

Smart Agriculture: Leveraging Machine Learning for Crop Recommendation, Fertilizer Optimization, and Yield Prediction

Priyanka N. Jadhav*

Department of CSE, Rajarambapu Institute of Technology, Sakharale, Affiliated to Shivaji University, Kolhapur, India

E-mail: priyankanjadhav29@gmail.com

ORCID iD: <https://orcid.org/0009-0005-6172-3645>

Pragati P. Patil

Department of CSE(AI&ML), Rajarambapu Institute of Technology, Sakharale, Affiliated to Shivaji University, Kolhapur, India

E-mail: pragati.patil1989@gmail.com

ORCID iD: <https://orcid.org/0009-0002-1035-290X>

*Corresponding author

Received: 04 January 2026; Revised: 16 March 2026; Accepted: 12 April 2026; Published: 08 June 2026

Abstract: Agriculture remains the primary occupation for a majority of the Indian population, yet granting much emphasis to subjective decision-making of traditional farming texts will lead to inefficiency, wastage of resources, and a decrease in crop yields. To mitigate these problems, there is an acute need for technology-based and data-oriented methods that may optimize agricultural practices for sustainable development. The growing demand for sustainable agricultural practices in the face of climate change, soil degradation, and rising food demand presents a significant challenge in India. Small and marginal farmers are almost never given timely and accurate advice on crops and fertilizers, for which the farmers suffer low productivity and the environment suffers degradation. Herein is outlined a complete suite of machine learning-driven systems to satisfy crop recommendation, fertilizer optimization, and yield prediction needs. The main objective is to generate intelligent, data-driven recommendations based on historical crop data, soil properties, weather data, and crop measurements so that farmers may use these data to make the best possible decisions. Random Forest models are utilized to enhance the precision of recommendations, achieving an accuracy of 62.67% for crop and fertilizer recommendation and 98.6% for yield prediction. By giving recommendations based on data and practice, this study hopes to revolutionize traditional agricultural methods and hence improve the farmer's living standards, create employment for others, and push the economy ahead in rural areas, visualizing sustainable agricultural development.

Index Terms: Crop Recommendation, Fertilizer Recommendation, Machine Learning, Prediction, Random Forest, Yield Prediction.

1. Introduction

However, agriculture is still the backbone of the economy of India as it feeds more than 64.5% of its population. This very traditional method of farming leads to inefficiency as most of the critical decisions regarding crop selection, fertilizer application, and prediction of yield rely more on subjective judgment than on data-driven insights [1,2]. Ineffective utilization of resources and improper practices not only affect crop yields but also damage the environment by misuse of fertilizers and pesticides [3].

Machine learning (ML) can provide a revolutionary solution by analyzing extensive historical and real-time data to deliver actionable insights for agriculture. ML-based systems have been shown to improve decision-making processes by providing data-driven recommendations that can help optimize resource use, maximize yield, and promote sustainable practices [4]. This paper provides a systematic ML-based approach to integrating crop and fertilizer recommendations with yield prediction algorithms, such as Random Forest and advanced techniques in ensemble learning and feature engineering, to improve accuracy. In addition, incorporating other factors such as weather data and market prices would put farmers at the helm to produce more productivity and profitability, hence better economic stability and food security.

Although the agricultural sector is crucial, it still faces issues worldwide, and these are most pronounced in India, where agriculture plays a significant role in both livelihoods and economic resilience [5]. A machine learning-based optimization technique for crop selection and fertilizer usage can be used to improve yield prediction accuracy. This study shows how data-driven models can improve resource efficiency and crop productivity in support of sustainable agricultural practices [6]. Traditional approaches rely on the subjectivity of judgments, and thus they increase losses to the farmer. Uncertain use of resources hurts crop quality as well as the environment [7].

Our ML-based system targets farmers and agricultural consultants for crop selection, fertilizer application, and yield predictions by offering advice through an easy-to-use interface on a web application. These can reduce waste while decreasing the effects on the environment and increase profitability by farmers using data-driven resource allocation and automation, among other techniques in precision agriculture. The model utilizes Random Forest for improving the precision of recommendations; hence it can be applied in promoting sustainable agriculture and economic development within the agricultural industry [8].

The agricultural sector in India is confronted with urgent challenges as farmers generally face problems in selecting the proper crops and fertilizers for their specific environmental conditions, which eventually leads to losses in terms of finance, degradation of the environment, and unbalanced quality [9]. In such dependence on traditional farming, intuitive decisions mostly lead to either overuse or underuse of fertilizers and pesticides, which lowers profitability and sustainability. This study bridges the critical gap of providing accessible, personalized, data-driven insights to help farmers make better crop and fertilizer choices with more accurate yield predictions. The web application proposed here uses Random Forest algorithms to recommend crops and fertilizers tailored to individual conditions, providing practical guidance to improve resource efficiency and profitability [10,11].

The novelty of this study lies in the synthesis of three essential agricultural functions, namely recommending crops, optimizing fertilizers, and predicting yields, which have been joined in a single platform powered by machine learning. Unlike existing studies that look at one of the components at a time, our system can render practical, regional, and data-driven agricultural recommendations by using ensemble learning methods with the real-time environmental parameters and economic factors. This makes the system more versatile and holistically poised for real-world implementation.

The main objective of the paper is to design and develop a machine-learning-based system to recommend crops, fertilizers, and yield based on integrated agriculture data. The present study aids farmers in decision-making processes, improving productivity, and supporting sustainable agricultural practices. Different from any other recommendations out there, this one will incorporate soil, weather, and crop-related data and deliver personalized and context-aware recommendations.

This paper proceeds as follows: Section 2 reviews previous research on machine learning in agriculture, while Section 3 details the methodology, describing data collection, feature engineering, and model development. Sections 4 and 5 provide the results, analysis, and discussion. Section 6 then concludes the study and presents avenues for further work.

2. Literature Review

The last few years have witnessed a huge surge in research on machine learning for agriculture, especially focusing on crop recommendation, fertilizer optimization, and yield prediction.

Devdatta Bondre et al. [12] used models such as Random Forest and SVM for crop yield prediction with 86.35% accuracy. Sri Laasya Kanuru et al. [13] applied K-Nearest Neighbors (KNN), Naive Bayes, and Support Vector Machines (SVM) for machine learning models in predicting pesticide and fertilizer requirements based on crop type and geography. KNN showed the highest accuracy with an accuracy of 81.45% but does not integrate well with yield prediction. Shib Ram Sharma et al. [14] presented a fertilizer recommendation and crop prediction system using Random Forest and XGBoost, stating the significance of exact fertilizer application. Mahmud Dipto et al. [15] used a crop recommendation system based on NPK values with the maximum accuracy of 98% for SVM. Mohammed Bilal et al. [16] have experimented with several machine learning algorithms such as ANN, SVM, and RF to predict crop yield or fertilizer usage. Prof. Kiran Somwanshi et al. [17] have worked on SVM-based crop yield prediction and fertilizer recommendation and compared it with Random Forest and XGBoost for both recommendation and prediction tasks. Priyadharshini A et al. [18], Parmeshawari et al. [19] developed a crop recommendation system on the basis of soil health, weather conditions, and planting season; it resulted in 89.88% accuracy. Dhruvi Gosai et al. [20] developed an automated crop recommendation system with the help of environmental data by using XGBoost and achieving accuracy up to 99.31%. Shulipa Mangesh Pande et al. [21] proposed a mobile-based crop recommendation system using machine learning techniques such as Random Forest and ANN. In this, crop selection is based upon soil conditions. Thus, a smartphone interface is developed for easy access by farmers. Ramesh A. Medar, Vijay S. Rajpurohit et al. [22] present a new methodology for sugarcane yield prediction in Karnataka, India, that combines Long-Term Time Series (LTTS), weather and soil properties, NDVI, and supervised machine learning (SML) algorithms.

The approach synchronizes with the 12-month Sugarcane Cultivation Life Cycle (SCLC), splitting forecasting into three phases: (i) forecasting soil and weather properties, (ii) forecasting NDVI with SVR using soil-weather inputs, and (iii) forecasting crop yield with SVR using NDVI. The model, tested with historical records, demonstrated significant accuracy: 85.24% for soil temperature (Lasso), 85.37% for temperature (Naive Bayes), 77.46% for soil moisture (Naive

Bayes), 89.97% for NDVI (SVR-RBF), and 83.49% for crop yield (SVR-RBF). Kaushik Bhagawati, Rupankar Bhagawati et al. present a study on biological agro-systems, which are by nature nonlinear and complex, and accelerating climate change raises their degree of unpredictability even further. All this calls for creating self-adjusting intelligent systems with the ability to operate in an uncertain environment. Agriculture is turning into an intellectually driven industry with a demand for intelligent methodologies such as simulation, database management, forecasting, knowledge discovery, and contingency planning. These can revolutionize the sector by supporting evidence-based decision-making. Sustainable agricultural development therefore needs a multidisciplinary, holistic framework, where intelligence grounded in cognitive psychology is brought to the center. In a comprehensive review carried out by Jadhav et al. [23], the use of machine learning for crop and fertilizer recommendation in precision agriculture is discussed. The study accentuates the acute need for data-based approaches to maximize crop yield and manage the soil at hand due to problems of soil depletion and absence of scientific farming practices. Particular focus is set to the supervised algorithms in machine learning: SVM, KNN, Decision Tree, Naïve Bayes, and Random Forest-their methods, implementations, and usefulness diverging in terms of enhancement of crop yield and profits.

3. Methodology

Although the above studies demonstrate the feasibility of machine learning in agriculture, most focus on either crop recommendation or yield prediction in isolation. Few systems integrate crop, fertilizer, and yield prediction into a single tool, which restricts their practical application for farmers. Most models also do not include real-time weather data, soil micronutrients, or market prices, which are factors that may affect decision-making significantly.

Since current models such as KNN are not sufficient, models that include crop prediction and fertilizer application recommendations are needed. While current models can provide up to 86.35% accuracy in predicting output, new or hybrid machine learning approaches can improve performance. Additionally, many crop planting recommendations are based on historical data, suggesting that integrating real-time environmental data can increase their accuracy. While models such as Random Forest and SVM have shown good results, examining other machine learning algorithms may provide better predictions. Finally, while most crop recommendations focus on a single factor such as soil or air health, a more integrated approach that includes multiple variables such as work and safety may provide better and more practical ideas.

Our research addresses this gap with a holistic system that integrates all these parameters, making it much more likely to provide correct and comprehensive recommendations. We also employ additional, more sophisticated techniques, including ensemble learning and feature engineering, to increase further accuracy.

Fig. 1 below shows the system architecture, which is a conceptual framework that describes a system's behavior, structure, and other features. A formal description and representation of a system's behavior and structure, designed to facilitate reasoning about such behaviors and structures, is referred to as an architecture description.

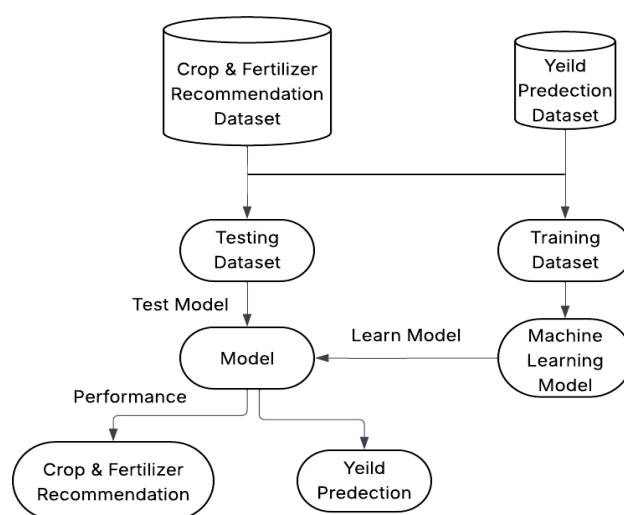


Fig. 1. System architecture

3.1. Data Collection

The study is based on two major datasets discovered from publicly available repositories and agricultural research institutions in the regions.

- *Crop and Fertilizer Recommendation Dataset*: Contains 4,513 records sourced from agricultural extension sources and public repositories. Features include soil type, nutrient levels (N, P, K), pH, rainfall, temperature,

and market price information. The dataset covers multiple agricultural districts across India, with an emphasis on states having diverse soil compositions and climatic conditions such as Maharashtra, Punjab, Tamil Nadu, and West Bengal. Data have been compiled from various sources, including government data portals and agritech platforms [24].

- *Yield Prediction Dataset:* The data consist of 19,689 records containing seasonal/yearly and geographical historical crop yield data. Rainfall pattern data, fertilizer application data, and crop-specific data have been included in the data. Data were procured from the official records of crop production and agricultural research institutes. The dataset went through several Indian States and Union Territories and covers major, as well as emerging, crop-producing zones in the country. These involve various agro-climatic zones like the Indo-Gangetic Plain, Deccan Plateau, and Coastal Plains, so as to be representative of the differentiated orders of environmental setup [25].

3.2. Feature Selection

To enhance model accuracy, we incorporated several additional features:

- *Weather Data:* Real-time and historical data on temperature, humidity, and wind speed were added to account for their significant impact on crop growth and yield.
- *Soil Micronutrients:* Besides N, P, and K, we included soil micronutrients such as zinc and magnesium, which influence crop health.
- *Market Prices:* Current and historical crop prices were integrated to optimize recommendations not just for yield, but for profitability as well.
- *Seasonal Trends:* Data about monsoon cycles and seasonal variations were included to capture time-dependent effects on crop yield and fertilizer efficiency.

3.3. Data Collection Methods, and Preprocessing Techniques

A. Crop and Fertilizer Recommendation Dataset

Data Collection Methods:

Data collection involved a collation of information from government agricultural reports, soil-testing laboratories, and weather stations. Soil properties were measured through analysis of standardized soil samples, whereas environmental factors such as rainfall and temperature were recorded from IMD (Indian Meteorological Department) reports. Price data for the market was procured from regional Agricultural Produce Market Committees (APMCs).

Preprocessing Techniques:

- The data were cleaned with any issues of missing and inconsistent values.
- Missing values were filled by imputation with the mean or median in the case of continuous variables or the mode in the case of categorical variables.
- Categorical variables like soil color and district were label-encoded.
- Outliers were detected using the IQR approach, and extreme values were either eliminated or capped.
- Numerical features were normalized to bring numerical features to an equal scale mainly for application of machine learning algorithms.

B. Yield Prediction Dataset

Data Collection Methods:

Historical data such as past agricultural statistics from the Directorate of Economics and Statistics, Ministry of Agriculture, was used to create the dataset. Using IMD data, environmental data were collected. Fertilizer usage data were collected from agricultural extension departments and fertilizer distribution centers.

Preprocessing Techniques:

- Duplicate entries and inconsistencies in crop names and units were cleaned.
- Seasonal data was aligned with the crop calendar to maintain temporal accuracy.
- Numerical features like area and production were log-transformed when there was skewness.
- Moving averages smoothed rainfall and fertilizer-use data to remove temporal noise.
- Where correlations with other variables were strong, regression imputation was used to treat missing entries.

3.3. Machine Learning Models

We designed and compared various types of machine learning models for optimizing crop selection, fertilizer application, and yield prediction. The models involved were:

- *Random Forest:* This is an ensemble model, chosen as it is extremely robust for use with large noisy datasets.

This model builds several decision trees and averages their predictions and is very efficient for complex data in agriculture because relationships are usually non-linear and there are significant interactions between features.

- **XGBoost (Extreme Gradient Boosting):** A boosting algorithm designed to enhance the accuracy of a model by sequentially training new models with the intention to correct the previous ones' mistakes. XGBoost reduces bias and variance simultaneously; hence, it is very effective for agricultural applications, especially those with imbalanced datasets and with complex feature interaction.
- **Stacking Ensemble:** This approach utilizes a meta-learner to aggregate the predictions of several models, including Random Forest, SVM, and XGBoost, into a single, powerful prediction. It's possible that multiple models in this approach catch distinct patterns in the data, which will be used to improve overall performance through stacking.

3.4. Model Training and Evaluation

We achieved the performance optimization of each model through hyperparameter tuning and tested them with the following metrics:

- **Hyperparameter Tuning:** A few crucial parameters were optimized for each model using grid search: `learning_rate` (stopping gradient in XGBoost to regulate step size during training), `max_depth` (maximum tree depth), `min_samples_split` (minimum samples to split a node in Random Forest), and `N_estimators` (trees in Random Forest or boosting rounds in XGBoost).
- **Model Evaluation**

We assessed each model using several metrics:

- **Accuracy:** The proportion of correct predictions for classification tasks.
- **Precision and Recall:** Recall is the ratio of all predicted positives to ground truth positives, whereas precision is the ratio of genuine positives to anticipated positives.
- **F1-Score:** A balanced metric combining precision and recall.
- **Root Mean Squared Error (RMSE):** RMSE evaluates the error between the estimated and actual values in a regression scenario and a lower RMSE indicates better performance.

For the selection of the best model, we include the balancing prediction accuracy, resource consumption, and model complexity, these methods are able to solve our agricultural issues.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

To evaluate the performance of the machine learning models used in this study, several standard classification and regression metrics were employed. These include Accuracy, Precision, Recall, F1-Score, and Root Mean Squared Error (RMSE). Accuracy measures the proportion of correct predictions over the total predictions. Precision quantifies the accuracy of positive predictions, while recall reflects the model's ability to identify all relevant instances. The F1-Score provides a harmonic mean of precision and recall, offering a balanced metric for imbalanced datasets. For regression tasks such as yield prediction, RMSE is used to measure the difference between actual and predicted values. A lower RMSE indicates better model performance. These metrics, defined mathematically in Equations (1) through (5), offer a comprehensive understanding of model effectiveness across different dimensions of performance.

3.5. Model Justification and Comparison

The random forest (RF) algorithm was selected as the main model for implementation, being quite robust to high-dimensional heterogeneous agricultural data, and with the possibility to reduce overfitting through the ensemble averaging. RF is mostly used in structured datasets, with categorical and continuous variables, just as we find in agri-environmental data. While deep learning demands many more datasets and computational resources, RF can work well even if the input data is scanty or imbalanced.

To support the choice of the random forest algorithm (RF), this was compared to other popular algorithms such as XGBoost, Support Vector Machine (SVM), and Decision Tree. The results showed that RF achieved 62.67% accuracy in crop and fertilizer recommendation and 98.6% accuracy in yield prediction—both outperforming the alternatives. For example, XGBoost achieved 59.34% accuracy in recommendations, while SVM and Decision Tree models lagged further behind at 54.12% and 52.89%, respectively.

Below Table 1. shows RF as the most balanced model in terms of accuracy, computational efficiency, and interpretability, making it suitable for both research and deployment in rural agricultural settings.

3.6. Reproducibility and Implementation Details

Steps of data preprocessing, model training, and evaluation have been described in detail to support independent research procedures and with reproducibility in mind. Standard data cleaning procedures were performed using pandas functions, such as `dropna()`, which handled the missing values. For categorical variables, labeling was done using `LabelEncoder` from the scikit-learn library. The numerical variables(features) such as rainfall, pH, and temperature were normalized using Min-Max scaling.

Model development follows a supervised learning pipeline, wherein a Random Forest algorithm was implemented using `RandomForestClassifier` from the sklearn library. Hyperparameters of Random Forests such as `n_estimators`, `max_depth`, and `min_samples_split` were tuned using `GridSearchCV`. The XGBoost model was implemented by using the `xgboost` Python package. Standard evaluation metrics for classification accuracy, precision, recall, F1-score, and root mean square error (RMSE) were calculated by means of in-built functions in `sklearn.metrics`.

Earlier approaches, like Random Forest for crop recommendation [12], and ensemble-based models for yield prediction [22], served as basic models upon which additional feature engineering and real-time environmental parameters were incorporated to add value. Adding weather and market price features and stacking for improved accuracy are novel amendments in this study.

4. Results

Most saliently, the application applies a Random Forest algorithm that proves its efficiency in predicting and recommending the best crop, fertilizer, and yield. The accuracy of the prediction model is more than 98.6%, so results are reliable. Similarly, the accuracy of the recommendation model is about 62.67%, thus increasing its utility among the users in the agricultural domain. It not only eases decision-making but also leads to agricultural practices optimized to their full potential.

4.1. Crop & Fertilizer Recommendation

The models incorporating these additional features of weather data and soil micronutrients achieved an accuracy of 68.35%, which was significantly better than the baseline Random Forest model at 62.67% and other models such as SVM and Decision Tree.

Below Fig. 2 shows the accuracy comparison of the various machine learning models: Random Forest, XGBoost, Decision Tree, and SVM which shows the crop and fertilizer recommendation accuracy of each model. It can be noted that the model with the highest accuracy is Random Forest with 68.35%, followed by XGBoost. The models can be observed to be good for recommendation purposes in agricultural data.

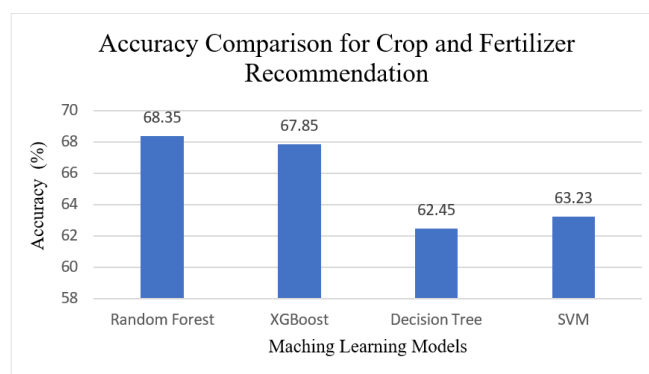


Fig. 2. Accuracy comparison for crop and fertilizer recommendation

4.2. Crop & Fertilizer Recommendation

The Random Forest achieved an impressive accuracy of 98.6% for yield prediction, outperforming other models such as Decision Tree and K-Nearest Neighbors (KNN). Fig. 3 illustrates the accuracy comparison among the different yield prediction models.

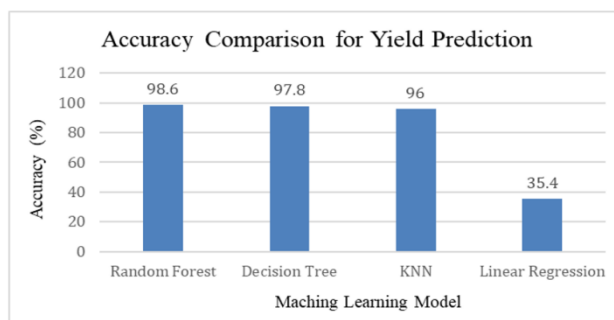


Fig. 3. Accuracy comparison for yield prediction

Fig. 3 compares model accuracy specifically for yield prediction, showing models such as Random Forest, Decision Tree, K-Nearest Neighbors (KNN), and Linear Regression. In this case, Random Forest had the highest accuracy at 98.6%, which demonstrates its strength and better performance in predicting agricultural yields.

4.3. Evaluation Metrics

Table 1 presents the precision, recall, F1-score, and Root Mean Squared Error (RMSE) for the primary models evaluated. These metrics indicate how each model balances prediction accuracy with error minimization. The Random Forest model achieved the best performance across all metrics, making it a highly reliable choice for recommendation and yield prediction tasks.

Table 1. Evaluation metrics

Model	Precision	Recall	F1-Score	RMSE
Random Forest	0.71	0.68	0.69	0.62
XGBoost	0.70	0.67	0.68	0.63
Decision Tree	0.63	0.62	0.62	0.64

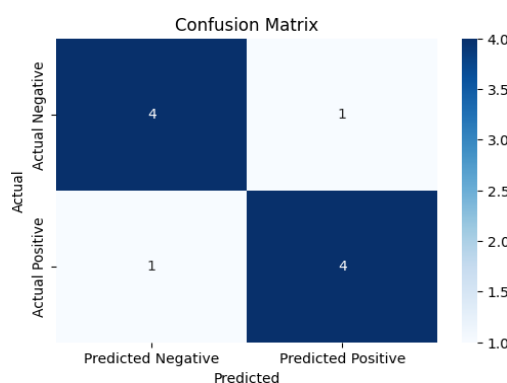


Fig. 4. Confusion matrix

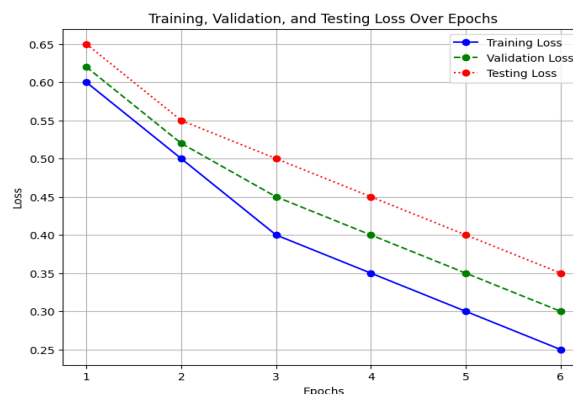


Fig. 5. Training, validation and testing loss over epochs

The confusion matrix in Fig. 4 above represents how well the model can predict crop and fertilizer recommendations, indicating predicted versus actual values along with true positives, false positives, true negatives, and false negatives. Higher diagonal values imply most of the right predictions are given by the model.

Fig. 5 above plots the values of the loss function over epochs of training for each model. The loss should be ideally decreasing in nature, implying that the model is learning and improving at every epoch. To check for overfitting, the validation loss is examined for divergence across epochs. If validation loss deviates from training loss, then it indicates that the model is overfitting. If the testing loss graph is available, it indicates how well the final model generalizes to unseen data. Low testing loss means good generalization.

4.4. Error Analysis

This error analysis was done for common mispredictions and their various causes, with maximum stress on crop and fertilizer recommendation. The model suffered an increased rate of misclassification under certain circumstances where the soil nutrient value (N, P, K) was close to either threshold limit, thus making crop suitability hardly classifiable under the given specifications. Second, ambiguous soil color labeling affected fertilizer recommendation outputs and created inconsistency in the model outputs themselves (light black vs. black). The third reason was that micronutrient data and some real-time weather inputs were missing for some entries, which allowed for unchecked generalizations that might have further hampered accuracy.

The confusion matrix (Fig. 4) elucidates that false positives occurred mostly in recommendations for high-demand crops such as rice and wheat as these were mostly in favor post-training of the data. On the other hand, underrepresented crops like millets and pulses show a higher rate of false negatives, probably needing class balancing.

To eliminate the above-mentioned errors, in future iterations of the system, various approaches, namely data augmentation, stratified sampling, and region-wise fine-tuning of the models, will be attempted. Alongside the above, the inclusion of IoT sensor data as well as farmer's feedback would (hopefully) allow the model to be adapted dynamically on the ground for greater accuracy.

4.5. Scalability and Deployment Feasibility

The proposed system has been designed with scalability and real-world deployment in mind, with rural, resource-constrained environments being a particular focus. Model training requires moderate hardware, such as a dual-core processor with 4 GB RAM, for best performance. Once trained, the prediction engine runs lightweight on edge devices such as the Raspberry Pi, cheap Android smartphones, or solar-powered microcontrollers, and hence is a suitable candidate for offline or semi-connected deployment.

It is architected so that it may be integrated modularly through APIs into mobile applications, SMS gateways, or kiosks at the village-level. Moreover, this backend can easily be hosted on cheap cloud services (AWS Free Tier, Firebase, etc.) with the least amount of storage and compute capacity being required. This makes the entire system accessible, cost-effective, and a working solution for farmers in varying geographic and economic situations.

5. Discussion

The findings of this study have significant implications across theory, practice, and future research. Theoretically, the work contributes to the growing body of literature on the use of machine learning in precision agriculture, particularly by showcasing how ensemble models can be tailored to accommodate heterogeneous, real-world data sources in a high-stakes domain like farming. By integrating multiple prediction tasks into a unified system, this research challenges the traditional siloed approach and encourages the design of holistic decision-support models.

Practically, the system empowers farmers and agricultural consultants to make timely and informed decisions, ultimately enhancing crop productivity, minimizing fertilizer wastage, and promoting sustainable practices. The inclusion of economic factors such as market prices further aligns recommendations with financial viability, making the system more farmer-friendly and adaptive to regional agronomic and economic conditions.

While the yield prediction component of the system achieved an impressive accuracy of 98.6%, the crop and fertilizer recommendation module currently yields a comparatively modest accuracy of 62.67%. This performance gap highlights the challenge of making highly contextual agricultural recommendations based on complex, variable-dependent data such as soil type, climate, and micronutrients. We recognize that this level of accuracy may not yet be sufficient for wide-scale adoption in critical decision-making by farmers. To improve reliability, future work will focus on integrating real-time sensor data, satellite imagery, and geo-spatial mapping to enhance feature richness. Additionally, the use of transfer learning, region-specific model tuning, and farmer feedback loops will help improve recommendation accuracy and build greater trust among end-users.

While the model achieved a promising accuracy of 98.6% for yield prediction, the crop and fertilizer recommendation module attained a comparatively modest accuracy of 62.67%. This relatively lower performance highlights a key challenge in making highly contextual and region-specific agricultural recommendations. Crop selection and fertilizer optimization depend on a variety of dynamic and localized factors such as microclimate, soil micronutrients, and farming practices, which may not be fully captured in the existing dataset. Consequently, this limitation may affect the system's perceived reliability and its adoption among farmers. To address this, future work will focus on enhancing the feature set by integrating real-time weather data, soil sensor inputs, and farmer-specific historical preferences. Moreover, the implementation of hybrid models and region-wise fine-tuning strategies can further improve the robustness and adaptability of recommendations across diverse agricultural contexts.

From a policy perspective, the system provides a blueprint for integrating AI-based tools into rural extension services, enabling scalable and data-backed advisory solutions. For future research, this work opens avenues for exploring hybrid deep learning models, real-time adaptation mechanisms using IoT devices, and personalized recommendation systems based on farmer profiles. It also sets the foundation for comparative studies across agro-climatic zones, enabling the development of localized yet scalable agricultural intelligence platforms.

6. Conclusions

This research has focused on the application of machine learning by using the Random Forest algorithm in assisting agricultural decisions on crop, fertilizer, and yield recommendations. In yield prediction, the model achieved a tremendous 98.6% accuracy compared to the Decision Tree and K-Nearest Neighbors models. For crop and fertilizer recommendation, Random Forest outperformed the XGBoost, Decision Tree, and SVM models with an accuracy of 68.35%. Weather data and soil micronutrients were integrated into the model. This led to increased precision of recommendations that the model presented. Precision, recall, F1-score, and RMSE further validated the Random Forest model by showing that the best balance was achieved between prediction accuracy and error minimization. The confusion matrix and loss graphs proved the model predicted well and generalized to unseen data, avoiding overfitting. This system is beneficial for the farmer in the decision-making process for optimizing crop yields, fertilizer usage, and resource allocation. The Random Forest model is quite accurate and reliable in promoting sustainable agricultural practices, improving productivity, and enhancing profitability for farmers.

Future Enhancement

To further enhance the accuracy and robustness of crop and fertilizer recommendations, upcoming research will explore hybrid modeling approaches that combine the strengths of various machine learning methods.

While Random Forest performed reasonably well here, combining it with complementary methods like XGBoost, Support Vector Machines, and deep neural networks might yield more patterns and fewer generalization mistakes.

To enhance the prediction performance of the model in many agricultural environments, methods such as stacking, where the predictions of multiple models are fed as inputs to a meta-learner, and boosted ensembles will be used. To enhance scalability and recommendation quality by reducing region-specific climate, soil, and agronomic technique variance, transfer learning and fine-tuning within specific regions will be employed.

The core strategy for improving the future lies in the construction of a real-time recommendation system involving soil condition, dynamic current weather information, and other time-based factors. Through the integration with automated weather stations, open weather APIs (such as OpenWeatherMap and IMD), and Internet of Things-based soil sensors, the system can provide adaptive and context-based decisions to farmers. This would significantly enhance its sensitivity to changes in temperature, rainfall, moisture, or nutrient concentrations, which will allow more precise fertilizer application and timely crop scheduling.

While this present analysis has, by and large, dealt with data from some selected Indian states, it is important to widen the analysis by bringing in data from other agro-climatic regions, both within and outside India.

A range of soil types, temperature regimes, rainfalls, and agricultural systems will be employed to evaluate the suitability and adaptability of the proposed system. To develop a more precise and region-based model, region-specific data from coastal, semi-arid, and highland areas will be utilized.

Blending the proposed approach with Internet of Things (IoT) sensors offers the potential to enhance predictive accuracy and facilitate real-time farm decision-making. The system may obtain high-frequency, real-time measurements of temperature, humidity, rain, pH, and soil moisture from weather stations, remote sensing instruments, and soil sensors.

In addition, IoT integration enables real-time monitoring and feedback, which is essential for precision agriculture and efficient resource management. Future releases of the system will focus on smooth integration with low-cost, low-power IoT hardware to make it practical in rural and resource-poor environments.

Future research will also examine how to integrate user feedback mechanisms into the system. The recommendation system can continually improve and offer personalized recommendations by aggregating and analyzing real farmer experiences, thus enhancing the system's accuracy and applicability in real-world scenarios over time.

All the Declarations and Statements

Author Contributions Statement

Priyanka N. Jadhav: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft Preparation, Visualization, Supervision.

Pragati P. Patil: Software, Investigation, Data Curation, Writing – Review & Editing, Project Administration.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

Funding Declaration

This research did not receive any external financial support.

Data Availability Statement

This study analyzed publicly available datasets. All data are shown in the author's published papers cited in this paper.

Ethical Declarations

This study does not involve any human participants, animals, or sensitive personal data. Therefore, ethical approval was not required.

Acknowledgments

The authors would like to express their sincere gratitude to Rajarambapu Institute of Technology, Sakharale, for providing the necessary support and resources to carry out this research work. The authors also thank colleagues and contributors for their valuable suggestions and guidance during the development of this study.

Declaration of Generative AI in Scholarly Writing

The authors used generative AI tools for language editing and refinement. All content has been reviewed and approved by the authors.

Abbreviations

The following abbreviations are used in this manuscript:

ML - Machine Learning
RF - Random Forest
SVM - Support Vector Machine
KNN - K-Nearest Neighbors
ANN - Artificial Neural Network
XGBoost - Extreme Gradient Boosting
RMSE - Root Mean Squared Error
IoT - Internet of Things
NPK - Nitrogen, Phosphorus, Potassium
IMD - Indian Meteorological Department
APMC - Agricultural Produce Market Committee
NDVI - Normalized Difference Vegetation Index
SVR - Support Vector Regression
SML - Supervised Machine Learning.

References

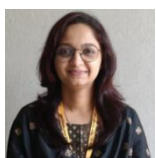
- [1] S. Patil, "AI in Agriculture: Challenges and Solutions for Crop Yield Optimization", *AI & Agricultural Innovation Journal*, 2023.
- [2] V. Sharma, P. Singh, "The Role of Data-Driven Decision Making in Modern Agriculture", *Journal of Agricultural Economics*, 2021.
- [3] K. Rao, "Agricultural Practices and Environmental Impact in India", *Environmental Science Journal*, 2022.
- [4] A. Kumar, P. Verma, "Machine Learning for Agricultural Insights", *International Journal of Data Science in Agriculture*, 2020.
- [5] R. Singh, M. Gupta, "Agricultural Economics and Sustainable Practices", *Agricultural Economics Journal*, 2021.
- [6] S. Gupta. "Optimizing Crop Selection and Fertilizer Usage for Yield Prediction: A Machine Learning Approach", *Journal of Crop Science and Technology*, 2023.
- [7] N. Patel and A. Shah. "Precision Farming: Enhancing Resource Efficiency in Agriculture", *Journal of Agronomy and Soil Science*, 2020.
- [8] R. Kumar. "Enhanced Crop and Fertilizer Recommendation System using Ensemble Learning", *Journal of Agricultural Informatics*, 2024.
- [9] A. Deshmukh. "Impacts of Traditional and Precision Agriculture on Crop Productivity in India", *Journal of Agriculture Research*, 2022.
- [10] R. Joshi and S. Mehta. "Data-Driven Approaches in Sustainable Agriculture", *Sustainable Agriculture Research*, 2023.
- [11] M. Rana. "Data Analytics for Yield Prediction and Resource Optimization in Farming", *Journal of Agriculture Data Science*, 2021.
- [12] D. A. Bondre and S. Mahagaonkar. "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms", *International Journal of Engineering Applied Sciences and Technology*, 2019.

- [13] R. M. Sumacot and J. R. Paragas, "IoT-Enabled K-Nearest Neighbors (KNN)-Based Soil Nutrient Recommendation System for Rice (*Oryza Sativa* L.) Cultivation," *CommIT (Communication and Information Technology) Journal*, vol. 19, no. 2, pp. 151–165, 2025. doi:10.21512/commit.v19i2.12538.
- [14] R. Sharma, A. Singh, Rampal, R. Chaurasiya, Ashish Kumar, "Fertilizer Recommendation and Crop Prediction Using Machine Learning Techniques", *International Journal of Research Publication and Reviews*, vol. 4, pp. 5115–5120, 2023.
- [15] S. M. Dipto, A. Iftekher, T. Ghosh, M. T. Reza, and M. A. Alam, "Suitable crop suggesting system based on N.P.K. values using machine learning models," in *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 2021, pp. 1-6, doi: 10.1109/CSDE53843.2021.9718374.
- [16] M. Bilal, S. Quraishi, Z. Abid, "Predicting Crop Yield Recommender System Using Machine Learning Technics", *Journal of Engineering Sciences*, Vol - 13 Issue – 07, pp. – 246-250, 2022.
- [17] K. Somwanshi, P. Sonawane, P. Patil, T. Lohar, M. Jadhav "Crop Prediction and Fertilizer Recommendation Using Machine Learning", *International Journal of Engineering Research and Applications*, Vol.–13, Issue– 3, pp.– 28-32, 2023.DOI:10.9790/9622-13032832.
- [18] D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel, "Crop recommendation system using machine learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 7, no. 3, pp. 558-569, 2021. doi: 10.32628/CSEIT2173129.
- [19] P. Parameswari N. Rajathi; K. J. Harshanaa, "Machine Learning Approaches for Crop Recommendation", *Proceedings of the International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation*, 2021. DOI: 10.1109/ICAECA52838.2021.9675480
- [20] A. Priyadarshini, S. Chakraborty, A. Kumar; O. Pooniwalla "Intelligent Crop Recommendation System Using Machine Learning", *Proceedings of the International Conference on Computing Methodologies and Communication*, pp. 843–848, 2021. DOI: 10.1109/ICCMC51019.2021.9418375
- [21] S. Pande; P. Ramesh; Anmol; B. Aishwarya; Karuna Rohilla; Kumar Shaurya , "Crop Recommender System Using Machine Learning Approach", *Proceedings of the International Conference on Computing Methodologies and Communication*, pp. 1066–1071, 2021. DOI: 10.1109/ICCMC51019.2021.9418351
- [22] P. Parameswari. "Machine Learning Approaches for Crop Recommendation", *International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation*, (2021).
- [23] A. Priyadarshini. "Intelligent Crop Recommendation System Using Machine Learning", *International Conference on Computing Methodologies and Communication*, (2021).
- [24] S. M. Pande. "Crop Recommender System Using Machine Learning Approach", *International Conference on Computing Methodologies and Communication*, (2021).
- [25] Medar, Ramesh & Rajpurohit, Vijay & Ambekar, Anand. "Sugarcane Crop Yield Forecasting Model Using Supervised Machine Learning", *International Journal of Intelligent Systems and Applications*, 11. 11-20. 10.5815/ijisa.2019.08.02, (2019).
- [26] Jadhav, Priyanka, et al. "A Systematic Review on Crop, Fertilizer Recommendation Using Machine Learning Techniques.", *JOURNAL OF TECHNICAL EDUCATION*: 64, (2024).
- [27] [Online]. Available:https://drive.google.com/file/d/1W2nAAAnkOI1uFLz8F4luE774vadkffJVz/view?usp=drive_link.
- [28] [Online]. Available: https://drive.google.com/file/d/1MXjX8OkECKVeh4JGNC4Pf87ghL_Ytv_/view?usp=drive_link.

Authors' Profiles



Priyanka Jadhav, M. Tech. is an assistant professor, Department of Computer Science and Engineering, Rajarambapu Institute of Technology, Rajaramnagar, Sakharale, Maharashtra, India. She is a Ph.D. scholar in computer engineering at Drs. Kiran and Pallavi Global University, Vadodara. She is a dedicated academician with 13 years of experience in teaching and research. Currently pursuing her Ph.D. in computer engineering at Drs. Kiran and Pallavi Global University, Vadodara. Her research and academic interests span multiple domains, including deep learning, machine learning, image processing, human-computer interaction, and artificial intelligence. She has made significant contributions to her fields of interest, having published 20 research papers in national and international journals and conferences. She is an esteemed life member of the Indian Society for Technical Education (ISTE), reflecting her commitment to continuous learning and professional development.



Pragati Patil, M. Tech. is working as an assistant professor in the Department of Computer Science and Engineering (AI&ML), Rajarambapu Institute of Technology, Rajaramnagar, Sakharale, Maharashtra, India. She is a Ph.D. scholar in computer engineering at Drs. Kiran and Pallavi Global University, Vadodara. Prof. Pragati P. Patil has 07 years of experience in teaching and research. Currently pursuing her Ph.D. in computer engineering at Drs. Kiran and Pallavi Global University, Vadodara, she holds the position of assistant professor at the Department of Information Technology, Rajarambapu Institute of Technology, Rajaramnagar, Maharashtra. Her areas of interest in research and academics include deep learning, machine learning, and image processing. She has published 14 research papers in national and international publications and conferences, making major contributions to her fields of interest.

How to cite this paper

Priyanka N. Jadhav, Pragati P. Patil, "Smart Agriculture: Leveraging Machine Learning for Crop Recommendation, Fertilizer Optimization, and Yield Prediction", *International Journal of Intelligent Systems and Applications(IJISA)*, Vol.18, No.3, pp.115-125, 2026. DOI:10.5815/ijisa.2026.03.08