A Comparative Study of Machine Learning Algorithms for AI-Based Crop Recommendation Using Soil and Climatic Parameters

B. Jampanna[1], Dr. Bonthu Kotaiah[2]
[1]*Department of Computer Science, Central Tribal University of Andhra Pradesh, Kondkarakam, Vizianagaram - 535005, India*
[2]*Associate Professor and Head, Department of Computer Science, Central Tribal, University of Andhra Pradesh, Vizianagaram², India*

***Abstract*—Choosing the right crop to grow can be one of the most important decisions a farmer makes. Usually, this decision is done out of habit. This paper treats crop selection as a multi-class classification task and compares four supervised learners: Decision Tree, K-Nearest Neighbours (KNN), Support Vector Machine (SVM) and Random Forest. Each model receives seven inputs: The models are given inputs of nitrogen, phosphorus, potassium, temperature, relative humidity, soil pH and rainfall. The output is the ccrop label. All experiments were run in Python with Scikit-learn on Google Colab, using one shared preprocessing pipeline and an identical stratified train/test split, so that score differences reflect the algorithms and not the setup. Performance is reported as accuracy, precision, recall, F1-score and a confusion matrix. In our runs Random Forest scored highest (accuracy 0.992), ahead of the Decision Tree (0.985), SVM (0.979) and KNN (0.978). The results suggest that machine learning can serve as a practical decision-support aid for farmers and advisers.**

***Index Terms*—crop recommendation, precision agriculture, machine learning, Random0 Forest, Decision Tree, KNN, SVM, soil parameters, climatic parameters**

I. INTRODUCTION

Farm output depends on many interacting conditions: what the soil contains, how much water arrives, and how warm and humid the season is. Among the choices a grower controls, the crop itself matters most, because each species has its own tolerance for nutrient levels, acidity, temperature and moisture. In practice, most farmers rely on what they have seen work on their own land, on seasonal patterns they remember, and on advice from extension workers. This knowledge is valuable, but it is hard to apply consistently when several measurements must be weighed together. A farmer would need to know the soil's nitrogen, phosphorus and potassium content, its pH, and the expected temperature, humidity and rainfall, and then judge how all of these combine for each candidate crop. Data-driven methods suit this problem well. If past records link soil and weather readings to crops that did well under them, a learning algorithm can pick up the pattern and apply it to new readings. Such a tool is best seen as a second opinion for farmers, researchers and advisory services.

This work frames the task as multi-class classification with seven numeric features and one categorical target, and compares four widely used classifiers under the same conditions. II. Related work, III.data , IV. Problem statement states the problem, objectives and research gap, Section V presents the methodology, Section VI reports results, and directions. VI. results VII. Conclusions limitations and Future directions.

II. RELATED WORK

A. Machine Learning in Precision Agriculture

The use of machine learning techniques has become quite widespread inprecision agriculture. Published work covers yield prediction, disease detection, soil classification, fertiliser advice and crop recommendation, with surveys such as Liakos et al. [7] and Kamilaris and Prenafeta-Boldú [8] summarising the field. Systematic reviews of yield prediction are given by van Klompenburg et al. [6]. On crop recommendation specifically, earlier studies have

applied ensemble and classical learners to soil and weather attributes [1]–[5]. Pudumalar et al. [1] proposed a crop recommendation system for precision agriculture built on an ensemble model. Kumar et al. [2] presented a machine learning based crop selection method aimed at maximising crop yield rate. Rajak et al. [3] developed a recommendation system that uses machine learning to maximise crop yield. Kulkarni et al. [4] applied an ensembling technique to improve crop productivity through crop recommendation. Bondre and Mahagaonkar [5] used machine learning algorithms to predict crop yield and recommend fertiliser. Together, these studies show growing interest in data-driven crop advice, but they differ in data and methods.

B. Classification Algorithms

The four algorithms studied here make different assumptions. A Decision Tree [11] splits the data on one feature at a time, producing rules that a person can read. KNN [12] labels a sample by looking at its closest neighbours, so it works well when samples of the same crop cluster together. SVM [10] searches for a boundary that separates classes with as wide a margin as possible. Random Forest [9] builds many trees on resampled data and combines their votes, which helps when relationships between features are non-linear.

C. Why Another Comparison

Reported accuracies depend heavily on the dataset, the number of crop classes, feature scaling, hyperparameters and the evaluation protocol, so results from different papers cannot be compared directly. This study addresses that by running all four models in a single controlled experiment.

III. DATASET

The dataset contains one record per observation, each with soil and climate measurements and the crop suited to those conditions. The experiments use the publicly available Crop Recommendation Dataset on Kaggle [14], which contains 2,200 records covering 22 crop classes with 100 records per class, so the classes are balanced. The seven features are:

- N, P, K - Nitrogen, Phosphorous and Potassium present in soil. 5 5 ,macronutrients, which crops need in different amounts.
- Temperature: ambient temperature.
- Humidity: relative humidity of the air.
- pH: soil acidity or alkalinity; each crop has a preferred range.
- Rainfall: water received from rain.

The target is the crop label. Formally, the model maps (N, P, K, temperature, humidity, pH, rainfall) to a recommended crop. Before modelling, the data was checked for missing values, duplicate rows, wrong data types and outliers. No missing values or duplicate records were found, and all seven features are numeric, so no imputation or encoding of inputs was needed. Features were then separated from the target.

IV. PROBLEM STATEMENT AND OBJECTIVES

Choosing a crop is difficult because several soil and climate factors act together, and many farmers cannot easily get the analysis needed to weigh them. A poor choice can mean low yield, wasted water and fertiliser, and financial loss. The problem addressed here is to build a model that recommends a suitable crop from the seven inputs above. It is intended to support, not replace, agricultural professionals.

The objectives are to:

1. train Decision Tree, KNN, SVM and Random Forest classifiers on one dataset with a common methodology;
2. compare them using; Accuracy Precision Recall F1-score Confusion Matrix Feature Importance Best Model
3. determine which model performs best on this dataset; and
4. examine which input parameters matter most.

Research gap. Many studies apply machine learning to crop recommendation, but they differ in data, features, preprocessing, models and metrics, which makes their results hard to compare. To the best of our knowledge, this work contributes a like-for-like comparison of four common algorithms on a single dataset under one preprocessing pipeline.

V. PROPOSED MODEL AND METHODOLOGY

A. System Overview

The workflow (Fig. 1) has these stages: load the dataset; clean it; select features and target; split into training and test sets (80/20, stratified); train each model separately; predict on the test set; compute

metrics; produce the confusion matrix and feature importances; and compare the models to choose the best.

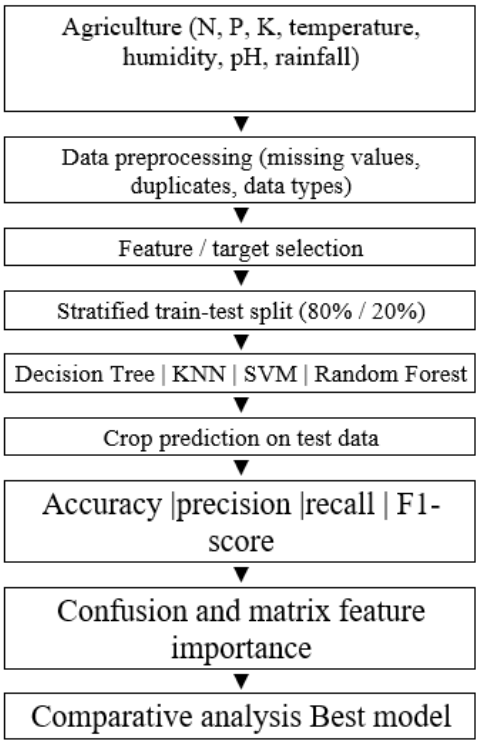


Fig. 1. Flow of the proposed crop recommendation system.

B. Implementation Environment and Data Split

Experiments ran on Google Colab, which provides a ready Python environment in the browser. Pandas and NumPy handled data manipulation, and Scikit-learn [13] provided the models and metrics. The data was split into 80% for training and 20% for testingwith stratified sampling, so each crop keeps its share in both parts (random state 42).

C. Machine Learning Models

Decision Tree. Builds a tree of single-feature tests until each leaf gives a crop label. Default parameters, random state 42.

K-Nearest Neighbours. Assigns the majority crop among the five closest training samples. Because distances depend on feature scale, a StandardScaler was applied, with k = 5.

Support Vector Machine. Finds class boundaries with maximum margin. As SVM is also scale-sensitive, a StandardScaler was placed in the pipeline; default SVC settings were used.

Random Forest. An ensemble of 100 trees with random state 42, chosen because soil and climate features can interact in complex ways.

Table I. Model Configuration

Model	Configuration
Decision Tree	Default parameters, random state = 42
KNN	StandardScaler, k = 5
SVM	StandardScaler, default SVC parameters
Random Forest	100 estimators, random state = 42

D. Evaluation

Accuracy gives the share of correct predictions, but with many crop classes it can hide weak spots. Precision, recall and F1-score were therefore also computed, using weighted averages across classes. A confusion matrix shows exactly which crops are mistaken for each other.

VI. RESULTS AND DISCUSSION

A. Comparison

The four algorithms were evaluated Performance of the models on test set (Table II).

Table II. Algorithm Performance on Test Set

Algorithm	Accuracy	precision	Recall	F1-score
Decision Tree	0.985	0.986	0.985	0.985
KNN	0.978	0.979	0.978	0.978
SVM	0.979	0.980	0.979	0.979
Random Forest	0.992	0.993	0.992	0.992

Random Forest was best on every metric. The Decision Tree came second, and SVM and KNN were close to each other at the bottom, though all four exceeded 97%.

B. Confusion Matrix

Fig. 2 shows, the confusion matrix for the best performing model is shown. Diagonal entries count correctly classified test samples, and off-diagonal entries show crops that were mistaken for one another. Because all four models score above 97%, only a few crops are misclassified, and any confusion is most likely between crops with similar nutrient and climate requirements.

C. Feature Importance

Fig. 3 shows the importance assigned to each of the seven inputs by the Random Forest model. Features with higher importance contribute more to separating the crops, so they are the measurements a farmer or adviser should prioritise collecting accurately.

D. Discussion

A plausible reason for Random Forest's lead is that averaging many trees handles non-linear feature interactions and reduces the variance of a single tree. This is a hypothesis, not something the experiment proves. The confusion matrix and feature importances in Figs. 2 and 3 indicate where the remaining errors occur and which inputs drive the predictions. Differences from numbers in other papers may come from the dataset, preprocessing, hyperparameters and evaluation method.

E. Limitations.

The models may not transfer to other regions. Real crop suitability also depends on soil texture, irrigation, crop variety, pests, disease and farming practice, none of which are inputs here. Scores this high may also reflect a clean, well-separated dataset, and results from a single split should be confirmed with cross-validation [15].

VII. CONCLUSION AND FUTURE WORK

This paper compared Decision Tree, KNN, SVM and Random Forest for recommending crops from soil nutrients, pH and climate. Using one pipeline in Python and Scikit-learn on Google Colab, and judging by Random Forest gave the highest accuracy, precision, recall and F1-score along with a confusion matrix. Performed best on the chosen dataset.

The results show that machine learning can help with agricultural decisions, but it should support, not replace, the knowledge of farmers and experts. Reliability depends on the quality, representativeness and regional relevance of the training data.

Future work could use live weather APIs for current temperature, humidity and rainfall, along with IoT sensors for real-time soil data. The system could be delivered as a web or mobile app giving crop, fertiliser and irrigation advice, with explainable AI to show why a crop was recommended. It should also be tested on region-specific datasets and real field data.

REFERENCES

[1]. S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in Proc. 8th Int. Conf. Advanced Computing (ICoAC), IEEE, 2016, pp. 32–36.

[2]. R. Kumar, M. P. Singh, P. Kumar, and J. P. Singh, "Crop selection method to maximize crop yield rate using machine learning technique," in Proc. Int. Conf. Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), IEEE, 2015, pp. 138–145.

[3]. R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod, and A. Devare, "Crop recommendation system to maximize crop yield using machine learning technique," Int. Res. J. Eng. Technol., vol. 4, no. 12, pp. 950–953, 2017.

[4]. N. H. Kulkarni, G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery, "Improving crop productivity through crop recommendation system using ensembling technique," in Proc. 3rd Int. Conf. Computational Systems and Information Technology for Sustainable Solutions (CSITSS), IEEE, 2018, pp. 114–119.

[5]. D. A. Bondre and S. Mahagaonkar, "Prediction of crop yield and fertilizer recommendation using machine learning algorithms," Int. J. Eng. Appl. Sci. Technol., vol. 4, no. 5, pp. 371–376, 2019.

[6]. T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," Comput. Electron. Agric., vol. 177, Art. no. 105709, 2020.

[7]. K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, Art. no. 2674, 2018.

[8]. A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Comput. Electron. Agric., vol. 147, pp. 70–90, 2018.

[9]. L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[10]. C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[11]. J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986.

[12]. T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, 1967.
[13]. F. Pedregosa et al., "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.
[14]. A. Ingle, "Crop Recommendation Dataset," Kaggle, 2020.
[15]. R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. 14th Int. Joint Conf. Artificial Intelligence (IJCAI), vol. 2, 1995, pp. 1137–1145.
[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, Art. no. 100880, 2023, doi: 10.1016/j.jafr.2023.100880.
[17]. 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, Art. no. e25112, 2024, doi: 10.1016/j.heliyon. 2024.e25112.
[18]. G. Tomar, G. Sharma, R. Rajkumar, A. Singh, S. S. Dhakare, and K. Kumar, "Development of a machine learning-based system for optimizing crop recommendations," in Proc. 1st Int. Conf. Advanced Computing, Communication and Networks (ICAC2N), IEEE, 2024, doi: 10.1109/ICAC2N63387.2024.10895212.
[19]. "Crop recommendation system based on soil quality and environmental factors using machine learning," in Proc. Int. Conf. Emerging Innovations and Advances in Computing (INNOCOMP), IEEE, 2024, doi: 10.1109/INNOCOMP63224.2024.00089.
[20]. M. S. B. Alam, V. Esichaikul, A. Lameesa, S. F. Ahmed, and A. H. Gandomi, "An approach for crop recommendation with uncertainty quantification based on machine learning for sustainable agricultural decision-making," Results Eng., vol. 26, Art. no. 105505, 2025, doi: 10.1016/j.rineng.2025.105505.
[21]. S. Tripathy, P. S. Rath, D. R. D. Adhikary, et al., "Leveraging nonlinear deep learning models for intelligent crop recommendation in precision agriculture," Discover Appl. Sci., 2026, doi: 10.1007/s42452-026-09248-y.
[22]. A. Kadu and K. T. V. Reddy, "Development of a machine learning based mobile and web crop recommendation system for precision farming in Wardha," J. Comput. Sci., vol. 22, no. 3, pp. 1045–1059, 2026, doi: 10.3844/jcssp.2026.1045.1059.
[23]. M. Murthy and S. Kulkarni, "Climate-smart agriculture using IoT-enabled machine learning for crop recommendation," Cogent Food Agric., vol. 12, 2026, doi: 10.1080/23311932.2026.2717436.
[24]. V. Balaji, K. S. Balamurugan, T. Thanarajan, and A. M. S. Sundarapandi, "Machine learning-based crop recommendation system integrating soil properties and weather conditions with IoT-driven data collection," Agrociencia, 2026, doi: 10.47163/agrociencia. v60i1.3428.
[25]. L. Narasimha Reddy and P. Kadiri, "Enhancing crop recommendation through graph attention and tabular deep learning on soil-climate datasets," Discover Comput., vol. 29, Art. no. 570, 2026, doi: 10.1007/s10791-026-10487-1.