

# Predictive modeling for crop suitability and productivity using machine learning techniques

Gulaganjihalli Ningegowda Shwetha<sup>1,2</sup>, Bhat Geetalaxmi Jairam<sup>1</sup>

<sup>1</sup>Department of Computer Science and Engineering, The National Institute of Engineering-Mysore, Affiliated to Visvesvaraya Technological University, Belagavi, India

<sup>2</sup>Department of Computer Science and Engineering, Sri Jayachamarajendra College of Engineering, JSS Science and Technology University, Mysuru, India

## Article Info

### Article history:

Received Feb 26, 2025

Revised Feb 16, 2026

Accepted May 12, 2026

### Keywords:

Decision tree

Extreme gradient boosting

Light gradient boosting machine

Random forest

Support vector classification

## ABSTRACT

With the increasing global population and rising food demand, improving agricultural productivity through data-driven decision support systems has become essential. This study proposes a cross-validated meta-stacking ensemble framework for multi-class crop suitability prediction using soil nutrient and environmental parameters. The dataset consists of 2,200 samples covering 22 crop types and seven predictor variables, including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. Six machine learning (ML) models—random forest (RF), decision tree (DT), light gradient boosting machine (LightGBM), extreme gradient boosting (XGBoost), support vector classifier (SVC), and k-nearest neighbors (KNN)—were trained and optimized using RandomizedSearchCV with k-fold cross-validation. A stacked ensemble model was then developed to combine heterogeneous learners and improve predictive robustness. Experimental results demonstrate that the RF model achieved an accuracy of 99.36%, while the proposed cross-validated meta-stacking ensemble achieved comparable performance with improved generalization stability. Precision, recall, and F1-score values of 0.99 indicate consistent classification across all crop classes. Feature importance analysis revealed N, K, and rainfall as the most influential predictors. Model robustness was evaluated using cross-validation and an independent test split to minimize overfitting risk. The findings highlight the effectiveness of ensemble learning for sustainable agricultural decision-making and crop recommendation systems.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## Corresponding Author:

Gulaganjihalli Ningegowda Shwetha

Department of Computer Science and Engineering, The National Institute of Engineering-Mysore

Affiliated to Visvesvaraya Technological University

Belagavi, India

Email: shwethagn@jssstuniv.in

## 1. INTRODUCTION

Though agriculture plays an important role in food security and economic development, problems such as low productivity, losses in the supply chain, and adverse effects on the environment still hinder its progress, especially in developing nations [1]. It should also be noted that agricultural processes contribute to climate change, making it hard to realize the sustainable development goal (SDG) target of Zero Hunger by 2030 [2], [3]. The correct choice of crop and estimation of yields are crucial for satisfying the growing demand. Classical approaches frequently neglect important aspects like water, weather, and geography, resulting in poor estimates [4]. Research has shown that taking into account various environmental factors,

including nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall, enhances the performance of algorithms [5]. Ensemble and tree-based machine learning (ML) algorithms have been proven efficient in predicting crop suitability and yields [6], [7]. Therefore, this paper aims to analyze the possibility of improving accuracy and generalizability through meta-stacking of various ML algorithms based on cross-validation.

The main contributions of this work are: i) development of a cross-validated meta-stacking ensemble integrating random forest (RF), light gradient boosting machine (LightGBM), extreme gradient boosting (XGBoost), support vector classifier (SVC), and k-nearest neighbors (KNN) models; ii) comparative analysis of bagging, boosting, and stacking approaches for multi-class crop suitability prediction; iii) feature importance analysis linking agronomic variables (N, P, and rainfall) with predictive performance; and iv) overfitting-aware evaluation using outlier removal, k-fold cross-validation, and independent test set validation. In contrast to traditional methods that mainly focus on the use of single learners or uniform ensembles, this study uses different types of learners such as bagging (RF), boosting (XGBoost and LightGBM), kernel-based learners (SVC), and instance-based learners (KNN). Furthermore, contamination-resilient pre-processing with the help of isolation forest and hybrid feature importance analysis have been adopted in this research. By combining heterogeneous learners within a meta-learning framework, this study seeks to enhance predictive reliability and support sustainable agricultural decision-making.

This work is organized as follows. Section 2 consists of a brief literature survey of related literature on ML in agriculture. Section 3 contains details concerning the study technique and basic data preprocessing including initial data elimination and correlation tests to evaluate the link between multiple environment parameters. The findings are discussed in section 4, followed by the implications of the study in section 5.

## 2. LITERATURE SURVEY

The evolution of agricultural technology through artificial intelligence (AI) and ML algorithms has helped to develop effective crop selection and yield predictions by using AI and ML approaches for soil analysis, crop recommendation, and productivity improvement [8], [9]. Various ML models have been employed by researchers for the purpose of disease detection, soil fertility analysis, fertilizer recommendation, and crop yield predictions. Akhter and Sofi [10] proposed an internet of things (IoT) based ML model for apple disease prediction, while Nishant *et al.* [11] used Kernel ridge, Lasso, and Stacking regression for crop yield predictions. The effectiveness of RF, gradient boost, and decision tree (DT) models for yield prediction was demonstrated by BanuPriya *et al.* [12], while Mahore and Gadge [13] presented the efficiency of image-based RF models. The authors in [14]–[16] successfully used ML and data mining techniques for predicting crop yields and soil fertility. Zhong and Lai [17] used a combination of kernel k-means and support vector machine (SVM) algorithms for a smart agriculture model that achieved 99.6% accuracy, while the effectiveness of LightGBM and XGBoost was demonstrated [18], [19] for soil and rice yield prediction. Fertilizer recommendation using SVM and XGBoost was achieved by Rao *et al.* [20], while Singha *et al.* [21] performed soil nutrient predictions using ML approaches with spectroscopy technique. A KNN-based android application for soil nutrient estimation was suggested by Solehatin *et al.* [22], while Devan *et al.* [23] used RF and logistic regression for fertilizer recommendation and crop yield predictions. Gopi and Karthikeyan [24] proposed a hybrid model using Red Fox optimization and RNN for crop prediction. The research in [25], [26] utilized XGBoost and voting ensemble classifiers for crop recommendations that achieved high accuracy. According to Nti *et al.* [27], XGBoost and RF models outperformed stacked-tree-based ensemble learning (TBEL) models. On the other hand, Hasan *et al.* [28] performed soil nutrient and crop yield prediction in Bangladesh using KNN, RF, and ridge regression models. Octavianto and Wibowo [29] used stacking classifier for enhancing prediction performance. However, most of the previous studies adopted the use of single model approaches with less efficient validation methods, which reduces the generalization capabilities and increases the risk of overfitting. Very few works exist on heterogenous ensemble approaches like bagging, boosting, and stacking in a combined approach for multi-class crop suitability predictions.

## 3. METHOD

### 3.1. Data collection and preprocessing

The data set used in this analysis was obtained from Kaggle, which contained 2,200 entries with seven predictor variables containing data on nutrients found in the soil, such as N, P, K, as well as environmental conditions like temperature, humidity, pH, and rainfall. The target variable contained 22 different types of crops, like 'apple', 'banana', 'blackgram', 'chickpea', with 100 samples each. The target

variable, 'label', contained crop types, which were encoded into integers by the LabelEncoder. A mapping was also created to map the encoded labels to their respective crop types, which is done for better interpretation. Outlier detection was also performed by using the isolation forest algorithm, which detected 5% of the data as outliers with a contamination level of 0.05. This data was then removed from the data set, which provided a clean data set. This clean data set was then scaled by the MinMaxScaler, which scales the data into a standard range, allowing ML models to perform better [18], [21], [27]. Correlation analysis was also performed by creating a correlation matrix, which was visualized as a heatmap, allowing us to see the correlation between environmental conditions.

### 3.2. Feature selection and importance analysis

The selection of features was performed through the implementation of both statistical and modeling techniques to analyze the effect of soil and environmental attributes on crop suitability. Correlation matrices were applied to study the relationship between predictor variables and detect any multicollinearity. The absence of redundancy means that all attributes were used for further investigation. The importance of features was estimated using tree-based algorithms [18], [19], [27]. In particular, the feature importance within the RF algorithm was derived from the Gini impurity measure as in (1).

$$Importance(f) = \sum_{t=1}^T \sum_{n \in N_f} \Delta Gini(n, t) \quad (1)$$

Here  $T$  represents the total number of trees,  $N_f$  represents nodes split based on feature  $f$ , and  $\Delta Gini$  represents impurity reduction due to splitting. XGBoost and LightGBM use gain-based feature importance to measure the significance of features. This method measures the gain in model performance due to a specific feature during building the trees. The study identified that the most impactful factors were N, K, and rainfall for crop classification. This is because these factors are major contributors to crop suitability and productivity based on agronomic knowledge. Unlike most studies that use correlation thresholds to filter out features, all features are used in this study to ensure interpretability of the model while using ensemble learning to implicitly handle feature contributions.

### 3.3. Model training

Six ML models—RF, DT, LightGBM, XGBoost, SVC, and KNN—were trained and evaluated to be their best and prevent overfitting with the use of the k-fold cross-validation. Hyperparameter values for each model were optimized using RandomizedSearchCV to improve predictive accuracy and minimize overfitting during training [26], [27]. Table 1 shows the hyperparameters for each model were tuned using RandomizedSearchCV to achieve optimal performance. Figure 1 shows steps involved in crop yield prediction.

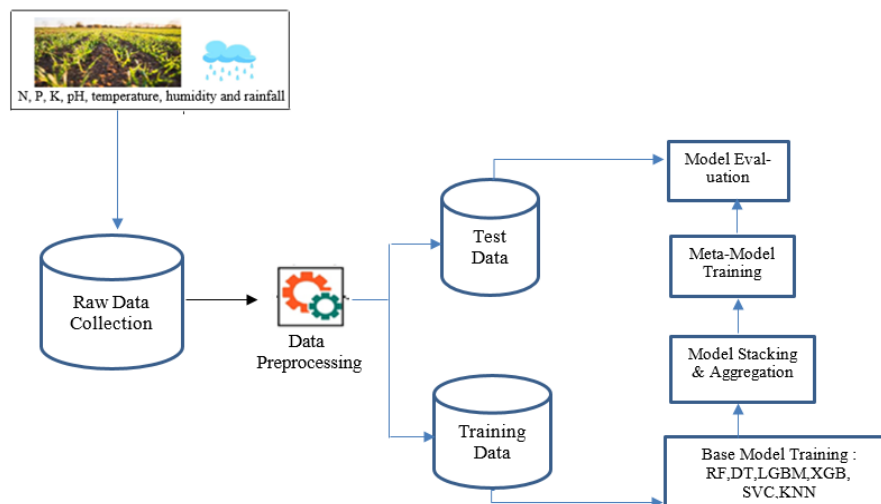


Figure 1. Data flow diagram of crop yield prediction

#### 3.3.1. Data preprocessing

The crop recommendation dataset is loaded which includes key environmental attributes such as N, P, K, temperature, humidity level, and rainfall. Several preprocessing steps were then performed:

- i) Label encoding: the target labels- crop types, are converted to numerical values with the use of the LabelEncoder method.
- ii) Feature scaling: the feature data undergoes normalization using the MinMaxScaler because all the features are assumed to be in different scales during feature extraction.
- iii) Data partitioning: the scaled dataset is split into training and testing sets using a 70-30 ratio via train\_test\_split, with 70% of the data used for model training and 30% reserved for testing.
- iv) Base model training: several ML models are trained, including RF, DT, LightGBM, XGBoost, SVC, and KNN. Hyperparameter tuning is conducted for each base model using RandomizedSearchCV, optimizing their performance through cross-validation.
- v) Model evaluation (base models): while bagging methods like RF reduce variance by averaging predictions from multiple models, and boosting methods like XGBoost focus on correcting errors of weaker models [26], [27], stacking synthesizes the strengths of these diverse approaches, aiming to improve predictive performance through a meta-learning framework [29]–[31].

Table 1. Tuned hyperparameters for each model

Model	Description	Tuned hyperparameters
RF	Robust for handling complex data and evaluating feature importance. Provides rankings for the significance of environmental and soil-related features.	n_estimators=100, max_depth=None, min_samples_split=2, min_samples_leaf=1
DT	Captures non-linear relationships and enables understanding of decision rules.	max_depth=10, min_samples_split=2
LightGBM	Optimized gradient boosting framework, suitable for big data.	n_estimators=100, learning_rate=0.1, max_depth=7
XGBoost	Scalable model with high predictive efficiency.	n_estimators=100, learning_rate=0.1, max_depth=6
SVC	Implemented with various kernels (linear, polynomial, RBF). Tuned regularization parameter C to handle dataset non-linearity.	Kernel types: linear, polynomial, RBF; C parameter tuned
KNN	Flexible for decision boundary complexity. Parameters like number of neighbors and distance measure were tuned.	n_neighbors and distance measure tuned
Meta-model (StackingClassifier)	Combines base models to enhance prediction. Utilizes stacked ensemble learning.	Uses outputs of RF, DT, LightGBM, XGBoost, SVC, and KNN as input features [30], [31]

### 3.4. Model stacking and aggregation

Stacking ensemble method was developed to utilize the advantages offered by several ML algorithms in predicting crop suitability. In the first place, several high-performing ML algorithms like RF, LightGBM, XGBoost, SVC, and KNN were chosen based on their accuracy, precision, recall, and F1-scores. The predictions obtained from these ML algorithms were then stacked using StackingClassifier method to generate meta-features. Stacking, unlike bagging and boosting techniques which are individual ensemble methods, makes use of meta-learning approach in capturing complementary learning from different classifiers [29]–[31].

### 3.5. Meta-model training

By employing a meta-model trained on the meta-features generated by the stacked base models, stacking allows the algorithm to learn the optimal way to combine the predictions of various models, potentially leading to improved accuracy and robustness. Unlike bagging and boosting, which emphasize single-objective learning, stacking can accommodate multiple algorithms and harness their collective insights. The meta-model is optimized to enhance the combined predictions from the base models, leading to an overall improvement in model accuracy [31].

### 3.6. Ensemble model evaluation

The stacked ensemble model is tested on the test set. This is where the combination of the results of different models is more helpful in providing a more reliable prediction compared to what is obtained by bagging or boosting alone. This is owing to the meta-learning nature of stacking. The stacking ensemble method enhances the performance of the model in making predictions [26], [27]. The detailed step-by-step procedure is standardized in the Algorithm 1, outlining the cross-validated meta-stacking methodology to enhance predictive accuracy.

Algorithm 1. Cross-validated meta-stacking for crop suitability

Step 1: Data loading  
Load dataset:

$D = \{(x_i, y_i)\}_{i=1}^N$ , where  $x_i \in \mathbb{R}^m$  represents the features (environmental factors),  
 $y_i \in \{C_1, C_2, \dots, C_k\}$  represents the crop labels.

Step 2: Data preprocessing

Encode target variable  $y$  into numerical format:

$$y_{\text{encoded}} = \text{LabelEncoder}(y)$$

Outlier detection using isolation forest:

$$\text{Fit model: } O = \text{IsolationForest}(x)$$

$$\text{Predict outliers: } \text{Outliers} = O.\text{predict}(x)$$

$$\text{Filter dataset: } D_{\text{clean}} = \{(x_i, y_i) | \text{outliers}_i = 1\}$$

Step 3: Feature scaling

Scale features using min-max scaling:

$$x_{\text{scaled}} = \frac{(x - x_{\min})}{(x_{\max} - x_{\min})}$$

Step 4: Exploratory data analysis (EDA)

$$\text{Correlation matrix R: } R = \text{corr}(x)$$

$$\text{Feature importance using RF: } \text{importance}_j = \text{RandomForest}(x, y_{\text{encoded}})_j$$

Step 5: Train–test split

Step 6: Define base models

Let  $M = \{M_1, M_2, M_3, M_4, M_5, M_6\}$  represents base models:

$M_1$ : RF,  $M_2$ : DT,  $M_3$ : LightGBM,  $M_4$ : XGBoost,  $M_5$ : SVC, and  $M_6$ : KNN

Step 7: K-fold cross-validation

For  $k$  folds, partition the training set:

$$D_{\text{train}} = \bigcup_{j=1}^k D_{\text{train}}^{(j)}$$

Step 8: Train base models

For each fold  $j$ : train each base model  $M_i$ :  $M_i.\text{fit}(D_{\text{train}}^{(j)})$

#### 4. RESULTS AND DISCUSSION

Features used for crop classification include N, P, and K as components of soil nutrients, as well as environmental features: temperature, humidity, pH, and rainfall. Figure 2 depicts the correlation matrix for all the chosen attributes; this shows how each attribute correlates with other attributes, thereby giving a good insight into the impact of such attributes on crop suitability. The correlation matrix depicts strong correlations between various environmental attributes without causing severe multicollinearity. Rainfall and humidity have positive correlations, while pH has a weak correlation with other attributes. Similar environmental and soil-related attributes have been widely used in previous crop recommendation and agricultural prediction studies [5], [25], [27].

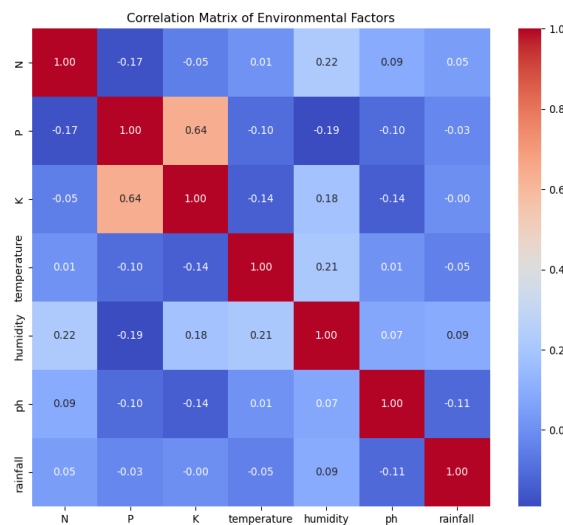


Figure 2. Correlation matrix of environmental factors

#### 4.1. Performance measures

Evaluation of the suggested models was performed using the classification measures. These include accuracy, precision, recall, and F1-score as in (2) to (5) [25]–[28]. These were employed to measure the accuracy, consistency, and robustness of the predictions made by the proposed models.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{Precision + Recall} \quad (5)$$

#### 4.2. Statistical validation and robustness analysis

Moreover, in order to test the reliability of the suggested framework, some cross-validation and agreement statistics were measured. It can be observed that the proposed stacking ensemble demonstrated reliable performance with low standard deviations for all folds. Cohen's Kappa statistic was used to measure agreement and classification accuracy beyond randomness. Similar validation strategies involving cross-validation and agreement statistics have been adopted in recent ensemble learning studies to evaluate predictive reliability [27], [29], [30]. The proposed stacking ensemble provided an approximate value of the Kappa statistic close to 0.99. Table 2 presents the mean cross-validation accuracy and standard deviation of the proposed models. Misclassifications mainly occurred among crops with similar environmental conditions and nutrient requirements. However, confusion matrices demonstrate that the proposed ensemble minimized inter-class prediction errors compared to individual models.

Table 2. Cross-validation performance stability

Model	Mean cross-validation accuracy	Standard deviation
RF	99.20	±0.31
Stacking	99.28	±0.24

#### 4.3. Feature importance analysis

In addition to model performance, the RF model also provided insights into feature importance. As shown in Figure 3, the most influential features for crop classification were N, K, and rainfall. Similar observations regarding the importance of soil nutrients and rainfall in crop recommendation systems have been reported in earlier agricultural ML studies [18], [25], [27], [28]. The comparative analysis shows that the cross-validated meta-stacking algorithm, which combined RF, LightGBM, and SVC, outperformed other models like XGBoost, KNN, and DT. With an accuracy of 99.36% and balanced metrics (precision, recall, and F1-score at 0.99), this approach effectively leverages the strengths of multiple models. Table 3 and Figure 4 shows the performance metrics of different models. From this experiment result, we came to know that the base models such as RF, DT, LightGBM, XGBoost, SVC, and KNN were evaluated using a host of metrics ranging from accuracy, precision, recall, to F1-score. The stacked ensemble model was significantly better, with an overall accuracy of 99.36%. The confusion matrices in Figure 5 depicts that the ensemble and RF classifiers had similar performance with a 99.36% high classification accuracy due to their efficiency in predicting crops. Moreover, from the confusion matrices, it is noted that the majority of crops were accurately classified without having overlapping amongst the categories. There were some minor inaccuracies in classification, which involved crops with similar climatic and soil nutrient needs.

#### 4.4. Overfitting analysis and model generalization

Although the accuracy level of 99.36% is high, evaluation was done to assess the possible risks associated with overfitting. The number of samples in the dataset was 2,200, which were evenly distributed across 22 crop classes, with each class having 100 samples. During training, the model was trained with k-fold cross-validation during hyperparameter tuning via RandomizedSearchCV. The dataset was also split into a train and test set with a 70-30 ratio for evaluating the performance on new, unseen data. The use of k-fold cross-validation and independent testing improved the generalization capability of the proposed framework and reduced the possibility of overfitting [26], [27], [30]. Furthermore, outlier detection using isolation forest (5% contamination level) reduced noise and improved model robustness. The comparable performance of the stacking model and RF classifier, along with stable confusion matrices across classes,

suggests that the model captures meaningful feature relationships rather than memorizing training samples. However, testing on more extensive agricultural data sets that cover different geographical regions is advised in future studies for further verification of its generalization ability. While the data set comprises 2,200 instances with 22 crop categories, k-fold cross-validation and independent testing have been applied to enhance robustness and prevent overfitting. The future work will involve testing the proposed architecture with multi-regional agricultural data sets under varying climate and soil conditions.

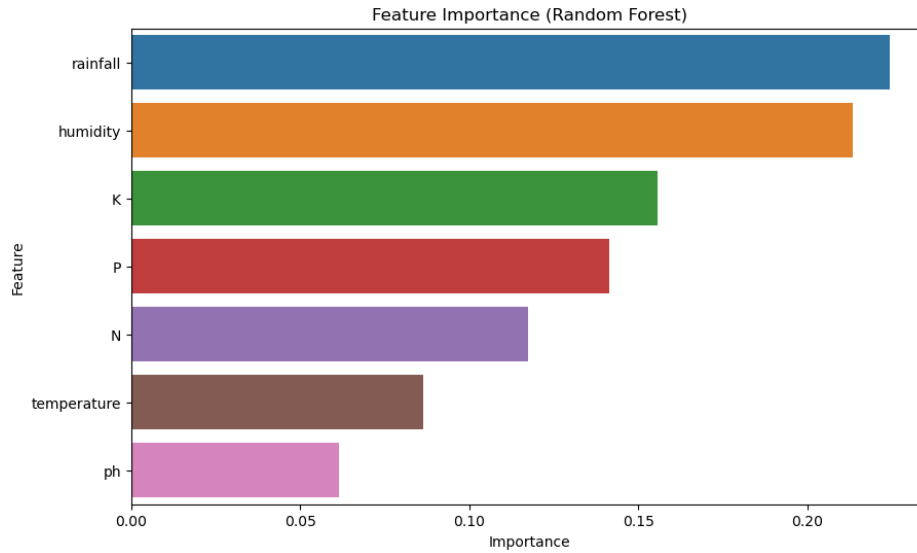


Figure 3. Feature importance of RF

Table 3. The performance metrics of different models

Model	Accuracy	Precision	Recall	F1-score
RF	99.36	0.99	0.99	0.99
DT	97.77	0.98	0.98	0.98
LightGBM	99.20	0.99	0.99	0.99
XGBoost	98.25	0.98	0.98	0.98
SVC	98.88	0.99	0.99	0.99
KNN	97.45	0.98	0.97	0.97
Cross-validated meta-stacking model	99.36	0.99	0.99	0.99

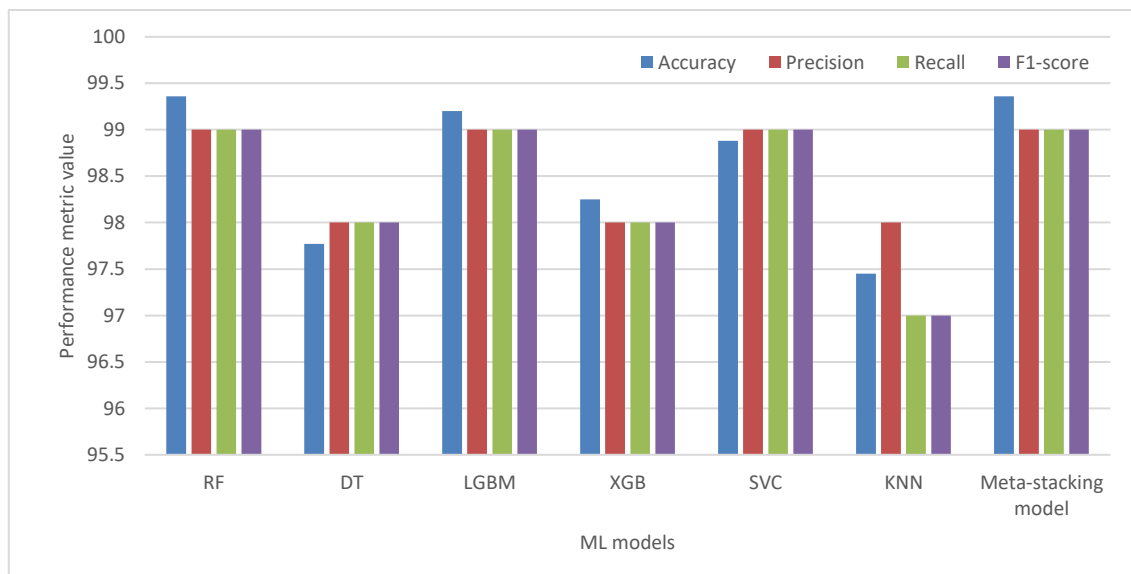


Figure 4. Comparative analysis of accuracy, precision recall, and F1-score across ML models

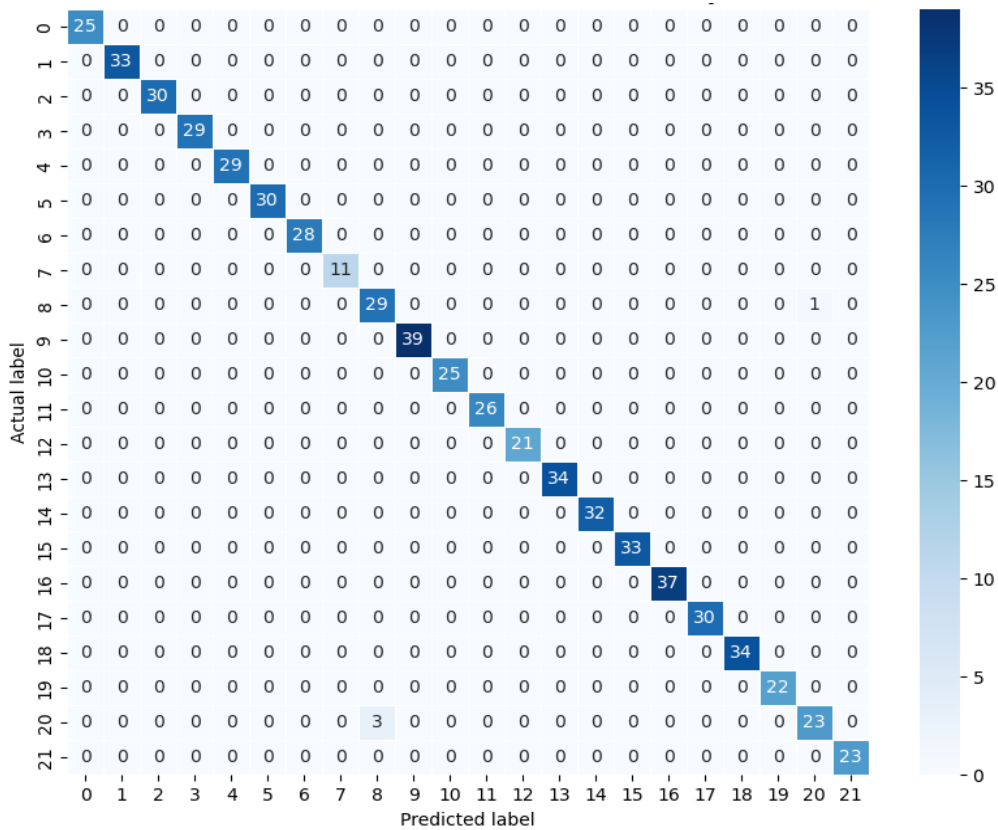


Figure 5. Confusion matrix-ensemble model–accuracy: 99.36%

**5. CONCLUSION**

From this study, a number of ML models have been reported successful. The model that topped out was the basic model: RF model with an accuracy of 99.36%, precision, recall, and F1-score of 0.99. The other models include LightGBM, XGBoost, DT, SVC, and KNN with an accuracy of 97.45%, 98.33%, 97.92%, 97.62%, and 96.67% respectively. With a final accuracy of 99.36% and a precision, recall, and F1-score of 0.99, the stacked ensemble model combines strengths of all the base models but leads to further overall stability and reliability in the predictions, reduces overfitting, and actually leverages diverse learning patterns while having matched the performance of the best base model RF. Future aspects of this research can consist of focusing on crucial areas: first, further environmental factors, such as soil composition and weather forecasts, along with real-time data from sensors, will improve the accuracy and adaptability of the model in changing conditions. In addition, it is important to be able to develop user-friendly applications so that farmers are able to receive crop recommendations from this model and incorporate the insights into their decision-making processes. The use of explainable AI techniques would result in increased model interpretability and help the user understand the reasons for the crop recommendation being made. Moreover, the presented framework may also be used in mobile-based crop advisory systems, IoT-based smart farming systems, and agricultural extension programs to provide real-time crop recommendations and help implement sustainable farming techniques. Additionally, the integration of this framework with soil sensors and environmental monitoring systems would aid the farmer in decision-making related to the crops planted and fertilizers used.

**FUNDING INFORMATION**

Authors state no funding involved.

**AUTHOR CONTRIBUTIONS STATEMENT**

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Gulaganjihalli	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			
Ningegowda Shwetha														
Bhat Geetalaxmi Jairam		✓				✓		✓	✓	✓	✓			

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review &amp; Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are openly available in the Kaggle repository at <https://www.kaggle.com/datasets/aksahaha/crop-recommendation>.




## REFERENCES

- [1] H. C. J. Godfray *et al.*, "Food security: the challenge of feeding 9 billion people," *Science*, vol. 327, no. 5967, pp. 812–818, Feb. 2010, doi: 10.1126/science.1185383.
- [2] S. S. Myers *et al.*, "Climate change and global food systems: potential impacts on food security and undernutrition," *Annual Review of Public Health*, vol. 38, no. 1, pp. 259–277, Mar. 2017, doi: 10.1146/annurev-publhealth-031816-044356.
- [3] A. J. Atapattu, C. S. Ranasinghe, T. D. Nuwarapaksha, S. S. Udumann, and N. S. Dissanayaka, "Sustainable agriculture and sustainable development goals (SDGs)," in *Emerging Technologies and Marketing Strategies for Sustainable Agriculture*, IGI Global Scientific Publishing, 2024, pp. 1–27, doi: 10.4018/979-8-3693-4864-2.ch001.
- [4] E. Elbasi *et al.*, "Crop prediction model using machine learning algorithms," *Applied Sciences*, vol. 13, no. 16, Aug. 2023, doi: 10.3390/app13169288.
- [5] P. S. Kiran, G. Abhinaya, S. Sruti, and N. Padhy, "A machine learning-enabled system for crop recommendation," *Engineering Proceedings*, vol. 67, no. 1, Sep. 2024, doi: 10.3390/engproc2024067051.
- [6] S. Gawdiya *et al.*, "Field scale wheat yield prediction using ensemble machine learning techniques," *Smart Agricultural Technology*, vol. 9, Dec. 2024, doi: 10.1016/j.atech.2024.100543.
- [7] T. Chakraborty, A. Bose, A. Samanta, A. Ghosh, A. Samanta, and K. C. Mondal, "Ensemble machine learning model for better crop production," in *Recent Advances in Artificial Intelligence and Smart Applications (RAISA 2023)*, Singapore: Springer, 2024, pp. 129–141, doi: 10.1007/978-981-97-3485-6\_10.
- [8] M. I. Sadiq, S. M. P. Rahman, S. Kayes, A. H. Sumaita, and N. A. Chisty, "A review on the imaging approaches in agriculture with crop and soil sensing methodologies," in *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Oct. 2021, pp. 1–7, doi: 10.1109/ICDS53782.2021.9626711.
- [9] Y. Akkem, S. K. Biswas, and A. Varanasi, "Smart farming using artificial intelligence: a review," *Engineering Applications of Artificial Intelligence*, vol. 120, Apr. 2023, doi: 10.1016/j.engappai.2023.105899.
- [10] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 8, pp. 5602–5618, Sep. 2022, doi: 10.1016/j.jksuci.2021.05.013.
- [11] P. S. Nishant, P. S. Venkat, B. L. Avinash, and B. Jabber, "Crop yield prediction based on Indian agriculture using machine learning," in *2020 International Conference for Emerging Technology (INCET)*, Jun. 2020, pp. 1–4, doi: 10.1109/INCET49848.2020.9154036.
- [12] N. BanuPriya, D. Tejasvi, and P. Vaishnavi, "Crop yield prediction based on Indian agriculture using machine learning," *International Journal of Modern Agriculture*, vol. 10, no. 3, pp. 73–82, 2021.
- [13] R. C. Mahore and N. G. Gadge, "Design a model for crop prediction and analysis using machine learning," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 9, no. 1, pp. 90–95, Jan. 2023, doi: 10.32628/CSEIT228681.
- [14] Y. Han *et al.*, "Prediction of maize cultivar yield based on machine learning algorithms for precise promotion and planting," *Agricultural and Forest Meteorology*, vol. 355, Aug. 2024, doi: 10.1016/j.agrformet.2024.110123.
- [15] M. Kuradusenge *et al.*, "Crop yield prediction using machine learning models: case of irish potato and maize," *Agriculture*, vol. 13, no. 1, Jan. 2023, doi: 10.3390/agriculture13010225.
- [16] R. Jayalakshmi and M. S. Devi, "Predictive model construction for prediction of soil fertility using decision tree machine learning algorithm," *INFOCOMP Journal of Computer Science*, vol. 20, no. 1, pp. 49–55, 2021.
- [17] J. Zhong and Q. Lai, "Smart agriculture system based on internet of things using kernel K-means with support vector machine," in *2024 Second International Conference on Data Science and Information System (ICDSIS)*, May 2024, pp. 1–4, doi: 10.1109/ICDSIS61070.2024.10594566.
- [18] R. Jaichandran, T. M. Krishna, S. H. Arigela, R. Raman, N. Dharani, and A. Kumar, "Light GBM algorithm based crop recommendation by weather detection and acquired soil nutrients," in *2022 International Conference on Power, Energy, Control and Transmission Systems (ICPECTS)*, Dec. 2022, pp. 1–5, doi: 10.1109/ICPECTS56089.2022.10047765.
- [19] E. B. Wijayanti, D. R. I. M. Setiadi, and B. H. Setyoko, "Dataset analysis and feature characteristics to predict rice production based on extreme gradient boosting," *Journal of Computing Theories and Applications*, vol. 1, no. 3, pp. 299–310, Feb. 2024, doi: 10.62411/jcta.10057.




- [20] G. S. M. Rao, S. Dangeti, and S. S. Amiripalli, "An efficient modeling based on XGBoost and SVM algorithms to predict crop yield," in *Advances in Data Science and Management*, Singapore: Springer, 2022, pp. 565–574, doi: 10.1007/978-981-16-5685-9\_55.
- [21] C. Singha, K. C. Swain, S. Sahoo, and A. Govind, "Prediction of soil nutrients through PLSR and SVMR models by VIs-NIR reflectance spectroscopy," *The Egyptian Journal of Remote Sensing and Space Sciences*, vol. 26, no. 4, pp. 901–918, Dec. 2023, doi: 10.1016/j.ejrs.2023.10.005.
- [22] S. Solehatin, H. Hadiq, and D. A. A. Pertiwi, "Application of the KNN method to check soil compatibility using a microcontroller for android-based banyuwangi citrus fruit plants," *Journal of Soft Computing Exploration*, vol. 4, no. 3, pp. 123–132, Sep. 2023, doi: 10.52465/josce.v4i3.197.
- [23] K. P. K. Devan, B. Swetha, P. U. Sruthi, and S. Varshini, "Crop yield prediction and fertilizer recommendation system using hybrid machine learning algorithms," in *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, Apr. 2023, pp. 171–175, doi: 10.1109/CSNT57126.2023.10134662.
- [24] P. S. S. Gopi and M. Karthikeyan, "Red fox optimization with ensemble recurrent neural network for crop recommendation and yield prediction model," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 13159–13179, 2023, doi: 10.1007/s11042-023-16113-2.
- [25] A. M. Lad, K. M. Bharathi, B. A. Saravanan, and R. Karthik, "Factors affecting agriculture and estimation of crop yield using supervised learning algorithms," *Materials Today: Proceedings*, vol. 62, pp. 4629–4634, 2022, doi: 10.1016/j.matpr.2022.03.080.
- [26] M. Mancera, L. S. Terrissa, S. Ayad, H. Laouz, and N. Zerhouni, "Advancing crop recommendation systems through ensemble learning techniques," in *Innovations in Smart Cities Applications Volume 7*, Cham, Switzerland: Springer, 2024, pp. 45–54, doi: 10.1007/978-3-031-54376-0\_4.
- [27] I. K. Nti, A. Zaman, O. N. -Boateng, A. F. Adekoya, and F. Keyeremeh, "A predictive analytics model for crop suitability and productivity with tree-based ensemble learning," *Decision Analytics Journal*, vol. 8, 2023, doi: 10.1016/j.dajour.2023.100311.
- [28] M. Hasan *et al.*, "Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation," *Frontiers in Plant Science*, vol. 14, Aug. 2023, doi: 10.3389/fpls.2023.1234555.
- [29] N. R. Octavianto and A. Wibowo, "Stacking classifier method for prediction of human body performance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 34, no. 3, pp. 1832–1839, 2024, doi: 10.11591/ijeecs.v34.i3.pp1832-1839.
- [30] M. Lee, "Comparison of bagging, boosting, and stacking ensemble models for airline customer satisfaction analysis," *FaST - Jurnal Sains dan Teknologi (Journal of Science and Technology)*, vol. 8, no. 1, p. 76, May 2024, doi: 10.19166/jstfast.v8i1.8166.
- [31] R. Dey and R. Mathur, "Ensemble learning method using stacking with base learner: a comparison," in *Proceedings of International Conference on Data Analytics and Insights, ICDAI 2023*, Singapore: Springer, 2023, pp. 159–169, doi: 10.1007/978-981-99-3878-0\_14.

## BIOGRAPHIES OF AUTHORS



**Gulaganjihalli Ningegowda Shwetha**    holds a B.E. in Information Science and Engineering and an M.Tech. in Computer Network Engineering from Visvesvaraya Technological University, Belgaum. She has 15 years of teaching experience and is currently an assistant professor in the Department of Computer Science and Engineering at Sri Jayachamarajendra College of Engineering, JSS Science and Technology University, Mysuru. She has worked on a Department of Science and Technology project titled "Efficacy of computer-based training module on central processing disorder". Her research interests include computer networks, internet of things, AI, and ML. She can be contacted at email: shwethagn22@gmail.com.



**Dr. Bhat Geetalaxmi Jairam**    is a professor in the Department of Information Science and Engineering at the National Institute of Engineering, Mysuru. She holds a B.E. in Electronics from Dr. Babasaheb Ambedkar University, Aurangabad, an M.Tech. from Visvesvaraya Technological University (II Rank in CNE-FT), and a Ph.D. from Visvesvaraya Technological University. With 15+ years of teaching experience, her research focuses on wireless sensor networks, AI, and theory of computation. She has received a VGST (K-FIST-Level-I) grant, published 20+ research papers (11 Scopus-indexed), and serves as a Visvesvaraya Technological University research guide and reviewer. She has also won best paper and most innovative paper awards at international conferences. She can be contacted at email: bgj@nie.ac.in.