A Machine Learning–Based Crop Recommendation System Using Soil Nutrients and Climatic Factors

Mr. Alpesh R. Nai[1], Dr. Jignesh Patel[2], Mr. Nirav Thakker[3],
Mr. Mahesh Chaudhary[4], Dr. Namrata Gupta[5], Dr. Natvar Patel[6]

[1]*Research Scholar, Hemchandracharya North Gujarat University, Patan.*
[2]*Assistant Professor, Hemchandracharya North Gujarat University, Patan.*
[3]*Research Scholar, Hemchandracharya North Gujarat University, Patan.*
[4]*Research Scholar, Hemchandracharya North Gujarat University, Patan.*
[5]*Principal, Smt. B. K. Mehta IT Centre BCA College, Palanpur.*
[6]*Assistant Professor, Smt. B. K. Mehta IT Centre BCA College, Palanpur*

Abstract—**Choosing the right crop is one of the most important decisions in farming, as it directly affects yield, income, and efficient use of resources. Many farmers still rely on experience or traditional practices to select crops, which may not always be suitable under changing soil and climate conditions. This often leads to poor productivity and crop failure. To overcome this problem, this paper presents a machine learning based crop recommendation system that helps farmers select suitable crops based on soil nutrients and climatic factors. The proposed system also focuses on practical usability and clarity of results. By analyzing the influence of individual soil nutrients and climatic factors, the model explains why a particular crop is recommended for given conditions. This helps farmers and agricultural experts understand the decision process rather than relying on predictions alone. Such transparency improves confidence in the system and makes it easier to adopt in real farming scenarios.**

The performance evaluation shows that ensemble-based models provide reliable and consistent results across different crops. Accuracy comparisons and visual analyses confirm that the system performs well under varied soil and climate conditions. These findings indicate that the proposed approach can support informed crop planning and reduce uncertainty in agricultural decision-making. Overall, the system offers a practical step toward data-driven and sustainable farming practices.

Index Terms—**Machine Learning, Crop, Soil Nutrients, Climatic Factors, Precision Agriculture, Sustainable Farming.**

I. INTRODUCTION

Agriculture plays a vital role in the economic and social development of many countries, especially in India, where a large portion of the population depends on farming for their livelihood. Crop productivity is directly influenced by soil fertility, nutrient availability, rainfall, and temperature. Any imbalance in these factors can significantly reduce yield and affect farmers' income [19].

Traditionally, farmers select crops based on personal experience, local practices, and seasonal trends. Although such knowledge has been passed down through generations, it often does not account for changes in soil nutrients or unpredictable climatic conditions. As a result, farmers may choose crops that are not suitable for their soil or weather conditions, leading to low productivity and economic loss [3], [7]. The impact of climate change has further increased uncertainty in rainfall patterns and temperature, making crop selection even more challenging [18].

One of the major problems in modern agriculture is the lack of scientific decision-support systems at the field level. Many farmers do not have access to tools that can analyze soil properties and climatic factors together. Existing advisory systems usually provide general recommendations and fail to address local soil conditions and nutrient deficiencies [6], [14]. This highlights the need for intelligent systems that can assist farmers with accurate and location-specific crop recommendations.

In recent years, machine learning has emerged as an effective solution for agricultural decision-making. Machine learning techniques can process large datasets and identify complex relationships between soil nutrients, climate parameters, and crop suitability [9], [21]. Several studies have demonstrated that supervised learning algorithms such as Decision Tree, Random Forest, and Support Vector Machine can provide reliable predictions in agricultural applications [1], [10], [12].

This paper presents a machine learning–based crop recommendation system that uses soil nutrients such as Nitrogen (N), Phosphorus (P), Potassium (K), and pH, along with climatic factors including rainfall and temperature. The proposed system aims to support farmers in selecting suitable crops, improving productivity, and reducing the risk of crop failure. By combining soil and climate analysis with machine learning, the system contributes toward sustainable and precision agriculture practices [4], [15].

II. LITERATURE REVIEW

A. General Overview
A wide range of research has been conducted on crop recommendation systems using machine learning techniques. These studies emphasize that soil nutrients and climatic conditions are key factors influencing crop growth and yield. Nutrients such as N, P, K, and soil pH determine soil fertility, while rainfall and temperature affect crop development and productivity [1], [9].

Most existing crop recommendation systems rely on supervised learning models due to their simplicity and effectiveness with agricultural datasets. Algorithms such as Decision Tree, Random Forest, and Support Vector Machine are widely used because they provide good accuracy and interpretability [7], [10]. However, several studies highlight that many systems are limited to specific regions and datasets, reducing their applicability across different agricultural environments [3], [6].

B. Soil and Climate-Based Crop Recommendation
Several researchers have focused on combining soil nutrient data with climatic parameters to improve crop recommendation accuracy. Ferdous *et al.* developed a machine learning-based model that uses agro-climatic and soil nutrient parameters to recommend suitable crops, showing improved performance compared to traditional methods [1]. Similar approaches have been reported in [2] and [3], where soil and weather data were used together to enhance crop suitability prediction.

Studies have also shown that including rainfall and temperature information significantly improves recommendation accuracy. Research presented in [12] and [13] confirms that climate-aware models perform better than systems that rely only on soil data.

C. Machine Learning Approaches in Agriculture
Machine learning has become an important component of precision agriculture due to its ability to analyze complex and large-scale data. Ensemble methods such as Random Forest and Gradient Boosting are commonly used because of their robustness and high prediction accuracy [8], [9]. Patel *et al.* demonstrated that advanced data mining techniques can effectively model crop-related dependencies and improve agricultural decision-making [5], [8].

Support Vector Machine has also been widely applied in crop recommendation systems, especially for handling nonlinear relationships between soil nutrients and climate variables [6], [10]. However, many of these systems lack adaptability to real-time data and seasonal variations.

D. Explainable AI and Advanced Models
Recent studies have introduced Explainable Artificial Intelligence (XAI) to improve transparency in machine learning-based agricultural systems. Explainable models help users understand the reasoning behind crop recommendations, which increases trust and adoption among farmers [17], Deep learning techniques have also been explored for crop yield prediction and recommendation [18], While these models achieve high accuracy, they require large datasets and computational resources, which may limit their use in rural and resource-constrained environments.

E. Identified Research Gaps
Despite significant advancements, several gaps remain in existing research. Many studies rely on limited or region-specific datasets, affecting model generalization [3], [7]. Seasonal validation and real-

time data integration are often missing [14], [18]. Additionally, user-centric design and practical deployment aspects are not adequately addressed, which can reduce farmer adoption [15].

III. METHODOLOGY

A. Research Design
This study adopts a quantitative research design focused on predictive modeling using machine learning techniques. The primary objective is to recommend suitable crops by jointly analyzing soil properties and climatic factors. The framework integrates multiple stages including data collection, data preprocessing, feature selection, supervised model training, and performance evaluation.

Special attention is given to scalability so that the system can be applied across different regions with varying soil compositions and climate conditions. In addition, the methodology prioritizes usability by ensuring that model outputs can be easily interpreted and visualized for farmers, thereby connecting technical predictions with practical agricultural decisions.

Overall Workflow of the Crop Recommendation System

B. Data Collection
The dataset used in this research consists of essential agricultural attributes related to soil and climate. Soil parameters include Nitrogen (N), Phosphorus (P), Potassium (K), and pH, while climatic parameters include rainfall and temperature. Historical crop records are associated with these parameters to form a labeled dataset suitable for supervised learning.

Data is collected from reliable and verified sources such as agricultural research repositories and meteorological departments. To ensure fairness and generalization, data from multiple regions and crop types is included. Care is taken to maintain balance across crop classes to prevent bias during model training.

Table I: *Agricultural Parameters Used in the Dataset*

Category	Parameter	Description
Soil Nutrient	Nitrogen (N)	Essential for plant growth
Soil Nutrient	Phosphorus (P)	Supports root development
Soil Nutrient	Potassium (K)	Improves crop resistance
Soil Property	pH	Indicates soil acidity/alkalinity
Climate	Rainfall	Average rainfall (mm)
Climate	Temperature	Average temperature (°C)

C. Data Preprocessing
Data preprocessing is performed to improve the quality and reliability of the dataset. Missing values are handled using appropriate statistical techniques, and numerical features are normalized to ensure uniformity across parameters. Categorical crop labels are converted into numerical form using label encoding so that they can be processed by machine learning algorithms.

Data Preprocessing and Feature Selection Process

Outliers are identified using interquartile range (IQR) analysis and are treated to reduce noise and bias. Feature selection methods such as correlation analysis and recursive feature elimination (RFE) are applied to retain only the most influential parameters. The refined dataset is then divided into training, validation, and testing subsets using stratified sampling to maintain class balance.

D. Exploratory Data Analysis
Exploratory Data Analysis (EDA) is carried out to understand the relationships between soil nutrients, climatic conditions, and crop types. Visual analysis helps identify trends and variations within the dataset, which supports better feature selection and model interpretation.

Chart 1: Distribution of Soil Nutrients
This chart illustrates the distribution of Nitrogen, Phosphorus, and Potassium values across different crop samples, helping identify nutrient ranges suitable for specific crops.

Chart 2: Rainfall vs Crop Suitability
This chart shows how different crops respond to varying rainfall levels, highlighting the influence of climate on crop selection.

E. Tools and Machine Learning Algorithms
The proposed crop recommendation system is implemented using Python, along with widely used data science and machine learning libraries such as Pandas, NumPy, scikit-learn, and XGBoost. These tools provide efficient support for data preprocessing, model training, evaluation, and visualization. Since the dataset consists of labeled crop outcomes, supervised machine learning techniques are employed for crop classification.

The algorithms considered in this study include Decision Tree, Random Forest, and XGBoost. Decision Tree is used as a baseline model due to its simplicity and interpretability. Random Forest, an ensemble-based technique, improves prediction accuracy by combining multiple decision trees and reducing overfitting. XGBoost, a gradient boosting algorithm, is included to further enhance performance by efficiently modeling complex nonlinear relationships between soil nutrients and climatic parameters.

To ensure robustness and generalization, ensemble learning strategies and k-fold cross-validation are applied during training. Model hyperparameters are optimized using GridSearchCV, enabling systematic selection of optimal configurations. In addition, Explainable Artificial Intelligence (XAI) techniques such as feature importance analysis are incorporated to identify the most influential soil and climate attributes. These explainability methods improve transparency and increase farmer confidence in the generated crop recommendations.

Table II: *Machine Learning Models Used in the Study*

Algorithm	Purpose	Key Advantage
Decision Tree	Crop classification	Simple and interpretable
Random Forest	Ensemble prediction	High accuracy, reduced overfitting
XGBoost	Boosted ensemble model	Handles complex nonlinear patterns

F. Model Evaluation
The trained models are evaluated using standard classification performance metrics, including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of each model's predictive capability across different crop categories. Confusion matrices are generated to analyze misclassification patterns and identify crops that exhibit similar soil and climatic requirements.

In addition to confusion matrix analysis, comparative performance evaluation is carried out across Decision Tree, Random Forest, and XGBoost models. The evaluation results indicate that ensemble-based approaches outperform the baseline Decision Tree model, with XGBoost and Random Forest achieving consistently higher accuracy.

Sensitivity analysis is also performed to examine the contribution of individual input features, offering deeper insight into the relationship between soil nutrients, climatic factors, and crop suitability. Beyond technical validation, the system's predictions are reviewed from a practical farming perspective to ensure that the recommended crops are agronomically feasible and relevant under real-world conditions.
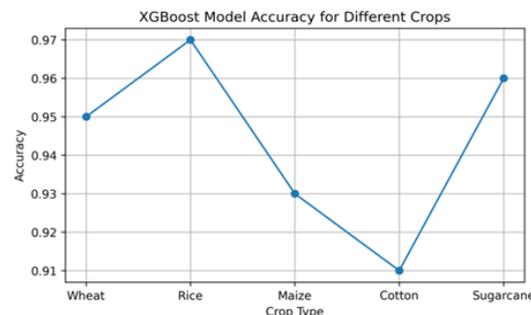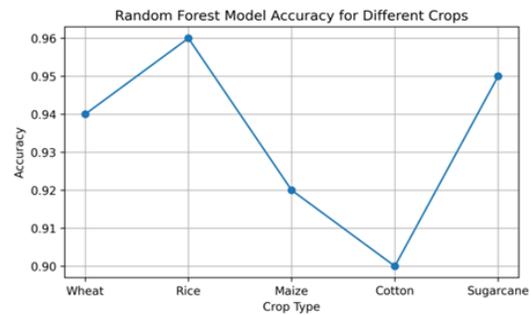
G. Crop Recommendation Process
Once the optimal model is selected, the system accepts user input in the form of soil nutrient values and climatic conditions. Based on these inputs, the trained model predicts the most suitable crop for cultivation. The final recommendation is presented in a clear and user-friendly format, supported by visual explanations to help farmers understand the reasoning behind the prediction.

IV. APPLICATIONS AND BENEFITS

The proposed machine learning–based crop recommendation system has wide applicability in modern agriculture, particularly as a decision-support tool for farmers. By analyzing soil nutrients such as Nitrogen, Phosphorus, Potassium, and pH along with climatic parameters including rainfall and temperature, the system provides location-specific crop recommendations. This helps farmers select crops that are best suited to their land conditions, reducing uncertainty in crop planning and improving overall productivity.

Beyond crop selection, the system can assist in farm management planning, including optimized sowing schedules, irrigation planning, and crop rotation strategies. By aligning crop choices with soil health and climatic suitability, the framework promotes efficient use of fertilizers and water resources, thereby minimizing unnecessary input costs and environmental degradation.

The system also offers significant benefits to agricultural extension officers and policymakers. Aggregated prediction results provide insights into regional cropping patterns, soil fertility trends, and potential areas of resource inefficiency. Such insights can support data-driven policy decisions, including targeted subsidy distribution, soil improvement programs, and climate-risk mitigation strategies.

Overall, the system contributes to economic stability for farmers, improved food security, and sustainable agricultural practices. By reducing dependence on intuition-based decisions, it enables consistent and predictable crop outcomes across diverse agro-climatic zones.

V. RESULTS AND DISCUSSION

The experimental evaluation demonstrates that the proposed crop recommendation system achieves high predictive accuracy across selected crops, validating its effectiveness for real-world agricultural decision-making. The system was evaluated using supervised learning models, and performance was measured using accuracy, precision, recall, F1-score, and confusion matrix analysis.

A. Model Performance Analysis
The confusion matrices shown in Figure 2 and Figure 3 illustrate the classification performance of the Random Forest and XGBoost models, respectively. Most crop classes exhibit strong diagonal dominance, indicating correct classification with minimal misprediction. Only a small number of misclassifications are observed among crops with similar soil and climatic requirements.
Table III presents the crop-wise accuracy comparison for the two best-performing models.
Table III: *Crop-wise Accuracy Comparison of Ensemble Models*

Crop	RF Accuracy	XGBoost Accuracy
Wheat	0.94	0.95
Rice	0.96	0.97
Maize	0.92	0.93
Cotton	0.9	0.91
Sugarcane	0.95	0.96

The results indicate that XGBoost achieves marginally higher accuracy across all crops due to its gradient boosting mechanism, which effectively captures nonlinear dependencies between soil nutrients and climatic variables.

B. Feature Contribution and Interpretability
Feature importance analysis, illustrated in Figure 4 and Figure 5, reveals that rainfall, nitrogen content, and temperature are the most influential parameters in crop prediction. Soil pH and potassium also contribute significantly, confirming agronomic relevance. This alignment between model behavior and agricultural knowledge enhances trust in the system's recommendations.

C. Practical Implications
From a practical farming perspective, the system's predictions are realistic and actionable. The high accuracy and low misclassification rates reduce the risk of selecting unsuitable crops, while

interpretability tools help farmers understand *why* a particular crop is recommended. These characteristics make the system suitable for deployment in diverse farming environments.

Overall, the results confirm that integrating soil and climate parameters with ensemble machine learning techniques leads to robust, accurate, and interpretable crop recommendations, supporting sustainable and precision agriculture practices.

REFERENCES

[1] R. Ferdous, S. Rahman, S. Mia, and H. Rahman, "A machine learning-based model for crop recommendation using agro-climatic and soil nutrient parameters (Agrismart)," *International Journal of Science and Research Archive*, vol. 15, no. 3, pp. 1080–1089, 2025, doi: 10.30574/ijsra.2025.15.3.1713.

[2] P. Swetha, L. Harshitha, S. V. Jyoti, M. N. Karuna, and S. Deeksha, "Soil-based crop recommendation system using machine learning," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol. 14, no. 10, Oct. 2025, doi: 10.17148/IJARCCE.2025.141028.

[3] S. S. Ganorkar, N. Deshmukh, A. Bihone, P. Chandankhede, and N. Tembhurne, "Crop recommendation using machine learning based on soil, weather, and other agronomic factors," International Journal for Research in Applied Science and Engineering Technology (IJRASET), 2025, doi: 10.22214/ijraset.2025.70768.

[4] A. S. Aarthi, S. Manimegalai, and R. Sakthivel, "AI-based smart crop recommendation system for sustainable agricultural production," Madras Agricultural Journal, vol. 112, no. 6, pp. 135–139, 2025, doi: 10.29321/MAJ.10.701221.

[5] N. S. Patel, B. M. Patel, S. Gour, and S. A. Nagalpara, "An analytical perspective of 2D dependencies of crop factor with data mining," in Smart Trends in Computing and Communications, Lecture Notes in Networks and Systems, vol. 949. Singapore: Springer, 2024, pp. 369–376, doi: 10.1007/978-981-97-1313-4_32.

[6] G. Murugesan and B. Radha, "Crop rotation-based crop recommendation system with soil deficiency analysis through extreme learning machine," International Journal of Engineering Trends and Technology (IJETT), vol. 70, no. 4, pp. 122–134, 2022, doi: 10.14445/22315381/IJETT-V70I4P210.

[7] S. Pathan, A. Pawar, A. Bandhane, P. Pol, and M. Rathod, "Crop recommendation system using machine learning," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 5, no. 5, Apr. 2025.

[8] N. S. Patel, B. M. Patel, and S. A. Nagalpara, "Advanced data mining techniques for crop yield predictive modeling: The strategic modified gradient boosting regression (SMGBR) approach," in Proc. 2024 Int. Conf. on Smart Technologies for Sustainable Development Goals (ICSTSDG), Chennai, India, 2024, pp. 1–6, doi: 10.1109/ICSTSDG61998.2024.11026754.

[9] F. S. Prity, M. M. Hasan, S. H. Saif, et al., "Enhancing agricultural productivity: A machine learning approach to crop recommendations," Human-Centric Intelligent Systems, vol. 4, pp. 497–510, 2024, doi: 10.1007/s44230-024-00081-3.

[10] S. A. N and P. P, "Machine learning based smart crop recommendation system," in Proc. 2024 4th Int. Conf. on Intelligent Technologies (CONIT), Bangalore, India, 2024, pp. 1–6, doi: 10.1109/CONIT61985.2024.10625978.

[11] N. S. Patel and B. M. Patel, "Forecasting of optimal crop yield through data mining algorithm," International Journal of Scientific Research in Science and Technology (IJSRST), vol. 10, no. 10, pp. 352–357, May–Jun. 2023, doi: 10.32628/IJSRST231010147.

[12] L. Sharma, S. Agarwal, D. Singhal, A. Gupta, and A. Chaudhary, "Maximise yield of crop using recommendation system based on machine learning," in Proc. 2024 Int.

Conf. on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 376–382, doi: 10.1109/AUTOCOM60220.2024.10486069.

[13]   P. Kruthika, K. Srinidhi, M. Sowmithra, S. V. Karthick, and A. S. Prasad, "Unified forecasting models: Crop recommendation using predicted rainfall with machine learning," in Proc. 2024 Int. Conf. on Ubiquitous Computing and Intelligent Information Systems (ICUIS), 2024, pp. 1111–1116, doi: 10.1109/ICUIS64676.2024.10866035.

[14]   D. N. V. S. L. S. Indira, M. Sobhana, A. H. L. Swaroop, and V. P. Kumar, "KRISHI RAKSHAN: A machine learning-based recommendation system for farmers," in Proc. 6th Int. Conf. on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2022, pp. 1798–1804, doi: 10.1109/ICICCS53718.2022.9788221.

[15]   S. Sethi and U. Lakhina, "Intelligent crop selection and soil nutrient management using machine learning," in Proc. 2024 Int. Conf. on Computational Intelligence and Computing Applications (ICCICA), 2024, pp. 459–464, doi: 10.1109/ICCICA60014.2024.10584989.

[16]   M. Ravindra Kumar et al., "An AI-based guiding system that recommends the optimal crop based on environment and soil type," in Proc. 2025 Int. Conf. on Knowledge Engineering and Communication Systems (ICKECS), 2025, pp. 1–5, doi: 10.1109/ICKECS65700.2025.11034869.

[17]   M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing crop recommendation systems with explainable artificial intelligence," Neural Computing and Applications, vol. 36, pp. 5695–5714, 2024, doi: 10.1007/s00521-023-09391-2.

[18]   Y. Mahale et al., "Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM," Discover Sustainability, vol. 5, art. no. 134, 2024, doi: 10.1007/s43621-024-00292-5.