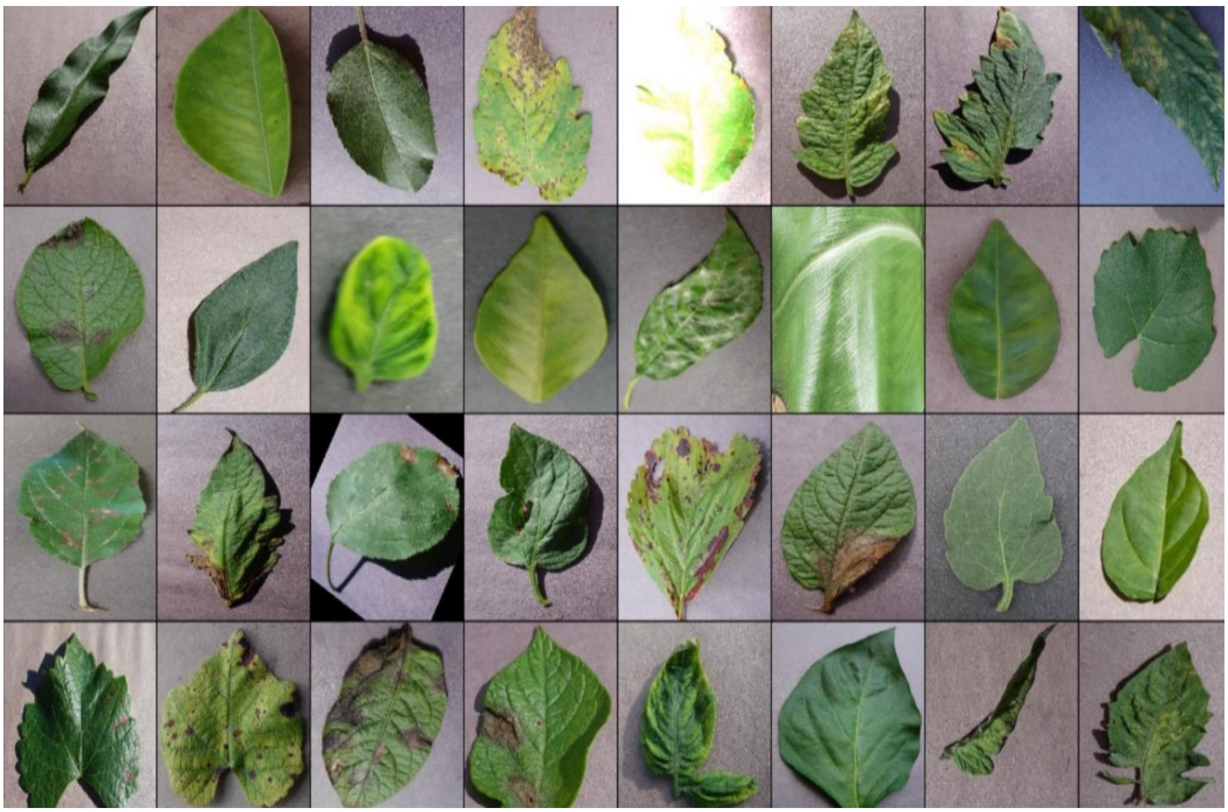

Fig. 4: Plant Disease Detection Dataset

1.2 Crop Prediction Dataset

- Source: Kaggle Open Data Platform
- Data_Description:
  The CSV document contains organized data fields that relate to crop yield assessment and recommendation tasks. Key attributes include:
    - Soil Nutrient Information
    - Average Temperature
    - Rainfall Levels
    - Soil Type
    - Historical Crop Yields
- Preprocessing Summary:
    - Missing Value Handling:

      Continuous variables: Mean imputation

      Categorical variables: Mode imputation
    - Feature Scaling:

      The model received transformed features through min-max normalization because it allowed all input data to exist on the [0,1] scale for consistent processing.
    - Data Balance:The model used SMOTE to create synthetic crop examples that balanced class distribution among different categories.

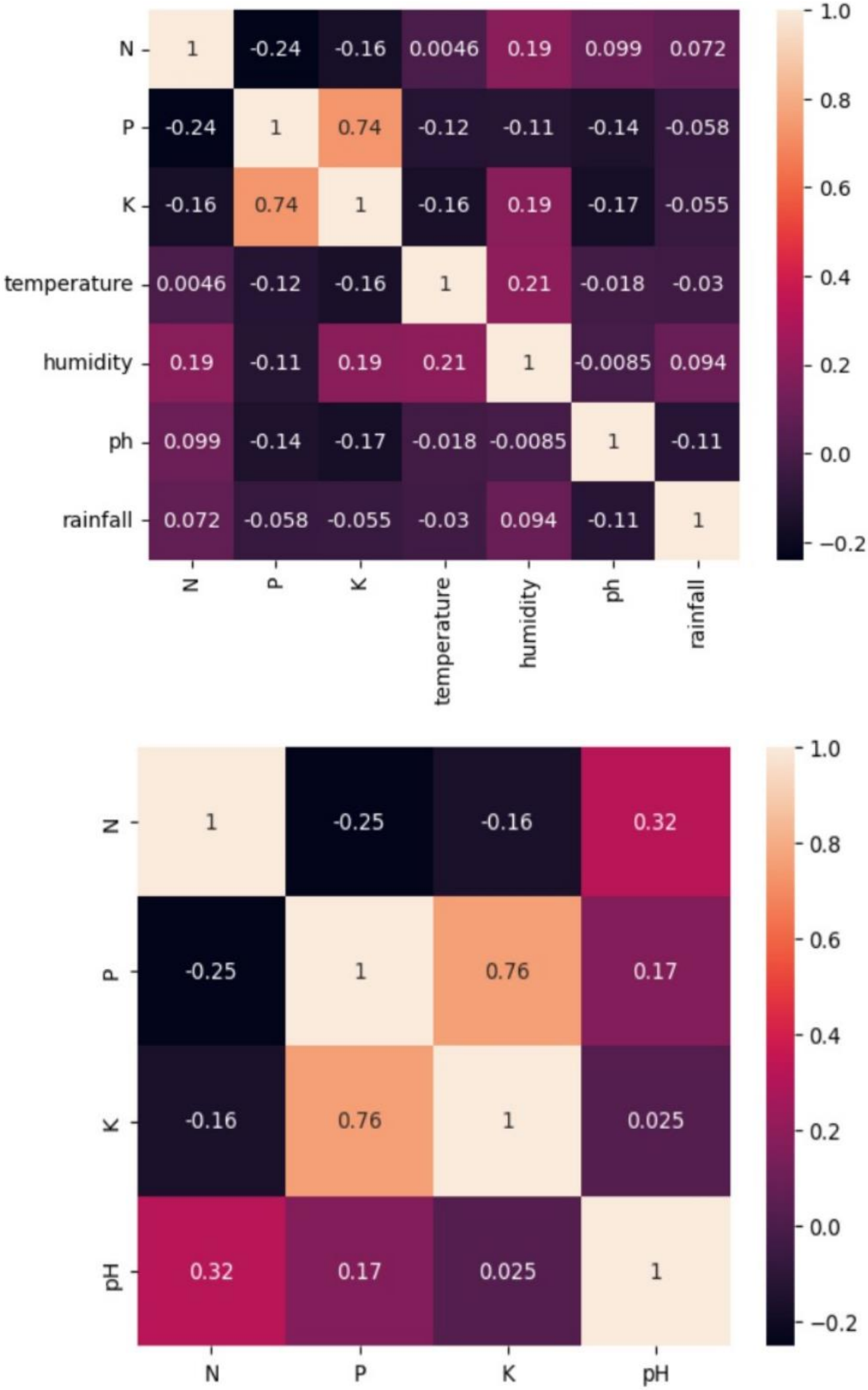

Fig. 5: Example of Crop Detection Dataset

1.3 Fertilizer Recommendation Dataset

- Source: Kaggle
- Data_Description:
  The CSV dataset includes records which support fertilizer decision making through various fields of information: The database includes specific readings of soil nutrient content composed of N, P and K.
  - Soil nutrient levels (N, P, K)
  - Crop types and requirements
  - Fertilizer properties and compatibility
- Preprocessing Summary:
  - Categorical_Encoding:
    All categorical data underwent One-Hot Encoding before transformation because this method enabled numerical machine learning algorithms to work with the encoded data.
  - Outlier Detection and Correction:

    Statistical IQR filtering methods and Z-score analysis detected outliers in soil nutrients which were later corrected by applicable winsorization or median substitution techniques.
  - Normalization:

    The normalization process standardized all numerical features so different features would not affect model learning through their magnitude values.

2. Preprocessing Techniques

2.1 Plant Disease Dataset Preprocessing

Deep learning model performance received an enhancement through these processing techniques that the image data received for optimization:

- Image-Resizing:
  All images received a uniform reformatting to 224 × 224 pixel size which matched the input requirements of ResNet-9 and related CNN models.

- Pixel-Normalization:
  Pixel values were normalized from 0 to 255 to the range of 0 to 1 which improved training convergence speed.

- Data Augmentation Techniques:
  - Rotation: Random rotations up to ±25 degrees.
  - Flipping:The process involved both horizontal and vertical flips for adding spatial randomness to the image data..
  - Brightness/Contrast Adjustment: The simulated application of different lighting levels through brightness/contrast adjustment made the model more capable of functioning in realistic environmental settings.

The deep learning model obtained improved generalization capabilities through these transformations that helped prevent overfitting because the training dataset was artificially enlarged.

2.2 Crop and Fertilizer Dataset Preprocessing

- Handling Missing Data:
    - Applied mean imputation for numerical fields (e.g., temperature, rainfall).
    - The appropriate imputation method for categorical features (such as soil type) included median or mode techniques according to each variable's distribution.
- Feature Normalization:
    - Min-Max Scaling in combination with Standard Scaling provided a solution to normalize features while decreasing training vulnerabilities. SVMs and Logistic Regression require careful input magnitude management because these models demonstrate high sensitivity to such variations.
- Synthetic Data Generation (SMOTE):
    - In order to resolve dataset class imbalance in the crop and fertilizer datasets researchers applied the SMOTE (Synthetic Minority Over-sampling Technique).
    - SMOTE creates synthetic labeled examples in feature space to advance minority class learning thus improving both F1-Score and Recall metrics significantly.

Report

ORIGINALITY REPORT

11%

SIMILARITY INDEX

PRIMARY SOURCES

1	www.coursehero.com Internet	230 words — 2%
2	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publications	98 words — 1%
3	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publications	83 words — 1%
4	idoc.pub Internet	78 words — 1%
5	fastercapital.com Internet	53 words — < 1%
6	pdfcoffee.com Internet	50 words — < 1%
7	www.mdpi.com Internet	45 words — < 1%
8	balisage.net Internet	36 words — < 1%
9	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelligent Computing	34 words — < 1%

and Communication Techniques - Volume 2", CRC Press, 2025
Publications

10 Anurag Tiwari, Manuj Darbari. "Emerging Trends in Computer Science and Its Application - Proceedings of the International Conference on Advances in Emerging Trends in Computer Applications (ICAETC-2023) December 21–22, 2023, Lucknow, India", CRC Press, 2025 — 29 words — < 1%
Publications

11 philarchive.org — 24 words — < 1%
Internet

12 www.researchsquare.com — 19 words — < 1%
Internet

13 K. Venkata Murali Mohan, M. Suresh Babu. "Disruptive Technologies in Computing and Communication Systems - Proceedings of the 1st International Conference on Disruptive technologies in Computing and Communication Systems", CRC Press, 2024 — 17 words — < 1%
Publications

14 www.engineeringmix.com — 17 words — < 1%
Internet

15 sairamce.edu.in — 16 words — < 1%
Internet

16 www.jetir.org — 16 words — < 1%
Internet

17 www.numberanalytics.com — 16 words — < 1%
Internet

18 1filedownload.com — 15 words — < 1%
Internet

19 datacenterfrontier.com — 15 words — < 1%
Internet

20	imsec.ac.in Internet	14 words — < 1%
21	www.itonlinelearning.com Internet	14 words — < 1%
22	Aditya Menon, James A. Thompson-Colón, Newell R. Washburn. "Hierarchical Machine Learning Model for Mechanical Property Predictions of Polyurethane Elastomers From Small Datasets", Frontiers in Materials, 2019 Crossref	13 words — < 1%
23	fjs.fudutsinma.edu.ng Internet	13 words — < 1%
24	net-informations.com Internet	12 words — < 1%
25	pmc.ncbi.nlm.nih.gov Internet	12 words — < 1%
26	www.researchgate.net Internet	12 words — < 1%
27	Sowmya B.J., Chetan Shetty, Seema S., Srinivasa K.G.. "An Image Processing and Machine Learning Approach for Early Detection of Diseased Leaves", International Journal of Cyber-Physical Systems, 2019 Crossref	11 words — < 1%
28	ruomoplus.lib.uom.gr Internet	11 words — < 1%
29	www.oiirj.org Internet	11 words — < 1%
30	cgspace.cgiar.org Internet	10 words — < 1%
31	home.inf.unibe.ch Internet	10 words — < 1%

32	link.springer.com Internet	10 words — < 1%
33	Huseyin Coskun. "chapter 1 AI-Based Mobile Application for Personalized Monitoring of Healthy Sitting Posture", IGI Global, 2025 Crossref	9 words — < 1%
34	Ram Kaundinya. "Chapter 14 Gaps, Opportunities and Impact of Indian Seed Industry", Springer Science and Business Media LLC, 2025 Crossref	9 words — < 1%
35	www.amazon.jobs Internet	9 words — < 1%
36	www.publishresearch.com Internet	9 words — < 1%
37	arxiv.org Internet	8 words — < 1%
38	eprints.utm.my Internet	8 words — < 1%
39	hackmd.io Internet	8 words — < 1%
40	managementjournal.usamv.ro Internet	8 words — < 1%
41	medinform.jmir.org Internet	8 words — < 1%
42	pubmed.ncbi.nlm.nih.gov Internet	8 words — < 1%
43	pure.au.dk Internet	8 words — < 1%

44	www.economyprime.com Internet	8 words — < 1%
45	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelligent Computing and Communication Techniques - Volume 3", CRC Press, 2025 Publications	6 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF

AKANKSHA TIWARI <akanksha.2125it1158@kiet.edu>

6th International Conference on Data Science and Applications : Submission (626) has been created.

Microsoft CMT <noreply@msr-cmt.org>
To: akanksha.2125it1158@kiet.edu

Wed, Apr 16, 2025 at 1:58 PM

Hello,

The following submission has been created.

Track Name: ICDSA2025

Paper ID: 626

Paper Title: Leveraging AI for Plant Disease Detection and Agricultural Predictions

Abstract:
The agriculture industry stands as an economic backbone of India by offering employment to numerous workers while simultaneously making major contributions to national GDP statistics. Various persistent agricultural obstacles including aquatic weather, depleted soil quality and regular plant disease occurrences reduce both crop productivity and farmer profits. The successful resolution of interconnected sustainable food safety issues depends on modern technological solutions which this research examines through an advanced machine learning (ML) operation system focused on crop suggestion and plant health monitoring and fertilizer management. The platform reaches its analysis results through Convolutional Neural Networks (CNNs) together with Random Forest and XGBoost models and soil information and weather data and crop health monitoring to establish actionable conclusions. XGBoost scored the highest among these algorithms at 99.3% accuracy in its prediction results and ensemble models showed high ability to recommend both crops and fertilizers. CNNs excel as image-based plant disease detectors.

Created on: Wed, 16 Apr 2025 08:28:27 GMT

Last Modified: Wed, 16 Apr 2025 08:28:27 GMT

Authors:
- akanksha.2125it1158@kiet.edu (Primary)

Primary Subject Area: Data Models and Algorithms

Secondary Subject Areas:
Data Science Applications

Submission Files:
Leveraging AI for Plant Disease Detection and Agricultural.docx (82 Kb, Wed, 16 Apr 2025 08:27:24 GMT)

Submission Questions Response:
1. Conflict of interest
   Agreement accepted
2. Status of using third-party material in your article.
   I am not using third-party material for which formal permission is required.
3. Certificate of originality
   Agreement accepted
4. Corresponding Author's Contact number
   +91 6392280347
5. Country Name
   India
6. Authors Contributions
   Author_1 - Akanksha

Akanksha took charge of creating the abstract along with the proposed methodology section. The members of our team first performed a literature research on multiple papers before collecting appropriate information. The literature review served as a foundation for creating both the abstract and associated keywords. The proposed method emerged after observing selected methods used to manage the recognized issue.

Author_2 – Abhishree
Abhishree played a role in drafting the introduction while leading the identification of classification techniques. She dedicated her effort to understand the modern technological landscape and analyze the present situation. The evaluation of her research produced suitable classifiers which were chosen for implementation.

Author_4 – Navya
Navya dedicated her effort to write up the "Materials and Methods" portion. She handled the data cleaning along with preprocessing activities because the project involved multiple methods to achieve model accuracy improvement. Her work played a vital role to establish the reliability and quality standards of the utilized dataset.

Author_5 – Mudita
Mudita developed the research conclusion and organized the work that would be conducted next. She researched new possibilities for innovation which could become effective in future academic research. She effectively condensed all essential research discoveries and offered well-structured outlines that directed future study projects.

Author_3 and Author_6 – Miss Ayushi and Mr Himan
Miss Ayushi together with Mr. Himan maintained their mentor roles during the entire research project. The researchers provided essential guidance throughout the entire project duration which proved crucial for both conception and completion of all the project stages. We received beneficial guidance and suggestions from them that resulted in enhancing the overall quality of our research work.

Thanks,
CMT team.