*As per UGC guidelines an electronic bar code is provided to secure your paper*

*International Journal for Modern Trends in Science and Technology*
*Volume 12, Special Issue 01*, pages 274-279.
ISSN: 2455-3778 online
Available online at: http://ijmtst.com/ICRCSET2026.html
DOI: https://doi.org/10.5281/zenodo.19561846

# Soil and Climatic Parameters-based Intelligent Crop Recommendation System Using ML

**Ch. Sravanthi, K. Kavya, A. Sai, B. Sudhakar Reddy, K. Kalpana**

*Department of ECE, Bapatla Engineering College, Bapatla, Andhra Pradesh, India.*

**To Cite this Article**
Ch. Sravanthi, K. Kavya, A. Sai, B. Sudhakar Reddy & K. Kalpana (2026). Soil and Climatic Parameters-based Intelligent Crop Recommendation System Using ML. International Journal for Modern Trends in Science and Technology, 12(SI01), 274-279. https://doi.org/10.5281/zenodo.19561846

**Article Info**
Received: 02 March 2026; Revised: 01 April 2026; Accepted: 04 April 2026.

**Copyright** © The Authors ; This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*KEYWORDS*

*Machine Learning, Multi-class classification problem, Crop recommendation system, CatBoost classifier, Soil nutrients, Precision agriculture*

## ABSTRACT

*Agriculture is one of the important sectors contributing to the service of the region's economy, and one of the important sectors is the area in which the productivity of the crops is based on the selection of the crops suitable for self-dependent conditions. In general, crop selection models are based on experience and intuition in the decision-making process, which does not ensure the attainment of the optimal crop yield. With the availability of more agricultural data, the decision-making process can be optimized using machine learning algorithms. Smart Crop Recommendation System Using Machine Learning is a project that optimizes the selection of the crop suitable for the region based on the soil nutrient availability and climatic conditions. The project is implemented using the Flask Framework , allowing the system to be used in a user-friendly manner.*

## I. INTRODUCTION

Agriculture is the backbone of human civilization, providing the world with food, raw materials, and a robust economy for human beings. In the case of developing countries, agriculture is the primary source of livelihood for the people, and this aspect plays an important role in the economy of the country [8]. The productivity of crops depends on the quality of the soil, weather, and type of crop. Farmers used to depend on their feelings, knowledge, and experience while choosing the crops, but these methods are not proving to be helpful due to the changing weather, poor quality of soil, and lack of resources [5]. However, these methods have shown poor results, which is why there is a need to use a science-based approach to farming, which would help farmers make profitable decisions [3]. In this regard, machine learning is considered a revolutionary technology that can analyze data, detect complex patterns, and make predictions to assist farmers in making decisions [2]. The history of machine learning is traced back to Samuel's study on 'checkers,' which paved the way for machine learning, showing that computers could learn from their own experiences [1]. Another study by Sutton and Barto, 'reinforcement learning paradigm,' gave a formal view of adaptive decision-making, which led to the creation of machine learning technology, which is being used to solve many problems, including those related to farming. Among these applications, crop recommendation systems are considered to be of critical importance, as these have a direct impact on farmers, which further influences productivity, CatBoost is considered to be more appropriate for use with categorical data related to crops, where there is no need to use label encoding, which is required by other machine learning models, CatBoost is highly effective in solving crop recommendation system problems [15], especially because of its ability to handle complex feature interactions, various studies between 2018 and 2025 have emphasized the importance of machine learning in crop recommendation, Liakos et al. (2018) carried out a comprehensive study on the application of machine learning in agriculture, where they noted its ability to change methodologies, Kulkarni et al. (2021) demonstrated how ensemble methods can help crops grow better, Doshi et al. (2018) and Gosai et al. (2021) designed smart systems that recommended crops and highlighted how soil nutrients and weather changes affect low crop yield [9][10]. Dahiphale et al.

The latest development in the field of DSS is the Smart Crop Recommendation System based on ML. The accuracy and reliability of the system are ensured by the use of modern algorithms such as CatBoost [15]. Web interfaces are used for the development of the system, thus ensuring its accessibility. This is an important factor for DSS [11]. The potential impact of ML on the field of agriculture and the position of the paper in the overall body of work published between 2018 and 2025 are demonstrated. The use of ML in agriculture is expected to become more imperative in the future as the situation with population, resource, and climate issues deteriorates. Crop recommendation systems show the potential of data in agriculture and its use for increasing yields and development.

## II. RELATED WORK

Some researchers have presented models that can be utilized to improve the precision and use of crop recommendations. The use of machine learning in agriculture is one area that has been researched extensively. Agronomic information was incorporated using decision trees or heuristic models, such as rule-based expert systems and statistical models. Even though these models were able to provide a certain level of information, they were not successful in generalizing across different regions, nor were they flexible enough to accommodate changes in environmental factors. More flexible models were made possible by the introduction of machine learning models that could utilize past data to identify relationships between crop yield, nutrients found in the soil, and environmental factors. Decision tree classifiers were one of the first machine learning models to be utilized in crop recommendation. Due to their ability to easily understand how predictions were made, decision tree classifiers were considered to be one of the most viable options to be utilized in this area. However, the major issues of overfitting and low accuracy were major challenges associated with the simple decision trees. This prompted the increased adoption of the Random Forest algorithm in an attempt to overcome the challenges associated with the decision trees. The algorithm was perceived as an efficient solution in the attempt to improve the robustness and ensure the achievement of higher accuracy in the predictions of the crops. When the algorithm was employed in the prediction of data with nutrient values like N, P, and K, and other parameters like pH, rainfall, and temperature, accuracies above 95% were recorded.

## III. PROPOSED METHODOLOGY

### 1. *Dataset Description:*

The Crop Recommendation dataset is a well-structured dataset related to agriculture, which is appropriate for machine learning techniques to predict the best crop to grow. This dataset contains a number of records with seven essential features: nutrients present in the soil (Nitrogen, Phosphorus, Potassium), environmental factors (temperature, humidity, rainfall), and pH value of the soil. This dataset is considered a multi-class classification problem, where the name of the crop is mentioned against each record. Various intricate relationships between soil fertility, environmental factors, and crop growth can be learned using machine learning techniques due to this dataset, which includes various scenarios related to different crops. This dataset is appropriate for training, testing, and evaluating any crop recommendation system, as it is well-balanced with respect to class representation and feature representation.

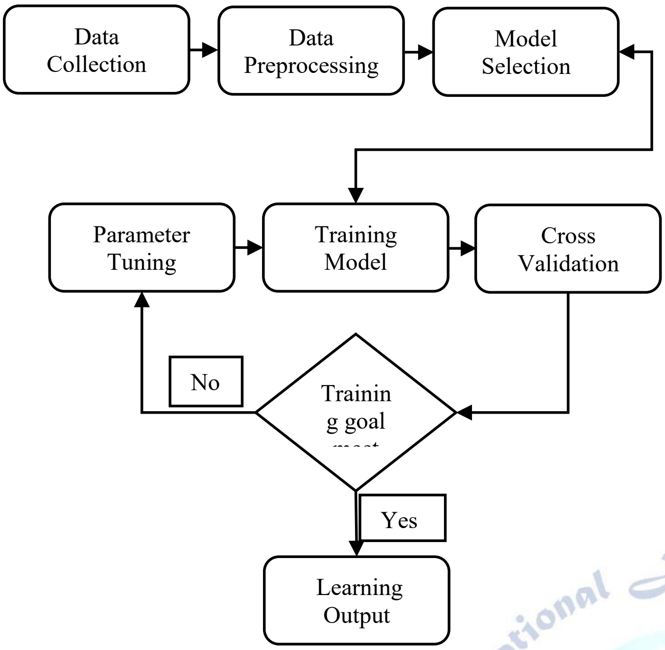

Figure 1. System Architecture

2. *Evaluation Procedure and Applied Models:*

I. Definition of the dataset: Let D = { ($x_1$, $y_1$), ($x_2$, $y_2$),…, ($x_n$, $y_n$) } represent the dataset, where each $x_i \in \Psi^2$ corresponds to the feature vector made up of soil nutrients and climatic parameters: $x_i$ = [N, P, K, temperature, humidity, pH, rainfall and $y_i \in$ C, where C stands for the set of crop label.

II. Split Training and Testing: The dataset D is divided into two disjoint subsets, D_train and D_test, so that D_train and D_test, such that D_train ∪ D_test = D and D_train ∩ D_test = ∅. Usually 20% of the data is used for testing and 80% is used for training.

III. Training model: D_train was subjected to several classifier models:

a) A gradient boosting algorithm tailored for categorical features is called the CatBoost Classifier. It minimizes multiclass loss by building an ensemble of decision trees one after the other.

b) Random Forest Classifier: A group of decision trees with random feature selection at each split that were trained on bootstrapped samples to minimize variance.

c) Support Vector Machine (SVM): A margin based classifier that divides the crop classes by locating hyperplanes in high dimensional space. d) K-Nearest Neighbors (KNN): A non-parametric classifier that labels the k nearest training samples according to the majority classes.

e) Logistic Regression: A linear model that uses the logistic function to estimate the class probabilities; softmax is used to extend it to multi class classification.

| Index | Model Name |
|---|---|
| 1 | CatBoost Classifier |
| 2 | Random Forest |
| 3 | SVM |
| 4 | K-Nearest Neighbour |
| 5 | Logistic Regression |
| 6 | Decision Tree |

TABLE 1: Models used

IV. Performance evaluation: To generate a hypothesis function h_M(x), each model M {CatBoost, Random Forest, SVM, KNN, Logistic Regression} is trained on D_train.

Forecast based on Test Data

The test is used to generate predictions for each model M:

$$S = h\_M(x_i),\ S(x_i,\ y_i) \in D\_test$$

The true labels $y_i$ are compared to the Predicted labels S.

Metrics for Evaluation

Classification metrics are calculated on D_test for every model:

a. Precision (P) is calculated as follows:

P = TP / (TP + FP), where TP stands for true positives and FP for false positives.

b. The accuracy of positive predictions is measured by precision.

c. Recall (R) is calculated as follows:

IJMTST

R = TP / (TP + FN), where FN stands for false negatives.

d. F1 Score (F1): F1 score is calculated as F1 = 2 * (P * R) / (P + R), which is the harmonic mean of precision and recall. The accuracy (A) formula is as follows: A = (correct predictions) / (total predictions made). The area under the receiver operating characteristic curve, i.e., ROCAUC, is used as a measure of the classifier's capability to distinguish between classes.

e. Comparative Evaluation: To compare the performance, the computed metrics {P, R, F1, A, ROC-AUC} for each of the models are tabulated. It is observed that CatBoost performed better in terms of generalization and robustness, achieving the highest values for all the computed metrics when compared with Random Forest, SVM, KNN, and Logistic Regression.

For the crop recommendation, the proposed methodology makes use of an organized machine learning pipeline. The training and evaluation of the CatBoost classifier are performed after the acquisition of the dataset, data preprocessing, and train-test split. For the precise crop recommendation, the optimized model is saved, and the implementation is performed using the Flask-based web application.

The PCA-based crop distribution figure also points to the importance of interactions between features and class separability. In aggregate, these results point to two important benefits that this system possesses: it is extremely accurate, and it is usable in real-time for sustainable agriculture.

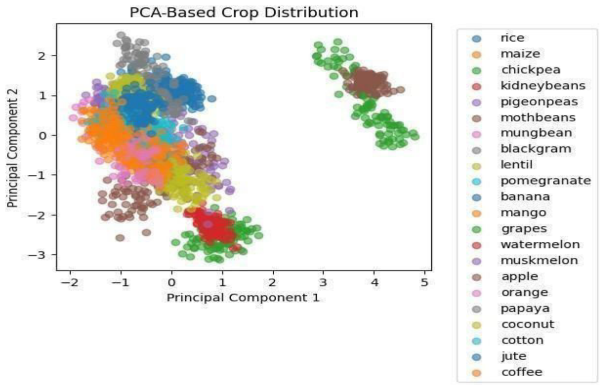

FIGURE 1: PCA-Based Crop Distribution

Here, a two-dimensional scatter plot of the crop samples is shown, projected onto the first two principal components, i.e., PC1 and PC2, using Principal Component Analysis (PCA). Each colored group of data points represents a different type of crop, showing how these crops are distributed and separated according to reduced feature dimensions.

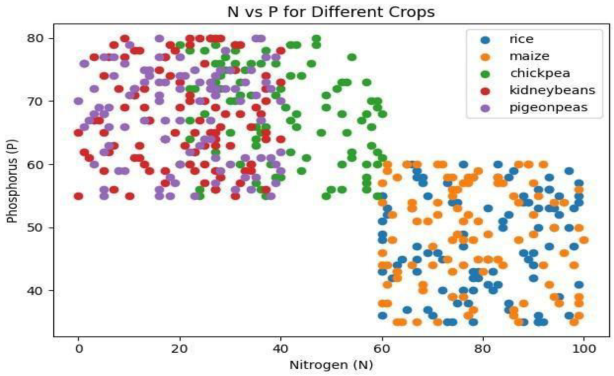

FIGURE 2: N vs P

In the scatter plot (FIGURE. 2), the level of nitrogen and phosphorus is indicated for five crops: rice, maize, chickpea, kidney beans, and pigeon peas. Each crop forms its own clusters, indicating the nutrient requirement for the particular crop.

## IV. RESULTS AND EVALUATION

The supremacy of the CatBoost classifier in dealing with the multiclass crop prediction problem was verified with the exceptional accuracy of the Smart Crop Recommendation System, which was 99.77%. It is evident that the CatBoost classifier outperforms Random Forest, SVM, KNN, and Logistic Regression in all the evaluation metrics, i.e., precision, recall, F1 score, and ROCAUC, as compared in the comparative analysis in Table II. The accuracy of the proposed approach is near perfect, reflecting the robustness of the approach. Agricultural and computational knowledge are reflected in the diagrams above. The average nitrogen content in crops varies from one crop to another, hence facilitating the management of the same for sustainability. The content of coffee and cotton is the highest, while in lentil and mango, the content is the lowest. The comparison of the performance of the models shows that CatBoost and Random Forest are better than SVM and Logistic Regression, hence facilitating accurate predictions in agriculture decision-making. The rainfall, humidity, and

International Journal for Modern Trends in Science and Technology IJMTST

temperature distributions, as illustrated below, show the main aspects of the climate. The rainfall distribution shows right-skewed variations, with the peaks occurring in the moderate range.

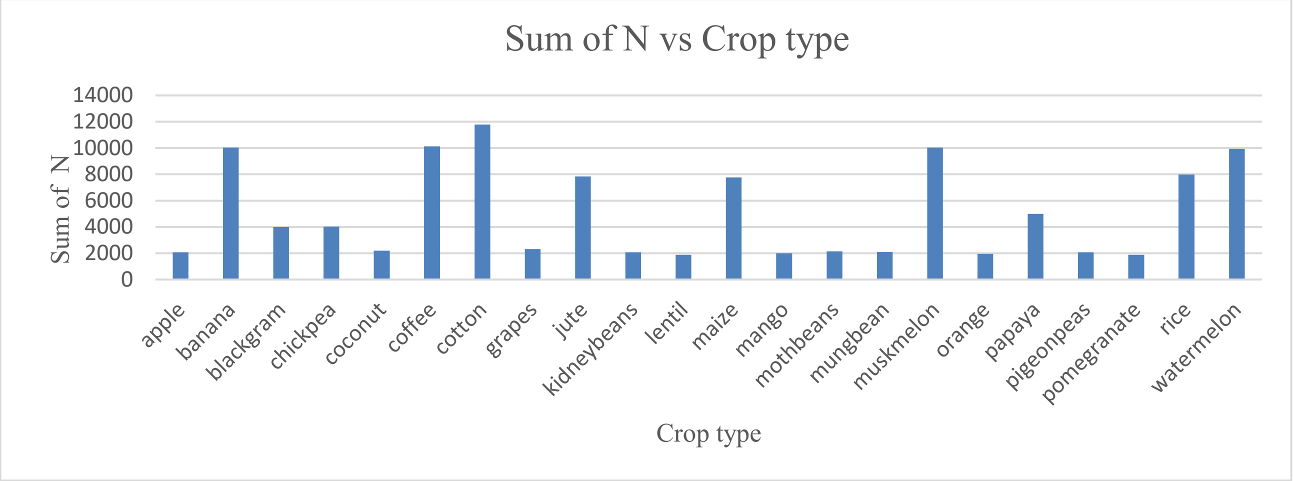

FIGURE 3: Sum of Nitrogen (N) content per Crop type

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) | ROC AUC (%) |
|---|---|---|---|---|---|
| Catboost | 99.77 | 99.78 | 99.77 | 99.77 | 100 |
| Random Forest | 99.55 | 99.57 | 99.55 | 99.55 | 100 |
| SVM | 98.86 | 98.96 | 98.86 | 98.87 | 100 |
| KNN | 98.18 | 98.23 | 98.18 | 98.17 | 99.88 |
| Logistic Regression | 97.05 | 97.25 | 97.05 | 97.07 | 99.94 |
| Decision Tree | 97.9 | 97.95 | 97.9 | 97.88 | 99.7 |

TABLE II: Comparison of models

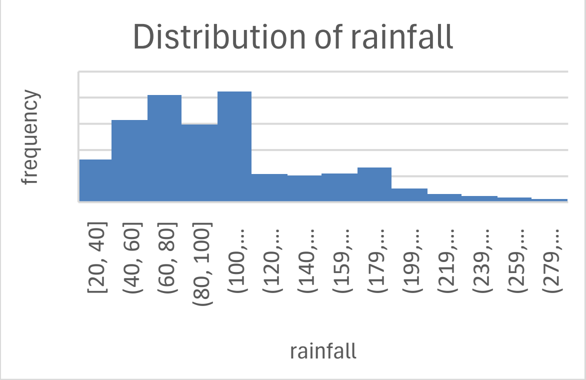

Fig.: 4(a)

These variables help in providing information for the planning of resources, sustainable agriculture, and environmental protection.

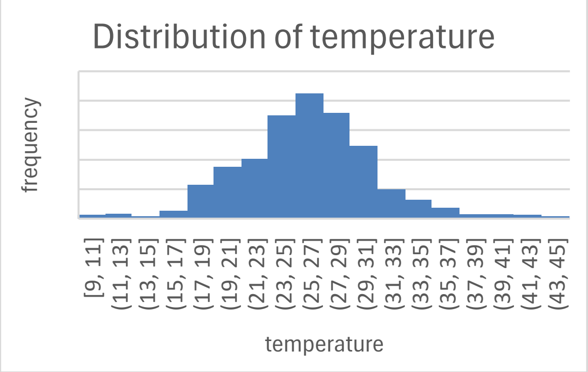

Fig.: 4(b)

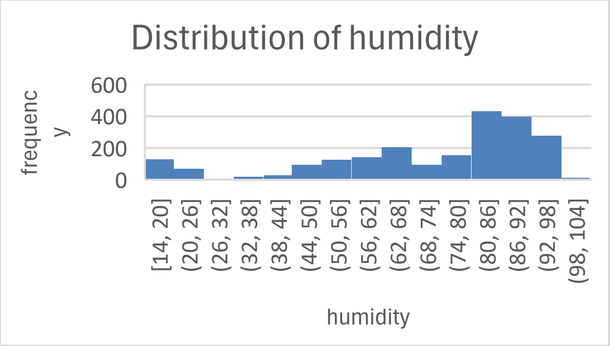

Fig.: 4(c)

FIGURE 4: Distribution of atmospheric parameters

## V. CONCLUSION

The project proposed a Smart Crop Recommendation System that uses categorical boosting, an advanced form of gradient boosting, for the prediction of the best crop based on nutrients and climate conditions. The system performed better than the conventional approach of other classifiers, such as Random Forest, SVM, KNN, and Logistic Regression, with an effective accuracy of 99.77%. The system was able to make predictions based on science by incorporating characteristics such as temperature, humidity, pH, rainfall, nitrogen, phosphorus, and potassium. The system has been extended to a real-time application using a Flask interface, and users can input soil and climate parameters to obtain timely recommendations from the system. Preprocessing and evaluation were aided by tools provided by Scikit-learn, and visualization was simplified by Google Colab, while the entire system was coded in Visual Studio Code for reproducibility. This system provides farmers with a reliable and convenient tool for selecting crops that will not only increase yield and reduce crop failure but will also be sustainable in use.

IJMTST

## Conflict of interest statement

Authors declare that they do not have any conflict of interest.

## REFERENCES

[1] Arthur Samuel, "Some studies in machine learning using the game of checkers," IBM Journal of Research and Development, vol. 3, no. 3, pp. 210–229, 1959.

[2] Richard S. Sutton and Andrew G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[3] Konstantinos G. Liakos, Patrizia Busato, Dimitrios Moshou, Simon Pearson, and Dionysis Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, 2018.

[4] N. Gosai and H. Patel, "Crop recommendation system using machine learning," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 7, no. 3, pp. 245–250, 2021.

[5] R. Doshi, P. Patel, and M. Shah, "Intelligent crop recommendation system using machine learning algorithms," International Journal of Engineering Research & Technology, vol. 7, no. 12, pp. 45–50, 2018.

[6] V. Kulkarni, S. Patil, and A. Deshmukh, "Improving crop productivity through a crop recommendation system using ensembling technique," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2021.

[7] S. Dahiphale and A. Deshpande, "Smart farming using machine learning techniques: Challenges and future directions," Procedia Computer Science, vol. 218, pp. 345–352, 2023.

[8] Food and Agriculture Organization, The State of Food and Agriculture, Rome, Italy: FAO, 2020.

[9] R. Reddy and K. Ramesh, "Logistic regression-based crop prediction model for agricultural datasets," International Journal of Data Science and Machine Learning, vol. 5, no. 2, pp. 55–61, 2022.

[10] H. Patel and R. Mehta, "K-nearest neighbors in agricultural prediction systems," International Journal of Data Science and Machine Learning, vol. 5, no. 2, pp. 55–61, 2021Y. Yaghmazadeh et al.

[11] T S. Sharma, P. Verma, and A. Singh, "Hybrid machine learning approaches for crop recommendation and yield optimization," IEEE Access, vol. 13, pp. 11234–11245, 2025.

[12] N. Gosai, H. Patel, and M. Shah, "Real-time crop recommendation using weather API and machine learning," Computers and Electronics in Agriculture, vol. 210, p. 108021, 2024.

[13] R. Doshi, P. Patel, and M. Shah, "Explainable AI framework for crop recommendation systems," IEEE Transactions on Artificial Intelligence, vol. 6, no. 2, pp. 300–310, 2024.

[14] H. Patel, R. Mehta, and S. Kulkarni, "Hybrid crop recommendation model using CatBoost and deep learning," Journal of Agricultural Informatics, vol. 15, no. 1, pp. 50–60, 2025.

[15] K. Reddy, P. Ramesh, and S. Kumar, "Transfer learning for region-independent crop recommendation systems," IEEE Access, vol. 12, pp. 84520–84532, 2024.