FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 22, November, 2026, pp 288 – 295
DOI: https://doi.org/10.33003/fjs-2026-1022-6172

A Machine Learning Framework for Precision Soil Quality Characterization for Crop-Specific Fertilizer Recommendation in Heterogeneous Agro-Ecosystems

***[1]Olajide Blessing Olajide, [2]Oladipo Ayodeji Bamidele, [3]Olabiyi Olatunji Coker, [4]Olayinka Dare Emmanuel, and [5]Omolayo Abegunde**

[1]Department of Computer Engineering, University of Ilesa, Ilesa, Osun State, Nigeria.
[2]Department of Cybersecurity, University of Ilesa, Ilesa, Osun State, Nigeria.
[3]Institute of Ecology and Environmental Studies, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.
[4]Department of Crop Production and Soil Science, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria.
[5]Department of Computer Engineering, Redeemers University, Ede, Osun State, Nigeria.

*Corresponding authors' email: blessing_olajide@unilesa.edu.ng Phone: +2348136207643

Received: 06 Oct. 2026
Accepted: 27 Oct. 2026
Published: 02 Nov. 2026

ABSTRACT
Accurate fertilizer recommendation is fundamental to precision agriculture, yet many machine learning models rely on static soil nutrient measurements without adequately incorporating environmental factors that influence nutrient availability. This study proposes a hybrid machine learning framework that integrates soil chemical, physical, and climatic variables into a Principal Component Analysis (PCA)-derived Soil Quality Index (SQI) for crop-specific fertilizer recommendation. The framework combines PCA-based feature extraction with Random Forest, Gradient Boosting, and Multilayer Perceptron (MLP) classifiers. Experiments were conducted on an 8,000-record dataset comprising 11 crop types, five soil classes, and seven fertilizer categories. Model development included standardized preprocessing, stratified train-test partitioning, and exhaustive hyperparameter optimization. Despite this rigorous framework, all models achieved only 12.4–14.0% test accuracy, equivalent to the 14.29% random baseline for a seven-class classification problem. Statistical auditing revealed that fertilizer labels were not significantly associated with soil or crop types (Chi-square, $p = 0.584$ and $p = 0.100$, respectively), while mutual information between continuous predictors and fertilizer labels was negligible. Learning-curve analysis further showed that the models have excellent training performance but no improvement in validation accuracy, indicating poor predictor-label dependency rather than inadequate model capacity. These findings demonstrate that dataset quality, rather than algorithm selection, constrained predictive performance. The primary contribution of this work is a robust and reusable machine learning pipeline complemented by a systematic predictor-label dependency auditing protocol. This framework provides a practical methodology for validating agricultural datasets before model deployment, thereby improving the reliability, reproducibility, and trustworthiness of precision agriculture and digital soil mapping applications.

Keywords: Precision Agriculture, Soil Quality Index, Principal Component Analysis, Machine Learning, Fertilizer Recommendation, Digital Soil Mapping, Predictor-Label Dependency

INTRODUCTION

Precision agriculture relies on accurate soil quality assessment to optimize irrigation, fertilizer application, and crop management. Soil quality is evaluated using measurable physical, chemical, and biological indicators, including soil texture, water dynamics, pH, nutrient availability, and microbial activity, which collectively determine soil productivity and sustainability (Lal, 2020). The selection and combination of appropriate indicators underpin any robust soil quality classification system (Vishal 2026). Soil quality is commonly classified as high, moderate, or low based on key soil indicators. High-quality soils exhibit balanced nutrients, favorable structure, high organic matter, active microbial processes, and adequate moisture, supporting sustainable crop production. Moderate-quality soils show partial limitations, whereas low-quality soils are affected by degradation factors such as erosion, salinity, compaction, acidity, or contamination, often requiring remediation (Vishal, 2026).

Conventional soil quality assessment relies on laboratory analysis of physical, chemical, and biological properties, providing accurate reference data for calibrating remote sensing and machine learning models. Nevertheless, these conventional approaches involve extensive human intervention and prolong processing time. This method also introduces observer subjectivity and is difficult to standardize across sites (Vullaganti et al., 2025). Machine learning and deep learning approaches have emerged as dominant methods for contemporary soil quality classification, outperforming traditional statistical approaches in accuracy, scalability and capacity to handle high-dimensional and multimodal datasets (Vishal, 2026; Olajide et al., 2025). Their ability to model nonlinear relationships and automatically extract meaningful representations has significantly improved classification accuracy, robustness and scalability across diverse applications. Consequently, they have become indispensable techniques for developing intelligent decision support systems (Olajide et al., 2024).

Quite a number of machine learning soil quality classification model have been developed in literature by authors using numerous methodologies. As an example, Diaz-Gonzalez et al., (2022) reviewed the use of machine learning and remote sensing to estimate soil indicators such as organic matter, texture, and moisture, concluding that ensemble tree-based methods generally outperform linear regression and single-tree models when covariates from multispectral or hyperspectral sensors are combined with laboratory ground-truth data. Adekunle et al., (2026) developed a hybrid Random Forest-Particle Swarm Optimization (RFPSO) model. The Hybrid RF-PSO models have improved soil-related prediction by combining feature selection and Random Forest optimization, achieving 91.4% accuracy and outperforming standalone Random Forest (87.0%) and Artificial Neural Networks (80.97%). Similarly, the study by Shahare et al., (2024) used Random Forest and XGBoost where Random Forest reduced prediction variance but remained sensitive to class imbalance, while XGBoost achieved 96% accuracy, demonstrating the superiority of boosting methods for imbalanced soil fertility classification. The research of Muthilakshmi et al., (2023) reviewed machine learning applications for soil fertility assessment and management, reporting that Random Forest achieved up to 99% accuracy in soil organic carbon prediction while XGBoost and AdaBoost reached up to 99.3% accuracy in pH prediction and potassium classification respectively. Their review, however, cautioned that such near-perfect accuracies are frequently obtained on small, geographically narrow datasets and may not generalize across agro-ecological zones. Alaboz et al., (2023) assessed soil quality for cultivated lands in the semi-humid Black Sea region using machine learning, confirming that RF-based models captured non-linear relationships between soil physical properties and quality indices better than classical multiple linear regression.

Micronutrient-focused work by Blesslin et al., (2022) applied machine learning algorithms within an IoT-enabled automated farm setting to classify soil micronutrients, reporting that ensemble classifiers outperformed single classifiers such as Naïve Bayes and k-Nearest Neighbours when working with sensor-streamed data. Sahour et al., (2021) used machine learning (RF, SVM, and boosted regression trees) for modelling and mapping water-induced soil erosion, finding that Random Forest again offered the best trade-off between accuracy and interpretability, although computational cost rose substantially with the addition of high-resolution topographic covariates. Collectively, these studies show that Random Forest and boosting ensembles (XGBoost, AdaBoost, Gradient Boosting) are the most frequently reported top performers for tabular soil-quality classification tasks, with reported accuracies ranging from roughly 85% to above 99% depending on dataset size, feature quality, and class balance, but with recurring weaknesses in generalizability and class-imbalance handling. Where soil data include spatial, spectral, or image-based structure, deep learning architectures have increasingly displaced or supplemented classical ML. The research of Escorcia et al., (2022) proposed an Intelligent Soil Nutrient and pH Classification technique using a Weighted Voting Ensemble of three deep learning models including Gated Recurrent Units (GRU), Deep Belief Networks (DBN), and Bidirectional LSTM with hyperparameters tuned via the Manta Ray Foraging Optimization (MRFO) metaheuristic. Their ensemble outperformed each individual deep model, illustrating how deep architectures benefit from optimisation-driven hyperparameter search, a theme that recurs strongly in the hybrid-model literature. Sujatha et al., (2023) developed a one-dimensional CNN (1D-CNN) for soil fertility classification and fertilizer recommendation using sequential soil chemical attributes. Combined with MinMax normalization and SMOTE oversampling, the model efficiently handled class imbalance and outperformed Extreme Learning Machines and Multilayer Perceptrons, particularly under limited training data.

Image-based deep learning has also been applied directly to soil classification. Gyasi and Purushotham (2023) introduced Soil-MobiNet, a lightweight CNN using depthwise-pointwise convolutions, to classify soil morphology and geospatial location from smartphone images, reporting strong accuracy while remaining deployable on resource-constrained mobile devices directly addressing the accessibility gap facing smallholder farmers in the Global South. Similarly, the study of Liu et al., (2020) applied deep learning to hyperspectral imaging for soil classification, and the study of Kim et al (2023) used a CNN on soil surface RGB images to predict water content and dry density, demonstrating that convolutional architectures can substitute for some laboratory soil-physical measurements when image acquisition is cheap and repeatable. Having reviewed previous machine learning soil quality classification attempts, Table 1 shows a summary of related frameworks, strength and weakness of the methodologies used.

Table 1: Key Findings from Review of Related Works on Machine Learning Models for Soil Quality Classification

S/N	Authors	Title	Year	Methods/Algorithms	Strength	Weakness
1.	Adekunle *et al.*	Hybrid RF-PSO model for species distribution prediction under soil temperature conditions	2026	Random Forest + Particle Swarm Optimization (RFPSO)	PSO-tuned RF (91.40%) outperforms standalone RF (87%) and ANN (80.97%)	PSO increases training time; risk of overfitting to optimisation objective
2.	Shahare *et al.*	ASUR: Agriculture soil fertility assessment using random forest classifier and regressor	2024	Random Forest (classifier + regressor)	Robust to noisy agronomic data; interpretable feature importance	Degraded performance on minority fertility classes
3	Muthulakshmi *et al.*	Insights of ML approach for soil fertility assessment and	2024	Review (RF, XGBoost, ANN, CNN, AdaBoost)	Reports very high accuracies (up to 99.3%) across nutrient types	Near-perfect accuracies may reflect narrow, non-generalisable datasets

S/N	Authors	Title	Year	Methods/Algorithms	Strength	Weakness
		management strategy				
4.	Alaboz *et al.*	Soil quality assessment using ML for cultivated lands in semi-humid Black Sea region	2023	RF-based regression/classification	Captures non-linear soil-quality relationships better than linear regression	Regionally specific; limited cross-climate validation
5	Sujatha *et al.*	1D-CNN-based soil fertility classification and fertilizer prescription	2023	1D-CNN + SMOTE + MinMax normalization	Compact, efficient on small datasets; links classification to fertilizer dosage	1D structure loses spatial context; sensitive to preprocessing choices
6	Kim *et al.*	CNN-based soil water content and density prediction using soil surface images	2023	CNN (image-based regression)	Low-cost alternative to lab moisture/density testing	Surface images may not reflect subsurface soil properties
7	Gyasi and Purushotham	Soil-MobiNet: CNN model for soil classification via morphology and geospatial location	2023	Lightweight CNN (depthwise-pointwise convolution)	Mobile-deployable; accessible for smallholder farmers	Accuracy dependent on image quality and lighting conditions
8	Díaz-González et al.	Machine learning and remote sensing techniques applied to estimate soil indicators: A review	2022	Review (RF, SVM, ensembles)	Wide coverage of soil indicator estimation methods	Lacks standardised benchmark across studies reviewed
9	Blesslin-Sheeba *et al.*	Ensemble learning algorithms combined with IoT sensor networks for micronutrient identification	2022	Ensemble classifiers + IoT sensors	Real-time micronutrient classification via IoT integration	Sensor noise and calibration drift affect reliability
10	Escorcia-Gutierrez *et al.*	AI-enabled agricultural modelling for soil nutrient analysis and pH classification	2022	GRU, DBN, BiLSTM ensemble + MRFO optimization	Weighted ensemble outperforms single DL models; optimised hyperparameters	High computational cost of multi-model ensemble with metaheuristic tuning
11	Sahour *et al.*	ML applications for water-induced soil erosion modelling and mapping	2021	RF, SVM, boosted regression trees	RF offers best accuracy-interpretability trade-off	High computational cost with high-resolution topographic data
12	Liu *et al.*	Deep learning-based soil classification using hyperspectral imaging	2020	CNN (hyperspectral)	Effective use of spectral bands for fine-grained classification	Requires expensive hyperspectral sensors; large data volume

This work has developed a machine learning framework that combined physical, chemical and weather-related soil data into a single and coherent Soil Quality Index (SQI) using Principal Component Analysis. It evaluated three different types of AI to find the best approach for crop-specific fertilizer recommendations: Random Forest, Gradient Boosting, and Multilayer Perceptron

MATERIALS AND METHODS

Dataset Description

The study utilized an 8,000-record soil–climate–crop dataset comprising six continuous variables (temperature, humidity, moisture, nitrogen, phosphorus, and potassium), two categorical variables (five soil types and eleven crop types), and seven fertilizer classes as the prediction target. Duplicate records were removed to ensure data integrity, while the balanced distribution of soil and fertilizer classes minimized class bias during model training. Continuous variables were standardized using z-score normalization, and One-hot encoding was applied to represent categorical variables as binary vectors. The dataset contained no missing values; therefore, imputation was unnecessary. These preprocessing steps ensured data consistency and supported robust, unbiased model evaluation.

Nonlinear Feature Engineering Using the Soil Quality Index (SQI)

Principal Component Analysis (PCA) was used to derive a Soil Quality Index by integrating six standardized soil and climatic variables into uncorrelated components. This reduced data redundancy while preserving the dominant variability, providing a compact representation of soil quality based on the minimum data set concept. *Let* $X \in R^{n \times 6}$ denote the standardized feature matrix, where $n = 8000$ soil samples and the six columns correspond to the selected physical, chemical, and climatic variables. PCA projects the original variables onto orthogonal principal components by solving the eigenvalue decomposition of the covariance matrix:

$$\Sigma \mathbf{v}_k = \lambda_k \mathbf{v}_k, \Sigma = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} \quad (1)$$

Where:

$\sum$ = covariance matrix of the standardized variables,

v_k = eigenvector defining the kth = principal component

λ_k = corresponding eigenvalue representing the variance explained by that component.

The first principal component, which explains the largest proportion of variance, was selected as the Soil Quality Index (SQI). The SQI score for each soil sample is computed as

$$\mathrm{SQI}_j = \sum_{i=1}^{6} v_{i1}\, x_{ji}, j = 1,2, \ldots, n \quad (2)$$

Where:

X_{ji} is the standardized value of the i^{th} variable for the j^{th} soil sample,

v^{i1} is the loading of the i^{th} variable on the firest principal component

The second principal component was retained solely for visualization of soil-profile distributions in two-dimensional feature space. Components satisfying the minimum data set criterion (eigenvalue $\lambda \geq 1$ and absolute loading $|v_{ik}| \geq 0.5$ were considered informative.

Model Architecture

Random Forest

Random Forest (RF) employs an ensemble learning strategy by generating multiple decision trees from bootstrap samples of the training data. At each node, a random subset of features is considered for splitting, thereby increasing model diversity, reducing tree correlation, and improving generalization performance. Node splitting is based on the Gini impurity, defined equation (3)

$$G(D) = 1 - \sum_{i=1}^{C} p_i^2 \quad (3)$$

Where $G(D)$ is the Gini impurity of node D, C is the number of classes, and p_i is the proportion of samples belonging to class i. The final prediction is obtained by majority voting across all T decision trees. This is shown by equation (4)

$$\hat{y}(x) = \mathrm{mode}\{h_1(x), h_2(x), \ldots, h_T(x)\} \quad (4)$$

Where $\hat{y}(x)$ is the predicted class of sample x, $h_t(x)$ is the prediction of the t^{th} tree $(t=1,2,\ldots,T)$ and T is the total number of trees. Bootstrap sampling also provides out-of-bag (OOB) samples for estimating model generalization without additional cross-validation

Gradient Boosting

Gradient Boosting (GB) constructs an ensemble sequentially by fitting each new decision tree to the negative gradient (residual) of the loss function generated by the current model. This iterative process progressively reduces prediction error. The residual at boosting stage m is computed as shown in equation (5)

$$r_{im} = -\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\Big|_{F(x)=F_{m-1}(x)}, i = 1,2, \ldots, n \quad (5)$$

Where r_{im} is the residual of sample i at iteration m, $L(\cdot)$ is the loss function, y_i is the true class label, $F_{m-1}(x_i)$ is the prediction from the previous ensemble, and n is the number of traning samples. The ensemblr is updated as shown in equation (6)

$$F_m(x) = F_{m-1}(x) + \nu\, \gamma_m\, h_m(x) \quad (6)$$

Where $F_m(x)$ is the updated ensemble prediction, $h_m(x)$ is the weak learner at iteration m, γ_m is the optimal step size, and ν is the learning rate $(0 < \nu \leq 1)$. In this study, ν was tuned over $\{0.05, 0.10\}$ to improve generalization and reduce overfitting.

Multilayer Perceptron

To model complex non-linear relationships, a Multilayer Perceptron (MLP) architecture comprising L hidden layers was employed. For a given hidden layer $l \in \{1, \ldots, L\}$ the activation vector $\mathrm{a}^{(l)}$ is computed through an affine transformation followed by a Rectified Linear Unit (ReLU) activation function in equation (7)

$$\mathbf{a}^{(l)} = \phi\left(\mathrm{W}^{(l)}\mathrm{a}^{(l-1)} + \mathrm{b}^{(l)}\right) \quad (7)$$

Where $W^{(l)}$ and $b^{(l)}$ represent the weight matrix and bias vector for layer l, $a^{(0)} = x$ is the input feature vector, and $\phi(\cdot) = \max(0,\cdot)$ denotes the element-wise ReLU function. For output layer $L+1$, Softmax activation function converts raw logits into a class-conditional probability distribution across $C=7$ distinct fertilizer categories as shown in equation (8)

$$\hat{y}_c = P(y = c|\mathbf{x}) = \frac{\exp\left(\mathbf{w}_c^{(L+1)\top}\mathbf{a}^{(L)} + b_c^{(L+1)}\right)}{\sum_{k=1}^{C}\exp\left(\mathbf{w}_k^{(L+1)\top}\mathbf{a}^{(L)} + b_k^{(L+1)}\right)} \quad (8)$$

Where $\hat{y}_c$ denotes the predicted probability for class $c \in \{1, \ldots, C\}$

The network parameters $\Theta = \{W, b\}$ were optimized by minimizing the categorical cross-entropy loss augmented with an L_2 weight-decay penalty (tuned over $\alpha \in \{10^{-4}, 10^{-3}\}$. This is expressed by equation (9)

$$\mathcal{L}_{\mathrm{reg}}(\Theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log \hat{y}_{i,c} + \frac{\alpha}{2}\sum_{l=1}^{L+1}\|\mathrm{W}^{(l)}\|_F^2 \quad (9)$$

Where N is the batch size, $y_{ic} \in \{0,1\}$ is the ground-truth binary indicator, and $||W^{(l)}||_F$ denotes the Frobenius norm. Optimization was executed using the Adam optimizer with default hyperparameters $(\eta, \beta_1, \beta_2, \in)$. To prevent overfitting, early stopping was enforced based on validation loss performance.

Hyperparameter Optimization and Evaluations Metrics

Each model was incorporated into a scikit-learn Pipeline combining the preprocessing transformer and the classifier, which was optimized by RandomizedSearchCV with 3-fold stratified cross-validation and macro-averaged F1-score as optimization objective. For each model, four candidate configurations were sampled from the following search

spaces. The Random Forest hyperparameters values are: n_estimators (150, 250), max_depth (None, 14), min_samples_leaf (1,3). The Gradient Boosting hyperparameter values are: n_estimators (100, 150), learning_rate (0.05, 0.10), max_depth (2, 3). MLP hyperparameter values are: hidden_layer_sizes ((64,32), (100,), alpha-L2 penalty (1e-4, 1e-3).

Evaluation Metrics and Statistical Auditing

Model performance was evaluated using a stratified hold-out test set comprising 20% of the dataset. Six complementary metrics were employed to assess classification performance, discrimination ability, probability calibration, and agreement beyond chance, namely Accuracy, Macro-F1 Score, Macro AUC-ROC, Log-Loss, and Cohen's Kappa.

Overall classification accuracy measures the proportion of correctly classified samples and is defined as expressed by equation (10)

$$Accuracy = \frac{1}{n}\sum_{i=1}^{n} 1(\hat{y}_i = y_i) \quad (10)$$

Where n is the number of test samples, y_i is the true class label, $\hat{y}_i$ is the predicted label, and $1(\cdot)$ is the indicator function that equals 1 when the prediction is correct and 0 otherwise.

To evaluate class-wise performance, Precision and Recall were computed for each class k as in equation (11)

$$\text{Precision}_k = \frac{TP_k}{TP_k+FP_k}, \text{Recall}_k = \frac{TP_k}{TP_k+FN_k} \quad (11)$$

Where TP_k, FP_k, FN_k denote the numbers of true positives, false positives, and false negatives, respectively.

The F1-score, which balances Precision and Recall, is expressed in equation (12)

$$F1_k = \frac{2\times\text{Precision}_k\times\text{Recall}_k}{\text{Precision}_k+\text{Recall}_k} \quad (12)$$

The overall Macro-F1 score is obtained by averaging the class-wise F1 values across all c classes. This is shown in equation (13)

$$\text{Macro-}F1 = \frac{1}{C}\sum_{k=1}^{C} F\,1_k \quad (13)$$

Where $c = 7$ represents the number of fertilizer classes.

The discriminative capability of each classifier was further assessed using the one-vs-rest Area Under the Receiver Operating Characteristic Curve (AUC-ROC) as shown in equation (14)

$$AUC_k = \int_0^1 TP\,R_k(\tau)\,dFPR_k(\tau) \quad (14)$$

Where TPR_k and FPR_k denote the true positive rate and false positive rate for class k at decision threshold τ. The reported Macro AUC-ROC is the average of AUC_k over all classes. To evaluate probabilistic predictions, the categorical cross-entropy (Log-Loss) was computed as expressed in equation (15)

$$\text{LogLoss} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{C} y_{ik}\ln(\hat{y}_{ik}) \quad (15)$$

Where $y_{ik} \in \{0,1\}$ is the one-hot encoded true label and $\hat{y}_{ik}$ is the predicted probability that sample i belongs to class k.

Finally, Cohen's Kappa was used to measure agreement between predicted and actual labels while accounting for chance agreement in equation (16)

$$\kappa = \frac{p_o - p_e}{1-p_e} \quad (16)$$

Where p_o is the observed accuracy and p_e the expected agreement estimated from the marginal class distributions. Higher values of k indicate stronger agreement, with $k = 1$ representing perfect agreement and $k = 0$ indicating agreement equivalent to random classification.

For statistical auditing, the data was divided into two sets: 80% (6,400 records) for training the models and 20% (1,600 records) for testing. To ensure a fair comparison, a stratified sampling was used, which kept the balance of fertilizer types the same in both groups. Prior to model training, Pearson's chi-square test was used to evaluate the relationship between the predictor variables (soil and crop types) and fertilizer recommendations. The entire machine learning pipeline was implemented in Python 3 using pandas, NumPy, SciPy, and scikit-learn.

RESULTS AND DISCUSSION

Exploratory Data Analysis

Figure 1 presents the Pearson correlation matrix of the six continuous variables. Temperature and humidity exhibited a moderate positive correlation (r = 0.53), while nitrogen showed moderate negative correlations with phosphorus (r = -0.64) and potassium (r = -0.45). Soil moisture exhibited negligible correlations with the remaining variables (|r| ≤ 0.09), indicating that it contributes large independent information to the predictive model.

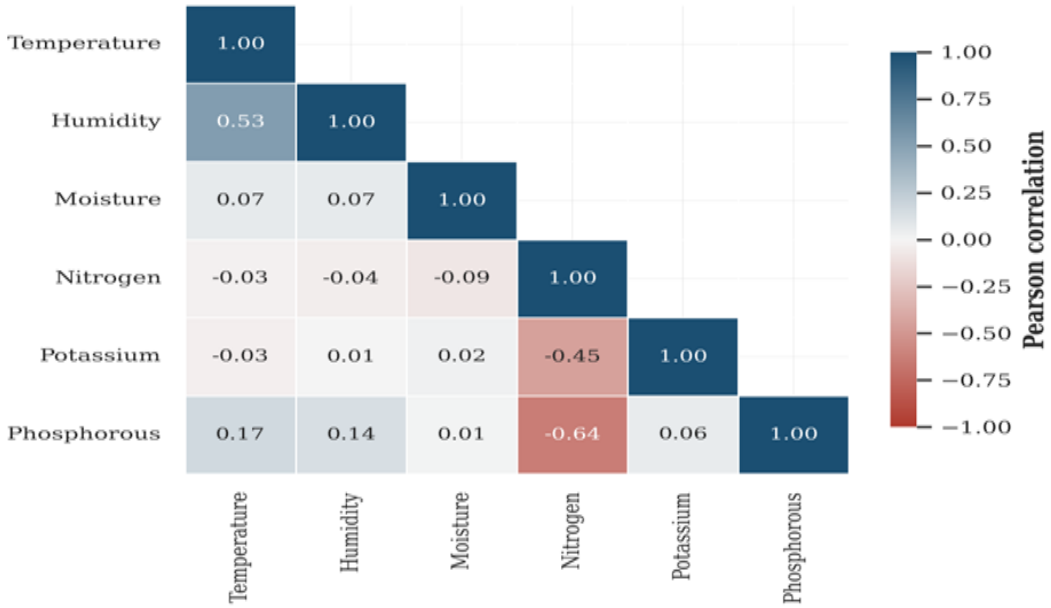


Figure 1: Correlation structure among physical–chemical soil and climatic variables

Figure 2 presents the distribution of nitrogen, phosphorus and potassium across the seven fertilizer classes. The visual distributions of macronutrients across classes overlap substantially, with broadly similar medians and inter-quartile ranges, a first qualitative indication, that macronutrient levels alone do not cleanly separate the recommendation classes in this dataset.

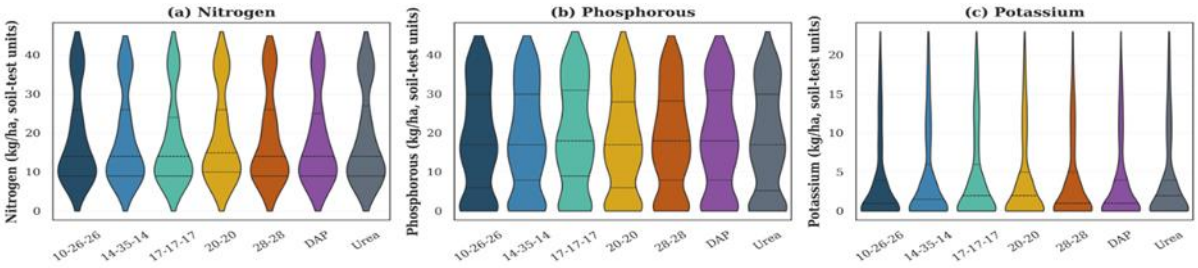


Figure 2: Distribution of nitrogen, phosphorus and potassium across the seven fertilizer classes.

Figure 3 shows temperature, humidity, and moisture stratified by soil type. Soil type had a modest but visible association with the climatic covariates (e.g., Black soils skewing toward marginally higher humidity), consistent with the crop-climate co-occurrence patterns typical of the represented growing regions.

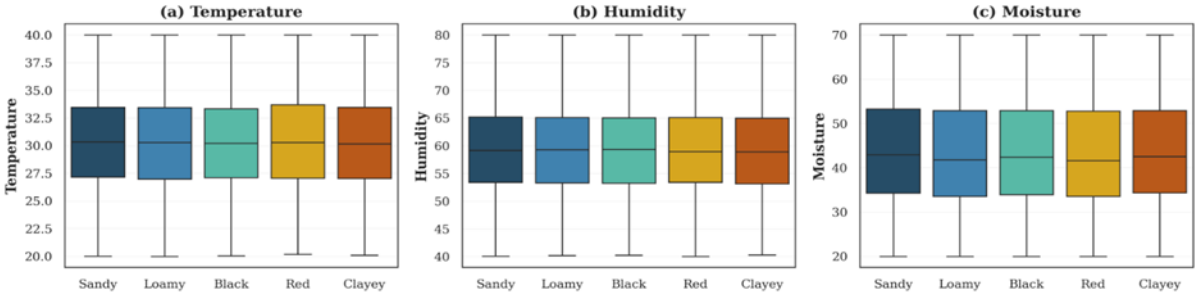


Figure 3: Environmental variable distributions across soil types.

Model Performance Comparison

The cross-validated and held-out test performance of the three tuned models across the full six-metric suite defined in Section 3.6 are summarized in Table 2. Random Forest achieved the highest test accuracy (14.00%) and macro-F1 (0.140), followed by MLP (13.25% / 0.126) and Gradient Boosting (12.38% / 0.121). For a balanced 7-class problem, the chance-level accuracy is 1/7 = 14.29% so all three models performed at or slightly below chance level. The macro-AUC values (0.492–0.510) were statistically indistinguishable from the null value of 0.500 expected under no discriminative power. The chance-corrected Cohen's Kappa is the most decisive single statistic in Table 2: all three models return $\kappa \in [-0.023, -0.004]$, i.e., agreement with the true label that is statistically indistinguishable from or fractionally worse than what would be expected from random labeling alone. Log-loss values (1.964–2.020) cluster tightly around the $\ln(7) = 1.946$ benchmark expected for a uniformly random 7-class predictor, confirming that even the models' probability estimates, not merely their arg-max decisions, carry no meaningful class-discriminative information.

Table 2: Cross-validated and Held-out Test Performance of the Three Tuned Models across Six Complementary Metrics, Alongside the Theoretical Chance Baseline (Accuracy = 1/7, AUC = 0.5, κ = 0, Log-loss = ln 7)

Model	CV Macro-F1	Test Acc.	Test Macro-F1	Test AUC (OvR)	Cohen's κ	Log-Loss
Random Forest	0.15	0.14	0.14	0.51	0	2.02
Gradient Boosting	0.14	0.12	0.12	0.49	-0.02	1.99
MLP (Neural Net)	0.14	0.13	0.13	0.49	-0.01	1.96
Chance baseline (C = 7)	—	0.14	0.14	0.5	0	1.95

Confusion Matrix Analysis

Figure 4 shows the row-normalized confusion matrix for the best-performing model (Random Forest). Every true class was predicted at a rate close to $1/7 \approx 0.143$ across all seven predicted classes, with no diagonal dominance — the signature of a classifier that has learned, at best, the marginal class distribution rather than any conditional structure linking covariates to the fertilizer label. The corresponding per-class precision, recall, and F1-score are reported in Table 3.

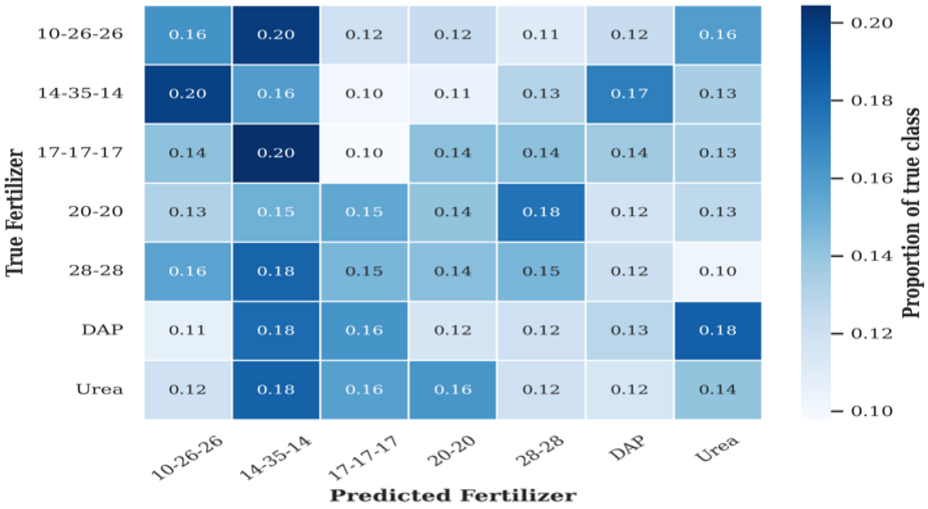


Figure 4: Normalized confusion matrix — Random Forest (best-performing model, test set)

Table 3: Per-class Precision, Recall, and F1-score for the Random Forest Model on the Held-out Test Set

Fertilizer class	Precision	Recall	F1-score	Support
10-26-26	0.16	0.16	0.16	226
14-35-14	0.13	0.16	0.14	238
17-17-17	0.1	0.1	0.1	225
20-20	0.15	0.14	0.14	220
28-28	0.15	0.15	0.15	224
DAP	0.14	0.13	0.14	233
Urea	0.15	0.14	0.14	234
Macro average	0.14	0.14	0.14	1600

Conclusion

The machine learning framework for precision soil quality characterization for crop-specific fertilizer recommendation implemented in this research have integrated PCA-based feature extraction with Random Forest, Gradient Boosting, and Multilayer Perceptron classifiers for crop-specific fertilizer recommendation. The study evaluated the performance of the models with multiple diagnostic analyses confirming negligible dependency between the predictor variables and fertilizer labels. These findings indicate that the observed limitation originates from dataset quality rather than model capability. The principal contribution of this work is the development of a robust and reusable machine learning pipeline, complemented by a systematic diagnostic protocol for assessing predictor–label dependency before deploying precision agriculture models. This framework provides a practical approach for identifying unreliable datasets and minimizing misleading decision-support systems. Future research should validate the proposed framework using agronomically verified datasets, incorporate additional soil fertility indicators such as pH, organic carbon, and cation exchange capacity, and evaluate model performance against actual crop yield outcomes to improve fertilizer recommendation reliability.

REFERENCES

Adekunle, T. A., Ogunwande, J. M. O., Ogundoyin, I. K., Akanbi, C. O., Babalola, A., Ayoola, T. A., & Roselin, O. F. (2026). A novel hybrid model for species distribution prediction of soil-transmitted helminthiasis (STH) under soil temperature conditions using Random Forest and Particle Swarm Optimization Algorithm. *Scientific Reports*, *16* Article 31604. https://doi.org/10.1038/s41598-025-31604-y

Alaboz, P., Demir, S., Işıldar, A. A., Tümsavaş, Z., & Cebeci, İ. (2023). Soil quality assessment based on machine learning approach for cultivated lands in semi-humid environmental condition part of Black Sea region. *Archives of Agronomy and Soil Science*, *69*, 2246–2259. https://doi.org/10.1080/03650340.2023.2246656

Blesslin Sheeba, T., Anand, L. D. V., Manohar, G., Selvan, S., Wilfred, C. B., Muthukumar, K., Padmavathy, S., Ramesh Kumar, P., & Asfaw, B. T. (2022). Machine learning algorithm for soil analysis and classification of micronutrients in IoT-enabled automated farms. *Journal of Nanomaterials*, *2022* Article 5343965. https://doi.org/10.1155/2022/5343965

Díaz-González, F. A., Vuelvas, J., Correa, C. A., Vallejo, V. E., & Patiño, D. (2022). Machine learning and remote sensing techniques applied to estimate soil indicators: A review. *Ecological Indicators*, *135* Article 108517. https://doi.org/10.1016/j.ecolind.2021.108517

Escorcia-Gutierrez, J., Gamarra, M., Soto-Diaz, R., Pérez, M., Madera, N., & Mansour, R. F. (2022). Intelligent agricultural modelling of soil nutrients and pH classification using ensemble deep learning techniques. *Agriculture*, *12*, Article 977. https://doi.org/10.3390/agriculture12070977

Gyasi, E. K., & Purushotham, S. (2023). Soil-MobiNet: A convolutional neural network model base soil classification to determine soil morphology and its geospatial location. *Sensors*, *23*, Article 6709. https://doi.org/10.3390/s23156709

Kim, D., Kim, T., Jeon, J., & Son, Y. (2023). Convolutional neural network-based soil water content and density

prediction model for agricultural land using soil surface images. *Applied Sciences*, *13*, Article 2936. https://doi.org/10.3390/app13052936

Lal, R. (2020). Soil organic matter and water retention. *Agronomy Journal*, *112*, 3265–3277. https://doi.org/10.1002/agj2.20282

Liu, F., Ji, W., Zhang, F., Zhang, Q., & Song, X. (2020). Deep learning-based soil classification using hyperspectral imaging. *Computers and Electronics in Agriculture*, *175* Article 105580. https://doi.org/10.1016/j.compag.2020.105580

Muthulakshmi, S., Backiyavathy, M. R., Gopalakrishnan, M., Kalpana, R., Thangamani, C., & Indhu Pavithra, S. (2024). Insights of machine learning approach for soil fertility assessment and management strategy. *Communications in Soil Science and Plant Analysis*, *56*, 436–457. https://doi.org/10.1080/00103624.2024.2416920

Olajide, B. O., Adeogun, A. E., Olayinka, T. O., Balogun, E., Areola, R. I., Falade, M. B., & Kemi, O. O. (2025). Analysis of computer network users activities using Support Vector Machine and Long Short-Term Memory Network. *Vallis Aurea*, *11*, 56–65.

Olajide, B. O., Sobowale, S. O., Kemi, O., Elijah, A. A., Olayinka, T. O., Oluwabunmi, O. B., & Adeosun, T. H. (2024). Diagnosing malaria and jaundice using selected machine learning methods. *International Journal of Research Publication and Reviews*, *5*, 3078–3087.

Sahour, H., Gholami, V., Vazifedan, M., & Saeedi, S. (2021). Machine learning applications for water-induced soil erosion modelling and mapping. *Soil and Tillage Research*, *211* Article 105032. https://doi.org/10.1016/j.still.2021.105032

Shahare, Y. R., Singh, M. P., Singh, S. P., Singh, P., & Diwakar, M. (2024). ASUR: Agriculture soil fertility assessment using a random forest classifier and regressor. *Procedia Computer Science*, *235* 1732–1741. https://doi.org/10.1016/j.procs.2024.04.164

Sujatha, M., Jaidhar, C. D., & Lingappa, M. (2023). 1D convolutional neural networks-based soil fertility classification and fertilizer prescription. *Ecological Informatics*, *78* Article 102295. https://doi.org/10.1016/j.ecoinf.2023.102295

Vishal, R. (2026). A modified multilayer perceptron model for soil quality classification and nutrient prediction in precision agriculture systems. *Biosystems Engineering*, *243* 118–131. https://doi.org/10.1016/j.biosystemseng.2026.01.009

Vullaganti, N., Mridha, N., & Garg, R. D. (2025). Site-specific soil quality classification for precision management using geospatial and machine learning approaches. *Precision Agriculture*, *26*, 847–869. https://doi.org/10.1007/s11119-025-10108-4

©2026 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via https://creativecommons.org/licenses/by/4.0/ which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.