



Comparative Analysis of Machine Learning Approaches for Crop Recommendation in Sustainable Agriculture in India

Shivani Yadav¹, *Hari Kumar Singh², Akriti Garg³, Mohd Mustafa Khan⁴

^{1,2,3}Department of ECE, F.E.T, M.J.P. Rohilkhand University, Bareilly, India

⁴Department of CS&IT, F.E.T, M.J.P. Rohilkhand University, Bareilly, India

shivaniydv10802@gmail.com, *hks@mjprou.ac.in, akritistudy@gmail.com, mustafamohd178@gmail.com

Abstract—This study analyzes machine learning methods for crop recommendation using an agricultural dataset that contains soil nutrients (N, P, and K), pH value, temperature, humidity, and rainfall. Several machine learning models were used and assessed, including Decision Tree, Random Forest, Gradient Boosting, Extra Tree, Ada Boost, and XG Boost. The accuracy for XG-Boost is 99.60%, AdaBoost is 99.55%, Extra Trees is 99.40%, Random Forest is 99.32%, Decision Tree is 98.64%, and Gradient Boosting is 98.18%, according to experimental results. The findings show how ML-based methodologies can assist farmers in making data-driven crop selections, boosting output and promoting sustainable farming practices.

Keywords—Crop recommendation, Machine Learning, Sustainable Agriculture, Classification, Random Forest, Data-Driven Farming

1. INTRODUCTION

Agriculture is not only considered the backbone of India's economy but also employs nearly half of the nation's workforce, providing ample opportunities in rural livelihood [1]. Although its share in Gross Domestic Product (GDP) has gone down with continuous industrialization, the sector still holds great importance for ensuring food security, providing raw material to industries and enhancing exports [2].

Some of the recent developments in computational intelligence have shown great promise, especially in Machine Learning (ML) and Deep Learning for precision agriculture and data-driven decision-making. This work proposes an advanced Python-based crop prediction system with multiple ML algorithms including Decision Tree, Naïve Bayes, Random Forest, K-Nearest Neighbor (KNN), AdaBoost and SVM, while also using a DL-based Multilayer Perceptron model. The proposed system analyses various agricultural parameters such as nutrients in the soil, pH of the soil, temperature, rainfall, and humidity to improve the accuracy in the prediction. By utilizing these models, the system offers trustworthy suggestions for the best crop choices and resource distribution, increasing agricultural productivity, facilitating sustainability, and guaranteeing that India meets its long-term economic and environmental objectives [3]. When deciding which crops to plant, farmers in many regions of India still primarily rely on traditional knowledge, experience and intuition. Such data may be useful, but it frequently ignores rapidly shifting environmental conditions, deteriorating soil health or market demand, which results in lower crop yield and increased economic risk.

According to a recent study, decision-making can be significantly improved by data-driven crop recommendation system that use soil parameters (such as nitrogen, phosphorus, potassium and pH), meteorological and climate data and economic factors [4]. As an illustration of the significance of time-series data in recommendation, 10-fold cross-validation yield very high accuracy (Random Forest ~99.96%) but performance decreases when temporal ordering (real-world scenario) is respected [5]. Similarly, farmers' adoptions of predictive system are hampered by their lack of access to trustworthy data, digital tools and algorithmic transparency. The infrastructure constraints like data scarcity, data quality and lack of sensor deployment undermine the utility of ML-based system in rural settings [6].

Precision agriculture (PA) is revolutionizing traditional farming by substituting information for intuition at every level of decision-making. Optimizing crop selection is one of the main uses of PA and it is essential for sustainable farming, risk reduction and yield maximization. Because most conventional procedures are based on experience, they lack complete and dynamic soil fertility, climate change, and market demand. However, depending on variables like temperature, humidity, rainfall, pH and soil nutrient profiles, PA uses a prediction algorithm to choose which crops are most appropriate for a particular plot. Using ML and data-driven methods, precision crop prediction searches for intricate connections between the nutritional, climatic and environmental characteristics of the soil. Algorithms like K-Nearest Neighbor, Random Forest, Support Vector Machines and Decision Trees, Naïve Bayes and ensemble approaches are used to analyze large agricultural datasets to identify pattern that may enhance crop recommendation with better yield prediction. PA which integrates past and present data to increase

© The Author(s) 2026

R. Vasanth Kumar Mehta et al. (eds.), *Proceedings of the International Conference on Intelligent Systems for a Sustainable Future (ISSF 2026)*, Atlantis Highlights in Intelligent Systems 16,

https://doi.org/10.2991/978-94-6239-693-7_57

resource efficiency, decrease input waste and enable farmers to make adaptable and ecologically conscious decision is a major step towards more intelligent and resilient agriculture.

Recent studies have shown that machine learning holds promise in this area. To provide trustworthy crop selection recommendations, Random Forest, Support Vector Machines, and ensemble algorithms will be applied to multidimensional agricultural datasets [8]. When these models take into account weather, land history, and soil nutrients like N, P, and K, they perform more accurately and consistently than a single algorithm approach. An optimal ensemble learning framework for using numerous algorithms to make strategic decisions in a range of field scenarios has been created by additional research in this field [9].

Another significant advancement in agriculture is the application of Explainable AI, which helps farmers comprehend why specific crops are recommended to them, fostering trust and increasing their adoption [10]. IoT-based sensors provide real-time data for continuous crop recommendation updates based on soil and weather conditions when combined with PA platform. In fact, these technologies are most effective in regions where severe resource shortages and climate variability have been significant problems.

PA plays a crucial role in crop selection optimization by combining data analytics, predictive modeling, and real-time sensing. Given that it provides farmers with useful insights, boosts resource efficiency, reduces risks, and fosters long-term sustainability, it is a crucial part of contemporary agriculture [11].

Crop recommendation has emerged as one of the major focusses related to precision agriculture whereby ML models are applied to predict the most appropriate crop to grow based on specific soil and climatic conditions. Amongst the various algorithms tested on agricultural dataset, Logistic Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), Naïve Bayes, K-Nearest Neighbors (KNN) and Gradient Boosting are most in use. Studies have confirmed that tree-based ensemble algorithms like Random Forest often outperform traditional classifiers because of their capabilities for capturing nonlinear feature interaction and handling heterogeneous agricultural data [12]. Decision Trees are simple to interpret but they usually have low accuracy compared to ensemble model [13]. Support Vector Machines and Logistic Regression perform well on linearly separable datasets whereas KNN may face scalability issues in large datasets [14].

Furthermore, recent research has shown that even Naïve Bayes classifiers typically exhibit competitive accuracy when applied to a well-balanced dataset, outperform complex models and making them suitable for rapid deployment [15]. However, the Random Forest and Gradient Boosting models consistently demonstrated superior predictive accuracy and noise robustness under a variety of agricultural conditions, making them the best model for crop recommendation. According to prior machine learning-based crop recommendation, algorithms like Random Forest, Gradient Boosting and ensemble models are helpful for making the best crop prediction based on soil nutrient, climate and environmental characteristics.

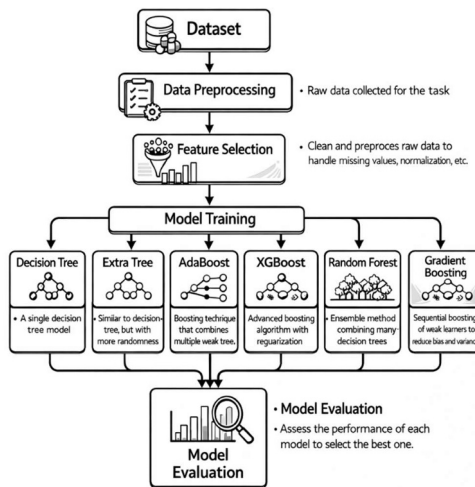


Fig. 1. Block Diagram of Crop Recommendation Model.

When compared to conventional techniques, accuracy is greatly improved. This led to the robustness and decision support which improves yield prediction and sustainability in precision agriculture.

2. METHODOLOGY

The fig 1 shows a typical machine learning workflow from raw data to model evaluation. It starts with a dataset, followed by data pre-processing and feature selection to clean and prepare the data. The prepared data is then used to train different tree-based models such as Decision Tree, Random Forest, Extra Trees, AdaBoost, and XG-Boost. Finally, model evaluation compares their performance to select the most effective model for the task.

2.1 Dataset Description

In the Precision Agriculture (PA) framework, soil nutrients nitrogen (N), phosphorus (P) and potassium (K) as well as soil parameters like pH and meteorological variables like temperature, humidity, and rainfall—are the main inputs used to predict the best crop. These components comprise the input features. Examples of the anticipated crop type that is the output include rice, maize, chickpeas, kidney beans, moth beans, mung beans, lentils, bananas, mangoes, grapes, watermelon, apples, oranges, papayas, and other crops. 2,200 examples are included in the dataset, which is arranged in rows with input and output variables indicated in 8 columns [21].

2.2 Data Pre-processing

To guarantee quality and data consistency, the dataset is preprocessed. To avoid bias and cut down on noise in the inconsistencies, learning process, and missing values were found and eliminated. To maintain uniformity throughout all numerical attributes, feature scaling is performed using standardization, confirming that variables with different units and ranges equal involvement to the model. Following preprocessing, the dataset was divided into two parts training and testing subsets in an 80:20 ratio to enable accurate model training and unbiased performance evaluation.

2.3 Model Development

In this work six machine learning methods Random Forest [17], Decision Tree [13], Gradient Boosting [14], AdaBoost [19], Extra Trees [18], and XG-Boost [20] is applied to crop prediction. The preprocessed training of the dataset was used to train every model. To improve the model performance and free from overfitting, hyper-parameters were adjusted using Parameter Grid Optimization and cross-validation. The generalization capacity and prediction accuracy of the modified models were then assessed using the testing dataset.

2.4 Performance Evaluation

The testing dataset is used to find the generalization capabilities and performance of each improved model. To examine the anticipated efficacy of each model, Classification performance indicators were employed, including accuracy, precision, recall, and F1-score. The results made it possible to compare individual tree-based and ensemble-based different techniques. Based on the model's overall performance across evaluation metrics, the Precision Agriculture framework identifies the model best way for crop prediction.

2.5 Interpretation of the Result

The output of each method was compared for accuracy and robustness. Then, by considering data-driven decision making in agricultural practices, the optimal model for crop recommendation was identified.

3 MODEL TRAINING

Machine learning (ML) techniques can help farmers choose crops based on data, increasing productivity and encouraging sustainable farming methods. Each selected model was trained on the designated training dataset in order to ascertain the mapping between the input parameters and the associated Crop Recommendation. Below is a list of each model's working principles and complete description.

3.1 Decision Tree

It works by breaking a dataset into smaller subsets while forming a tree-like structure of decisions. Each decision is based on answering a simple question about the data [13]. Each internal node represents a condition on a feature, each branch represents an outcome, and each leaf node gives the final prediction. Decision trees are easy to

interpret and visualize which makes them useful for understanding decision-making logic. However, they can over fit the data if not properly pruned or regularized. Entropy ($H(S)$) is a measure of impurity or uncertainty in a dataset S .

$$H(S) = - \sum (p_i * \log_2(p_i)) \quad (1)$$

Where p_i is the proportion of samples belonging to class i in set S . Information Gain ($IG(S, A)$) The reduction in entropy achieved by splitting a dataset S on an attribute A .

$$IG(S, A) = H(S) - \sum ((|S_v| / |S|) * H(S_v)) \quad (2)$$

Where, S_v is the subset of S for which attribute A has value v , and $|S|$ and $|S_v|$ are the number of samples in S and S_v respectively. Expected Value (EV): Used in decision analysis to calculate the weighted average of potential outcomes.

$$EV = \sum (\text{Probability_outcome} * \text{Value_outcome}) \quad (3)$$

3.2 Random Forest

Random Forest is an ensemble machine learning algorithm used for classification and regression that builds multiple decision trees during training. Each tree is trained on a random subset of the data and features, which reduces overfitting and improves generalization. The final prediction is obtained by majority voting (for classification) or averaging (for regression) across all trees. Random Forest successfully manages huge datasets with high-dimensional characteristics and is resilient to noise. But compared to a single decision tree, it is harder to understand [17]. For a regression problem, the final predicted value $\hat{Y}(x)$ for an input x is the average (mean) of the predictions from B individual trees.

$$\hat{Y}(x) = \frac{1}{B} \sum_{b=1}^B \hat{y}_b(x) \quad (4)$$

3.3 Gradient Boosting

It is an ensemble machine learning method for regression and classification that builds models sequentially. Gradient descent is used to minimize a loss function to train each new model to correct the mistakes made by the earlier models. This method produces high predictive accuracy by concentrating more on samples that are challenging to predict. Complex patterns and interactions in data can be captured using gradient boosting. However, it is computationally intensive and prone to overfitting if not properly tuned [14]. The model is updated iteratively and is formulated as follows:

$$F_M(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (5)$$

Where:

$F_M(x)$: Final prediction of the model after M iterations. M : Total number of weak learners (trees). $h_m(x)$: Weak learner (usually a decision tree) at iteration m . γ_m : Weight/step size of the m th weak learner. x : Input feature vector.

3.4 Extra Trees

Similar to Random Forest, Extra Trees builds many trees, but introduces more randomness by choosing split points randomly, not optimally. This high randomization reduces model variance and computation time, providing highly accurate and generalized classification [18]. Mathematically, for a given feature j and its range $[a_j, b_j]$, a split threshold s might be drawn from a uniform distribution. Random threshold selection:

$$s \sim \text{Uniform}(a_j, b_j)$$

Where:

s : Random split threshold for feature j . a_j, b_j : Minimum and maximum values of feature j . The final prediction (x) of an Extra Tree model (for regression) is the average of predictions from M individual trees:

$$H(x) = \frac{1}{M} \sum_{m=1}^M h_m(x) \quad (6)$$

$H(x)$: Predicted output for input x . M : Number of trees in the ensemble. $h_m(x)$: Prediction of the m th tree.

3.5 AdaBoost

It sequentially trains weak classifiers (e.g., shallow trees) and heavily weights the training samples that were misclassified by previous learners. This adaptive process forces subsequent classifiers to focus on challenging data, combining the weighted votes of all learners for a robust final recommendation.[15] [19]This equation calculates the weight (α_t) of a weak classifier (h_t) based on its total error (ϵ_t):

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1-\epsilon_t}{\epsilon_t} \right) \tag{7}$$

The final prediction is a weighted sum of all weak classifiers:

$$H(x) = \text{sign}(\sum_{t=1}^T \alpha_t h_t(x)) \tag{8}$$

3.6 XG-Boost

XG-Boost (Extreme Gradient Boosting) is an advanced ensemble learning algorithm based on the gradient boosting framework. It builds decision trees sequentially while optimizing a regularized objective function to reduce overfitting. XG-Boost is known for its high speed and performance due to features such as parallel processing and efficient handling of missing values. It is widely used in classification and regression problems because of its strong predictive accuracy. However, it requires careful hyper-parameter tuning to achieve optimal results [20]. It consists of two main parts the Training Loss and the Regularization Term.

$$\text{Obj}(T) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{9}$$

XG-Boost uses a second-order Taylor expansion to approximate the loss function, resulting in an objective that can be optimized for each new tree fit.

$$\widetilde{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \tag{10}$$

Where, g_i and h_i are the first and second order derivatives (gradient and hessian) of the loss function, respectively. Regularization Term: This term penalizes model complexity.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \tag{11}$$

Where T is the number of leaf nodes, w_j is the score of leaf j , γ is a penalty for the number of leaves, and λ is an L2 regularization parameter for the leaf weights. Optimal Leaf Weight After the approximation and some algebraic manipulation, the optimal weight w_j^* for a specific leaf node j is found to be:

$$w_j^* = - \frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{12}$$

Where, I_j is the set of instances in leaf j .

4 EXPERIMENTAL RESULTS

All machine learning models produced extremely good classification results, with an overall accuracy of above 98%, according to the performance comparison table. With the highest accuracy (99.60%), precision (99.62%), recall (99.61%), and F1-score (99.61%) of any model, XG-Boost outperformed the others, demonstrating exceptional predictive power and well-rounded performance. Ada-Boost came in second, with somewhat worse but still excellent results (accuracy 99.55%). With accuracy exceeding 99%, Extra Trees and Random Forest both showed excellent ensemble learning performance; Extra Trees outperformed Random Forest by a small margin. The decision tree model performed rather poorly (accuracy of 98.64%), demonstrating the benefit of ensemble approaches over a single tree model. Despite having the lowest metrics of all the models on the list (98.18% accuracy), Gradient Boosting continued to perform well in terms of prediction. Based on all evaluation metrics, ensemble boosting techniques in particular, XG-Boost and Ada-Boost proved to be the most successful models overall as shown in Table 1.

Table1 Showing Performance of all Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
XG-Boost	99.60	99.62	99.61	99.61
Ada-Boost	99.55	99.56	99.55	99.55
Extra Trees	99.40	99.42	99.38	99.39
Random Forest	99.32	99.35	99.30	99.30
Decision Tree	98.64	98.70	98.60	98.60
Gradient Boosting	98.18	98.20	98.10	98.10

The histogram analysis of soil nutrients such as N, P and K, pH, temperature, humidity and rainfall, shows their distribution pattern across the data set. While rainfall has a broader dispersion, temperature and humidity follow seasonal fluctuations, nitrogen, phosphate, and potassium exhibit modest variability, and pH levels stay within an acceptable farming threshold as shown in figure2.

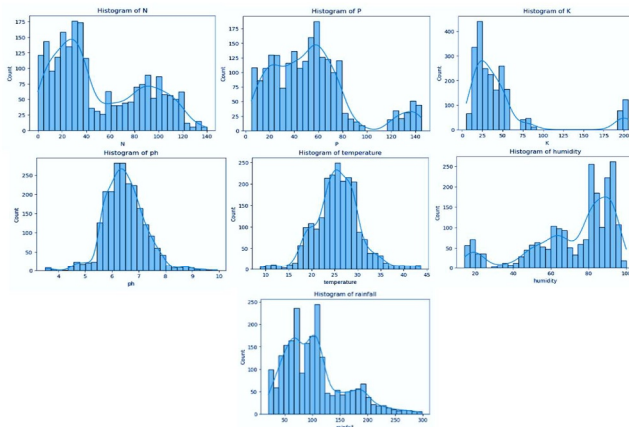


Fig. 2. Histogram of different soil nutrients (N, P, K), pH value, temperature, humidity and rainfall.

The graph in Figure 3 clearly ranks the models based on performance, with XG-Boost scoring highest and AdaBoost and Extra Trees following closely behind. It has been confirmed that XG-Boost is the best-performing algorithm for this crop recommendation task. The superior predictive power of complex, regularization-aware boosting techniques is demonstrated by the clear distinction between the top ensemble models XG-Boost and AdaBoost and the simpler Decision Tree and Gradient Boosting models.

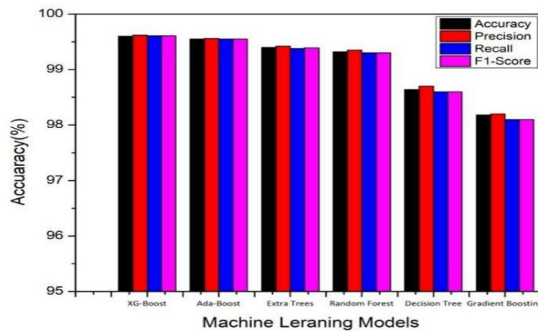


Fig. 3. Performance of Advance Tree- Based ML Models

Each of the four metrics (Accuracy, Precision, Recall, and F1-Score) has bars that are nearly the same length. This indicates that there is no appreciable trade-off between false negatives (Recall) and erroneous positives (Precision), highlighting the models' superior robustness and balance. To effectively distinguish the performance of the top-tier models—which is essential when all scores are close to perfection—the x-axis is zoomed in from 98% to 100%

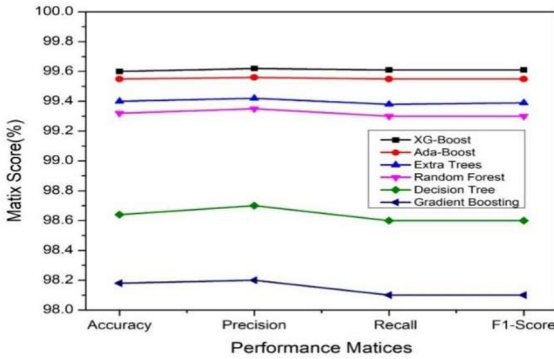


Fig. 4. Enhanced Performance Consistency Across Metrics

The data labels XG-Boost in blue and Gradient Boosting in brown are added to the points for the top and bottom-performing models, respectively. The very high and flat consistency of XG-Boost (approximately 0.996) contrasts with the lower, flatter scores of Gradient Boosting (approximately 0.981), as shown in Fig. 4.

The comparative analysis in Table 2 demonstrates that the performance of the proposed model is consistent with and slightly higher than recent studies in crop recommendation systems. Prior research such as Dey et al. (2024) and Hasan et al. (2023) reported Random Forest and ensemble frameworks achieving accuracies around 99.2–99.3%. Similarly, Afzal et al. (2025) and Prity et al. (2024) demonstrated the significance of soil parameters in conjunction with 1`lboosting algorithms, attaining 98–99% accuracy. In contrast, the present study achieves a highest accuracy of **99.60% using XGBoost**, along with highly balanced precision, recall, and F1-score values.

Table2 Comparison with Previous Studies on Crop Recommendation

Study	Dataset Size	Algorithms Used	Best Model	Reported Accuracy (%)	Key Observation
(Dey et al., 2024)	NPK + Climate variables	RF, SVM, KNN	Random Forest	~99.2	Ensemble models outperform single classifiers
(Hasan et al., 2023)	Multidimensional agricultural dataset	Ensemble ML	Ensemble Framework	~99.3	Improved robustness using ensemble learning
(Afzal et al., 2025)	Soil + ML integration	RF, GB	Random Forest	~98–99	Soil information significantly improves prediction
(Prity et al., 2024)	Agricultural dataset	GB, RF	Gradient Boosting	~98.7	Boosting improves nonlinear pattern learning
Proposed Work (2026)	2200 samples (N, P, K, pH, Temp, Humidity, Rainfall)	DT, RF, GB, Extra Trees, AdaBoost, XG-Boost	XG-Boost	99.60	Regularized boosting with hyperparameter tuning gives highest balanced performance

5 CONCLUSION

The study's findings, which highlights the significance of machine learning in precision farming (PA), are very consistent with earlier conversations and investigations. Several recent studies have shown that data-driven models can substantially outperform experience-driven or traditional rule-based crop selection methods when significant and soil and environmental factors are considered. This study focuses on integrating factors such soil nutrients (N, P, and K), pH, humidity, temperature, and rainfall these 8 results in a very relevant dataset for accurate crop classification, which is

uniform with other research. Results from previous comparative studies, where ensemble techniques often outperformed single classifiers like KNN, Decision Trees, or Naïve Bayes, are supported by the exceptionally better performance of ensemble learning algorithms, especially AdaBoost, XG-Boost and Extra Trees. Researchers attribute this improved performance to ensemble learning methods capacity to manage nonlinear relationships in agricultural datasets while optimizing the bias–variance balance. The slight accuracy changes of the study's best ensemble models offer more proof of the approach's stability and adaptability across a variety of data distributions.

ACKNOWLEDGEMENT

Through research grant number SUR/2022/000940 under the SERB-SURE initiative, Anusandhan National Research Foundation (previously Science and Engineering Research Board).

REFERENCES

1. P. Sharma, P. Dadheech, N. Aneja and S. Aneja, "Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning," in *IEEE Access*, vol. 11, pp. 111255-111264, 2023.
2. Kothamasu Venkata Jaya Saiteja, Uday Kiran Kasi, 2024, Crop Prediction Model using Machine Learning and Deep Learning Methods, *International Journal of Engineering Research & Technology (IJERT)* Volume 13, Issue 11 (2024)
3. Afzal, H., Amjad, M., Raza, A. *et al.* Incorporating soil information with machine learning for crop recommendation to improve agricultural output. *Sci Rep* 15, 8560(2025) <https://doi.org/10.1038/s41598-025-88676-z>
4. Prity, F.S., Hasan, M.M., Saif, S.H. *et al.* Enhancing Agricultural Productivity: A Machine Learning Approach to crop Recommendations Hum-Cent Intell Syst 4,497–510 (2024) <https://doi.org/10.1007/s44230-024-00081-3>
5. Sam, Steven, and Silima Marshal DAbreo. "Crop recommendation with machine learning: leveraging environmental and economic factors for optimal crop selection." *arXiv preprint arXiv:2505.21201* (2025).
6. Rahul Satheesan Nair, Shreeram Sanjay Sawant, Dr. Ujjwala V. Gaikwad, "Smart Crop Selection: A Machine Learning-Based Decision Support System for Farmers", *Journal for Research in Applied Science and Engineering Technology* 2025 DOI Link: <https://doi.org/10.22214/ijraset.2025.72308>
7. Pokhariyal, S., Patel, N.R. and Govind, A., "Machine learning-driven remote sensing applications for agriculture in India: A systematic review. *Agronomy*", 13(9), p.2302.
8. B. Dey, J. Ferdous, R. Ahmed, "Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables," *Heliyon*, vol. 10, no. 3, e25112, Feb. 2024. doi:10.1016/j.heliyon.2024.e25112
9. M. Hasan, M. Marjan, P. Uddin, M. Ibn Afjal, S. Kardy, S. Ma, Y. Nam, "Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation," *Frontiers in Plant Science*, vol. 14, 1234555, 2023. doi:10.3389/fpls.2023.1234555
10. Gunasekaran, D. Kanmani, R. Krishnamoorthi, "Optimized ensemble learning for enhanced crop recommendations: Leveraging ML for smarter agricultural decision-making," *Engineering Proceedings*, vol. 82, 2024. doi:10.3390/ecsa-11-20366
11. S. Khatibi, J. Ali, "Harnessing the power of machine learning for crop improvement and sustainable production," *Frontiers in Plant Science*, vol. 15, 1417912, 2024. doi:10.3389/fpls.2024.1417912
12. Kumar, C., Dhillon, J., Huang, Y., Reddy, K.N., 2025. Explainable machine learning models for corn yield prediction using uav multispectral data. *Computers and Electronics in Agriculture* 231, 109990.URL: <https://doi.org/10.1016/j.compag.2025.109990>, doi:10.1016/
13. Kandiwal, P., Kulkarni, S., Shinde, S., 2020. Crop yield prediction using data mining techniques: A survey. *International Journal of Advanced Science and Technology* 29, 1530–1536.
14. Gunasekara, N., Pfähringer, B., Gomes, H.M., Bifet, A., 2024. Gradient boosted trees for evolving data streams. *Machine Learning* 113, 3325–3352. URL: <https://doi.org/10.1007/s10994-024-06517-y>, doi:10.1007/s10994-024-06517-y.
15. Feng, D.C., Liu, Z.T., Wang, X.D., Chen, Y., Chang, J.Q., Wei, D.F., Jiang, Z.M., 2020. Machine learning-based compressive strength prediction for concrete: An adaptive boosting approach. *Construction and Building Materials* 230, 117000.
16. Mienye, I.D., Jere, N., 2019. A survey of decision trees: Concepts, algorithms and applications. *IEEE Access* 7, 151962–151989. doi:10.1109/ACCESS.2019.2945232.
17. Farhadi, Z., Bevrani, H., Feizi-Derakhshi, M.R., Kim, W., Ijaz, M.F., 2022. An ensemble framework to improve the accuracy of prediction using clustered random-forest and shrinkage methods. *Applied Sciences* 12, 10608. URL: <https://doi.org/10.3390/app122010608>, doi:10.3390/app122010608.
18. Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. *Machine Learning* 63, 3–42. doi:10.1007/s10994-006-6226-1. the foundational algorithm description.
19. Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. <https://doi.org/10.1006/jcss.1997.1504>
20. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. <https://doi.org/10.1145/2939672.2939785>
21. Suyash Gholap, Harsh Singh, Rohan Rajesh, Shweta Barshe, "Crop Recommendation", *IEEE Dataport*, February 11, 2024, doi:10.21227/3ays-5h79

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

