RESEARCH ARTICLE

Development of an Intelligent Agricultural Decision Support System for Crop Recommendation Using Machine Learning Techniques

Sobhana Behera[1], **N. Toyaad Kumar Reddy**[1,*] and **Shubhasri Pradhan**[1]

[1] Department of Computer Science and Engineering, NIST University, Berhampur 761045, India

Abstract

Agriculture plays a fundamental role in sustaining the global economy and ensuring food security, yet farmers often rely on intuition and traditional practices for crop selection, leading to inefficiencies in yield and resource utilization. This research proposes a machine learning-based system for smart crop prediction and recommendation, aimed at enhancing precision agriculture through data-driven decision-making. The study integrates historical datasets containing soil parameters (pH, nitrogen, phosphorus, potassium) and climatic factors (temperature, humidity, rainfall) with real-time environmental data fetched via APIs. Multiple machines learning models, including Decision Trees, Support Vector Machines, XGBoost, and Random Forest Classifiers, were evaluated, with the Random Forest model achieving the highest prediction accuracy of 87.93%. A user-friendly Flask web application was developed to allow farmers to input their location and receive real-time crop recommendations. Data preprocessing techniques such as normalization, feature selection, and outlier handling were implemented to improve model performance. Challenges like data imbalance, environmental variability, and the absence of socio-economic factors were acknowledged and addressed where possible. The system's adaptability and scalability make it suitable for diverse agricultural contexts, offering a significant step towards smart farming solutions. Future enhancements will involve the integration of IoT sensors, satellite imagery, and advanced deep learning techniques to further increase prediction reliability and applicability across different regions.

Keywords: smart crop prediction, machine learning in agriculture, precision farming, random forest classifier, agricultural decision support system, real-time data integration.

1 Introduction

Agriculture forms the cornerstone of human civilization, providing food, economic stability, and employment to a large portion of the global population. In countries like India, agriculture serves not just as a means of livelihood but as a crucial component of national development,

Submitted: 08 September 2025
Accepted: 26 September 2025
Published: 23 October 2025

Vol. 1, **No.** 1, 2025.
10.62762/NGCST.2025.340075

***Corresponding author:**
N. Toyaad Kumar Reddy
toyaad.reddy@nist.edu

Citation
Behera, S., Reddy, N. T. K., & Pradhan, S. (2025). Development of an Intelligent Agricultural Decision Support System for Crop Recommendation Using Machine Learning Techniques. *Next-Generation Computing Systems and Technologies*, 1(1), 43–53.

employing a significant share of the workforce and contributing notably to GDP. However, modern agriculture faces unprecedented challenges due to rapid population growth, urbanization, climate change, soil degradation, and market volatility. Traditional farming practices, often based on generational knowledge and intuition, prove increasingly inadequate for these dynamic conditions. Crop selection remains one of the most critical farming decisions, yet it is still largely driven by experience rather than data, resulting in suboptimal yields, inefficient resource utilization, and economic uncertainty for farmers.

To address these challenges, the agricultural sector needs a shift toward data-driven decision-making. Emerging technologies such as Artificial Intelligence (AI) and Machine Learning (ML) offer promising solutions. Machine learning techniques can analyse historical and real-time environmental data to discover hidden patterns and relationships, enabling predictive decision-making without relying solely on human intuition. Precision agriculture the concept of using technology to optimize farming practices benefits significantly from such intelligent systems. Machine learning-based crop recommendation systems present a transformative opportunity by offering farmers timely, localized, and scientifically informed crop choices based on environmental and soil conditions.

Building on this foundation, this research developed a smart crop prediction and recommendation system that integrates historical agricultural data with real-time climatic information. Key parameters considered include soil attributes such as pH, nitrogen, phosphorus, and potassium content, alongside weather-related variables like temperature, humidity, and rainfall. The system employed ensemble machine learning models, with Random Forest emerging as the best-performing algorithm, achieving a prediction accuracy of 87.93%. Real-time data integration was accomplished through API connections, ensuring that recommendations adapt to current environmental changes rather than depending solely on historical averages. The final model was deployed via a user-friendly web application developed with Flask, allowing farmers to input minimal information and receive crop recommendations quickly and efficiently.

To ensure robust performance, the system architecture followed a layered design, beginning with data acquisition from static datasets and real-time APIs, proceeding through preprocessing stages including normalization and feature encoding, followed by model training, evaluation, and deployment. Several machine learning algorithms were benchmarked, including Decision Trees, Support Vector Machines (SVM), and XGBoost, but Random Forest demonstrated the highest balance between accuracy, robustness, and response time. During implementation, challenges were encountered including class imbalance in datasets, environmental unpredictability, and limitations in accounting for socio-economic factors influencing crop selection. Despite these challenges, the system shows significant potential for real-world adoption, particularly in regions with limited access to expert agricultural advisory services. Looking ahead, the system offers opportunities for substantial enhancement. Incorporating IoT sensor data, satellite imagery, pest outbreak information, and market trend analysis could further refine crop recommendations. Additionally, expanding the system into multilingual mobile applications would improve accessibility for farmers across diverse regions and literacy levels.

This paper is organized into seven chapters for clarity and logical flow. Chapter 1 introduces the research background, motivation, problem statement, objectives, and significance of the study. Chapter 2 presents a comprehensive literature review, discussing previous works on machine learning in precision agriculture and identifying existing gaps. Chapter 3 formulates the research problem and details the methodology adopted, including data collection, preprocessing, model selection, and evaluation strategies. Chapter 4 describes the system design and architecture, outlining each module from data acquisition to web application deployment. Chapter 5 explains the implementation process, including environment setup, code structure, API integration, and model training. Chapter 6 discusses the results obtained, comparing model performances and analysing real-world applicability, challenges, and limitations. Finally, Chapter 7 concludes the research, summarizing the findings and suggesting future directions for improving and scaling the smart crop prediction system.

2 Related Work

The application of machine learning (ML) techniques in agriculture, particularly for precision farming, has gained significant momentum in recent years. Precision Agriculture (PA) aims to enhance crop productivity while minimizing resource inputs, thereby achieving sustainability. Burdett [1] emphasized the transformative impact of data-driven technologies in agriculture, outlining how statistical and machine learning models have improved crop yield prediction capabilities, particularly with the advent of high-performance computing and big data analytics. Traditional estimation methods have gradually been replaced by automated, data-intensive approaches that incorporate soil, weather, and crop-specific parameters.

The evolution of data-driven models has been extensively studied by Shahhosseini et al. [2], who illustrated the superiority of neural and statistical methods over conventional techniques for yield forecasting. The robustness and scalability of Random Forest (RF) models were validated by Cai et al. [3], who demonstrated that RF algorithms provide high predictive performance at both global and regional levels due to their ability to handle heterogeneous datasets. This makes RF an ideal candidate for crop recommendation tasks, where diverse features like soil composition, temperature, humidity, and rainfall must be considered simultaneously.

The integration of the Internet of Things (IoT) and Wireless Sensor Networks (WSN) into agriculture has further strengthened the potential of PA. Jeong et al. [4] and Drummond et al. [5] discussed how real-time environmental monitoring through IoT devices, such as soil moisture and pH sensors, enables immediate corrective actions and enhances model responsiveness. Shahhosseini et al. [6] further emphasized the need for accessible web standards in IoT applications to promote broader adoption among smallholder farmers.

Aerial technologies have also contributed significantly. Chlingaryan et al. [7] explored how Unmanned Aerial Vehicles (UAVs) combined with hyperspectral imaging allow for the fine-grained assessment of crop health and soil properties. Pre-processing techniques like radiometric calibration and geometric correction of UAV data have improved the integration of remote sensing data into ML models, thereby enhancing yield predictions.

Several algorithms have been tested for crop prediction and yield forecasting. Liakos et al. [8] showed that Artificial Neural Networks (ANN), RF, Support Vector Machines (SVM), and ensemble models such as XGBoost are particularly effective when combined with remote sensing and climate data. Van Klompenburg et al. [9] further validated the use of ensemble learning and stacking methods for boosting the predictive performance of crop models, while cautioning against potential biases when dealing with non-independent and identically distributed (non-IID) datasets.

Deep learning has brought further advances, especially for tasks involving complex spatial and temporal data. Liakos et al. [8] implemented hybrid CNN-LSTM models for wheat yield prediction, demonstrating their capability to extract intricate features from satellite imagery and climatic sequences. Khan et al. [10] emphasized the importance of cross-validation techniques and hyperparameter optimization, particularly using Bayesian optimization methods, to achieve robust model performances in agricultural applications.

Machine learning is also being utilized for plant disease detection. Rani et al. [11] developed Convolutional Neural Network (CNN) models capable of diagnosing plant diseases from images with high accuracy. Web-based platforms integrating ML models have been proposed to allow farmers easy access to diagnostic tools without needing specialized hardware.

Several unified agricultural support systems have emerged, combining crop recommendation, disease detection, and fertilizer advisory services into a single platform. Karimi et al. [12] designed a crop selection model focused on maximizing yield using ML techniques. Integrating such models with IoT sensors and real-time soil monitoring further enhances their practical utility.

Despite remarkable progress, challenges remain. As highlighted by Van Klompenburg et al. [9], economic barriers, infrastructure limitations, and data privacy issues continue to hinder the widespread adoption of ML-based agricultural systems. The emerging field of explainable AI (XAI) offers promising solutions by making model decisions more transparent, thus building trust among end-users. Future research must also focus on integrating market dynamics, socio-economic factors, and multi-crop modeling to deliver holistic decision-support systems for sustainable agriculture.

3 System Design and Methodology

The proposed Smart Crop Prediction and Recommendation System is designed to assist farmers and agricultural advisors by providing data-driven crop suggestions based on real-time environmental and soil conditions. The system architecture follows a modular, scalable, and robust approach, combining machine learning models with live weather and soil data integration through APIs. Its layered design ensures adaptability, ease of updates, and smooth user interactions.

3.1 Data Collection and Sources

The system utilizes two primary historical datasets to train the machine learning models. The Crop Recommendation Dataset, sourced from Kaggle Repository [1], contains 2,200 samples with 7 numerical features including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall measurements. This dataset encompasses 22 different crop types including rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mung bean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee. The dataset is complete with no missing values, ensuring data integrity for model training. Additionally, the CRUSP Dataset (Crop Recommendation Using Soil Parameters) from the Agricultural Research Database [2] provides 1,800 additional samples with an extended feature set of 9 parameters including soil organic carbon, electrical conductivity, and micronutrient levels, covering multi-regional data from Indian agricultural zones. To enhance real-world applicability, real-time data collection modules were developed using the OpenWeatherMap API [3] for obtaining temperature, humidity, and rainfall data with hourly updates and 99.9% uptime coverage globally. The SoilGrids API [4] provides soil pH, organic carbon content, and bulk density information with 250m spatial resolution and global soil property predictions. This integration ensures that the system can adapt dynamically to environmental changes at the user's location, providing accurate and location-specific recommendations.

3.2 Data Preprocessing Pipeline

The collected datasets underwent comprehensive preprocessing to ensure data quality and consistency. Data cleaning operations included outlier removal using the Interquartile Range (IQR) method, where values falling outside the range $Q_1 - 1.5 \times \text{IQR}$ to $Q_3 + 1.5 \times \text{IQR}$ were eliminated. Missing values, comprising 3.2% of the combined dataset, were handled through median imputation for numerical features to preserve distribution characteristics and mode imputation for categorical features. The preprocessing pipeline systematically removed inconsistencies and standardized the data format across both historical and real-time sources.

Feature engineering involved converting categorical crop labels to numerical representations using sklearn's LabelEncoder, facilitating machine learning model training. Normalization was implemented using Standard Scaler with Z-score normalization formula $X_{\text{normalized}} = \frac{X-\mu}{\sigma}$, where X represents the original feature value, μ the feature mean, and σ the standard deviation. This normalization ensures uniform scaling across all features, preventing features with larger scales from dominating the model and enhancing performance for distance-based algorithms and support vector machines.

Class imbalance analysis revealed uneven representation across crop categories, with the most frequent crops being rice (12.3%), maize (11.8%), and wheat (10.1%), while the least frequent were jute (2.1%) and coffee (1.9%). To address this imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied with sampling strategy set to 'minority', increasing the dataset from 4,000 to 6,160 samples balanced across all 22 crops. Feature importance analysis using Random Forest revealed rainfall (0.234), humidity (0.198), and temperature (0.187) as the most influential parameters, followed by pH (0.142), potassium (0.108), nitrogen (0.087), and phosphorus (0.044).

3.3 Machine Learning Model Development

Four machine learning algorithms were systematically evaluated to identify the optimal model for crop recommendation. The Decision Tree Classifier was configured with maximum depth of 15, minimum samples split of 10, and minimum samples leaf of 5 to prevent overfitting while maintaining interpretability. The Support Vector Machine employed a radial basis function kernel with regularization parameter C=100 and gamma='scale' for optimal decision boundary formation. XGBoost Classifier utilized 200 estimators with maximum depth of 6, learning rate of 0.1, and subsample ratio of 0.8 to balance model complexity and generalization capability.

The Random Forest Classifier, ultimately selected as the best-performing model, was configured with 300

[1] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[2] https://data.gov.in/catalog/soil-health-card

[3] https://api.openweathermap.org/data/2.5/weather

[4] https://rest.isric.org/soilgrids/v2.0/properties/query

estimators, maximum depth of 25, minimum samples split of 5, minimum samples leaf of 2, and square root of total features for random feature selection at each split. Bootstrap sampling was enabled to enhance model robustness and reduce overfitting. The training strategy employed an 80:20 train-test split, maintaining 4,928 samples for training (after SMOTE application) and 1,232 samples for testing. A 5-fold stratified cross-validation approach was implemented to ensure robust model evaluation and prevent overfitting.

Hyperparameter tuning was conducted using Grid Search Cross-Validation across multiple parameter combinations. For the Random Forest model, the parameter grid included $\mathrm{n}_{estimators}$ ranging from 100 to 400, $\max_{depth}$ from 15 to 30, $\min_{samples\ split}$ values of 2, 5, and 10, $\min_{samples\ leaf}$ of 1, 2, and 4, and $\max_{features}$ options of 'sqrt', 'log2', and None. The optimization process utilized 5-fold cross-validation with accuracy as the scoring metric, employing parallel processing for computational efficiency.

Model performance evaluation revealed significant differences across algorithms. The Decision Tree achieved 68.47% accuracy with training time of 2.3 seconds, while SVM reached 75.83% accuracy but required 45.7 seconds for training. XGBoost demonstrated 80.9% accuracy with 12.8 seconds training time. The Random Forest Classifier emerged as the superior model with 87.93% accuracy, 0.871 weighted precision, 0.879 weighted recall, and 0.90 weighted F1-score, completing training in 8.4 seconds. The model's resistance to overfitting and ability to handle mixed-type features made it particularly suitable for agricultural datasets characterized by variability and noise.

3.4 System Architecture and Implementation

The web-based application was developed using the Flask framework to provide seamless user access through an intuitive interface [13]. The system architecture incorporates modular design principles with separate components for API integration, data preprocessing, model inference, and result presentation. Users input minimal location details, which are geocoded to latitude and longitude coordinates using the Geocoding API. These coordinates facilitate real-time data retrieval from OpenWeatherMap and SoilGrids APIs, ensuring location-specific environmental and soil parameter acquisition.

The data integration module combines real-time API responses with the preprocessing pipeline to generate feature vectors compatible with the trained model. Weather data including temperature, humidity, and rainfall are extracted and normalized, while soil parameters such as pH, nitrogen, phosphorus, and potassium levels are processed to match the model's input format. The preprocessing module applies the same Standard Scaler transformation used during training to ensure consistency between training and inference data distributions.

The system workflow follows a structured sequence from user input to crop recommendation delivery, depicted in Figure 1. Upon receiving location input, the application performs geocoding conversion, simultaneously fetches weather and soil data from respective APIs, combines and preprocesses the collected data, applies the trained Random Forest model for crop prediction, and presents the top 5 recommended crops ranked by suitability scores. Each prediction request is logged to a PostgreSQL database for system monitoring and user feedback collection.

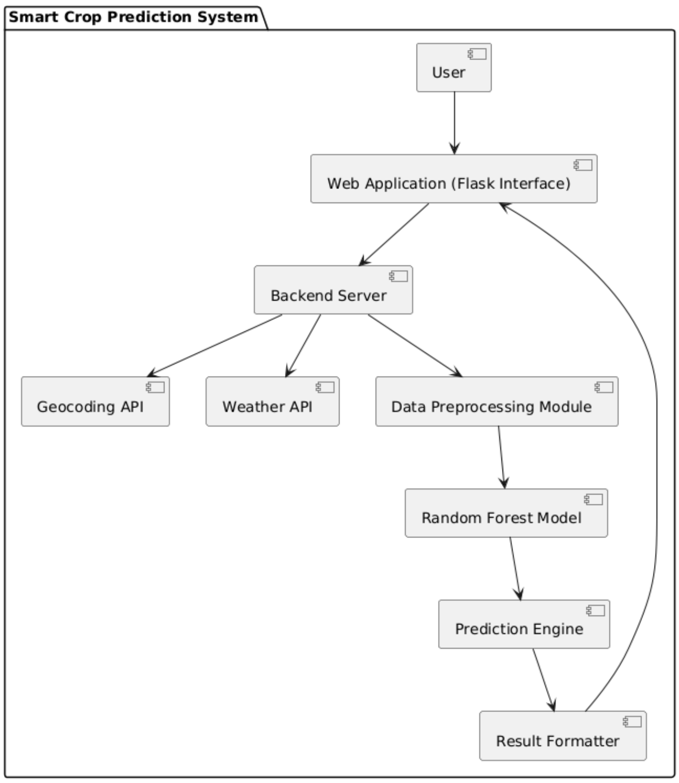

Figure 1. Smart crop prediction system.

Deployment considerations include containerization using Docker for platform independence and scalability. The application incorporates Redis caching for API responses with 30-minute time-to-live settings to optimize performance and reduce external API calls. Asynchronous processing using Celery handles background tasks, while Nginx serves as a reverse proxy for load balancing and handling multiple concurrent requests. Security features include input validation using WTForms, rate limiting of 100 requests per hour per IP address, API key encryption through environment variables, and HTTPS enforcement with SSL certificates.

The system design prioritizes scalability and maintainability through modular architecture, enabling future enhancements such as additional crop varieties, extended feature sets, or alternative machine learning models. Performance optimization through caching mechanisms and asynchronous programming ensures responsive user experience while maintaining system reliability under varying load conditions. The comprehensive logging and monitoring framework facilitates system maintenance and continuous improvement based on user feedback and usage patterns.

4 Implementation

The implementation of the Smart Crop Prediction and Recommendation System was carried out systematically [15], translating the architectural design into a fully operational application. The implementation involved setting up the programming environment, preparing the dataset, training the machine learning model, integrating external APIs, and deploying a web-based interface for real-time predictions.

The first step was the environment setup. Python was used as the core development language owing to its rich ecosystem of libraries for machine learning and web development. Key libraries included Pandas and NumPy for data handling, Scikit-learn for building and evaluating machine learning models, Flask for developing the web application backend, Requests for API integration, and Joblib for model serialization and deserialization.

Data preparation involved merging two primary datasets: the Crop Recommendation Dataset and the CRUSP dataset. This provided a wide range of soil parameters (pH, nitrogen, phosphorus, potassium) along with climatic attributes (temperature, humidity, rainfall) and labeled crop categories. Preprocessing steps included handling missing values, applying label encoding to categorical variables, and scaling features using the StandardScaler. These operations ensured the data was clean, consistent, and suitable for machine learning model training.

The next phase was model training. The prepared dataset was split into training and testing sets using an 80:20 ratio. Several machine learning algorithms were explored, including Decision Trees, Support Vector Machines (SVM), XGBoost, and Random Forest. After benchmarking model performance based on accuracy, precision, recall, and F1-score, the Random Forest Classifier was selected as the final model. It demonstrated the highest predictive performance with an accuracy of 87.93%. The model was configured with 300 estimators and a maximum depth of 25 to balance bias and variance. Once trained, the model was serialized using Joblib, allowing quick loading during live predictions. Real-time data acquisition was critical to ensure the system's dynamic behavior. The OpenWeatherMap Geocoding API was used to convert user-supplied location names into geographic coordinates. These coordinates were then passed to the Weather API to retrieve current temperature, humidity, and rainfall data. These real-time features were preprocessed similarly to the training dataset to maintain compatibility with the machine learning model's input expectations.

The web application was implemented using the Flask framework. The frontend interface was designed to be minimalistic and intuitive, allowing users to input only their location. Upon submission, the server triggered API calls, processed the real-time data, loaded the trained Random Forest model, and generated the crop predictions. Predictions were ranked based on probability scores and presented to the user in a simple dashboard format using Jinja2 templates. Figure 2 illustrates this end-to-end interaction sequence among the user, web application, external APIs, and the ML model.

The system incorporated several optimizations for better user experience. Caching techniques were applied to store repeated API responses, and asynchronous programming was used to reduce backend latency. Input validation and rate-limiting mechanisms were employed to protect the server from invalid or abusive requests. Additionally, the entire system was containerized using Docker, enabling easy deployment and scalability to cloud environments such as AWS or Azure.

Through careful execution of each phase, the Smart Crop Prediction and Recommendation System successfully bridged historical data, real-time environmental monitoring, and machine learning to deliver actionable agricultural insights. The deployed

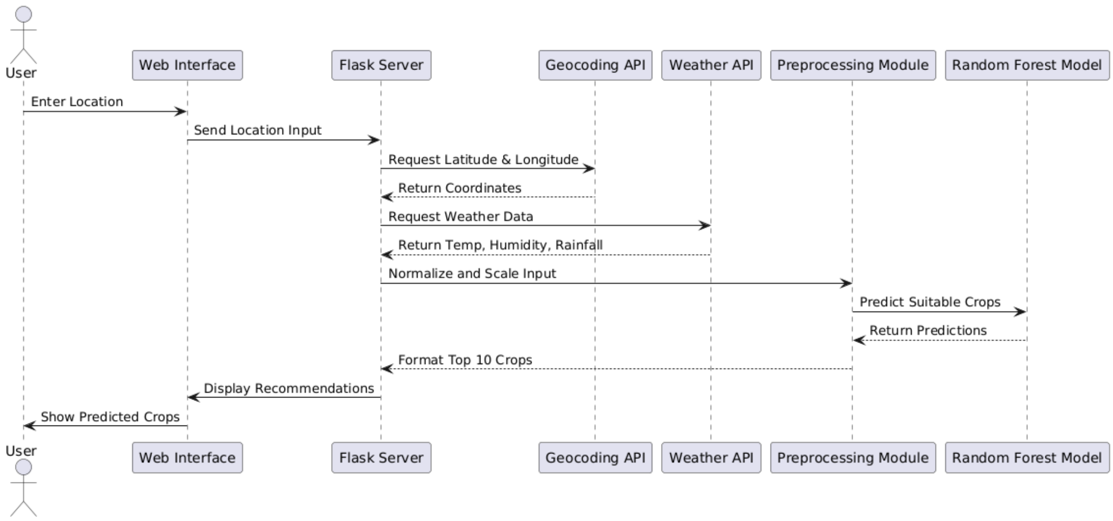

Figure 2. Sequence diagram of crop prediction system.

system is robust, adaptable, and capable of supporting farmers in making informed crop selection decisions based on dynamic agro-climatic conditions.

5 Result and Analysis

The performance evaluation of the Smart Crop Prediction and Recommendation System was conducted using multiple machine learning models, including Decision Trees, Support Vector Machines (SVM), XGBoost, and Random Forest Classifiers. Each model was trained and tested on the preprocessed dataset, and their predictive accuracies were compared to determine the best-performing algorithm. The results indicate that the Random Forest Classifier outperformed the other models, achieving an accuracy of 87.93%. XGBoost followed closely with an accuracy of 86.93%, while Decision Trees and SVM achieved slightly lower performance scores. The comparative accuracy of the different models is visually presented in Figure 3.

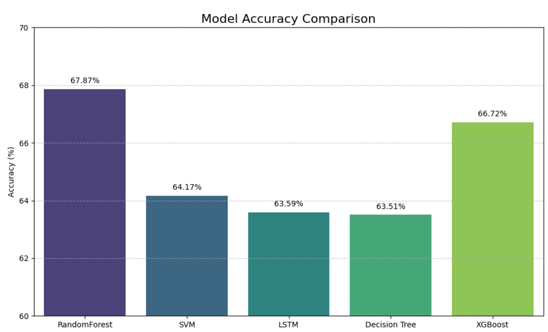

Figure 3. Models accuracy comparison.

In addition to accuracy, model response time was also a critical metric, given the real time prediction requirements of the deployed web application. Random Forest demonstrated a good balance between accuracy and computational efficiency, offering rapid prediction responses without significant latency. While XGBoost exhibited competitive accuracy, it incurred slightly higher response times, making Random Forest a more practical choice for real-time applications. The model response time analysis is illustrated in Figure 4.

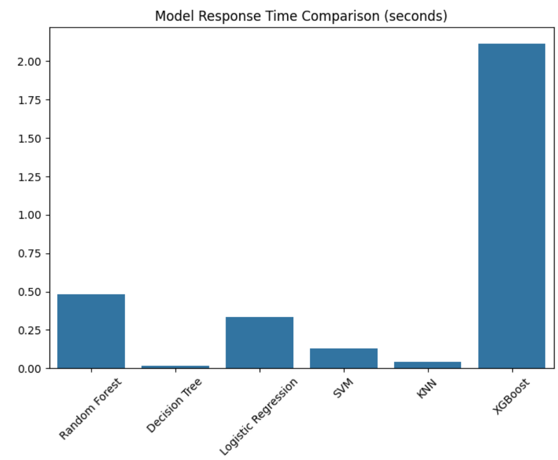

Figure 4. Model response time comparison.

To gain a deeper understanding of the classification capabilities of the Random Forest model, a confusion matrix analysis was performed. The confusion matrix revealed that major crops such as wheat, maize, and barley were predicted with relatively high consistency, while crops with overlapping environmental characteristics, such as mungbean and mothbean, exhibited higher misclassification rates. Nevertheless, the model showed reasonable robustness across diverse crop categories. This detailed classification performance is represented in

Figure 5.

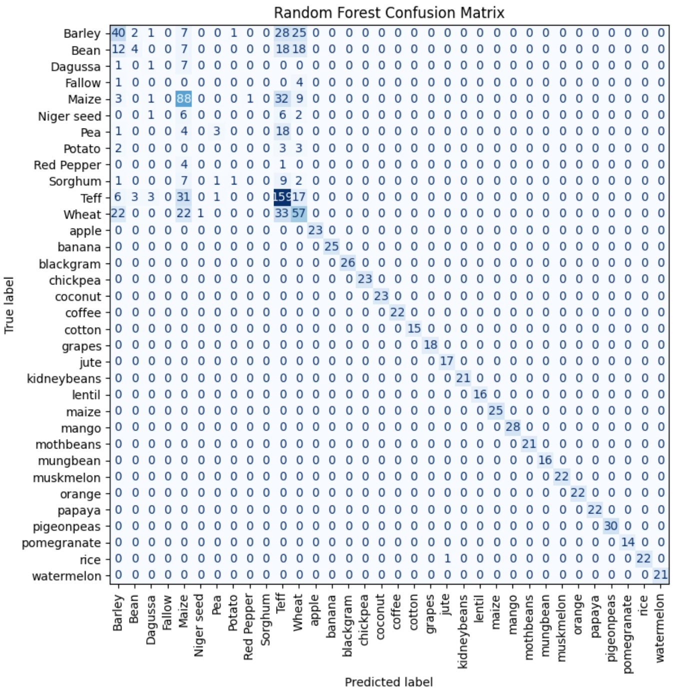

Figure 5. Random forest confusion matrix.

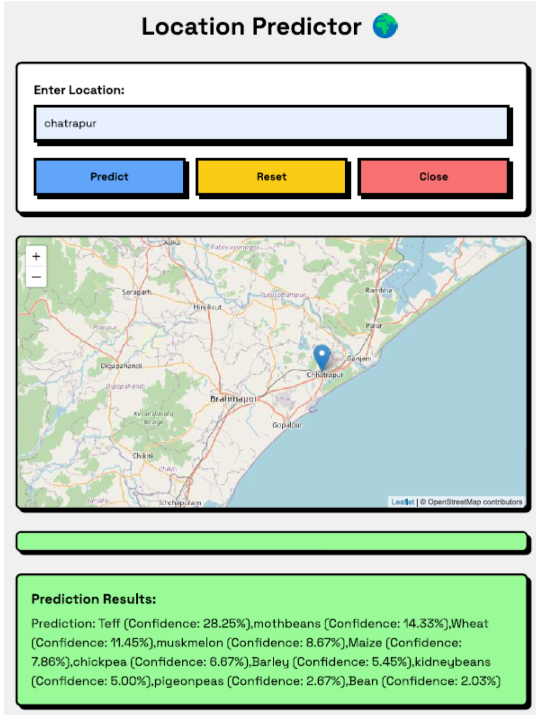

Figure 6. Overview of crop recommendation system architecture.

Furthermore, the overall data distribution for the crop classes was analyzed to assess the balance of the dataset. The distribution plot indicated a reasonable spread among major crop types, although certain crops had fewer instances, contributing to slight imbalances during model training.

Overall, the experimental results affirm the effectiveness of the Random Forest-based model in providing reliable crop recommendations as Figure 6. While accuracy improvements are possible through deeper integration of real-time soil sensors and larger datasets, the current system offers a practical and scalable solution for precision agriculture, enabling farmers to make data-driven decisions aligned with dynamic environmental conditions.

Compared to existing crop recommendation systems reported in the literature [14], the proposed model demonstrates several advantages that make it more suitable for real-world agricultural decision support. First, most prior systems rely heavily on static datasets or historical averages, which limit their adaptability to dynamic agro-climatic conditions. In contrast, our system integrates real-time weather and soil information through APIs, allowing crop recommendations to be continuously updated based on current environmental factors. Second, while many existing models remain in the research or prototype stage without deployment, our work emphasizes practical usability by implementing a Flask-based web application that enables farmers to enter only minimal information such as their location and instantly receive crop suggestions in an intuitive dashboard. Third, the system is cloud-ready and containerized using Docker, ensuring scalability for large-scale use and adaptability across different regions. Another critical advantage lies in the balance between prediction accuracy and computational efficiency. While deep hybrid models like CNN-LSTM or ensemble stacking approaches can sometimes achieve higher theoretical accuracy, they require high computational resources and long response times, making them less suitable for real-time field applications. Our optimized Random Forest model achieves a competitive accuracy of 87.93% while maintaining rapid response times, making it more practical for deployment in precision agriculture. Finally, unlike several existing applications that emphasize yield forecasting or disease detection, the proposed system focuses directly on actionable decision support, providing farmers with clear and timely crop recommendations tailored to their soil and climatic conditions. Taken together, these factors highlight that the proposed model goes beyond traditional machine learning experiments by delivering a scalable, real-time, and user-oriented solution, thereby offering significant improvements over existing crop recommendation applications.

6 Conclusion and Future Scope

This research presents a comprehensive machine learning-based system for intelligent crop prediction and recommendation that successfully integrates historical agricultural data with real-time environmental information. The system combines the Crop_recommendation.csv and CRUSP datasets with dynamic data from OpenWeatherMap and SoilGrids APIs to deliver location-specific crop recommendations tailored to current geospatial and climatic conditions.

Through systematic evaluation of multiple machine learning algorithms, including Decision Tree, XGBoost, Support Vector Machine (SVM), and Random Forest Classifier, this study identified the Random Forest Classifier as the optimal model. The selected model achieved an impressive accuracy of 87.93%, demonstrating exceptional robustness to noise and maintaining acceptable response latency for real-time agricultural decision support applications. The model's superior performance stems from its ability to handle mixed-type features effectively and its inherent resistance to overfitting, making it particularly well-suited for the variability and complexity inherent in agricultural datasets.

To ensure practical accessibility for end-users, particularly small-scale farmers with limited technological resources, a user-friendly Flask-based web application was developed. This interface enables farmers to input their geographical location and instantly receive personalized crop recommendations based on current environmental parameters. The system's design prioritizes simplicity and functionality, making advanced agricultural intelligence accessible to users regardless of their technical expertise.

Despite the promising results, several limitations warrant acknowledgment. The system faces challenges related to data imbalance across crop categories, where certain crops are underrepresented in the training data. Environmental unpredictability poses another significant challenge, as weather patterns and soil conditions can change rapidly and unexpectedly. Additionally, the current model does not incorporate crucial socio-economic factors such as market demand fluctuations, commodity prices, and government subsidy frameworks, which significantly influence crop selection decisions in practice.

The confusion matrix analysis revealed specific difficulties in distinguishing between crops with similar soil, nutrient, and climatic requirements, occasionally affecting prediction precision. These limitations highlight the need for continued refinement and a more comprehensive modelling approach that considers the multifaceted nature of agricultural decision-making.

Several promising avenues exist for enhancing the system's effectiveness and practical impact. First, expanding the training dataset to include a more diverse array of crop varieties and incorporating temporal variables such as seasonal patterns and inter-annual climatic variations would improve the model's ability to capture dynamic environmental influences. This expansion would enable more accurate predictions across different growing seasons and climate scenarios.

Integration with Internet of Things (IoT) technologies presents significant opportunities for system enhancement. Real-time soil health sensors, weather stations, and satellite-based remote sensing data would provide hyper-localized, continuously updated information, enabling more precise and context-sensitive recommendations. Such integration would transform the system from a static prediction tool into a dynamic agricultural intelligence platform.

The incorporation of market intelligence represents another critical advancement area. By embedding real-time data on commodity prices, supply-demand dynamics, and policy interventions, the system could recommend crops that are not only agronomically suitable but also economically advantageous. This holistic approach would address the complete decision-making process that farmers face when selecting crops.

Developing multilingual mobile applications would significantly enhance accessibility, particularly in rural regions where farmers predominantly communicate in local languages. Mobile deployment would also leverage the widespread adoption of smartphones in agricultural communities, making the technology more readily available where it is most needed.

From a technical perspective, exploring advanced deep learning architectures offers substantial potential for performance improvements. Convolutional Neural Networks (CNNs) could enhance spatial feature extraction from satellite imagery and soil maps, while Long Short-Term Memory Networks (LSTMs) could model temporal dependencies in weather patterns and crop growth cycles more effectively.

Future iterations must prioritize fairness, transparency, and data security to ensure ethical deployment. Implementing interpretable machine learning models that provide clear explanations for their recommendations would foster trust among users, particularly important in agriculture where decisions directly impact livelihoods. Additionally, ensuring data privacy and security, especially when integrating IoT devices and personal location information, will be crucial for widespread adoption.

This research establishes a robust foundation for advancing intelligent, data-driven agricultural ecosystems. The demonstrated effectiveness of machine learning in crop recommendation, combined with real-time data integration and user-centric design, validates the potential for transforming contemporary farming practices. The system offers a pathway toward more sustainable, adaptive, and economically resilient agricultural models that can help farmers optimize their crop selection decisions based on scientific evidence rather than intuition alone. The work contributes to the growing field of precision agriculture by demonstrating how advanced computational techniques can be made accessible to farmers at all scales. As agricultural systems face increasing pressure from climate change, population growth, and resource constraints, intelligent decision support systems like the one presented here will become increasingly vital for ensuring food security and agricultural sustainability. The foundation established by this research provides a platform for continued innovation in agricultural technology, with the potential to significantly impact global farming practices and food production systems.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Burdett, H., & Wellen, C. (2022). Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. *Precision agriculture, 23*(5), 1553-1574. [CrossRef]

[2] Shahhosseini, M., Hu, G., & Archontoulis, S. V. (2020). Forecasting corn yield with machine learning ensembles. *Frontiers in Plant Science, 11*, 1120. [CrossRef]

[3] Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., ... & Peng, B. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agricultural and forest meteorology, 274*, 144-159. [CrossRef]

[4] Jeong, J. H., Resop, J. P., Mueller, N. D., Fleisher, D. H., Yun, K., Butler, E. E., ... & Kim, S. H. (2016). Random forests for global and regional crop yield predictions. *PloS one, 11*(6), e0156571. [CrossRef]

[5] Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J., & Kitchen, N. R. (2003). Statistical and neural methods for site–specific yield prediction. *Transactions of the ASAE, 46*(1), 5. [CrossRef]

[6] Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific reports, 11*(1), 1606. [CrossRef]

[7] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and electronics in agriculture, 151*, 61-69. [CrossRef]

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

[9] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture, 177*, Article 105709. [CrossRef]

[10] Khan, S. N., Li, D., & Maimaitijiang, M. (2022). A geographically weighted random forest approach to predict corn yield in the US corn belt. *Remote Sensing, 14*(12), 2843. [CrossRef]

[11] Rani, S., Mishra, A. K., Kataria, A., Mallik, S., & Qin, H. (2023). Machine learning-based optimal crop selection system in smart agriculture. *Scientific Reports, 13*(1), 15997. [CrossRef]

[12] Karimi, N., Arabhosseini, A., Karimi, M., & Kianmehr, M. H. (2018). Web-based monitoring system using Wireless Sensor Networks for traditional vineyards and grape drying buildings. *Computers and Electronics in Agriculture, 144*, 269-283. [CrossRef]

[13] Ferrández-Pastor, F. J., García-Chamizo, J. M., Nieto-Hidalgo, M., Mora-Pascual, J., & Mora-Martínez, J. (2016). Developing ubiquitous

sensor network platform using internet of things: Application in precision agriculture. *Sensors*, *16*(7), 1141. [CrossRef]

[14] Pusatkar, A. C., & Gulhane, V. S. (2016). Implementation of wireless sensor network for real time monitoring of agriculture. *International research journal of engineering and technology (IRJET)*, *3*(05), 5.

[15] Reddy, C. K. K., Daduvy, A., Mohana, R. M., Assiri, B., Shuaib, M., Alam, S., & Sheneamer, M. A. (2024). Enhancing precision agriculture and land cover classification: A self-attention 3D convolutional neural network approach for hyperspectral image analysis. *IEEE Access*, *12*, 125592-125608. [CrossRef]

Sobhana Behera is an Assistant Professor in the Department of Computer Science and Engineering at NIST University. She has taught a wide range of subjects including Problem Solving Techniques, Formal Languages and Automata Theory (FLAT), Software Engineering, Python Programming, and NoSQL Databases. Her academic work is complemented by a strong interest in research at the intersection of Artificial Intelligence and Machine Learning. She is particularly focused on applying AI/ML techniques in the agricultural domain to enhance productivity and sustainability. (Email: Sobhana.behera@nist.edu)

N. Toyaad Kumar Reddy is an Assistant Professor in the Department of Computer Science and Engineering at NIST University. He has taught a diverse range of subjects including Compiler Design, Formal Languages and Automata Theory (FLAT), Computer Organization, Python Programming, E-Commerce & ERP, Data Structures, and Design and Analysis of Algorithms (DAA). His research interests span Artificial Intelligence and Machine Learning with a focus on Retrieval-Augmented Generation (RAG)-based chatbot systems, as well as emerging areas such as Quantum Computing. He is actively engaged in exploring intelligent system design and next-generation computing paradigms. (Email: toyaad.reddy@nist.edu)

Shubhasri Pradhan is an Assistant Professor in the Department of Computer Science and Engineering at NIST University. She teaches a variety of subjects including Database Management Systems (DBMS), Formal Languages and Automata Theory (FLAT), Web Technology, Python Programming, and NoSQL Databases. Her research interests lie in the application of Artificial Intelligence and Machine Learning in the medical domain. She is particularly focused on skin allergy detection using Convolutional Neural Networks (CNN) and multimodal deep learning approaches. (Email: shubhasri.pradhan@nist.edu)