Vol. 7 No. 1, 2026, pp. 16-25 **Jurnal Pepadun**

Model Comparison and Feature Selection for Crop Recommendation

[*1]Harnan Malik Abdullah and [2]Imam Fahrurrozi

[1]Dept. of Creative and Digital Industries, Faculty of Vocational Studies, Universitas Brawijaya, Malang, Indonesia
[2]Dept. of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Yogyakarta, Indonesia
e-mail: [*1]**harnan_malik@ub.ac.id** , [2]imam.fahrurrozi@ugm.ac.id

Abstract *- Selecting crops appropriate to soil and environmental conditions is a crucial component of decision-making in smart agriculture. This study aims to evaluate the effect of feature selection on the predictive performance, stability, and computational efficiency of several machine learning models in crop recommendation tasks. The dataset employed in this study is a publicly available crop recommendation dataset, encompassing attributes such as nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall. Six distinct models were evaluated: Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, XGBoost, and LightGBM, under two distinct conditions: utilizing all available features and employing a selection of features. The models' performance was assessed through various metrics, including accuracy, precision, recall, F1-score, the mean and standard deviation of cross-validation accuracy, as well as the inference time per sample. Random Forest outperformed other models, achieving high accuracies (0.993–0.995) across both full and selected feature scenarios. The model input was simplified with minimal performance impact by the feature selection, which left the temperature and pH unselected. These results indicate that environmental factors, in addition to soil nutrients, substantially affect crop recommendations. Consequently, this research underscores that the evaluation of models for crop recommendation should prioritize not only accuracy but also stability, inference efficiency, and feature relevance to facilitate practical application within smart agricultural systems.*

Keywords**:** *Crop Recommendation; Feature Selection; Machine Learning; Smart Agriculture.*

1. INTRODUCTION

Agriculture remains a crucial sector for ensuring food security, efficient resource utilization, and economic sustainability, particularly in developing countries. As pressures from population growth, climate change, and limited land resources increase, the agricultural sector is required to transform toward more precise, data-driven systems [1][2][3][4][5][6]. In this context, smart agriculture has evolved as an approach that integrates soil, weather, sensor, and intelligent computing data to improve the quality of cultivation decision-making [1][7][8][9][10].

One key issue in smart agriculture is crop recommendation, which involves determining the most suitable crop types based on land and environmental conditions. Various studies have shown that attributes such as nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall can be used to predict the most suitable crops under specific conditions [11][12][13][14][15]. Research by Dey *et al*. demonstrated that models such as Support Vector Machine, Random Forest, K-Nearest Neighbors, and XGBoost can provide very high performance in crop recommendation tasks based on agro-environmental attributes [11]. Similar findings were also reported by Prity *et al.* and Senapaty *et al.*, who showed that ensemble and boosting-based models are very competitive on agricultural tabular data [12][13].

However, some previous research still focuses on the application of a single model or on reporting accuracy as the primary metric. This approach does not fully address the needs of more practical implementations. The most accurate model might not be the most practical choice if computational costs are high, stability is a problem, or the number of required features is too large. Recent studies emphasize the importance of a more thorough analysis, including comparisons between models, the relevance of features, how efficiently inferences are made, and how easy it is to understand the results [12][14][15][16][17][18].

 | doi: 10.23960/pepadun.v7i3.345

In addition, research using public crop recommendation datasets has moved toward more complex models and more practical systems. For example, Islam *et al.* combined IoT-based soil nutrient monitoring with crop recommendation [16], while Bouni *et al.* highlighted the importance of a more understandable approach to crop recommendation systems [17]. Aryanti *et al.* also demonstrated that the Gradient Boosting approach can achieve high accuracy on a similar dataset, but their focus remains on a single primary model [18]. Thus, there is still room for research to answer the more fundamental question of which model is most appropriate when considering accuracy, stability, and efficiency simultaneously.

Based on this background, this study focuses on comparing several machine learning models for crop recommendation tasks by evaluating two scenarios: the use of all features and the use of selected features. In contrast to single-model studies, this research evaluates predictive performance, stability, and efficiency simultaneously. Thus, the contribution of this study lies in providing a comprehensive evaluation framework that balances predictive performance, model stability, and computational efficiency. This research moves beyond identifying the model by analyzing the trade-offs between input complexity and operational reliability, offering broader insights for developing lightweight and cost-effective smart agricultural systems.

2. RESEARCH METHODOLOGY

This investigation employs a quantitative experimental methodology to assess the efficacy of various machine learning models in crop recommendation scenarios. The dataset under examination is a publicly available crop recommendation dataset, encompassing 2,200 instances, seven numerical attributes, and 22 distinct crop classes [18][19][20]. The attributes incorporated in the analysis are nitrogen, phosphorus, potassium, temperature, humidity, pH, and rainfall, with the target label representing the recommended crop type. This particular dataset was selected due to its prevalent utilization in crop recommendation research and its provision for open replication of the experimental procedures [11][12][16].

The experimental phase begins with an assessment of data quality, which involves checking for missing values, duplicate entries, and how the classes are distributed. Next, the data is separated into features and targets, while the target labels are converted to numeric format using label encoding. The data is then divided into training and test data with a ratio of 80:20 [18][21]. Because some models are sensitive to data scale, standardization was performed using StandardScaler on Logistic Regression, K-Nearest Neighbors, and Support Vector Machine [21][22].

To reduce input complexity and evaluate feature relevance, this study implemented feature selection using *SelectKBest* based on *mutual_info_classif*. Statistical relevance-based feature selection is widely used to assess the contribution of individual attributes to the target and support more concise models [17][23]. In this experiment, five selected features were used in the second scenario, while the first scenario retained all seven original features. The selection of k=5 features was based on the distribution of Mutual Information (MI) scores, where a clear gap was observed between the fifth feature (Nitrogen) and the sixth (Temperature). By selecting the top 5 features namely humidity, rainfall, K, P, and N, this study aimed to evaluate a significantly more compact model (reducing input by ~28%) while still retaining the most informative agro-environmental variables for classification.

This investigation compared six different models namely Logistic Regression, K-Nearest Neighbors, Support Vector Machine, Random Forest, XGBoost, and LightGBM. Linear, distance-based, kernel-based, ensemble, and boosting methodologies were selected for representation. XGBoost and LightGBM were incorporated due to their established efficacy in handling tabular data and their prevalent application in predictive modeling across diverse domains [24][25]. Model efficacy was assessed through accuracy, precision, recall, and F1-score metrics, utilizing the test dataset. Additionally, cross-validation accuracy mean and standard deviation were used to assess the models' ability to generalize. Unlike many previous studies that use total execution time, this study separates training time and inference time per sample to make model efficiency more meaningful for implementation contexts [15][24][25][26]. Model stability was rigorously assessed using Stratified 5-Fold Cross-Validation. This strategy was chosen to ensure that each fold maintains the same proportion of the 22 crop classes as the original dataset, providing a more reliable estimate of the models'

© 2026 The Authors | doi: 10.23960/pepadun.v7i3.345

generalization ability across different data partitions. Table 1 exhibits the summary of the experimental configuration setup.

Table 1. Summary of experimental setup.

Components	Configuration
Dataset	Crop Recommendation Dataset
Number of instances	2200
Number of initial features	7
Number of classes	22
Data splitting	80% train, 20% test
Cross-validation	Stratified 5-Fold
Feature selection	SelectKBest with mutual information
Number of selected features	5
Model	LR, KNN, SVM, RF, XGBoost, LightGBM
Evaluation metrics	Accuracy, Precision, Recall, F1-score, CV Accuracy, Inference Time
Experimental environment	Python on Google Colab

To ensure reproducibility, all models were implemented using scikit-learn, XGBoost, and LightGBM libraries with consistent hyperparameters. For instance, Random Forest was configured with 300 estimators and Gini impurity, while XGBoost and LightGBM used a learning rate of 0.1 and a maximum depth of 6. These sensible defaults were chosen to provide a fair baseline for comparison. The experiments were conducted in a Python environment on Google Colab, utilizing standard libraries such as Scikit-Learn for traditional models and the official XGBoost and LightGBM packages for boosting methods.

3. RESULTS AND DISCUSSION

3.1. Predictive Performance on All Features

Experimental results on the all-feature scenario are depicted in Tabel 2. The results show that all models have high predictive performance. This indicates that the soil and environmental attributes in the dataset have strong representation of crop classes. In contrast, the models showed some differences. In this experiment, the Random Forest model performed the best when all features were included, followed by XGBoost and LightGBM. Conversely, Logistic Regression yielded the least favorable results, albeit within an acceptable range. This observed trend implies that the correlation between agro-environmental characteristics and crop classifications is not exclusively linear. Therefore, decision tree-based and ensemble models are more effective in capturing nonlinear interactions between variables. This finding aligns with previous research that also demonstrated the dominance of ensemble and boosting models in crop recommendation tasks [11][12][13][14][15][19][26].

Table 2. Experimental results on all feature scenarios.

Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9932	0.9935	0.9932	0.9932
XGBoost	0.9909	0.9915	0.9909	0.9908
LightGBM	0.9886	0.9894	0.9886	0.9886
SVM	0.9841	0.9856	0.9841	0.9840
KNN	0.9795	0.9804	0.9795	0.9793
Logistic Regression	0.9727	0.9740	0.9727	0.9725

The superior performance of Random Forest can be attributed to its ensemble architecture, which effectively captures the complex, non-linear dependencies between agro-environmental features and crop classes. Unlike

 | doi: 10.23960/pepadun.v7i3.345

linear models that struggled with the multi-dimensional nature of the dataset, Random Forest's bootstrap aggregating (bagging) mechanism ensures high stability and robust generalization, as evidenced by its low cross-validation standard deviation [27][13].

3.2. Feature Selection Results

Feature selection was performed using the *SelectKBest* method with a mutual information score function to identify features with the strongest relationship to plant labels. This approach was used to evaluate whether a more compact feature subset could maintain model performance while improving computational efficiency. The mutual information score results are shown in Table 3.

Table 3. Mutual information score as the result of *SelectKBased.*

No.	Feature	MI Score
1.	**humidity**	**1.732005207**
2.	**rainfall**	**1.632814726**
3.	**K**	**1.631451886**
4.	**P**	**1.296750167**
5.	**N**	**1.014665380**
6.	temperature	0.995991162
7.	ph	0.709289646

Based on the selection results, the five features with the highest scores were humidity, rainfall, potassium (K), phosphorus (P), and nitrogen (N). Temperature and pH features were not included in the selected subset because they have relatively lower mutual information values compared to other features. The high relevance of humidity and rainfall suggests that climatic factors act as the primary physiological constraints for crop suitability in this dataset. This empirical evidence supports the integration of real-time environmental sensors as a priority over periodic soil sampling in smart farming frameworks. Based on the mutual information scores within this specific dataset, these results suggest that environmental factors, particularly humidity and rainfall, potentially play a more significant role in distinguishing plant types compared to the recorded soil nutrients. Therefore, the selected feature subset retains important information relevant to the classification task. The prioritization of environmental factors is empirically supported by the higher MI scores for humidity (1.73) and rainfall (1.63) compared to Nitrogen (1.01), as detailed in Table 3.

3.3. Predictive Performance on Feature Selection

Based on the selected feature subset, further experiments were conducted to evaluate the model performance on reduced feature scenarios. The performance of all models remained generally high as shown in the Table 4. This indicates that most of the key discriminatory information can still be retained even when the number of features is reduced. Random Forest remained the best model, with a very small performance decrease compared to the full-feature scenario. This finding indicates that feature selection does not cause significant performance distinction or degradation and also opens up opportunities for implementing more lightweight systems.

Table 4. Experimental results on selected feature scenarios.

Model	Accuracy	Precision	Recall	F1-score
Random Forest	0.9955	0.9957	0.9955	0.9955
XGBoost	0.9909	0.9915	0.9909	0.9910
KNN	0.9886	0.9890	0.9886	0.9887
LightGBM	0.9864	0.9866	0.9864	0.9863
SVM	0.9795	0.9821	0.9795	0.9794
Logistic Regression	0.9750	0.9771	0.9750	0.9749

 | doi: 10.23960/pepadun.v7i3.345

This result is important because many previous studies have emphasized model accuracy but have not explicitly compared full-feature and reduced-feature conditions systematically [18]. In this study, feature selection was shown to simplify input without substantially reducing prediction quality. Thus, the practical value of a model is measured not only by prediction accuracy but also by its ability to work with more efficient input structures.

3.4. Model Stability and Inference Efficiency

Model stability was observed through the cross-validation accuracy mean and cross-validation accuracy standard deviation as depicted in Figure 1 and Table 5, respectively. The results show that the best models, particularly Random Forest, XGBoost, and LightGBM, not only produced high accuracy but also had low performance variation.

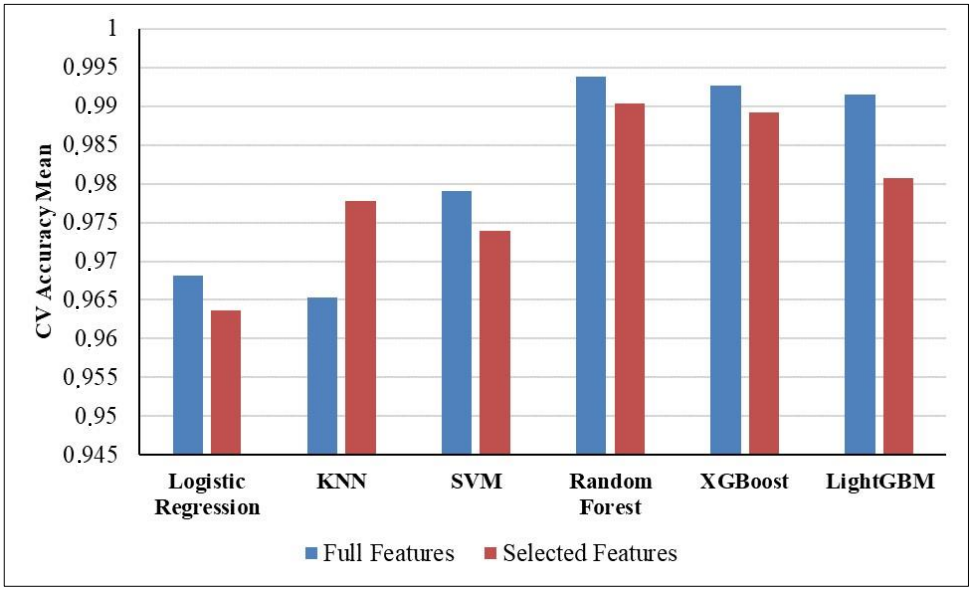

Figure 1. Cross-validation accuracy mean.

Table 5. Cross-validation accuracy standard deviation.

Model	CV Accuracy Standard Deviation	
	Full Features	Selected Features
Logistic Regression	0.0066	0.0116
KNN	0.0121	0.0092
SVM	0.0103	0.0133
Random Forest	0.0055	0.0080
XGBoost	0.0023	0.0055
LightGBM	0.0031	0.0055

Inference time evaluation was measured for each trained model, the results of which are shown in Figure 2. Logistic regression showed the shortest inference time, followed by KNN. The next shortest inference times were XGBoost, SVM, LightGBM, and Random Forest, which required the longest inference time. Selected features did not always show better inference time performance compared to full features. This was the case for the LightGBM and Random Forest models. While refers to the accuracy metric as well as the inference time, XGBoost indicate as the moderate model with the second high accuracy and the third inference time performance. However, a critical trade-off is observed. While Random Forest achieves the highest accuracy, it also exhibits the longest inference time per sample. This finding implies that for real-time applications on resource-constrained IoT devices, XGBoost might serve as a more balanced alternative, providing competitive accuracy with significantly higher computational efficiency.

 | doi: 10.23960/pepadun.v7i3.345

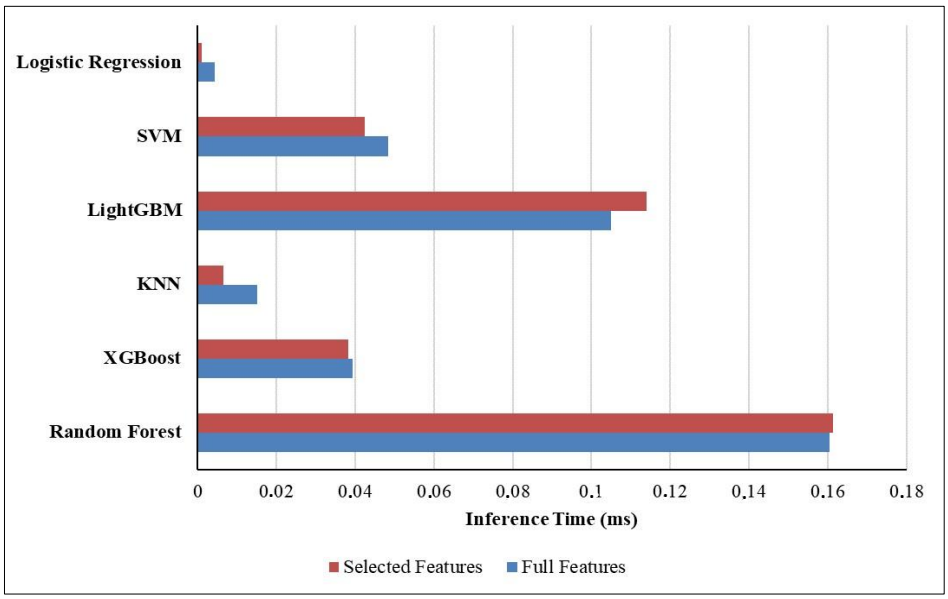

Figure 2. Inference time.

3.5. Comparison to Previous Works

In contrast to earlier studies that used a single Gradient Boosting model on a similar dataset [18], this research offers a different contribution. The focus of this research is not on creating a single recommendation system, but on a more rigorous comparative evaluation of multiple machine learning models, the impact of feature selection, and inference efficiency. Thus, this research broadens the discussion from simply *whether a model works* to *which model is most appropriate when performance, stability, and efficiency are simultaneously considered*. This direction is more aligned with the needs of real-world implementations such as in smart agriculture, which require not only accurate models but also lightweight, stable, and explainable models [26][28][29][30][31][32].

3.6. Practical Implications

The findings of this study offer several practical suggestions for improving crop recommendation systems used in smart agriculture. Initially, the observation that Random Forest demonstrates superior performance implies that ensemble-based models should be emphasized as the foundational element of decision support systems designed for agricultural tabular data. These models possess the capacity to effectively represent nonlinear associations among features and exhibit a degree of stability across various data partitions, thereby rendering them appropriate for recommendation systems that necessitate a high degree of reliability [11][12][13][14][15][19].

Second, the feature selection results indicate that crop recommendation systems can be built with more compact inputs without significant performance degradation. In an implementation context, this is crucial because field systems often face sensor limitations, data acquisition costs, and the need for rapid inference. By retaining only, the most relevant features, the system becomes simpler, more computationally efficient, and more realistic for implementation on digital agricultural platforms or lightweight computing devices [16][26][33][34].

Third, the dominance of rainfall and humidity features provides practical guidance that environmental factors, particularly rainfall and humidity, deserve primary attention in crop recommendation systems. This means that in real-world implementations, integrating weather data, local environmental data, or microclimate sensors can improve recommendation quality. This study provides a broader scientific insight: crop recommendation is more sensitive to dynamic environmental variables (humidity and rainfall) than to relatively stable soil chemical properties (N, P, K). This fundamental finding suggests that future smart agriculture frameworks

 | doi: 10.23960/pepadun.v7i3.345

should prioritize the integration of real-time weather sensors or microclimate data over frequent and costly soil sampling, thereby improving the sustainability and scalability of the system [3][8][16].

Overall, the practical implication of this research is that the development of a crop recommendation system should consider three aspects simultaneously: an accurate model, relevant features, and efficient inference. This approach is better suited to the needs of smart agriculture than systems that prioritize maximum accuracy without considering how easy they are to use.

3.7. Research Limitations

Although the results of this study demonstrate excellent performance, there are several limitations that should be noted. First, the dataset used was relatively clean, structured, and balanced across classes, so the experimental conditions tended to be more ideal than actual field conditions. This phenomenon can lead to strong experimental performance, but it might not translate to the same level of generalization when the model encounters data that is noisy, imbalanced, or affected by a wider range of environmental factors [1][4][11].

Additionally, the dataset used didn't include spatial or agroclimatic information, such as location coordinates, elevation, regional soil type, or climate zone classification. As a result, the model's recommendations are still based on conditions, not fully incorporating location. This means the recommendations are best understood as being based on a combination of numerical soil and environmental factors, rather than a system that specifically considers geographic differences [16][18].

Third, the feature selection in this study used only one approach, namely *SelectKBest*, based on mutual information. While this approach is representative enough to assess the relevance of individual features, the results are not necessarily identical to other feature selection approaches, such as recursive feature elimination, tree-based feature selection, or explainable AI-based methods. Therefore, the results of the selected features in this study should be understood as empirical results consistent with the approach used, not as a final conclusion regarding the universally best features [14][29].

Fourth, the present experiment did not encompass exhaustive hyperparameter optimization for all models under consideration. Instead, model configurations were selected based on sensible and consistent parameters, thereby facilitating equitable comparisons and simplifying the process of replication. As a result, the findings of this investigation are best interpreted as a comparative benchmark, rather than an exhaustive search for the optimal configuration for each individual model.

4. CONCLUSIONS

This study establishes that high-performance recommendation systems can be streamlined through strategic feature selection and model choice, balancing reliability with the operational constraints of smart farming. Beyond identifying Random Forest as a robust model, this research proves that input complexity can be reduced by 28% (from 7 to 5 features) with minimal impact on stability, providing a benchmark for the development of resource-efficient agricultural decision support systems. Furthermore, applying feature selection using *SelectKBest* demonstrated that using a more compact feature subset i.e., humidity, rainfall, potassium, phosphorus, and nitrogen, still maintained predictive performance with minimal degradation. These findings confirm that not all features are necessary for optimal performance, and therefore feature selection can be an important strategy for simplifying models without sacrificing accuracy.

From a practical perspective, the results of this study indicate that an effective crop recommendation system depends not only on accuracy but also on model efficiency and simplicity. A smaller number of features makes it more suitable for implementation in device-based smart agriculture systems or real-world operational environments. However, this study still has limitations because it used a relatively clean dataset and did not consider spatial context or more complex variations in field conditions. Therefore, future research can be directed at using more heterogeneous datasets, integrating location-aware attributes, and exploring a wider range of feature selection methods to improve model robustness. In addition, evaluation in real implementation

 | doi: 10.23960/pepadun.v7i3.345

environments, such as IoT systems or edge devices, is also an important step to ensure that the proposed model is truly ready for use in the smart agriculture ecosystem.

LITERATURE

[1] J. Botero-Valencia, V. Garcia-Pineda, A. Valencia-Arias, J. Valencia, E. Reyes-Vera, M. Mejia-Herrera, and R. Hernandez-Garcia, "Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives," Feb. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agriculture15040377.

[2] Food and Agriculture Organization of the United Nations, "The challenge," Rome, 2009. Accessed: Mar. 28, 2026. [Online]. Available: https://www.fao.org/fileadmin/templates/wsfs/docs/Issues_papers/HLEF2050_Global_Agriculture.pdf

[3] S. O. Araújo, R. S. Peres, J. C. Ramalho, F. Lidon, and J. Barata, "Machine Learning Applications in Agriculture: Current Trends, Challenges, and Future Perspectives," Dec. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/agronomy13122976.

[4] S. A. Frimpong, M. Han, W. Zheng, X. Li, E. Akpaku, and A. P. Obeng, "Machine and Deep Learning in Agricultural Engineering: A Comprehensive Survey and Meta-Analysis of Techniques, Applications, and Challenges," Oct. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/computers14100438.

[5] P. Velusamy, S. Rajendran, R. K. Mahendran, S. Naseer, M. Shafiq, and J. G. Choi, "Unmanned aerial vehicles (Uav) in precision agriculture: Applications and challenges," Jan. 01, 2022, *MDPI*. doi: 10.3390/en15010217.

[6] A. Indryani, U. Khaira, and M. F. Putri, "Rainfall Prediction Using Long Short-Term Memory Method (Case Study: Jambi City)," *Jurnal Pepadun*, vol. 6, no. 1, pp. 57–70, Apr. 2025, doi: 10.23960/pepadun.v6i1.256.

[7] L. Benos, A. C. Tagarakis, G. Dolias, R. Berruto, D. Kateris, and D. Bochtis, "Machine learning in agriculture: A comprehensive updated review," Jun. 01, 2021, *MDPI AG*. doi: 10.3390/s21113758.

[8] T. Ayoub Shaikh, T. Rasool, and F. Rasheed Lone, "Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming," *Comput. Electron. Agric.*, vol. 198, p. 107119, 2022, doi: https://doi.org/10.1016/j.compag.2022.107119.

[9] I. Attri, L. K. Awasthi, and T. P. Sharma, "Machine learning in agriculture: a review of crop management applications," *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 12875–12915, 2024, doi: 10.1007/s11042-023-16105-2.

[10] H. M. Abdullah, J. E. Istiyanto, A. Z. K. Frisky, and E. A. Suyono, "Feature enhanced multistream RNN for growth phase prediction of Euglena sp. microalgae in an IoT-based outdoor cultivation environment," *Smart Agricultural Technology*, vol. 11, Aug. 2025, doi: 10.1016/j.atech.2025.100939.

[11] B. Dey, J. Ferdous, and 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, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25112.

[12] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture (Switzerland)*, vol. 14, no. 8, Aug. 2024, doi: 10.3390/agriculture14081256.

[13] F. S. Prity, MD. M. Hasan, S. H. Saif, Md. M. Hossain, S. H. Bhuiyan, Md. A. Islam, and Md. T. H. Lavlu, "Enhancing Agricultural Productivity: A Machine Learning Approach to Crop Recommendations," *Human-Centric Intelligent Systems*, vol. 4, no. 4, pp. 497–510, Dec. 2024, doi: 10.1007/s44230-024-00081-3.

[14] M. Y. Shams, S. A. Gamel, and F. M. Talaat, "Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making," *Neural Comput. Appl.*, vol. 36, no. 11, pp. 5695–5714, Apr. 2024, doi: 10.1007/s00521-023-09391-2.

 | doi: 10.23960/pepadun.v7i3.345

[15] M. A. Bakr, A. J. Khan, S. D. Khan, M. H. Zafar, M. Ullah, and H. Ullah, "Evaluation of Learning-Based Models for Crop Recommendation in Smart Agriculture," *Information (Switzerland)*, vol. 16, no. 8, Aug. 2025, doi: 10.3390/info16080632.

[16] M. R. Islam, K. Oliullah, M. M. Kabir, M. Alom, and M. F. Mridha, "Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation," *J. Agric. Food Res.*, vol. 14, Dec. 2023, doi: 10.1016/j.jafr.2023.100880.

[17] M. Bouni, B. Hssina, K. Douzi, and S. Douzi, "Interpretable Machine Learning Techniques for an Advanced Crop Recommendation Model," *Journal of Electrical and Computer Engineering*, vol. 2024, 2024, doi: 10.1155/2024/7405217.

[18] A. Aryanti and N. Iryani, "A Smart Recommendation System for Crop Seed Selection Using Gradient Boosting Based on Environmental and Geospatial Data," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC

[19] H. Gunasekaran, K. Ramalakshmi, S. Debnath, and D. K. Swaminathan, "Physics-Aware Ensemble Learning for Superior Crop Recommendation in Smart Agriculture," *Sensors*, vol. 25, no. 19, Oct. 2025, doi: 10.3390/s25196243.

[20] A. Ingle, "Crop Recommendation Dataset," Kaggle. Accessed: Mar. 28, 2026. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[21] T. B. Shahi, C. Y. Xu, A. Neupane, and W. Guo, "Recent Advances in Crop Disease Detection Using UAV and Deep Learning Techniques," May 01, 2023, *MDPI*. doi: 10.3390/rs15092450.

[22] R. P. Sishodia, R. L. Ray, and S. K. Singh, "Applications of remote sensing in precision agriculture: A review," *Remote Sens. (Basel).*, vol. 12, no. 19, pp. 1–31, Oct. 2020, doi: 10.3390/rs12193136.

[23] A. Triantafyllou, P. Sarigiannidis, and S. Bibi, "Precision agriculture: A remote sensing monitoring system architecture," *Information (Switzerland)*, vol. 10, no. 11, Nov. 2019, doi: 10.3390/info10110348.

[24] I. Anam, N. Arafat, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A systematic review of UAV and AI integration for targeted disease detection, weed management, and pest control in precision agriculture," Dec. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.atech.2024.100647.

[25] D. G. Pai, R. Kamath, and M. Balachandra, "Deep Learning Techniques for Weed Detection in Agricultural Environments: A Comprehensive Review," 2024, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/ACCESS.2024.3418454.

[26] M. Baishya and L. Dutta, "Tiny ML based crop recommendation system for precision agriculture 5.0," *Smart Agricultural Technology*, vol. 12, Dec. 2025, doi: 10.1016/j.atech.2025.101247.

[27] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture (Switzerland)*, vol. 14, no. 8, Aug. 2024, doi: 10.3390/agriculture14081256.

[28] R. Guebsi, S. Mami, and K. Chokmani, "Drones in Precision Agriculture: A Comprehensive Review of Applications, Technologies, and Challenges," Nov. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/drones8110686.

[29] S. Abdumalikov, J. Kim, and Y. Yoon, "Performance Analysis and Improvement of Machine Learning with Various Feature Selection Methods for EEG-Based Emotion Classification," *Applied Sciences (Switzerland)*, vol. 14, no. 22, Nov. 2024, doi: 10.3390/app142210511.

[30] C. S. K. Dash, A. K. Behera, S. Dehuri, and A. Ghosh, "An outliers detection and elimination framework in classification task of data mining," *Decision Analytics Journal*, vol. 6, Mar. 2023, doi: 10.1016/j.dajour.2023.100164.

[31] I. Izonin, R. Tkachenko, N. Shakhovska, B. Ilchyshyn, and K. K. Singh, "A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain," *Mathematics*, vol. 10, no. 11, Jun. 2022, doi: 10.3390/math10111942.

[32] A. Rejeb, K. Rejeb, and A. Hassoun, "The impact of machine learning applications in agricultural supply chain: a topic modeling-based review," Dec. 01, 2025, *Springer Nature*. doi: 10.1007/s44187-025-00419-1.

 | doi: 10.23960/pepadun.v7i3.345

[33] G. Gupta and S. Kumar Pal, "Applications of AI in precision agriculture," *Discover Agriculture*, vol. 3, no. 1, Apr. 2025, doi: 10.1007/s44279-025-00220-9.

[34] M. Padhiary, D. Saha, R. Kumar, L. N. Sethi, and A. Kumar, "Enhancing precision agriculture: A comprehensive review of machine learning and AI vision applications in all-terrain vehicle for farm automation," Aug. 01, 2024, *Elsevier B.V.* doi: 10.1016/j.atech.2024.100483.