A

Project Report

on

Agri-Go

submitted as partial fulfilment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2024-25

in

Information Technology

by

Akanksha Tiwari (2100290130015)

Abhishree Bisht (2100290130008)

Navya Rajvanshi (2100290100108)

Mudita Shakya(2100290100101)

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

TABLE OF CONTENTS

Page No.

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: Akanksha Tiwari

Roll No.:2100290130015

Date:30/11/24

CERTIFICATE

This is to certify that the project report entitled "Agri-Go" which is submitted by Akanksha Tiwari, Abhishree Bisht, Navya Rajvanshi and Mudita in partial fulfilment of the requirement for the award of degree B. Tech. in the department of Information Technology 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. Adesh Kumar Pandey**
(Designation)	**(Head of Department)**

Date: 30/11/24

ACKNOWLEDGEMENT

It gives us great pleasure to present the report of the B. Tech project undertaken during B. Tech. Final Year. We owe a special gratitude to Ms. Ayushi , Department of Information Technology, KIET, Ghaziabad, for his constant support and guidance throughout our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us.

We also take the opportunity to acknowledge the contribution of Dr. Adesh Kumar Pandey, head of the Department of Information Technology, 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 department's faculty members 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: 30/11/24
Signature:
Name : Akanksha Tiwari
Roll No.: 2100290130015

ABSTRACT

The agriculture sector, a pillar of India's economy, employs a sizable section of the workforce and contributes significantly to the country's GDP. However, it faces chronic obstacles such as erratic weather patterns, deteriorating soil fertility, and frequent plant diseases, all of which have a detrimental influence on crop yields and farmer revenue. Using sophisticated technologies is critical for tackling these concerns and guaranteeing sustainable agriculture and food security.This paper describes an advanced machine learning (ML) platform used for crop recommendation, plant disease diagnostics, and fertilizer optimization. The platform uses complex algorithms such as Convolutional Neural Networks (CNNs), Random Forest, and XGBoost, as well as soil, weather, and crop health data, to provide actionable insights. These algorithms provide for accurate predictions and recommendations, with XGBoost having the highest scoring 99.3%, CNNs excel at detecting plant diseases through images, whereas ensemble models such as Random Forest and XGBoost provide accurate crop and fertilizer recommendations.

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
KNN	K-Nearest Neighbors
ML	Machine Learning
NPK	Nitrogen, Phosphorus, Potassium
SMOTE	Synthetic Minority Oversampling Technique
SVM	Support Vector Machine
VGG	Visual Geometry Group (used in CNN architecture)

Abbreviation	Full Form
RGB	Red, Green, Blue (color model for images)
F1 Score	Harmonic Mean of Precision and Recall
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Agriculture is an important part of India's economy, employing roughly half of the country's population and accounting for 16% of its GDP.[1] Despite its importance, the industry faces ongoing challenges such as unpredictable weather patterns, deteriorating soil fertility, and frequent outbreaks of plant diseases. These challenges have a considerable impact on agricultural production and farmer livelihoods, highlighting the urgent need for sustainable farming practices. Furthermore, with global hunger projected to triple by 2050, there is a rising need for precision agriculture to effectively manage natural resources and avoid overexploitation.

In recent years, machine learning (ML) and artificial intelligence (AI) have emerged as disruptive technologies for modernising old agricultural

techniques. These tools allow farmers and stakeholders to make data-driven decisions about crop selection, disease diagnosis, soil analysis, and resource management. Decision Trees, Random Forest, XG- Boost, [2][3]and Convolutional Neural Networks (CNNs) have all proven to be quite effective in handling these difficulties. Research emphasizes the ability of CNNs to detect plant illnesses with high accuracy using image-based analysis, whereas ensemble methods like Random Forest and XGBoost have demonstrated remarkable performance in predictive analytics and crop recommendation tasks. T-
his study aims to fill current gaps by delivering a comprehensive platform driven by cutting-edge machine learning techniques. The platform combines several variables, such as soil factors, weather conditions, and plant health indicators, to give farmers with actionable insights and personalized suggestions. Its primary features include crop selection guidance, early detection of plant illnesses via picture analysis, and optimum fertilizer application based on soil and crop

requirements.[4] To provide exact and dependable results, the platform employs a variety of models such as Decision Trees, Naive Bayes, XGBoost, and CNNs. For example, CNNs thrive in diagnosing plant health through image-based analysis, while XGBoost and Random Forest are used for predictive analytics, guaranteeing farmers receive realistic solutions to complicated agricultural challenges.

Furthermore, the platform's strong performance is supported by the high accuracy of its models. XGBoost has an accuracy of 99.3%, Logistic Regression 94.7%, Naive Bayes 98.8%, and Decision Tree 91.5%, however CNNs routinely outperform in disease identification. This result demonstrates the potential of ML-powered solutions to improve agricultural output, reduce environmental hazards, and enable sustainable practices on a wide scale.

Furthermore, this method is consistent with worldwide initiatives aimed at developing resilient farming systems capable of meeting the simultaneous challenges of climate change and rising food demand. By providing farmers with cutting-edge tools and insights, the platform not only addresses India's severe agricultural concerns, but also helps to assure food security and sustainable development on a global scale.

Machine learning (ML) and artificial intelligence (AI) are emerging as important technologies for transforming traditional agriculture methods. These technologies give data-driven insights, enabling wiser decision-making in a variety of fields, including:

- Crop Recommendation: Identifying the best crops based on soil type, weather circumstances, and past performance.

- Disease Diagnosis: Detecting plant diseases in their early stages using image-based analysis to prevent extensive crop harm.

- Fertilizer optimization entails recommending exact fertilizer applications based on unique soil and crop requirements.

Soil analysis involves evaluating soil factors such as pH, moisture, and nutrient content to maintain fertility.

W

1.2 PROJECT DESCRIPTION

This project focuses on leveraging advanced machine learning (ML) and artificial intelligence (AI) technologies to address three critical areas in agriculture: crop prediction, plant disease detection, and fertilizer recommendation. By developing an integrated platform that provides precise, data-driven recommendations, the project aims to improve agricultural productivity, sustainability, and resource optimization.

The crop prediction module employs machine learning algorithms such as logistic regression, support vector machines (SVM), and decision trees to analyze environmental factors like soil type, temperature, and rainfall. This system predicts the most suitable crops for a given region, empowering farmers to make informed planting decisions based on scientific data.

The plant disease detection module utilizes a Convolutional Neural Network (CNN) trained on the PlantVillage Dataset, which contains over 87,000 labeled images of healthy and diseased crop leaves. This module enables early identification of diseases through image analysis, allowing farmers to take timely actions to prevent large-scale crop losses. The system is further enhanced using transfer learning techniques, employing pre-trained models like ResNet50 and VGG16 to boost accuracy and reduce computational requirements.

The fertilizer recommendation module combines predictive models, such as K-Nearest Neighbors (KNN), Gaussian Naive Bayes, and ensemble learning methods like Random Forest, to suggest optimal fertilizers based on soil nutrient profiles and crop requirements. This ensures balanced nutrient application, enhances soil health, and minimizes wastage of resources.

The project incorporates exploratory data analysis (EDA) to uncover patterns and trends in datasets while addressing issues like missing values and class imbalance using techniques such as SMOTE. Feature engineering is applied to extract meaningful information, enhancing model performance. Preprocessing methods, including data normalization, augmentation, and encoding, ensure consistency and improve the robustness of predictions.

The platform is designed with a user-friendly web interface, enabling farmers to input basic parameters like soil quality and crop health while receiving actionable insights. Additionally, the system is optimized for deployment in rural areas, considering constraints like low bandwidth and the need for multilingual support.

This project not only addresses pressing challenges in agriculture but also aligns with global efforts to promote sustainable farming practices. By integrating IoT data, satellite imagery, and climate projections in future iterations, the platform aims to revolutionize precision agriculture and contribute to food security in a changing world.

CHAPTER 2

LITERATURE REVIEW

The combination of technology and data analytics is propelling a revolution in agriculture, promising higher production and financial success. The convergence of Industry 4.0 and the digital agriculture revolution has offered new prospects for modernizing old agricultural systems, particularly in nations such as India, where agriculture is essential to livelihoods and accounts for 16% of GDP. Digital technologies, such as artificial intelligence (AI) and machine learning (ML), are rapidly being used to improve agricultural operations, with an emphasis on crop recommendations, plant disease diagnosis, and fertilizer optimization.[5]

Plant Disease Detection

The diagnosis of plant diseases has received a lot of attention in recent years, thanks to advances in machine learning. Simona E. Grigorescu et al. investigated the application of Gabor filter-based texture analysis to improve feature extraction for Convolutional Neural Networks (CNNs). These methods dramatically increase the accuracy of illness identification using image analysis. Similarly, Dheeb Al Bashish et al. created a neural network classifier that uses statistical classification methods to detect and classify disorders well. Advances in image processing enable deep learning models to detect pre-symptomatic diseases by studying minute changes in plant leaves, with accuracies exceeding 95% and frequently outperforming traditional manual assessments. However, scaling across different plant species and climatic conditions remains a barrier, restricting the broad use of these approaches.

Crop Recommendation Systems

Crop recommendation systems improve agricultural efficiency by recommending appropriate crops based on environmental and soil data. Taj et al. introduced a hybrid model that combines Artificial Neural Networks (ANNs) for regression and K-Nearest Neighbors (KNN) for classification, resulting in tailored suggestions for various agricultural situations. Building on this, Banavlikar et al. used neural networks to align crop selection with soil and temperature data, resulting in tailored advise for farmers. Despite these advances, there are still issues in adapting models to different agricultural contexts and incorporating real-time data, which are required for practical and scalable implementation in dynamic farming operations.

Fertilizer Optimization

Optimizing fertilizer use is crucial for increasing crop yields while reducing environmental effect. Hussain et al. investigated ML-based systems that customize fertilizer recommendations depending on soil parameters and crop needs. These systems prioritize data-driven decision-making to enhance nutrient application and productivity. However, present models frequently rely on limiting characteristics like pH and moisture, failing to account for the intricate interaction of factors influencing soil health and crop nutrition. Expanding these models to include a broader variety of data, such as micronutrient levels and past crop performance, may improve their usefulness.

Technology Integration and AI Applications

Recent breakthroughs indicate AI's potential to alter agricultural methods. For example, in 2021, ICAR began working with private companies to develop AI-powered disease forecasting models and precision agriculture solutions. These technologies use remote sensing and the Internet of Things (IoT) to optimize resource management, particularly water and pesticide use [6]. Similarly, in 2019, NITI Aayog partnered with IBM to develop AI-powered weather forecasting technologies, allowing farmers to make more educated planting and irrigation decisions [7]. In 2020, the World Bank-backed "Sustainable Agriculture in a Changing Climate" effort used AI to improve cropping patterns in response to changing climatic circumstances, allowing farmers to conserve resources while increasing yields. AI-powered systems such as eNAM (National Agriculture Market), which began in 2016, use machine learning algorithms to match supply and demand, assuring fair pricing for farmers [8]. This internet trading tool has helped millions gain improved market access. Furthermore, tailored agricultural advice systems, such as IBM Watson, use machine learning and environmental data to select appropriate crops. Studies show that these algorithms can anticipate crop yields and soil suitability with up to 90% accuracy, allowing farmers to make better decisions.[9]

CHAPTER 3

PROPOSED METHODOLOGY

The planned study employs a methodical approach to tackle significant agricultural issues. It starts with data collection on agricultural yields, weather patterns, and soil health from reputable sources such as ICAR, IMD, and commercial labs. Consistency is ensured by data preprocessing, which includes feature engineering to extract features like pest rates and drought indices, normalization (e.g., Min-Max scaling), and imputation techniques to handle missing values.

SVMs evaluate soil and climatic data for disease risks, Decision Trees and Random Forests are used for crop disease classification, and XGBoost is utilized for predictive crop recommendations because of its efficiency and scalability for model selection. Data splitting (80:20 ratio), hyperparameter tuning, and model evaluation using measures like accuracy, F1-score, and MSE are all part of the model training and evaluation process.

By combining forecasts with suggestions, the system provides useful information such as crop recommendations and disease alerts based on weather and soil data. Lastly, a user interface is created that enables farmers to enter information and get real-time guidance. The platform ensures accessibility and usefulness for a wide range of users by supporting many languages and offering user-friendly visualizations.

CHAPTER 4

RESULTS AND DISCUSSION

According to the study, ResNet-9's deep learning architecture, which effectively extracted intricate visual patterns from plant leaf photos, allowed it to perform exceptionally well in plant disease identification. By addressing training issues, the model's residual connections enabled it to efficiently handle complex datasets. ResNet-9 fared better than conventional models like Decision Trees and SVMs by automatically extracting pertinent features, which made it extremely accurate in identifying a variety of plant illnesses and providing insightful information for precision agriculture.

Various machine learning models' crop recommendation performance metrics are provided. The following algorithms were evaluated using F1-Score, Precision, and Recall: Naïve Bayes, XGBoost, Random Forest, SVM, Decision Trees, and Logistic Regression. XGBoost has the highest accuracy of 99.3%. Results were greatly enhanced by SMOTE and feature engineering, especially when it came to managing dataset imbalances and improving model correctness.

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

In order to address important agricultural issues like crop forecasting, plant disease detection, and fertilizer recommendations, this research offers an AI-based platform. The disease detection module correctly detects plant diseases using a CNN technique, allowing for early intervention. The crop prediction model recommends the best crops depending on environmental parameters by combining logistic regression, SVM, and decision trees. The precision of crop and fertilizer recommendations is further increased by Random Forest classifiers. Farmers can enter data and get useful insights thanks to an intuitive UI. Metrics like as accuracy, precision, recall, and F1-score were used to validate the platform's performance, guaranteeing its dependability for real-world applications.

For wider implementation, particularly in rural areas, the system should be scaled and optimized for low-bandwidth conditions. Simplifying the user interface and supporting other languages would increase the platform's accessibility to a wider audience. Collaboration with agricultural research institutions and government agencies, such as ICAR, may also help confirm the system's predictions on a broader scale and enhance the quality of data used to train models. Finally, implementing more tailored recommendations based on detailed data from farms could improve crop rotation, pest management, and resource utilization, resulting in more precise and sustainable agricultural methods.

In conclusion, while this study lays the groundwork for the integration of AI and machine learning in agriculture, ongoing improvements in real-time data integration, model accuracy, and system scalability are critical for realizing the platform's full potential and impact in the real world.

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. 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.

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

It provides an in-depth description of the datasets and preprocessing techniques employed in the study. It serves to complement the methodology section, offering comprehensive information to ensure the research can be replicated and understood fully.

1. Datasets

PlantVillage Dataset (Plant Disease Detection):

Source: PlantVillage Dataset Repository.

Data: 87,000 RGB images of crop leaves, categorized into 38 classes (healthy and diseased conditions).

Split: 80% for training and validation, 20% for testing.

Characteristics:

Image format: RGB, varying resolutions standardized for CNN processing.

Balanced class distributions were ensured after augmentation.

Crop Prediction Dataset:

Source: Kaggle.

Data: Structured CSV dataset with features including soil type, temperature, rainfall, and historical crop yields.

Preprocessing:

Missing values handled using mean/mode imputation.

Features scaled using normalization techniques.

Fertilizer Recommendation Dataset:

Source: Kaggle.

Data: CSV dataset including soil nutrient levels (NPK), crop requirements, and fertilizer properties.

Preprocessing:

Features encoded using one-hot encoding for categorical variables like soil type.

Outlier detection and correction.

2. Preprocessing Techniques

Plant Disease Dataset Preprocessing:

Images resized to a consistent resolution (e.g., 224x224 pixels).

Normalization scaled pixel values between 0 and 1.

Data augmentation included rotation, flipping, and brightness adjustments to increase model robustness.

Crop and Fertilizer Dataset Preprocessing:

Missing data filled using imputation techniques (mean, median).

Features normalized to ensure uniform scales.

Synthetic Minority Oversampling Technique (SMOTE) applied to balance imbalanced datasets.