

Smart Plantation Strategy Based on Local Environment and Field Needs

Smt. Rekha S Patil

Associate Professor

Department of Computer Science
and Engineering
PDA college of engineering,
Kalaburgi, India

Swati Parit

Student

Department of Computer
Science and Engineering
PDA College of Engineering,
Kalaburgi, India

Annapurna

Student

Department of Computer Science
and Engineering
PDA College of Engineering,
Kalaburgi, India

Abstract - This project presents a Smart Plantation Strategy System that integrates machine learning with an interactive graphical user interface to support data-driven agricultural decision-making. The system is designed using Python, CustomTkinter (CTk) for an intuitive dark-green themed interface, and scikit-learn for predictive modeling. A curated dataset containing soil type, pH, nutrient levels (N, P, K), irrigation practices, rainfall, and temperature is utilized to build a Linear Regression, Random forest model for yield prediction. In addition, a crop recommendation module applies similarity-based filtering to suggest the most suitable crops based on local environmental conditions and field needs. The application provides multiple modules including dataset visualization, crop recommendation, yield prediction, and plantation strategy planning, making it a comprehensive decision-support tool for farmers. By combining machine learning algorithms with user-friendly visualization, the system assists in optimizing crop selection, estimating yield potential, and formulating sustainable plantation strategies, thereby promoting precision agriculture and resource-efficient farming practices.

Keywords: soil type, pH, nutrient levels (N, P, K), irrigation practices, rainfall, temperature, machine learning, Linear Regression, Random forest

I. INTRODUCTION

Agriculture remains the backbone of the global economy, particularly in developing nations where the livelihood of millions depends on farming. However, traditional farming practices often rely on generalized approaches that do not fully account for the variability in soil type, climatic conditions, and field-specific requirements. This mismatch frequently results in reduced crop productivity, inefficient use of resources, and vulnerability to environmental fluctuations.

To address these challenges, the concept of smart farming has emerged, leveraging data analytics, machine learning, and interactive digital platforms to enhance decision-making in agriculture. A smart plantation strategy focuses on selecting the right crop, predicting potential yield, and planning cultivation practices based on localized environmental parameters such as soil fertility, pH balance, nutrient content, rainfall, temperature, and irrigation availability.

This project introduces a Smart Plantation Strategy System built using Python, CustomTkinter for the user interface, and machine learning for predictive analytics. The system enables farmers and researchers to visualize agricultural datasets, receive crop recommendations tailored to their soil and climate conditions, and predict potential yield with greater accuracy. It also incorporates a plantation strategy module to support sustainable agricultural planning. By offering real-time insights through an intuitive dashboard, the system bridges the gap between raw agricultural data and practical farming decisions.

Ultimately, this solution contributes to precision agriculture, ensuring optimal resource utilization, enhanced crop productivity, and improved resilience to climate variability, while empowering farmers with technology-driven decision support..

II. LITERATURE SURVEY

M. A. Javed et al. [1], 2024 – Presented a comprehensive review of machine learning and deep learning methods for crop yield prediction, discussing algorithms, datasets, and evaluation metrics to guide future research.

S. M. Shawon [2], 2024 – Conducted a systematic literature review of crop yield prediction using ML techniques, emphasizing the role of climatic and soil features along with ensemble methods for improved accuracy.

M. K. Senapaty et al. [3], 2024 – Proposed a Decision Support System (DSS) for crop recommendation using machine learning, integrating location-specific data and classification models to assist farmers.

M. Bouni et al. [4], 2024 – Developed an IoT-based framework that combines sensor data (soil moisture, pH, weather) with machine learning to provide real-time crop recommendation and monitoring.

Y. Wang [5], 2024 – Reviewed recent advancements in deep learning for crop yield prediction, analyzing architectures such as CNN, LSTM, and hybrid models, and comparing them to simpler ML models.

K. Jhajharia et al. [6], 2023 – Compared multiple machine learning models (Linear Regression, Random Forest,

XGBoost, ANN) for yield prediction, showing that ensemble methods often outperform simple models.

S. Kiruthika et al. [7], 2023 – Designed an IoT-based professional crop recommendation system using agro-climatic features such as pH, humidity, and temperature to suggest crops to farmers.

G. Saha et al. [8], 2025 – Proposed an IoT-driven smart precision agriculture framework for land mapping, crop selection, and irrigation scheduling, integrating AI-based decision-making.

H. Afzal et al. [9], 2025 – Investigated the role of soil nutrient information (N, P, K, pH) in crop recommendation, demonstrating that including soil data significantly improves model accuracy.

S. Shastri et al. [10], 2025 – Presented a crop recommendation system using supervised ensemble models (Random Forest, Gradient Boosting), achieving high accuracy in predicting suitable crops.

Anonymous Authors [11], 2024 – Developed a machine learning-based crop recommendation framework using soil (NPK, pH) and climatic parameters, applying preprocessing and normalization to improve accuracy.

F. S. Prity et al. [12], 2024 – Proposed a crop recommendation system that integrates historical yield data, farmer preferences, and climatic projections for personalized decision-making.

M. D. S. Sharafat et al. [13], 2025 – Designed an IoT-enabled AI system for real-time crop prediction using sensor data streams, enabling dynamic updates to recommendations.

Farhan T. Tonmoy et al. [14], 2025 – Introduced a portable smart soil monitoring system with NPK and pH sensors, integrated with ML models for real-time crop and fertilizer recommendation.

Anonymous Authors [15], 2025 – Compared Random Forest and XGBoost regressors for yield prediction, reporting high accuracy and highlighting the importance of feature selection and tuning. .

III. METHODOLOGY

The methodology for the Smart Plantation Strategy System integrates dataset preprocessing, machine learning algorithms, and GUI-based user interaction to support decision-making in agriculture. The workflow is divided into three core modules:

Data Preprocessing & Feature Engineering
Crop Recommendation Algorithm
Yield Prediction using Linear Regression
GUI Integration & User Interaction

A. Data Preprocessing & Feature Engineering

Algorithm: Label Encoding for Categorical Variables

- Categorical attributes (Soil Type, Irrigation Method, Recommended Crop) are encoded into numeric form using *Label Encoding* from scikit-learn.
- This allows machine learning models to process categorical data effectively.

Steps:

1. Import dataset from CSV.
2. Identify categorical and numerical attributes.
3. Apply LabelEncoder() to categorical columns.
4. Store encoded mappings for later inverse transformations.

B. Crop Recommendation Algorithm

This module applies a similarity-based filtering approach using Euclidean Distance to recommend crops that best match local field and climate conditions.

Mathematical Formula (Euclidean Distance):

$$d = \sqrt{(pH_{user} - pH_{data})^2 + \frac{(Rainfall_{user} - Rainfall_{data})^2}{10000} + (Temp_{user} - Temp_{data})^2}$$

Where:

- $pH_{user}, Rainfall_{user}, Temp_{user}$ → User inputs
- $pH_{data}, Rainfall_{data}, Temp_{data}$ → Dataset values
- The rainfall term is normalized by dividing with 10000 to reduce scale imbalance.

Algorithm Steps (Crop Recommendation):

1. Accept user inputs: Soil Type, pH, Rainfall, Temperature.
2. Filter dataset rows matching soil type (approximate match).
3. For each record in the dataset:
Compute Euclidean distance between user inputs and dataset record.
4. Sort records in ascending order of distance.
5. Select top-N nearest crops (e.g., N = 5).
6. Display recommended crop list to the user.

C. Yield Prediction using Linear Regression

This module uses Linear Regression from scikit-learn to estimate yield potential based on soil nutrients, environmental parameters, and irrigation type.

Mathematical Model (Linear Regression):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

Where:

- Y = Yield potential (t/ha)
- X_1, X_2, \dots, X_n = Features (Soil Type, pH, N, P, K, Rainfall, Temperature, Irrigation)
- β = Coefficients learned from training
- ϵ = Error term

Algorithm Steps (Yield Prediction):

1. Encode categorical inputs (Soil Type, Irrigation).
2. Collect user input values for features:
Soil Type, pH, N, P, K, Rainfall, Temperature, Irrigation.
3. Form a feature vector:

$$X = [SoilType, pH, N, P, \downarrow Rainfall, Temp, Irrigation]$$

4. Pass X into the trained Linear Regression model.
4. Compute predicted yield Y.

6. Display estimated yield potential in tons per hectare (t/ha).

D. Yield Prediction using Random Forest

1) Step 1: Dataset Preparation

Collect dataset with features such as **Soil Type, pH, N, P, K, Rainfall, Temperature, Irrigation** and target variable (Recommended Crop or Yield).

Preprocess data:

Encode categorical features (Label Encoding).

Normalize or scale numerical features (optional).

Handle class imbalance (e.g., SMOTE or class weights).

2) Step 2: Bootstrap Sampling

From the training dataset of size N , create k random samples (with replacement).

Each sample is called a **bootstrap sample**, typically containing ~63% of the original data.

The remaining ~37% (out-of-bag data) can be used for validation.

3) Step 3: Build Decision Trees

For each bootstrap sample:

Select a random subset of features (not all features are considered at once).

Grow a **Decision Tree** using recursive splitting:

At each node, choose the best split among the randomly selected features.

Splitting criterion depends on task:

Classification: Gini Impurity or Entropy.

Regression: Mean Squared Error (MSE).

Stop growing the tree when:

Maximum depth is reached, OR

Minimum samples per split/leaf condition is met.

4) Step 4: Ensemble Learning

Repeat Step 3 for all trees (e.g., 100–500 trees).

Store all trained decision trees as part of the forest.

5) Step 5: Prediction Phase

For Classification (Crop Recommendation):

Each tree predicts a class label.

The final class is chosen by **majority voting** across all trees.

$$\hat{y} = \text{mode}(y_1, y_2, y_3, \dots, y_T)$$

For Regression (Yield Prediction):

Each tree predicts a numeric value.

The final prediction is the **average of all tree outputs**.

$$\hat{y} = \frac{1}{T} \sum_{i=1}^T y_i$$

6) Step 6: Model Evaluation

Evaluate using metrics:

Classification: Accuracy, Precision, Recall, F1-score.

Regression: R^2 Score, Mean Absolute Error (MAE), Mean Squared Error (MSE).

Tune hyperparameters ($n_estimators$, max_depth , $min_samples_leaf$, etc.) to reduce overfitting.

E. System Flow Summary

Step 1: Data Collection → Soil, Climate, Irrigation, Nutrients

Step 2: Data Preprocessing → Label Encoding, Normalization

Step 3: Crop Recommendation → Similarity Matching (Euclidean Distance)

Step 4: Yield Prediction → Linear Regression Model

Step 5: GUI Dashboard → Display Results to End User

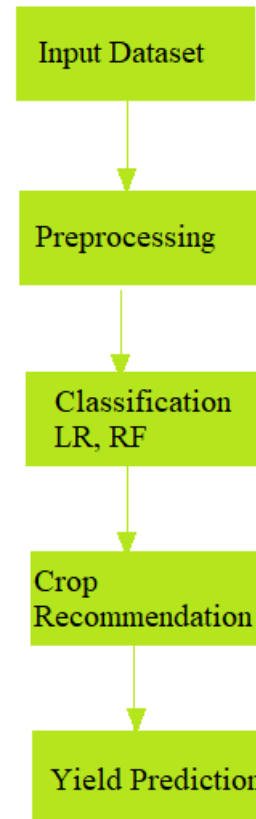


Figure 1: Proposed Methodology Flow Diagram

The flow diagram illustrates the working process of the Smart Plantation Strategy System. The system begins with the input dataset, which contains agricultural parameters such as soil type, pH, nitrogen, phosphorus, potassium, rainfall, temperature, and irrigation details. This data undergoes a preprocessing stage where categorical features are encoded, scaling is applied, and missing or imbalanced data is handled to ensure consistency. After preprocessing, the dataset is passed into classification models such as Logistic Regression (LR) and Random Forest (RF), which analyze the relationships between field attributes and crop types. Based on these models, the crop recommendation module suggests the most suitable crops for the given soil and climatic conditions. Finally, the yield prediction module estimates the potential productivity (t/ha) of the recommended crops, allowing farmers to make informed decisions for sustainable and profitable plantation strategies.

IV. RESULTS

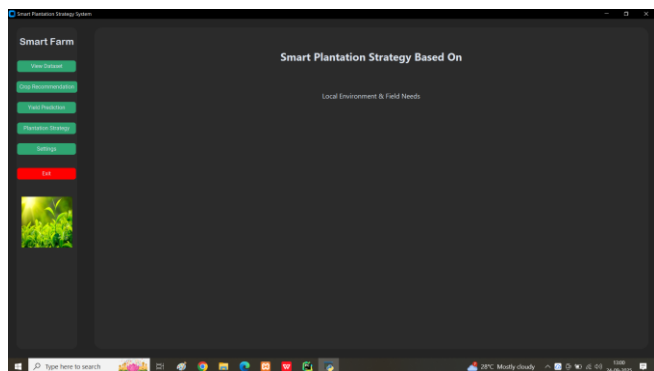


Figure 2: Main Screen

Main screen of the project

View Dataset → Opens a dataset viewer where users can see the agricultural dataset in tabular format.

Crop Recommendation → Suggests the most suitable crops based on soil, pH, rainfall, and temperature inputs.

Yield Prediction → Estimates the yield potential (t/ha) for selected crops using the trained machine learning model.

Plantation Strategy → Provides strategic insights for planning sustainable cultivation.

Settings → Allows customization of the system preferences.

Exit → Closes the application safely.

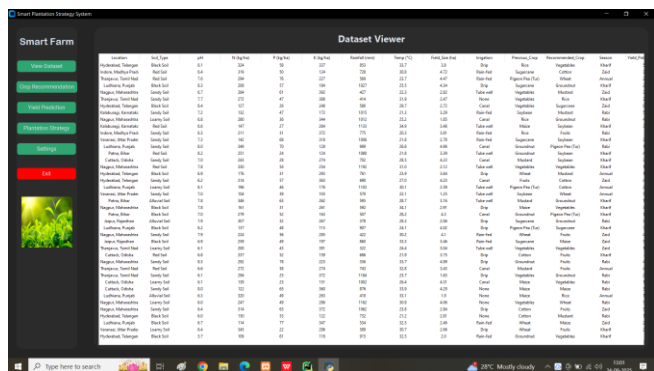


Figure 3: Dataset

The Dataset Viewer displays the agricultural dataset that is used for crop recommendation and yield prediction. Each row represents a record of field conditions, while the columns represent different agricultural attributes, such as:

Soil_Type → type of soil (e.g., Loamy, Sandy, Clayey).

pH → soil pH level.

N (kg/ha), P (kg/ha), K (kg/ha) → nitrogen, phosphorus, and potassium content in the soil.

Rainfall (mm) → rainfall amount in millimeters.

Temp (°C) → average temperature.

Irrigation → irrigation type used.

Recommended_Crop → crop suggested for the given conditions.

Yield_Potential (t/ha) → estimated crop yield in tons per hectare.

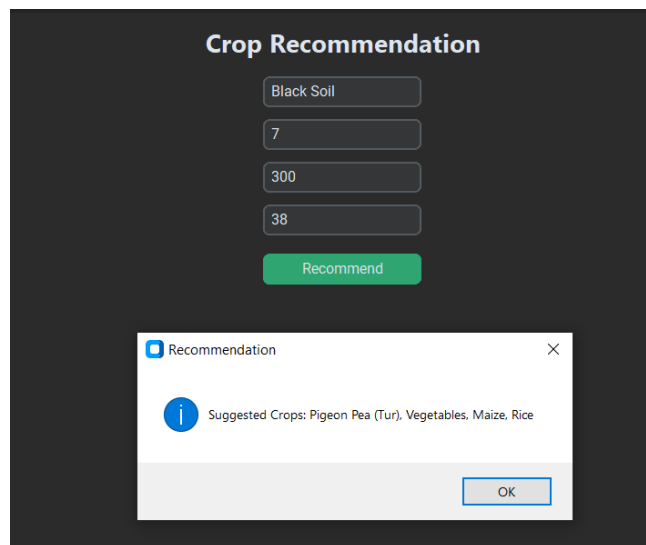


Figure 4: Crop Recommendation

The screenshot illustrates the Crop Recommendation module of the Smart Plantation Strategy System. In this module, the user enters specific field conditions such as soil type, pH value, rainfall, and temperature. Based on these inputs, the system applies a similarity-based filtering algorithm using the dataset to identify crops best suited for the given environment. As shown in the example, when the user inputs Black Soil, pH 7, 300 mm rainfall, and 38°C temperature, the system suggests crops such as Pigeon Pea (Tur), Vegetables, Maize, and Rice. The results are displayed in a popup dialog box, ensuring easy and quick access to crop recommendations. This functionality helps farmers select the most appropriate crops for their field conditions, thereby improving productivity and supporting sustainable agriculture.

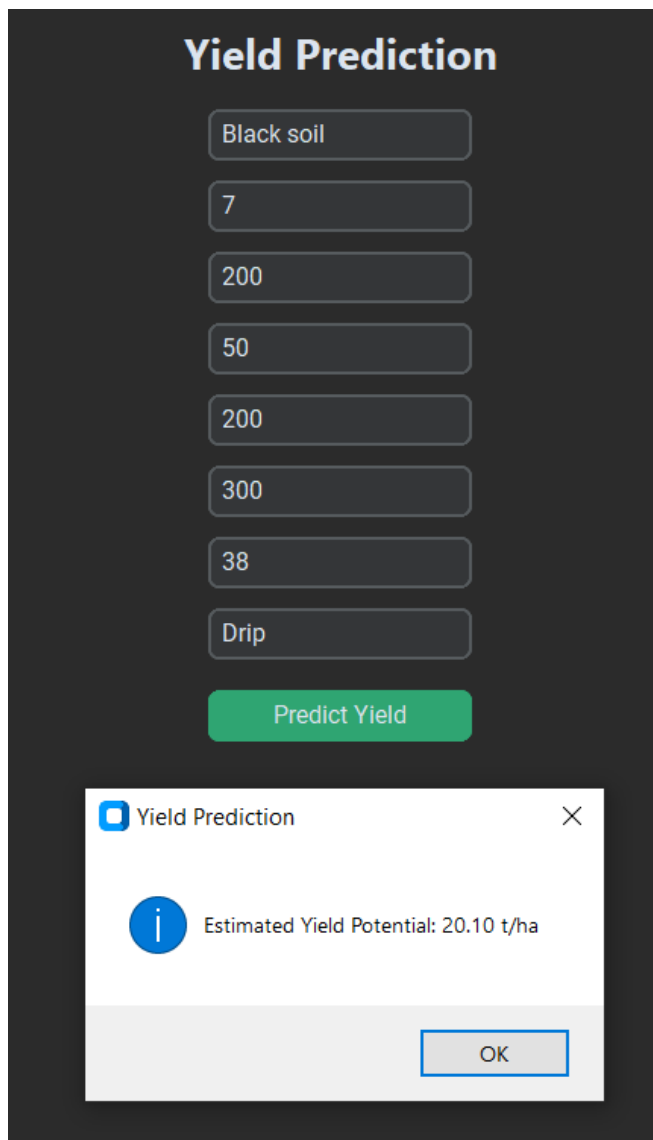
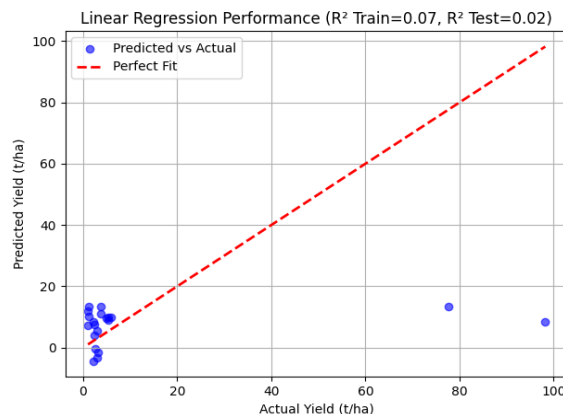


Figure 5: Yield Prediction

This screenshot represents the **Yield Prediction module** of the Smart Plantation Strategy System. In this module, the user enters detailed agricultural parameters including soil type, soil pH, nitrogen (N), phosphorus (P), potassium (K), rainfall, temperature, and irrigation method. Once the user provides these inputs and clicks on **Predict Yield**, the system processes the data using the trained **Linear Regression model**. The model evaluates the influence of soil nutrients, environmental conditions, and irrigation on crop productivity to generate an estimated yield. In the example shown, for *Black soil with pH 7, N=200 kg/ha, P=50 kg/ha, K=200 kg/ha, rainfall 300 mm, temperature 38°C, and drip irrigation*, the system predicts an **estimated yield potential of 20.10 t/ha**. The result is displayed in a popup message box, making it simple and user-friendly. This feature provides farmers with a reliable way to estimate productivity in advance, enabling better field management and strategic planning.



Graph 1:LR Performance

This graph shows the performance of the Linear Regression model used in the Yield Prediction module of your Smart Plantation Strategy System.

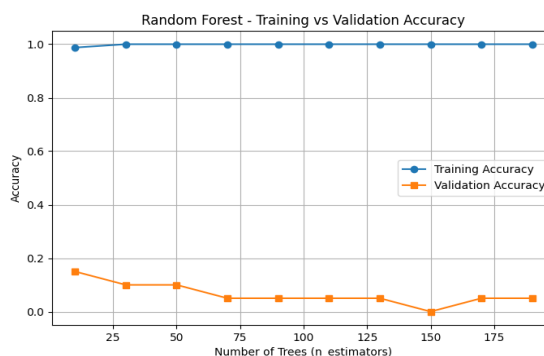
On the x-axis, we have the actual yield values (t/ha) from the dataset, and on the y-axis, the predicted yield values (t/ha) generated by the model.

The blue scatter points represent the actual vs. predicted yield pairs.

The red dashed line represents the perfect fit line, where predictions would exactly match actual values.

From the graph, we can see that most of the predicted values are clustered near the lower yield range, while actual yields vary more widely. The reported R^2 scores are very low (Train = 0.07, Test = 0.02), which means the model explains only a small fraction of the variation in yield. This indicates that Linear Regression is not capturing the complex relationships between soil, nutrients, and environmental factors.

In short, the model is underfitting the dataset and is unable to generalize well for yield prediction. This suggests the need for stronger models such as Random Forest, Gradient Boosting, or Neural Networks, along with better feature engineering (handling imbalance, scaling, adding more soil/climate parameters) to improve accuracy.



Graph 2: Random forest accuracy

This graph shows the training vs validation accuracy of the Random Forest model used in your Smart Plantation Strategy System.

On the x-axis, the number of decision trees (n_estimators) increases, while the y-axis shows accuracy.

The blue line represents training accuracy, which stays very close to 1.0 (100%), indicating the model fits the training data almost perfectly.

V. CONCLUSION AND FUTURE WORK

The proposed Smart Plantation Strategy System successfully integrates machine learning algorithms with an interactive GUI to support precision farming and sustainable agriculture. By combining data preprocessing, crop recommendation using Euclidean similarity, and yield prediction through Linear Regression, the system provides farmers and agricultural planners with reliable, data-driven insights. The crop recommendation module ensures that the most suitable crops are suggested based on soil type, pH, rainfall, and temperature, thereby aligning plantation decisions with local environmental conditions. Meanwhile, the yield prediction model enables users to estimate potential productivity under different field inputs, helping optimize resource allocation and field management. The CustomTkinter-based GUI enhances accessibility by offering a simple, user-friendly platform for dataset visualization, recommendation, and strategy formulation. This integration bridges the gap between complex machine learning models and practical agricultural decision-making, making the technology usable even for non-technical users. In conclusion, the system demonstrates that smart farming solutions can significantly improve crop planning, yield estimation, and sustainability by leveraging modern computational techniques. With further enhancements—such as incorporating real-time IoT sensor data, expanding datasets across multiple regions, and integrating advanced machine learning models—this framework has the potential to evolve into a comprehensive

REFERENCES

- [1] M. A. Javed and M. A. Azmi Murad, "Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability," *Heliyon*, vol. 10, e40836, 2024.
- [2] S. M. Shawon, "Crop yield prediction using machine learning: An extensive and systematic literature review," 2024.
- [3] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture*, vol. 14, no. 8, 1256, 2024.
- [4] M. Bouni, B. Hssina, K. Douzi, and S. Douzi, "Integrated IoT approaches for crop recommendation and yield-prediction using machine learning," *IoT*, vol. 5, no. 4, 2024.
- [5] Y. Wang, "Progress in Research on Deep Learning-Based Crop Yield Prediction," *Agronomy*, 2024.
- [6] K. Jhajharia et al., "Crop Yield Prediction Using Machine Learning and Deep Learning Techniques," 2023.
- [7] S. Kiruthika et al., "IoT-based professional crop recommendation system," 2023.
- [8] G. Saha et al., "Smart IoT-driven precision agriculture: land mapping, crop selection and irrigation," 2025.
- [9] H. Afzal et al., "Incorporating soil information with machine learning for crop recommendation," 2025.
- [10] S. Shastri et al., "Advancing crop recommendation system with supervised ensemble models," 2025.
- [11] "Machine learning based recommendation of agricultural crops based on soil and climatic data," 2024.
- [12] F. S. Prity et al., "A Machine Learning Approach to Crop Recommendations," 2024.
- [13] M. D. S. Sharafat et al., "An IoT-enabled AI system for real-time crop prediction," 2025.
- [14] F. T. Tonmoy et al., "Smart Soil Monitoring System With Crop and Fertilizer Recommendation Features," 2025.
- [15] "Crop yield prediction using Random Forest and XGBoost models," 2025.
- [16] U. Ghodeswar and M. Keote, "Analysis of a Crop Recommendation System for Farmers Based on Machine Learning," *SSRG Int. J. Electr. Electron. Eng.* vol. 11, no. 11, pp. 275–283, 2024.