Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2026, 2026:83
https://doi.org/10.28919/cmbn/9884
ISSN: 2052-2541

RANDOM FOREST BASED APPROACH FOR CROP RECOMMENDATION

MIRIAM E. RAMIREZ-SILVA[1], ROCIO A. LIZARRAGA-MORALES[2,*],
GEOVANNI HERNANDEZ-GOMEZ[1], SANTIAGO DAMIAN-MUNIZ[1]

[1]Department of Multidisciplinary Studies, University of Guanajuato, Yuriria, Mexico
[2]Department of Art and Enterprise, University of Guanajuato, Salamanca, Mexico

Abstract. The selection of crop species that are well-suited to the soil conditions, climate, and water availability is essential for maximizing resource efficiency, boosting yields, and supporting food security. Traditional methods for crop recommendation involve empirical methods and statistical analysis of large amounts of data, which limits the capacity to respond effectively to different challenges. The use of machine learning algorithms for crop recommendation has become a key tool to support farmers' decision-making. In this work, we propose an effective crop recommendation model based on the Random Forest algorithm, which is optimized for this agricultural context. The model is trained with inputs corresponding to soil composition and climate data, while the output corresponds to the best crop for a given input condition. The proposed methodology involves preprocessing the data and finding the best hyperparameters for the random forest-based classification. Additionally, a thorough evaluation of the resulting model is implemented. The results of this methodology show an improvement in the accuracy, precision, recall, and F1-score metrics for the recommendation of 22 different crop classes in comparison with other classical machine learning and state-of-the-art approaches. This work demonstrates that our approach can provide a practical and accessible solution in systems within agricultural environments with limited infrastructure.

Keywords: agriculture; classification; crop recommendation; random forest.
2020 AMS Subject Classification: 62H30, 68T05.

*Corresponding author
E-mail address: ra.lizarragamorales@ugto.mx
Received Mar. 17, 2026

1. INTRODUCTION

Crop selection is a fundamental task in sustainable agricultural systems. The selection of crop species that are better adapted to local soil type, climate and water availability plays a key role in optimizing resources, improving productivity, and ensuring food security. In vulnerable regions, farmers can improve yields, reduce input costs, and mitigate the risks associated with climate variability [10, 16]. In contrast, inadequate decisions can lead to soil degradation, increased need for fertilizers and pesticides, and reduced stability in the face of extreme weather events, ultimately compromising the food security of entire communities [33].

Crop selection must take into account a number of key factors such as drought tolerance, resistance to pests and diseases, nutrient use efficiency, and the integration of sustainable agricultural practices [17]. These considerations contribute to reducing the use of agrochemicals and mitigating environmental impacts, including soil and water pollution [16]. Even though the importance of the crop selection decision is recognized, farmers usually lack the necessary tools to optimize that decision. In many regions, crop selection remains largely empirical, which limits the capacity to respond effectively to climate variability and natural resource challenges [19]. Additionally, limited access to information and insufficient training in the use of technological tools hinder farmers' ability to adapt to such emerging environmental scenarios [22].

Historically, crop selection has relied on empirical methods and statistical analysis. Such historical methods have included simplified models based on empirical equations, regression analysis, and direct measurements [7]. Nabateregga et al. [8] have summarized methods through comparative tables, flowcharts, and statistical summaries of errors such as bias and variance. They noted that in Sub-Saharan Africa, the adoption of technology for accurate measurements remains very limited due to high costs, lack of training, and poor infrastructure. Therefore, it is essential to develop tools for crop selection capable of processing large amounts of data that are simple to use, do not require sophisticated methods, and are easy to implement in low-resource environments.

Recently, the use of machine learning (ML) techniques for agricultural decisions has attracted much attention, especially for decisions that promote resilient and sustainable agriculture [30]. In a closely related agricultural application, Bakhtiar et al. [1] compared several classical ML

algorithms for rice leaf disease classification (KNN, SVM, Random Forest, Decision Tree, and Naive Bayes) and reported Random Forest as the best performing model, highlighting that strong predictive performance can be achieved with practical and implementable approaches. ML has demonstrated more accurate estimations that support appropriate crop selection decisions [11, 31]. For example, He et al. [12] and Jabed and Murad [13] demonstrated how ML can process extensive datasets related to soil, climate, and agricultural practices, generating predictions better adapted to each context. More recently, Kiruthika and Karthika [15] proposed a system that combines an optimization algorithm with a long short-term memory (LSTM) network, achieving faster and more accurate recommendations. Umar et al. [21] developed a model ensemble that integrates seven algorithms into a single system, significantly improving precision by adapting to specific edaphoclimatic conditions. The method proposed by Stracqualursi [32] introduced an exponentially weighted ensemble that automatically adjusts each model's weight according to its performance, providing accurate recommendations through a simple and adaptable approach.

Specifically, several investigations emphasize diverse ML methods as highly suitable for crop recommendation. Dey et al. [6] analyzed various models and found that XGBoost achieved the best results in their experiments. Elbasi et al. [9] evaluated a wide range of algorithms and introduced a new feature combination scheme, highlighting the Bayesian Network model as the most effective. Raja et al. [25] explored different variable selection techniques and classifiers, demonstrating that the Random Forest model, with bagging, delivered the best performance. Bakthavatchalam et al. [2] developed a multilayer perceptron (MLP) for optimal crop selection. Ramzan et al. [26] proposed a system in which KNN, Random Forest, and SVM jointly contributed to the overall performance of their model, achieving accurate classification of their dataset. Nevertheless, many of these studies that employ complex models show limitations, as they require important computational resources and fail to adapt optimally to the variability of local conditions. These limitations hinder their practical implementation in regions with limited technological infrastructure and delay their adoption. Overcoming this challenge is essential to ensure that innovations reach those who need them most, enabling crop selection to be optimized in a sustainable and adaptable manner [14, 19].

In response to this issue, this study proposes the development of an accurate model for crop recommendation based on the Random Forest algorithm. It is demonstrated that highly reliable results can be achieved using a machine learning technique, offering a practical and accessible alternative to the more complex and demanding architectures recently proposed. The proposed model achieves optimal performance on a Crop Recommendation dataset, facilitating its practical adoption in precision agriculture systems. In this way, this work contributes a novel solution by demonstrating that a simple and optimized model can match or even exceed the performance of more sophisticated methods, significantly lowering the technological barrier for implementing agricultural recommendations in resource-constrained contexts.

The organization of this work is as follows: Section 2 provides the description of the Materials and Methods, section 3 presents the Results and Discussion, section 4 the Conclusion and finally the section 5 the Acknowledgments.

2. MATERIALS AND METHODS

This section is divided into four stages, as shown Figure 1. In the first stage, the dataset is standardized using Min–Max normalization to place all predictive variables on the same scale. The next step, with the data already normalized, is to process it using a Random Forest (RF) classifier, which is trained and evaluated through three validation strategies: Hold-Out 70/30, Hold-Out 80/20, and a ten-fold cross-validation. These approaches provide more robust and generalizable performance estimation. Performance was measured through widely used indicators, such as accuracy, precision, recall, and F1-score. This procedure allowed for a comprehensive assessment of both global and class-level performance, facilitating the systematic comparison presented in the Results section.

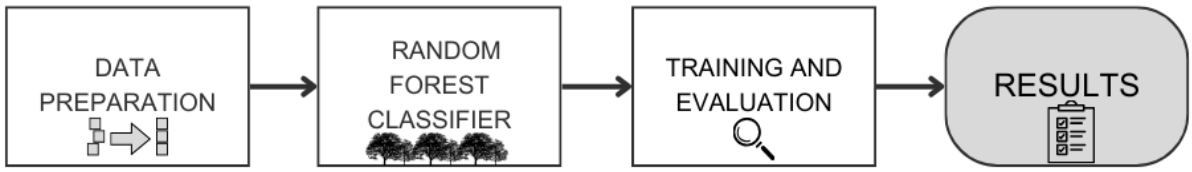

FIGURE 1. Methodology overview diagram.

2.1. Preprocessing. This work used the crop recommendation dataset compiled by the Food and Agriculture Chamber of India [20]. The dataset is composed of seven numerical attributes, namely nitrogen (N), phosphorus (P), potassium (K), temperature (°C), relative humidity (%), pH level, and rainfall (mm). The target variable includes twenty-two crop classes such as apple, banana, black gram, and chickpea, each represented by one hundred samples. In total, the dataset contains 2200 observations, making it a balanced and high-dimensional multiclass classification problem. Table 1 summarizes the variables together with their units and a brief description.

TABLE 1. Dataset characteristics.

Variable	**Description**
Nitrogen (N)	Amount of nitrogen available in the soil.
Phosphorus (P)	Phosphorus content in the soil.
Potassium (K)	Potassium concentration.
Temperature (°C)	Temperature in the crop region.
Humidity (%)	Relative air humidity.
pH	Level of acidity or alkalinity in the soil.
Rainfall (mm)	Rainfall in the agricultural region.

Considering that each variable appears in different ranges, a Min-Max normalization process was applied to all numerical variables in the dataset. This method transforms each variable x_i to the range $[0,1]$, as shown in Eq. 1.

$$x_i' = \frac{x_i - \min(X)}{\max(X) - \min(X)}, \tag{1}$$

where x_i is the original feature value for sample i, x_i' is the normalized value, X is the set of all values of a given feature, $\min(X)$ is the minimum value, and $\max(X)$ is the maximum value in the dataset. Min-Max normalization is suitable in scenarios where preserving the original data distribution is desired while scaling all features into a common range. This is beneficial for scale-sensitive algorithms. Applying the same preprocessing across models also helps maintain

consistency. As stated in [23], this technique is simple, effective, and widely recommended when severe outliers are not present.

2.2. Random Forest Classifier. Breiman [4] introduced Random Forest (RF) as a supervised learning method that constructs an ensemble of decision trees. Its performance is enhanced through bootstrap aggregating (bagging), which reduces variance and improves predictive accuracy. In this approach, each tree is trained on a subset of data obtained through sampling with replacement, which increases model diversity and reduces variance. At each split node, RF uses a random subset of predictors (m_{try}) instead of considering all variables, which promotes independence among trees and improves generalization capacity [28]. Among the key parameters are the minimum leaf size (MinLeafSize), which regulates tree depth, and the number of trees in the forest (n_{trees}), whose increase generally improves model stability until reaching a saturation point. To assess performance, it is common to apply k-fold cross-validation, which estimates the model's generalization ability across different data partitions. In classification problems, the evaluation metric most commonly used is accuracy, defined as the proportion of correctly classified observations with respect to the total [18, 27]. These characteristics make RF an effective and robust method for multiclass classification tasks, balancing model complexity with predictive power, which explains its good generalization in different areas [3, 5].

2.3. Training and Evaluation. The training and evaluation process implemented in this study is designed to objectively assess the performance of the Random Forest classification model. To ensure robustness, three different data partitioning strategies were employed, Hold-Out 70/30 split, Hold-Out 80/20 split, and 10-fold cross-validation. In the Hold-Out 70/30 scheme, 70% of the dataset is used to fit the model, while the remaining 30% is reserved for testing. This method is simple and efficient. However, for high-dimensional datasets, the reduced training portion may limit the model's ability to fully capture complex patterns [29]. The Hold-Out 80/20 scheme, in contrast, increases the training to four-fifths, whereas the last fifth is utilized for testing purposes. This partition offers a more appropriate balance between training and validation data, which contributes to improved predictive performance [29]. Finally, to obtain a more stable and less variance-prone estimate of the model generalization capability, 10-fold cross-validation was also employed. Cross-validation divides the dataset into 10-fold of equal

size. In each iteration, the model is trained using nine of these folds, while the remaining one is used for testing. This cycle continues until every fold has been employed once as the validation set. Although this method is computationally more intensive, it provides a more reliable estimate of the model's overall performance [34]. All models were trained on normalized data, using consistent configurations to ensure a fair comparison.

To evaluate the performance of the classifiers [24], widely-used metrics are considered: Precision, Recall, Accuracy and F1-score. Precision (Eq. 2) measures the proportion of correctly identified positive predictions for each class. Recall measures the model's ability to correctly identify all real instances of each class as is shown in Eq. 3. Accuracy represents the overall percentage of correct predictions across all samples Eq.4. In Eq. 5 is shown the F1-Score which provides a harmonic mean between precision and recall for each class.

$$\text{Precision} = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FP_i + \varepsilon}, \tag{2}$$

$$\text{Recall} = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FN_i + \varepsilon}, \tag{3}$$

$$\text{Accuracy} = \frac{\sum_{i=1}^{K} TP_i}{\sum_{i=1}^{K}\sum_{j=1}^{K} C_{ij}}, \tag{4}$$

$$F1 = \frac{1}{K}\sum_{i=1}^{K}\frac{2\cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i + \varepsilon}, \tag{5}$$

where K is the total number of classes, TP_i corresponds to the true positives for class i. FP_i represents the false positives for class i, and FN_i counts the false negatives for class i. C_{ij} is the entry in the confusion matrix for the actual class i and predicted class j, and ε a small constant introduced to avoid division by zero.

2.4. Hyperparameter Optimization. In this study, a Random Forest (RF) model was applied to classify the observations in the dataset. To fine-tune its performance, different configurations were tested, varying the number of trees from 1 to 500 using 10-fold cross-validation. The results from this evaluation, lead us to configure the RF classifier with 211 trees. Figure 2 shows the relationship between the number of trees and the achieved accuracy. As it can be seen in this figure, the curve rises quickly at the beginning, and accuracy becomes steady at approximately 100 trees. After that, the results stay almost constant, with only small variations. The best performance was obtained with 211 trees, yielding an accuracy of 99.55%, highlighted in yellow. This result indicates that adding more trees beyond this point does not provide further improvements and might add minor inconsistencies instead. Therefore, 211 trees were selected as the most suitable hyperparameter, ensuring both strong predictive accuracy and consistent model behavior.

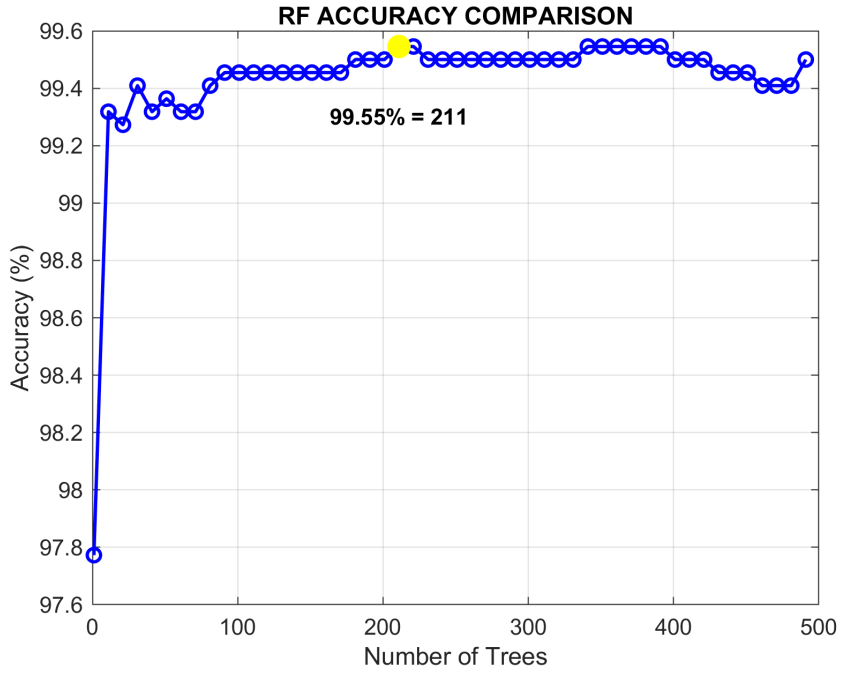

FIGURE 2. Accuracy Comparison (RF).

3. Results and Discussion

The results are presented in two experiments. The first experiment consisted of comparing the current proposal with three other classical machine learning classifiers: K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), and Support Vector Machines (SVM). The corresponding hyperparameters for each method were optimized using a 10-fold cross-validation scheme. Subsequently, these models were analyzed using Hold-Out partitions of 70/30, 80/20, and cross-validation, in order to determine which validation strategy offered the best predictive performance. The second experiment consists of comparing the results of the proposed model with two state-of-the-art approaches that used the same dataset, one of them based on an exponentially weighted ensemble, and the other on XGBoost. This allowed for assessing the competitiveness of the proposed approach against recent state of the art methods.

3.1. Performance comparison with other machine learning methods. The first experiment consisted of comparing our proposed Random Forest (RF), with other classical machine learning methods: K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), and Support Vector Machines (SVM). Hyperparameters for KNN, MLP, and RF were tuned through 10-fold cross-validation. For KNN, values of k from 1 to 15 were examined; for MLP, several two-layer architectures were tested ([10 10], [20 20], [50 50], [80 40], [100 50], [110 20]).

Table 2 presents the results of each approach for each evaluation scheme. Under cross-validation, the best performance of KNN was obtained with k = 7, reaching an accuracy of 98.73%. For MLP, the two-hidden-layer architecture with 100 and 50 neurons was the most accurate, achieving 98.64%. SVM, evaluated with its default configuration, showed consistent performance of 99.22%. In contrast, Random Forest achieved its best accuracy of 99.55% with 211 trees and reached optimal classification (100%) using the Hold-Out 80/20 strategy. Overall, RF outperformed the other models, while SVM showed the most consistent results across all validation schemes. It is also worth noting that MLP, KNN, and RF performed slightly better with the 80/20 partition, as the larger training set allowed them to capture more complex patterns.

TABLE 2. Comparative performance of KNN, MLP, RF, and SVM across evaluation metrics.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
		Hold-Out 70/30		
KNN	98.64	98.70	98.64	98.63
MLP	98.64	98.64	98.64	98.61
SVM	99.09	99.11	99.09	99.09
RF	**99.70**	**99.71**	**99.70**	**99.70**
		Hold-Out 80/20		
KNN	98.86	98.97	98.86	98.87
MLP	99.32	99.33	99.32	99.32
SVM	99.09	99.12	99.09	99.08
RF	**100**	**100**	**100**	**100**
		Cross-Validation (10-fold)		
KNN	98.73	98.87	98.73	98.72
MLP	98.64	98.77	98.64	98.63
SVM	99.22	99.26	99.22	99.22
RF	**99.55**	**99.59**	**99.55**	**99.54**

3.2. Comparison with state of the art models. In the second experiment, the proposed RF model was compared with two recent state-of-the-art methods. These methods were selected for comparison because they represent the most recent advances reported in the literature while relying on the same dataset, ensuring a fair and consistent evaluation. The first method, developed by Stracqualursi [32], is an exponentially weighted ensemble that integrates various base classifiers such as KNN, Decision Trees, Random Forest, SVM, Naive Bayes, Logistic Regression, and XGBoost. The evaluation scheme performed by the authors was to split into 80 percent for training and 20 percent for testing, with an additional 30 percent of the training set reserved for validation, which was used to determine the weights of each classifier. This procedure yielded an accuracy of 99.8%. The second method, presented by Dey et al. [6], implements the XGBoost classifier, a gradient boosting ensemble algorithm that combines multiple decision trees

sequentially with regularization. Their study employed a Hold-Out 70/30 validation scheme and reported a maximum accuracy of 99.09% on the crop recommendation dataset. As mentioned, both studies employed the same Crop Recommendation dataset, which enables a direct comparison under consistent conditions. Table 3 presents the accuracy values reported in their work alongside those obtained in this study, a dashed (-) indicates the validation schemes that were not considered by the authors.

TABLE 3. Comparison of Accuracy with state of the art model.

Method	**Accuracy**		
	Hold-Out 70/30	**Hold-Out 80/20**	**CV (10-fold)**
Exponential-Weighted Ensemble [32]	—	99.80%	—
XGBoost [6]	98.51%	—	—
Proposed (RF)	**99.70%**	**100%**	**99.55%**

As shown in Table 3, the proposed RF achieved the highest accuracy across all evaluation schemes, notably reaching optimal performance (100%) under the Hold-Out 80/20 strategy. In contrast, XGBoost reported an accuracy of 98.51% exclusively under the Hold-Out 70/30 strategy and did not provide results for other configurations. The Exponentially Weighted Ensemble reported 99.80% accuracy under Hold-Out 80/20, but no results were presented for 70/30 or cross-validation. It is also important to highlight that, unlike our study, neither Dey et al. [6], nor Stracqualursi [32], employed cross-validation, relying exclusively on Hold-Out partitions. This may limit the generalization of their results. By incorporating cross-validation, our study strengthens the statistical robustness and reliability of performance estimates, confirming the competitiveness of the proposed model against recent state-of-the-art methods.

4. CONCLUSION

This study demonstrates that machine learning techniques, when properly optimized, can deliver highly competitive results in multiclass crop recommendation tasks. Random Forest emerged as the most successful model, achieving flawless classification under the Hold-Out

80/20 configuration and outperforming advanced techniques like the Exponentially Weighted Ensemble and XGBoost.

These results highlight the importance of employing reliable and optimized models, particularly in environments with limited resources. Likewise, the strong performance of the Hold-Out 80/20 partition across all classifiers confirms that this validation strategy achieves an appropriate balance between training and evaluation reliability. The results also confirm the relevance of integrating essential data preparation techniques, such as Min-Max normalization, into model development, as this allows variables to be unified and supports a more stable learning process for the classifier.

Overall, this work demonstrates that a well-tuned Random Forest model can provide a practical and accessible solution for precision agriculture, positioning itself above state-of-the-art methods in terms of performance while offering greater computational efficiency. This makes it a solid candidate for implementation in crop recommendation systems within agricultural environments with limited infrastructure.

ACKNOWLEDGMENTS

The authors would like to thank the University of Guanajuato for the financial support. In addition, we would like to thank the Mexican SECIHTI for the financial support provided.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] R.H. Bakhtiar, Y. Dani, D. Suandi, Machine Learning Based Classification of Rice Leaf Diseases: A Comparative Study, Commun. Math. Biol. Neurosci. 2025 (2025), 43. https://doi.org/10.28919/cmbn/9198.

[2] K. Bakthavatchalam, B. Karthik, V. Thiruvengadam, S. Muthal, D. Jose, K. Kotecha, V. Varadarajan, IoT Framework for Measurement and Precision Agriculture: Predicting the Crop Using Machine Learning Algorithms, Technologies 10 (2022), 13. https://doi.org/10.3390/technologies10010013.

[3] M. Belgiu, L. Drăguţ, Random Forest in Remote Sensing: A Review of Applications and Future Directions, ISPRS J. Photogramm. Remote Sens. 114 (2016), 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[4] L. Breiman, Random Forests, Mach. Learn. 45 (2001), 5–32. https://doi.org/10.1023/A:1010933404324.

[5] D.R. Cutler, T.C. Edwards, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J.J. Lawler, Random Forests for Classification in Ecology, Ecology 88 (2007), 2783–2792. https://doi.org/10.1890/07-0539.1.

[6] 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 10 (2024), e25112. https://doi.org/10.1016/j.heliyon.2024.e25112.

[7] J. Dumanski, C. Onofrei, Techniques of Crop Yield Assessment for Agricultural Land Evaluation, Soil Use Manag. 5 (1989), 9–15. https://doi.org/10.1111/j.1475-2743.1989.tb00754.x.

[8] M. Nabateregga, S.Ø. Sølberg, J. van Etten, K. de Sousa, Trade-Offs of On-Farm Yield Estimation Approaches and Key Factors Affecting Yield Accuracy in Smallholder Farming Systems in Sub-Saharan Africa. A Review, preprint, (2024). https://doi.org/10.21203/rs.3.rs-3756160/v1.

[9] E. Elbasi, C. Zaki, A.E. Topcu, W. Abdelbaki, A.I. Zreikat, E. Cina, A. Shdefat, L. Saker, Crop Prediction Model Using Machine Learning Algorithms, Appl. Sci. 13 (2023), 9288. https://doi.org/10.3390/app13169288.

[10] FAO, The State of Food and Agriculture 2023: Revealing the True Cost of Food to Transform Agrifood Systems, FAO, Rome, 2023. https://doi.org/10.4060/cc7724en.

[11] G. Mariammal, A. Suruliandi, S.P. Raja, E. Poongothai, Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique with Various Classifiers, IEEE Trans. Comput. Soc. Syst. 8 (2021), 1132–1142. https://doi.org/10.1109/TCSS.2021.3074534.

[12] S. He, P. Peng, Y. Chen, X. Wang, Multi-Crop Classification Using Feature Selection-Coupled Machine Learning Classifiers Based on Spectral, Textural and Environmental Features, Remote Sens. 14 (2022), 3153. https://doi.org/10.3390/rs14133153.

[13] M.A. Jabed, M.A. Azmi Murad, Crop Yield Prediction in Agriculture: A Comprehensive Review of Machine Learning and Deep Learning Approaches, with Insights for Future Research and Sustainability, Heliyon 10 (2024), e40836. https://doi.org/10.1016/j.heliyon.2024.e40836.

[14] X. Jin, L. Kumar, Z. Li, H. Feng, X. Xu, G. Yang, J. Wang, A Review of Data Assimilation of Remote Sensing and Crop Models, Eur. J. Agron. 92 (2018), 141–152. https://doi.org/10.1016/j.eja.2017.11.002.

[15] S. Kiruthika, D. Karthika, IoT-Based Professional Crop Recommendation System Using a Weight-Based Long-Term Memory Approach, Meas. Sens. 27 (2023), 100722. https://doi.org/10.1016/j.measen.2023.100722.

[16] B.K. Kogo, L. Kumar, R. Koech, Climate Change and Variability in Kenya: A Review of Impacts on Agriculture and Food Security, Environ. Dev. Sustain. 23 (2021), 23–43. https://doi.org/10.1007/s10668-020-00589-1.

[17] M.S. Kukal, S. Irmak, Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production, Sci. Rep. 8 (2018), 3450. https://doi.org/10.1038/s41598-018-21848-2.

[18] A. Liaw, M. Wiener, Classification and Regression by randomForest, R News 2 (2002), 18–22. https://journal.r-project.org/articles/RN-2002-022/.

[19] S. Mardero, B. Schmook, S. Calmé, R.M. White, J.C. Joo Chang, G. Casanova, J. Castelar, Traditional Knowledge for Climate Change Adaptation in Mesoamerica: A Systematic Review, Soc. Sci. Humanit. Open 7 (2023), 100473. https://doi.org/10.1016/j.ssaho.2023.100473.

[20] Indian Chamber of Food and Agriculture, https://www.icfa.org.in.

[21] M. Olalere, G.I.O. Aimufua, M.U. Abdullahi, B.H. Egga, Ensemble-Based Predictive Model for Crop Recommendation, Dutse J. Pure Appl. Sci. 10 (2024), 391–409. https://doi.org/10.4314/dujopas.v10i2c.36.

[22] V. Pandith, H. Kour, S. Singh, J. Manhas, V. Sharma, Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis, J. Sci. Res. 64 (2020), 394–398. https://doi.org/10.37398/JSR.2020.640254.

[23] S.G.K. Patro, K.K. Sahu, Normalization: A Preprocessing Stage, Int. Adv. Res. J. Sci. Eng. Technol. 2 (2015), 20–22. https://doi.org/10.17148/IARJSET.2015.2305.

[24] D.M.W. Powers, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness and Correlation, Int. J. Mach. Learn. Technol. 2 (2011), 37–63. https://doi.org/10.48550/arXiv.2010.16061.

[25] S.P. Raja, B. Sawicka, Z. Stamenkovic, G. Mariammal, Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers, IEEE Access 10 (2022), 23625–23641. https://doi.org/10.1109/ACCESS.2022.3154350.

[26] S. Ramzan, Y.Y. Ghadi, H. Aljuaid, A. Mahmood, B. Ali, An Ingenious IoT Based Crop Prediction System Using ML and EL, Comput. Mater. Contin. 79 (2024), 183–199. https://doi.org/10.32604/cmc.2024.047603.

[27] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Springer New York, 2009. https://doi.org/10.1007/978-0-387-84858-7.

[28] H.A. Salman, A. Kalakech, A. Steiti, Random Forest Algorithm Overview, Babylonian J. Mach. Learn. 2024 (2024), 69–79. https://doi.org/10.58496/BJML/2024/007.

[29] B. Sekeroglu, Y.K. Ever, K. Dimililer, F. Al-Turjman, Comparative Evaluation and Comprehensive Analysis of Machine Learning Models for Regression Problems, Data Intell. 4 (2022), 620–652. https://doi.org/10.1162/dint_a_00155.

[30] M.K. Senapaty, A. Ray, N. Padhy, A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms, Agriculture 14 (2024), 1256. https://doi.org/10.3390/agriculture14081256.

[31] A. Sharma, A. Jain, P. Gupta, V. Chowdary, Machine Learning Applications for Precision Agriculture: A Comprehensive Review, IEEE Access 9 (2021), 4843–4873. https://doi.org/10.1109/ACCESS.2020.3048415.

[32] L. Stracqualursi, A Lightweight Exponential-Weighted Ensemble for Crop Recommendation, J. Agric. Biol. Environ. Stat. (2025), 1–17. https://doi.org/10.1007/s13253-025-00694-6.

[33] B. Sultan, M. Gaetani, Agriculture in West Africa in the Twenty-First Century: Climate Change and Impacts Scenarios, and Potential for Adaptation, Front. Plant Sci. 7 (2016), 1262. https://doi.org/10.3389/fpls.2016.01262.

[34] D. Wilimitis, C.G. Walsh, Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial, JMIR AI 2 (2023), e49023. https://doi.org/10.2196/49023.