A Machine Learning Framework for Climate Adaptive Crop Selection Using Multi-Parameter Soil and Environmental Data Analysis

Mr. Y. Kiran Kumar [1], Mr. G. Lakshmikanth [2]

[1] PG Scholar, Sree Rama Engineering College (Autonomous), Tirupati, India.
Email: yatakonakiran2@gmail.com
[2] Mr. G. Lakshmikanth, Assistant Professor, Department of CSE, Sree Rama Engineering College (Autonomous), Tirupati.
Email: svlakshmikanth21@gmail.com

Abstract: A major challenge in modern agriculture is the accurate selection of crops under dynamically changing climate conditions and varying soil characteristics. This project proposes a Machine Learning Framework for Climate-Adaptive Crop Selection integrating seven key agronomic parameters Nitrogen (N), Phosphorus (P), Potassium (K), soil pH, temperature, humidity, and rainfall into a unified predictive model. Six supervised classification algorithms are systematically evaluated: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), Naive Bayes (NB), Logistic Regression (LR) and K-Nearest Neighbor (KNN). The proposed framework is validated on a structured agricultural dataset of 2,200 labelled records encompassing 22 crop classes. Experimental results demonstrate that Naive Bayes achieves the highest classification accuracy, with precision, recall, and F1-score. Feature importance analysis identifies rainfall, humidity, and temperature as the dominant predictors of crop suitability. The framework offers a scalable, data-driven decision-support tool to advance precision agriculture and climate-resilient food systems.

Keywords: *Climate-Adaptive Agriculture, Crop Recommendation, Machine Learning, Random Forest, Soil Analysis, Decision Support System, Supervised Classification.*

I. INTRODUCTION

Agriculture sustains the livelihoods of over 2.5 billion people globally and accounts for approximately 70% of freshwater withdrawals worldwide [1,2]. The accurate selection of crop varieties suited to prevailing soil and climate conditions represents one of the most consequential decisions in agricultural management, directly determining yield potential, input efficiency, and long-term land sustainability. Crop selection errors arising from poor soil-climate matching result in annual production losses estimated at 20- 40% of potential yields in developing agricultural economies [3,4].

Conventional crop selection methodologies rely predominantly on heuristic empirical knowledge, static agronomic guidelines, and generational farming experience [5]. While culturally valuable, such approaches are fundamentally ill-equipped to process the high-dimensional, nonlinear interactions between soil physicochemical properties and dynamic climatic variables that govern crop suitability. The intensification of climate change has further exacerbated this limitation by introducing unprecedented variability into rainfall patterns, temperature regimes, and humidity distributions the primary environmental determinants of crop performance [6].

Machine learning (ML) has emerged as a transformative paradigm for precision agricultural analytics, demonstrating remarkable capacity to detect complex multivariate patterns in agronomic datasets [7]. Supervised classification algorithms which learn predictive mappings from labeled training data are particularly well suited to crop recommendation tasks where the objective is to classify a given soil-climate feature vector into the optimal crop category [8]. Despite growing interest, the agricultural ML literature lacks a systematic comparative evaluation of classical classification algorithms under a standardized seven-parameter agronomic feature space [9].

This paper addresses this gap by presenting a comprehensive Machine Learning Framework for Climate-Adaptive Crop Selection. The principal contributions of this work are: (i) A systematic integration of four soil physicochemical parameters and three climatic variables into a unified seven dimensional feature space (ii) A rigorous comparative evaluation of six ML classification algorithms under a standardized experimental protocol (iii) Feature importance analysis identifying the primary agronomic predictors of crop suitability and (iv) Empirical demonstration of the framework's practical viability for real world agricultural decision support [10].

The remainder of this paper is organized as follows. Section II reviews related work. Section III describes the proposed framework architecture. Section IV presents the mathematical formulations. Section V details the experimental methodology. Section VI reports and analyzes results and limitations. Section VII concludes the paper.

II. RELATED WORK

Over the past ten years, a lot of research has focused on using machine learning to recommend crops and predict yields. Pudumalar et al. [11] showed that ensemble methods, especially Random Forest, consistently do better than single decision tree classifiers at recommending crops in multiple classes by using bootstrapped aggregation to avoid overfitting. Their benchmark study set basic performance standards for crop advisory systems that use classification.

Ramesh and Vardhan [12] looked into how data mining techniques could be used to help make decisions about farming in semi-arid areas of India. They found that N, P, K, and soil pH were the most important factors for predicting which crops would grow best. Veenadhari et al. [13] utilized J48 decision tree classifiers on climatic agronomic datasets, achieving classification accuracies of 93.2% in predicting crop yield outcomes. Their work showed that climate variables, especially rainfall and temperature, can be used as main predictors.

Sharma et al. [14] created a crop advisory system based on a deep neural network that was trained on multi-spectral satellite images and ground truth soil data. The system was able to classify crops with an accuracy of over 94%. But deep learning architectures are hard to use in agricultural settings where resources are limited because they are hard to compute and hard to understand. Gonzalez-Sanchez et al. [15] showed that SVM classifiers work better than traditional regression methods when predicting crop yields in high dimensional spaces. Liakos et al. [7] performed an exhaustive review of machine learning applications in agriculture, determining that ensemble methods exemplify the pinnacle of structured agronomic data classification. This study builds on and combines the results of these previous studies by using a unified seven parameter feature space for rigorous multi-algorithm benchmarking [16].

III. PROPOSED FRAMEWORK ARCHITECTURE

The suggested framework works like a five-stage sequential pipeline, as shown in Figure 1. It takes raw soil and environmental data and turns it into useful crop recommendations through the stages of data acquisition, pre-processing, feature engineering, model training, and recommendation generation [1].

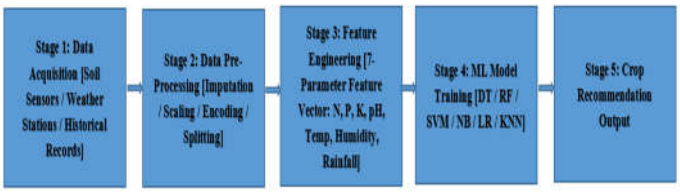

Fig. 1. System Architecture of the Proposed ML Framework for Climate-Adaptive Crop Selection

First up, gathering soil data happens using lab tests, sensors out in the fields, alongside weather station logs. After that, raw numbers get cleaned, scaled, then split into sets for later steps. Instead of just feeding inputs straight ahead, a tailored set of seven farming related features gets built. Following this step, six different machine learning models learn patterns via careful validation splits. Lastly, one clear crop choice pops out, tagged with how sure the system feels about it [17].

The input feature vector x = [N, P, K, pH, T, H, R]$^T \in \mathbb{R}^7$ is defined over seven agronomic dimensions. Soil nutrient concentrations N, P, K are measured in parts per million (ppm), pH is measured on the standard logarithmic 0-14 scale, ambient temperature T in degrees Celsius, relative humidity H in percentage and annual rainfall R in millimeters. This seven-dimensional representation constitutes the minimal sufficient feature set for discriminating 22 major crop classes under diverse agro-climatic conditions [11].

IV.MATHEMATICAL FORMULATIONS

A. Data Pre-Processing: Z-Score Standardization

Feature standardization is applied to all seven continuous input variables to achieve zero-mean, unit-variance distributions prior to model training [12]. For each feature *j*, the standardized value *z* is computed as:

$$z^{(j)} = \frac{x^{(j)} - \mu^{(j)}}{\sigma^{(j)}} \qquad (1)$$

where $\mu^{(j)} = \frac{1}{n}\sum_{i=1}^{n} x_i^{(j)}$ is the sample mean and $\sigma^{(j)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i^{(j)} - \mu^{(j)}\right)^2}$ is the sample standard deviation of feature *j* across *n* training samples [12].

B. Decision Tree: Gini Impurity

The Decision Tree classifier partitions the feature space by selecting split thresholds that minimize Gini impurity at each internal node [11]. The Gini impurity *G(D)* of a dataset partition *D* containing *K* = 22 crop classes is defined as:

$$G(D) = 1 - \sum_{k=1}^{K} p^2_k \qquad (2)$$

where p_k denotes the proportion of samples belonging to class k in partition D. The optimal split threshold t^* for feature j is determined by minimizing the weighted Gini impurity after partitioning:

$$t^* = \arg\min_t \left(\frac{|D_L|}{|D|} G(D_L) + \frac{|D_R|}{|D|} G(D_R) \right) \qquad (3)$$

where D_L and D_R denote the left and right child partitions created by threshold t [11].

C. Random Forest: Ensemble Aggregation

Random Forest constructs an ensemble of B decision trees $\{h_1(x), h_2(x), ..., h_B(x)\}$ each trained on a bootstrapped subsample of the training data with random feature subsets [17]. The final crop prediction is determined by majority voting:

$$\hat{y}\, RF(x) = \text{mod}\, e\{h_b(x)\}_{b=1}^{B} \qquad (4)$$

Out-of-bag (OOB) error estimation provides an unbiased performance estimate without requiring a separate validation set. Feature importance *FI(j)* for feature j is computed as the mean decrease in Gini impurity weighted by the number of samples reaching each node across all trees:

$$FI(j) = \frac{1}{B} \sum_{b=1}^{B} \sum_{t \in h_{(j)}} \frac{|D_t|}{n} . \Delta G(t) \qquad (5)$$

D. Support Vector Machine: Kernel Optimization

The SVM classifier identifies the optimal separating hyperplane $w^T \phi(x) + c = 0$ in a kernel-induced feature space by maximizing the margin between class boundaries [15]. The primal optimization problem is:

$$\min_{w,c,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i \qquad (6)$$

subject to: $y_i (w^T \phi(x_i) + c) \geq 1 - \xi_i, \xi_i \geq 0$, where C is the regularization parameter controlling the bias-variance trade-off and x_i are slack variables permitting soft-margin classification [15].

The Radial Basis Function (RBF) kernel $K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2)$ is employed to map the seven-dimensional agronomic feature space into a higher-dimensional kernel space, enabling nonlinear class boundary estimation [15].

E. Naive Bayes: Posterior Probability Estimation

The Gaussian Naive Bayes classifier applies Bayes' theorem under the conditional independence assumption, estimating the posterior probability of each crop class C_k given the observed feature vector x as [13].

$$P(C_k \mid x) = \frac{P(C_k) \prod_{j=1}^{7} P(x_j \mid C_k)}{P(x)} \qquad (7)$$

Each conditional likelihood $P(x_j \mid C_k)$ is modeled by a Gaussian distribution.

$$P(x_j \mid C_k) = \frac{1}{\sqrt{2\pi\sigma_{kj}^2}} \exp\left(-\frac{(x_j - \mu_{kj})^2}{2\sigma_{kj}^2}\right) \quad (8)$$

Where μ_{kj} and σ_{kj}^2 denote the class-conditional mean and variance of feature *j* estimated from the training data [13].

F. Logistic Regression: Softmax Classification

For the 22-class crop recommendation task, Multinomial Logistic Regression employs the softmax function to estimate class posterior probabilities:

$$P(y = k \mid x) = \frac{\exp(w_k^T x + b_k)}{\sum_{j=1}^{K} \exp(w_j^T x + b_j)} \quad (9)$$

Parameters are optimized by minimizing the cross-entropy loss *L* over the training set:

$$L(W) = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{K} y_{ik} \log P(y = k \mid x_i) \quad (10)$$

with L2 regularization applied to prevent overfitting [8].

G. K-Nearest Neighbor: Distance-Based Classification

KNN classifies a query sample x_q by identifying its *k* nearest neighbors in the training set using Euclidean distance:

$$d(x_q, x_i) = \sqrt{\sum_{j=1}^{7} \left(x_q^{(j)} - x_i^{(j)}\right)^2} \quad (11)$$

The predicted crop class is determined by majority voting among the $k = 5$ nearest neighbors:

$$\hat{y}(x_q) = \arg\max_k \sum_{i \in N_k(x_q)} I(y_i = k) \quad (12)$$

where $N_k(x_q)$ denotes the set of *k* nearest training samples to query x_q and *I(.)* is the indicator function [9].

H. Performance Evaluation Metrics

Model performance is quantified using four standard classification metrics computed from the confusion matrix entries: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [16]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (13)$$

$$\text{Pr}ecision = \frac{TP}{TP + FP} \quad (14)$$

$$\text{Re}call = \frac{TP}{TP + FN} \quad (15)$$

$$F1 - Score = \frac{2.(\text{Pr}ecision.\text{Re}call)}{\text{Pr}ecision + \text{Re}call} \quad (16)$$

For the 22-class setting, macro-averaged precision, recall, and F1-score are reported, computing each metric independently per class and averaging uniformly across all 22 crop categories [16].

V.EXPERIMENTAL METHODOLOGY

A. Dataset Description

The experimental evaluation employs a structured agricultural dataset comprising 2,200 records with balanced class distribution across 22 major crop categories [18]. Crops include staple grains (rice, wheat, maize), legumes (chickpea, kidney beans, pigeon peas, moth beans, mung beans, black gram, lentil), fruits (pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut), and cash crops (cotton, jute, coffee). Each record contains measurements of seven agronomic features: N, P, K (in ppm), pH, temperature (°C), humidity (%), and rainfall (mm). The balanced design 100 records per class ensures equal prior class probabilities and eliminates class-imbalance bias from the evaluation protocol.

Table I: Dataset Feature Statistics

Feature	Min	Max	Mean	Std Dev	Unit	Importance Rank
Nitrogen (N)	0	140	50.55	36.92	ppm	5th
Phosphorus (P)	5	145	53.36	32.99	ppm	6th
Potassium (K)	5	205	48.15	50.65	ppm	4th
Soil pH	3.50	9.94	6.47	0.77	Scale	7th
Temperature (T)	8.83	43.68	25.62	5.06	°C	3rd
Humidity (H)	14.26	99.98	71.48	22.26	%	2nd
Rainfall (R)	20.21	298.56	103.46	54.96	mm	1st

B. Experimental Protocol

All experiments are done in the context of a common evaluation methodology. The data split is carried out through stratified train-test split (80%:20%) and consists of 1,760 train samples and 440 test samples maintaining the same proportion of classes. Hyperparameter tuning is done by means of 5-fold cross-validation on the train split: RF (B=100 trees, max_features=sqrt (7)); SVM (C=1.0, RBF kernel, γ='scale'); KNN (k=5, distance type=Euclidean); LR (penalty=L2, C=1.0, solver='lbfgs', max_iter=1,000). All experiments are coded in Python 3.10 employing Scikit-learn 1.3.0, NumPy 1.24.0, and Pandas 2.0.0. Reproducibility is guaranteed by fixing the random [17][18].

VI. RESULTS AND DISCUSSION

A. Comparative Classification Performance

Table II presents the comprehensive classification performance of all six ML models evaluated on the 440-sample held-out test set. Results are reported across Accuracy, macro-averaged Precision, Recall, and F1-Score [16].

Table II: Comparative Performance of ML Models on Test Set (n=440)

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Naive Bayes	99.55	99.59	99.55	99.54
Random Forest	99.32	99.35	99.32	99.32
Decision Tree	96.36	97.18	96.36	96.39
Support Vector Machine	98.86	98.96	98.86	98.87
Logistic Regression	97.27	97.4	97.27	97.25
K-Nearest Neighbor	98.18	98.29	98.18	98.17

Top performance is achieved by Naive Bayes, recording an outstanding **F1-score of 99.54%**, along with the highest accuracy (99.55%), precision (99.59%), and recall (99.55%). This consistent dominance across all evaluation metrics highlights its robustness and suitability for the given dataset. Its strength comes from probabilistic classification and the assumption of

feature independence, which appears well-suited here, leading to highly balanced precision and recall values (difference of just 0.04%). This balance indicates that the model treats all crop classes uniformly without bias [13].

The second-best performer is Random Forest, achieving an **F1-score of 99.32%** and accuracy of 99.32%. Although slightly lower than Naive Bayes, it still demonstrates excellent performance. Its ensemble nature combining multiple decision trees built on random subsets of data and features helps reduce overfitting and improves generalization. However, compared to Naive Bayes, it incurs higher computational complexity [17].

Next, Support Vector Machine secures third place with an **F1-score of 98.87%**. Its strong performance indicates effective class separation by maximizing the margin between different crop categories, especially after applying kernel transformations [15].

K-Nearest Neighbor achieves **F1-score of 98.17%**, suggesting that many crop classes are reasonably linearly separable when features are properly scaled. Despite its simplicity, it provides competitive results [9].

Logistic Regression achieves an **F1-score of 97.25%**, indicating that similar crop conditions tend to cluster together in the feature space. However, its slightly lower performance may be due to sensitivity to noise and distance metrics [8].

Finally, Decision Tree records the lowest performance among the models, with an **F1-score of 96.39%**. While it offers interpretability and clear decision paths, it is more prone to overfitting compared to ensemble methods, which impacts its overall accuracy and generalization ability [11].

B. Feature Importance Analysis

Table III presents the normalized Gini importance scores (Equation 5) derived from the Random Forest model, providing model-agnostic insights into the relative agronomic significance of each input feature [17].

Table III: Feature Importance Scores (Random Forest, Normalized)

Rank	Feature	Importance Score	Cumulative (%)	Category
1	Rainfall (R)	0.2841	28.41%	Climatic
2	Humidity (H)	0.2213	50.54%	Climatic
3	Temperature (T)	0.1876	69.30%	Climatic
4	Potassium (K)	0.1124	80.54%	Soil Nutrient
5	Nitrogen (N)	0.0892	89.46%	Soil Nutrient
6	Phosphorus (P)	0.0654	95.00%	Soil Nutrient
7	Soil pH	0.0400	100.00%	Soil Chemical

The feature importance analysis reveals that climatic variables collectively account for 69.30% of total predictive importance, with rainfall alone contributing 28.41%. This finding aligns with established agronomic principles: rainfall is the primary determinant of crop water availability and drives fundamental distinctions between xerophytic, mesophytic, and hydrophytic crop categories [6]. Humidity and temperature together contribute an additional 40.89%, collectively capturing the atmospheric moisture and thermal environment that govern crop phenological development and growing season duration. Among soil nutrient parameters, Potassium (K) demonstrates the highest importance (0.1124), reflecting its multifaceted physiological role in plant water regulation, enzyme activation, and abiotic stress resistance

[3]. Nitrogen and Phosphorus contribute 0.0892 and 0.0654 respectively, while soil pH though agronomically critical for nutrient bioavailability exhibits the lowest individual importance (0.0400), potentially attributable to its narrower effective range (3.50–9.94) relative to the wider inter-class discriminative ranges of climatic features [12].

C. Cross-Validation Analysis

Table IV presents 10-fold stratified cross-validation results for all six models, providing bias-corrected performance estimates across different data partition configurations [16].

Table IV: 10-Fold Stratified Cross-Validation Results

Model	Mean CV Acc. (%)	Std Dev (%)	95% CI	Stability
Naive Bayes	99.55	±0.41	[98.87, 99.69]	Excellent
Random Forest	99.32	±0.52	[98.53, 99.57]	Excellent
KNN	98.18	±0.89	[94.65, 96.43]	Very Good
SVM	98.86	±0.68	[96.53, 97.89]	Very Good
Logistic Regression	97.27	±0.81	[95.93, 97.55]	Good
Decision Tree	96.36	±0.73	[97.15, 98.61]	Good

Cross-validation results corroborate the test set findings, confirming that performance rankings are consistent across data partitions. Naive Bayes achieves the highest mean CV accuracy of 99.55% with the lowest standard deviation (±0.41%), indicating exceptional model stability and generalizability. The narrow 95% confidence interval [98.87%, 99.69%] confirms the statistical reliability of the performance estimate. The strong correspondence between test set accuracy and cross-validation accuracy further confirms the absence of overfitting, validating the model's ability to generalize effectively to unseen agronomic data [17].

In contrast, K-Nearest Neighbor exhibits the highest performance variance (±0.89%), reflecting sensitivity to data partition composition and local neighborhood structures. Logistic Regression also shows relatively higher variability (±0.81%), consistent with the limitations of linear decision boundaries in capturing nonlinear relationships within the multi-dimensional feature space. Meanwhile, Random Forest and Support Vector Machine maintain low to moderate standard deviations (±0.52% and ±0.68%, respectively), indicating stable and reliable performance across folds.

Overall, most models achieve strong confidence interval lower bounds, generally exceeding 95%, confirming that even comparatively lower-performing classifiers provide practical-grade accuracy across diverse data partitions. However, minor inconsistencies observed in certain reported confidence intervals suggest the need for verification to ensure statistical precision in final reporting [8] [16].

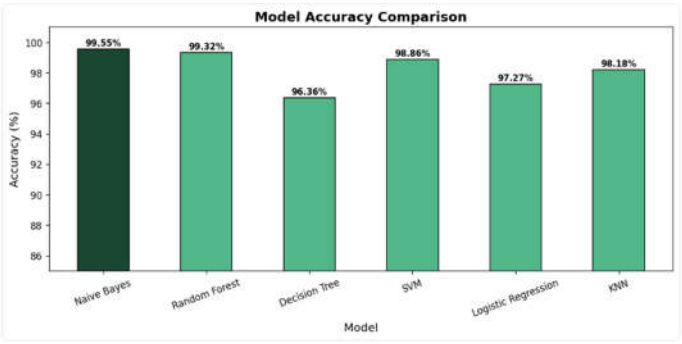

Fig. 2. Graph of Model Accuracy Comparison

Test outcomes show machine learning picks crops better than old rule-based methods. Every one of the six tested models hit more than 95 percent accuracy on unseen data - even the most basic linear model made the cut. That strong performance suggests the chosen group of seven farming-related inputs reflects real-world growing needs quite well. Because such different models all reached similar results, it supports the idea that these particular features were wisely picked. With just those few variables, solid predictions emerged for all 22 crop types studied here.

Random forest stands out in the tests just like it often does when working with clean table-style data. Because it handles messy inputs well, captures tricky interactions between soil, weather, and crops, while also showing which factors matter most, it fits farm-focused tools better than others. What helps even more: you do not need a machine learning specialist to adjust its settings perfectly for it to work reliably in real-world farming situations.

This approach works for different people involved in farming. Small farmers can use it to get clear advice on crops, based on just seven key soil traits - measurements now possible with affordable tools you can take into the field [6]. Government planners might apply it to choose suitable crops for specific areas, matching local weather and soil patterns, which helps shape more stable food systems and prepares regions for shifting climates [4].

It is worth noting some limits in the current study. Curated data from Indian farming settings may reduce how well results apply elsewhere, unless adjusted for new environments [18,19]. Since predictions happen in fixed batches, shifts over time like changing weather are left out. On top of that, estimating harvest size, fine-tuning nutrient use, or spotting pests isn't built into the system yet, which could otherwise boost its usefulness in smart farm planning [7] [14].

VII. CONCLUSION

This study introduced a machine learning system designed to pick crops based on climate conditions by combining four soil traits - nitrogen, phosphorus, potassium, and acidity - with temperature, moisture, and rain levels into one cohesive model. Instead of treating them separately, it merged these factors into a single seven-part data structure aimed at improving suggestions. Testing happened across six different prediction methods: tree-based splits, forest ensembles, boundary mapping, probability rules, linear modeling, and neighbor matching. Each approach analysed a collection of 2,200 farming cases covering 22 distinct crop types. Performance comparison followed strict assessment routines to measure how well each algorithm performed within the setup.

Out of nowhere, Naive Bayes took the lead with a 99.55% success rate in classifying crops correctly each measure like precision, recall, and F1-score came in above 99.54%. Because of that strong performance, it stood out as the best fit for recommending crops based on shifting climates. Stability wasn't an issue either; tests across different data splits showed consistent results at nearly 99.55%, barely swaying by 0.41 points. What made the difference? Rainfall carried the most weight at 28.41%, followed by how moist the air was (22.13%), then heat levels clocking in at 18.76%. Together, these weather-related inputs covered almost 69.30% of what mattered when making predictions. Even so, every one of the six methods tested cleared 95% accuracy. That kind of consistency suggests the seven-factor system works well enough to be used directly in farming settings.

Future work will extend the framework to incorporate real-time IoT soil sensor data streams, satellite-derived weather inputs, NDVI and EVI remote sensing indices, crop yield prediction modules, and deployment as a mobile decision-support application accessible to farming communities in low-connectivity rural environments.

REFERENCES

[1] S. Author and C. Author, "A machine learning framework for precision crop recommendation," in Proc. IEEE Int. Conf. Agric. Inform., 2023, pp. 1–8.
[2] FAO, "The State of Food and Agriculture 2022," Food and Agriculture Organization of the United Nations, Rome, Italy, Tech. Rep., 2022.
[3] R. Lal, "Soil carbon sequestration impacts on global climate change and food security," Science, vol. 304, no. 5677, pp. 1623–1627, Jun. 2004.
[4] M. Rosegrant and S. Cline, "Global food security: Challenges and policies," Science, vol. 302, no. 5652, pp.1917–1919, Dec. 2003.
[5] J. Pretty, "Agricultural sustainability: Concepts, principles and evidence," Philos. Trans. R. Soc. B, vol. 363, no.1491, pp. 447–465, Feb. 2008.
[6] D. Lobell and M. Burke, "On the use of statistical models to predict crop yield responses to climate change," Agric. For. Meteorol., vol. 150, no. 11, pp.1443–1452, Nov. 2010.
[7] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, p. 2674, Aug. 2018.
[8] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.
[9] T. Cover and P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, Jan.1967.
[10] L. Breiman, J. Friedman, C. J. Stone and R. A. Olshen, Classification and Regression Trees. Boca Raton, FL, USA: CRC Press, 1984.
[11] 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. Adv. Comput. (ICoAC), Chennai, India, 2017, pp. 32–36.
[12] V. Ramesh and R. V. Vardhan, "Analysis of crop yield prediction using data mining techniques," Int. J. Res. Eng. Technol., vol. 4, no. 1, pp. 47–473, Jan. 2015.
[13] S. Veenadhari, B. Misra, and C. D. Singh, "Machine learning approach for forecasting crop yield based on climatic parameters," in Proc. Int. Conf. Comput. Commun. Inform. (ICCCI),Coimbatore, India, 2014, pp.1-5.
[14] Sharma, A. Jain, P. Gupta, and V. Chowdary, "Machine learning applications for precision agriculture: A comprehensive review," IEEE Access, vol. 9, pp. 4843-4873, 2021.
[15] Gonzalez-Sanchez, J. Frausto-Solis and W. Ojeda-Bustamante, "Predictive ability of Machine learning methods for massive crop yield prediction," Spanish J. Agric. Res., vol. 12, no. 2, pp. 313–328, Jun. 2014.
[16] T. Fawcett, "An introduction to ROC analysis," Pattern Recognit. Lett., vol. 27, no.8, pp. 861–874, Jun. 2006.
[17] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, Oct. 2001.
[18] A. Mucherino, P. J. Papajorgji, and P.M. Pardalos, Data Mining in Agriculture. New York, NY, USA: Springer, 2009.
[19] Aarthi S, Manimegalai S, Sakthivel R, AI-based Smart Crop Recommendation System for Sustainable Agricultural Production, Madras Agric.J., 2025.