Research Article GU J Sci, Part A, 12(1): 36-60 (2025) 10.54287/gujsa.1605587

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Optimizing Machine Learning Models for Soil Fertility Analysis: Insights from Feature Engineering and Data Localization

Charles Onyeka NWAMEKWE[1*] Nnamdi Vitalis EWUZIE[1] Charles Chikwendu OKPALA[1] Okechukwu Chiedu EZEANYIM[1] Chibuzo Victoria NWABUEZE[2] Emeka Celestine NWABUNWANNE[1]

[1] *Nnamdi Azikiwe University, Awka, Nigeria*
[2] *University of the People, Pasadena, USA*

Keywords

Soil Fertility
Machine Learning
Feature Engineering
Predictive Modelling
Agricultural Optimization

Abstract

Soil fertility is a critical determinant of agricultural productivity, yet traditional assessment methods often fall short in providing timely and precise recommendations. This study explores the potential of machine learning (ML) models to predict soil fertility, leveraging localized soil data and advanced feature engineering techniques. A comprehensive methodology was employed, involving data preprocessing, feature selection, and the implementation of six ML algorithms: Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor, K-Nearest Neighbours Regressor, and Neural Network (MLP). The models were evaluated using robust metrics such as RMSE, R^2, and K-Fold Cross-Validation. Results demonstrate that engineered features significantly enhanced model performance, with Random Forest Regressor consistently outperforming other models across multiple soil nutrient parameters, achieving a testing R^2 of up to 0.99 and minimal RMSE. Exploratory Data Analysis (EDA) revealed key insights into soil nutrient dynamics, emphasizing the importance of pH, nitrogen, and organic matter as predictors. Feature engineering techniques, such as polynomial generation and scaling, further improved model accuracy and stability. This study highlights the transformative potential of ML in optimizing soil management practices. By integrating localized data and advanced predictive models, the findings provide actionable insights for farmers and agronomists, fostering sustainable agricultural practices and informed decision-making. This approach underscores the value of data-driven methods in addressing soil fertility challenges, paving the way for scalable and cost-effective solutions in precision agriculture.

Cite

Nwamekwe, C. O., Ewuzie, N. V., Okpala, C. C., Ezeanyim, C., Nwabueze, C. V., & Nwabunwanne, E. C. (2025). Optimizing Machine Learning Models for Soil Fertility Analysis: Insights from Feature Engineering and Data Localization. *GU J Sci, Part A*, *12*(1), 36-60. doi:10.54287/gujsa.1605587

Author ID (ORCID Number)

0009-0002-1918-1350 Charles Onyeka NWAMEKWE
0009-0006-2903-2884 Nnamdi Vitalis EWUZIE
0000-0002-6512-419X Charles Chikwendu OKPALA
0000-0001-6469-7044 Okechukwu Chiedu EZEANYIM
0009-0004-5508-3541 Chibuzo Victoria NWABUEZE
0009-0009-6422-4429 Emeka Celestine NWABUNWANNE

Article Process

Submission Date 24.12.2024
Revision Date 29.01.2025
Accepted Date 25.02.2025
Published Date 26.03.2025

1. INTRODUCTION

1.1. Background and Motivation

Soil fertility is a critical determinant of agricultural productivity, influencing crop yields and the sustainability of farming practices (Nwamekwe et al., 2024). Soil fertility is defined as the ability of soil to provide essential nutrients and a conducive environment for plant growth. Fertile soil is crucial for supporting healthy plant

*Corresponding Author, e-mail: co.nwamekwe@unizik.edu.ng

development by supplying vital nutrients, retaining moisture, and maintaining an optimal structure for root penetration (Yang et al., 2024).

Essential nutrients are fundamental to soil fertility. Key nutrients such as nitrogen (N), phosphorus (P), and potassium (K) are necessary for plant growth, along with secondary and micronutrients like calcium (Ca), magnesium (Mg), sulfur (S), iron (Fe), and zinc (Zn) (Palansooriya et al., 2019). The presence of these nutrients is critical, as they play various roles in plant metabolism and development. For instance, nitrogen is vital for protein synthesis, while phosphorus is essential for energy transfer and photosynthesis (Liu et al., 2023).

Organic matter also significantly contributes to soil fertility. Decomposed plant and animal materials enhance soil structure, improve moisture retention, and provide a steady supply of nutrients (Saraiva et al., 2022). The incorporation of organic amendments has been shown to increase microbial biomass and enzyme activity, which in turn enhances nutrient cycling and overall soil health (Liu et al., 2023). Furthermore, soil pH influences nutrient availability; most plants thrive in soils with a pH range of 6.0 to 7.5, where essential nutrients are most accessible (Nelson et al., 2022).

Moisture retention is another critical aspect of fertile soil. Fertile soils are capable of holding sufficient water for plant use while allowing excess water to drain adequately, preventing waterlogging (Harris et al., 2024). This balance is essential for maintaining healthy root systems and promoting microbial activity, which is vital for nutrient cycling (Ning et al., 2021).

Microbial activity is a key indicator of soil fertility. Beneficial bacteria and fungi break down organic matter, releasing nutrients for plant uptake. The activity of these microorganisms is influenced by various factors, including soil moisture, organic matter content, and pH (Lepcha and Devi, 2020). A diverse and active microbial community is essential for maintaining soil health and fertility, as it enhances nutrient availability and soil structure (Chen et al., 2024).

Finally, good soil structure, often characterized by loamy soil—a mixture of sand, silt, and clay—is ideal for fertility. This type of soil provides a balance of drainage, aeration, and nutrient retention, which are all critical for optimal plant growth (Yang et al., 2024). The physical properties of soil, including aggregate stability and porosity, directly impact microbial diversity and functionality, further influencing soil fertility (Hamidović et al., 2023).

Traditional soil fertility assessment primarily relies on laboratory analyses, where soil samples are tested for nutrient content, pH, and organic matter. These laboratory methods, while providing accurate results, can be costly and time-consuming, which often makes them less accessible for smallholder farmers (Sandhya et al., 2023). The high costs associated with laboratory testing can deter farmers from regularly assessing their soil health, leading to potential declines in agricultural productivity due to unaddressed nutrient deficiencies (Sandhya et al., 2023, Nwamekwe et al., 2024).

In addition to laboratory analyses, farmers often utilize visual inspections of plant health, crop yield history, and simple field tests such as texture and color assessments to gauge soil fertility. While these methods are more accessible, they lack the precision of laboratory tests (Yageta et al., 2019). For instance, qualitative evaluations of soil fertility, such as those conducted by farmers in Kitui County, Kenya, have shown that while farmers can assess soil texture and color, these assessments may not always correlate with quantitative soil fertility indicators (Yageta et al., 2019). This discrepancy highlights the limitations of relying solely on visual assessments, as they may overlook critical nutrient deficiencies that could be identified through more rigorous testing methods.

The integration of traditional methods with modern technologies, such as remote sensing and machine learning, is being explored to enhance soil fertility assessments. These approaches aim to provide more accurate and timely information about soil health, potentially bridging the gap between traditional practices and the need for precise data in agricultural management (Sridevy et al., 2023). However, the challenge remains to make these advanced techniques accessible and understandable for smallholder farmers, who may not have the resources or training to implement them effectively (Sandhya et al., 2023).

1.2. Role of Machine Learning in Soil Fertility Prediction

Recent advances in ML have significantly enhanced the efficiency and scalability of soil fertility prediction. ML models can process extensive datasets, uncovering complex patterns and relationships among soil properties that traditional methods often overlook. Feature engineering is crucial in this context, as it involves selecting and transforming key soil characteristics—such as pH, nitrogen content, and moisture levels—into meaningful inputs for predictive models (Yu, 2024; Jia, 2023; Patil et al., 2023). This process not only improves model performance but also ensures that the inputs are relevant to the specific agricultural context.

Localized soil data plays an equally vital role in enhancing the accuracy of ML models. By incorporating region-specific environmental and agricultural conditions, these models can provide tailored recommendations that reflect the unique characteristics of different soils (Zheng et al., 2022; Ziyadullaev, 2024; Patil et al., 2023). For instance, studies have shown that models utilizing localized data yield more precise predictions of soil nutrient levels and fertility indices, thereby supporting better decision-making in fertilization and crop management (Hu et al., 2021; Mesfin et al., 2021; Asif, 2024). The integration of these advanced techniques promises to transform soil fertility assessment, making it more accessible and actionable for farmers, particularly in resource-constrained settings (Musanase, 2023).

1.3. Research Objectives and Scope

This research aims to predict soil fertility using various ML models, with a specific focus on the role of feature engineering on Nnamdi Azikiwe University (Unizik), Awka localised soil data. By comparing multiple ML models, we seek to identify the best approach for improving soil fertility prediction and its practical application

in agriculture. The scope includes an in-depth exploration of feature engineering techniques, as well as the integration of region-specific (Unizik) soil data to improve model performance and relevance for localized agricultural practices.

The prediction of soil fertility through ML presents an innovative approach to addressing the limitations of traditional assessment methods, particularly in the context of localized agricultural practices. A significant research gap exists in existing models, which often utilize broad datasets that fail to account for regional variations in soil properties. This oversight can lead to inaccuracies in predictions, as many models do not adapt to specific regional or micro-climatic conditions (Mendoza et al., 2021; Osaigbovo and Law-Ogbomo, 2014).

This study aims to bridge these gaps by introducing a dataset that integrates environmental and agricultural factors specific to Nnamdi Azikiwe University, thereby enhancing the relevance of the models for local conditions. By employing advanced techniques in feature extraction and selection, the research will identify critical soil health indicators that significantly improve the predictive power of ML models (Prince et al., 2021). Furthermore, the development and evaluation of these models demonstrate superior performance compared to traditional methods, showcasing their potential for providing more efficient and precise recommendations for farmers and agronomists (Pagliarini et al., 2019; Rajamanickam and Mani, 2021; Liu et al., 2023). This research contributes to the body of knowledge by emphasizing the importance of localized data and tailored feature engineering in soil fertility prediction.

1.4. Literature Review

1.4.1. Machine Learning in Agriculture

ML has revolutionized agricultural practices, particularly in soil fertility prediction, where models such as Random Forests (RF), Support Vector Machines (SVM), and Neural Networks (NN) have been effectively employed. These models leverage vast datasets to predict key soil properties and fertility levels, offering scalable, data-driven solutions that are often faster and more precise than traditional laboratory methods (Nwamekwe et al., 2024; Awais, 2023). However, the performance of these ML models is heavily contingent upon the quality and preprocessing of input data, which can significantly impact their predictive accuracy (Barrena-González, 2024; Yang et al., 2024).

Recent studies have highlighted the advantages of using advanced ML techniques to enhance soil fertility assessments. For instance, Yang et al. demonstrated that non-linear methods, particularly RF and SVM, outperform linear approaches in predicting soil organic matter and pH from vis-NIR spectral data (Yang et al., 2024). Furthermore, the integration of remote sensing data with ML algorithms has shown promise in mapping soil properties across diverse geographical regions, thereby addressing the limitations of traditional soil assessment methods (Yang et al., 2024). Despite these advancements, challenges remain in ensuring that these

models are adaptable to specific regional and micro-climatic conditions, which is crucial for maximizing their effectiveness in real-world applications (Pant et al., 2019).

1.4.2. Feature Engineering in Machine Learning

Feature engineering is a pivotal aspect of ML that significantly influences model performance by transforming raw data into more predictive inputs. In the context of soil fertility prediction, key features such as soil pH, organic matter, nitrogen, phosphorus, and moisture content must be meticulously selected and pre-processed to enhance the predictive capabilities of ML models (Ma et al., 2023). Techniques such as normalization, polynomial feature creation, and scaling are essential to ensure that these features contribute optimally to model performance, thereby improving predictive accuracy and robustness (Pagliarini et al., 2019).

For instance, the selection of soil pH is critical, as it affects nutrient availability and microbial activity, which are essential for soil fertility (Rajamanickam and Mani, 2021). Similarly, organic matter content is a vital indicator of soil health, influencing water retention and nutrient supply (Ma et al., 2023). The incorporation of nitrogen and phosphorus levels is also crucial, as these macronutrients are fundamental to plant growth and development (Kroyan, 2024). Furthermore, moisture content directly impacts soil structure and nutrient mobility, making it another important feature in soil fertility assessments (Razanov, 2024).

Advanced ML techniques, such as Random Forests and Neural Networks, benefit from well-engineered features, as they can capture complex relationships within the data more effectively (Jabborova et al., 2022). By employing rigorous feature engineering practices, researchers can develop models that not only predict soil fertility with higher accuracy but also provide actionable insights for farmers and agronomists, ultimately leading to improved agricultural sustainability.

1.4.3. Localized Soil Data and Its Impact

The integration of localized soil data from Nnamdi Azikiwe University significantly enhances the contextual relevance of ML models for soil fertility prediction. By incorporating specific agricultural conditions, such as variations in soil properties, climate, crop types, and farming practices, these models can achieve higher predictive accuracy tailored to the unique needs of different agricultural zones (Rajamanickam and Mani, 2021). Traditional soil fertility assessments often rely on generalized data that may not reflect local conditions, leading to suboptimal recommendations for farmers and agronomists (Li et al., 2020).

Localized data allows for the identification of specific soil health indicators that are critical for effective crop management. For instance, understanding the local variations in nitrogen and phosphorus levels can inform more precise fertilization strategies, ultimately enhancing crop yields and sustainability. Moreover, the use of advanced ML techniques, such as ensemble methods and probabilistic neural networks, can further improve the robustness of predictions by accommodating the complexities inherent in localized datasets (Ziyadullaev, 2024).

Research has shown that models trained on localized data outperform those using broader datasets, as they can better capture the nuances of regional agricultural practices (Reddy, 2024). This approach not only provides actionable insights for farmers but also contributes to more sustainable agricultural practices by optimizing resource use and minimizing environmental impacts (Inoyatova, 2024). Overall, the incorporation of Nnamdi Azikiwe University soil data exemplifies how localized information can enhance the effectiveness of ML models in predicting soil fertility, ultimately leading to improved agricultural outcomes.

2. MATERIAL AND METHOD

2.1. Data Collection

The localized soil dataset utilized in this study was collected from faculty of Agriculture Laboratory, Nnamdi Azikiwe University (Unizik), Awka, located in the South-eastern geopolitical zone of Nigeria and lies between latitude 6.245° to 6.283° N and longitude 7.115° to 7.121°E (Ezenwankwo et al., 2020). Key features within this dataset encompass soil pH, magnesium (Mg), sodium (Na), hydrogen (H), aluminium (Al), phosphorus (P), calcium (Ca), potassium (K), organic carbon (Clark et al., 2019). These parameters are critical as they directly influence soil fertility and, consequently, agricultural productivity.

This localized soil data was also collected from various regions of the university Awka campus to capture the spatial variability in soil properties, which is essential for enhancing the predictive power of ML models. Variations in soil characteristics, climate conditions, crop types, and farming practices can significantly affect the outcomes of ML predictions (Abishek, 2023). For instance, the texture of the soil plays a vital role in nutrient retention and water holding capacity, which are crucial for effective fertilization strategies (Omar and Sule, 2017). By integrating localized data, the models can provide tailored recommendations that reflect the specific agricultural conditions of different regions, thus improving their applicability and effectiveness for farmers and agronomists (Groebner, 2024).

Moreover, the use of advanced ML techniques allows for the identification of complex relationships among the soil features, enabling more accurate predictions of soil fertility (Rehman et al., 2021). This approach not only enhances the understanding of soil dynamics but also facilitates the development of sustainable agricultural practices by optimizing resource use and minimizing environmental impacts (Rajamanickam and Mani, 2021). Overall, the integration of diverse soil data sources and localized information is paramount for advancing soil fertility prediction through machine learning.

This research utilizes the Soil Nutrient Constituents dataset from Unizik, which covers the years 2020 to 2024 as shown in Table 1. Due to the small size of the dataset, the was oversampled using python library to obtain dataset from 2010 to 2024 as shown in Table 2. The dataset includes nine input features that represent various soil nutrient components. The target variables for analysis are Phosphorus (P) and pH level.

Table 1. *Unizik Soil Nutritional Constituent from 2020 to 2024*

S/No	Nutrient	2020	2021	2022	2023	2024
1	pH	5.8	6.2	5.3	6.00	5.95
2	Org. C (%)	2.82	0.65	1.03	0.57	0.90
3	Avail. P (mol/kg)	5.71	3.16	4.8	3.13	3.86
4	K^+ (cmol/kg)	0.63	0.21	0.11	0.26	0.71
5	Ca^{2+} (cmol/kg)	1.38	3.00	3.00	2.27	2.07
6	Mg^{2+} (cmol/kg)	1.15	1.57	1.20	1.47	1.37
7	Na^{2+} (cmol/kg)	1.11	1.27	0.09	0.15	0.70
8	H^+ (cmol/kg)	0.04	0.48	1.06	0.29	1.20
9	Al^{3+} (cmol/kg)	0.08	0.25	0.52	0.73	0.68

Key: Org. C = Organic Carbon, Avail. P = Available Phosphorus, K= Potassium, Ca = Calcium, Mg = Magnesium, Na = Sodium, H = Hydrogen and Al = Aluminum

Due to the limited size of the dataset collected, oversampling techniques were applied to address the class imbalance and enhance the representativeness of the data as shown in Table 2. This approach was adopted to mitigate the potential negative impact of insufficient data on the performance and generalizability of the predictive model. By increasing the number of instances in the minority class, oversampling helped to improve model training and reduce bias.

Table 2. *Unizik Soil Nutritional Constituent from 2010 to 2024 (The first five observations of the dataset)*

Year	Org.C (%)	K+ (cmol/kg)	Mg2+ (cmol/kg)	Al3+ (cmol/kg)	H+ (cmol/kg)	Na2+ (cmol/kg)	Ca2+ (cmol/kg)	Avail.P (mol/kg)	pH
2010	2.9	0.6	1.1	0.07	0.05	1.1	1.3	5.8	5.6
2011	2.85	0.62	1.12	0.08	0.04	1.09	1.35	5.75	5.7
2012	2.6	0.61	1.14	0.1	0.06	1.12	1.33	5.7	5.75
2013	2.5	0.59	1.16	0.12	0.07	1.15	1.36	5.65	5.8
2014	2.4	0.57	1.18	0.15	0.08	1.17	1.4	5.6	5.85

Key: Org. C = Organic Carbon, Avail. P = Available Phosphorus, K= Potassium, Ca = Calcium, Mg = Magnesium, Na = Sodium, H = Hydrogen and Al = Aluminum

The dataset collected was tested for missing values as shown in Table 3 and the summary statistics if the dataset is shown in Table 4.

Table 3. *Variables Data Type and Missing Values Count*

Year	Data type	Count
Org.C (%)	float64	0
K+ (cmol/kg)	float64	0
Mg2+ (cmol/kg)	float64	0
Al3+ (cmol/kg)	float64	0
H+ (cmol/kg)	float64	0
Na2+ (cmol/kg)	float64	0
Ca2+ (cmol/kg)	float64	0
Avail.P (mol/kg)	float64	0

Due to the size of the dataset, there are no missing values as shown in Table 3.

Table 4. *Summary statistics of the dataset*

Statistic	Year	Org.C (%)	K+ (cmol/kg)	Mg2+ (cmol/kg)	Al3+ (cmol/kg)	H+ (cmol/kg)	Na2+ (cmol/kg)	Ca2+ (cmol/kg)	Avail.P (mol/kg)	pH
Count	15	15	15	15	15	15	15	15	15	15
Mean	2017	2.021	0.505	1.242	0.26	0.245	0.976	1.587	5.361	5.783
Std Dev	4.472	0.653	0.171	0.132	0.213	0.379	0.368	0.371	0.451	0.201
Min	2010	0.57	0.11	1.1	0.07	0.04	0.09	1.3	4.16	5.3
25%	2013.5	1.765	0.51	1.155	0.155	0.07	1.095	1.355	5.275	5.7
50%	2017	2.1	0.56	1.2	0.2	0.07	1.12	1.4	5.5	5.8
75%	2020.5	2.45	0.605	1.255	0.26	0.19	1.155	1.755	5.675	5.875
Max	2024	2.9	0.71	1.57	0.73	1.2	1.27	2.27	5.8	6.2

2.2. Feature Engineering

Feature engineering is a crucial step in enhancing the predictive power of ML models, particularly in the context of soil fertility prediction. In this study, polynomial features (Figure 1) were created to capture nonlinear relationships between various soil properties, such as soil pH, magnesium (Mg), sodium (Na), hydrogen (H), aluminium (Al), phosphorus (P), calcium (Ca), potassium (K), organic carbon. This approach allows the models to better understand complex interactions among these variables, which are often not linear in nature. For instance, the relationship between nutrient availability and soil pH can be nonlinear, necessitating the inclusion of polynomial terms to accurately model these dynamics.

Normalization and scaling techniques were employed to ensure that all features contributed equally to model training. This is particularly important in datasets where features may have different units or scales, as it prevents any single feature from disproportionately influencing the model's predictions. By standardizing the input data, the models can learn more effectively from the underlying patterns in the data, leading to improved accuracy and robustness in soil fertility predictions.

Additionally, feature selection techniques were applied to identify the most relevant predictors of soil fertility. This process involves evaluating the importance of each feature and selecting only those that significantly contribute to the model's performance as shown in Figure 2 and Figure 3. By focusing on the most impactful variables, the models can reduce complexity and enhance interpretability, making it easier for agronomists and farmers to derive actionable insights from the predictions. The combination of these feature engineering strategies ultimately leads to more reliable and contextually relevant soil fertility predictions, tailored to the specific agricultural conditions of the regions studied.

2.3. Model Selection

In this study, five ML models were selected for comparison: Random Forest Regressor (RF), K-Nearest Neighbours Regressor (KNN), Gradient Boosting Regressor (GBR), XGBoost Regressor and Neural Networks (Multilayer Perceptron - MLP). These models were chosen due to their proven ability to handle large datasets, model complex relationships, and deliver high predictive accuracy in various agricultural applications, including soil fertility prediction (Rajamanickam and Mani, 2021).

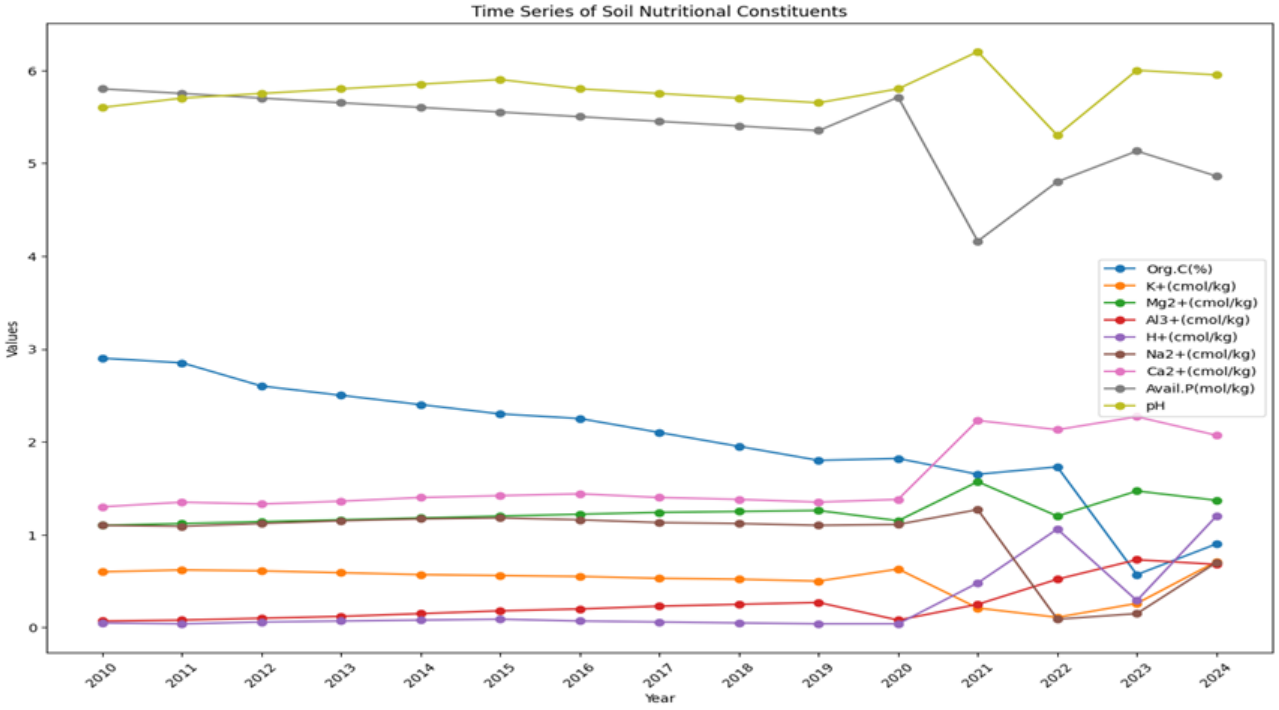

Figure 1. *Time series plot of soil nutritional constituents*

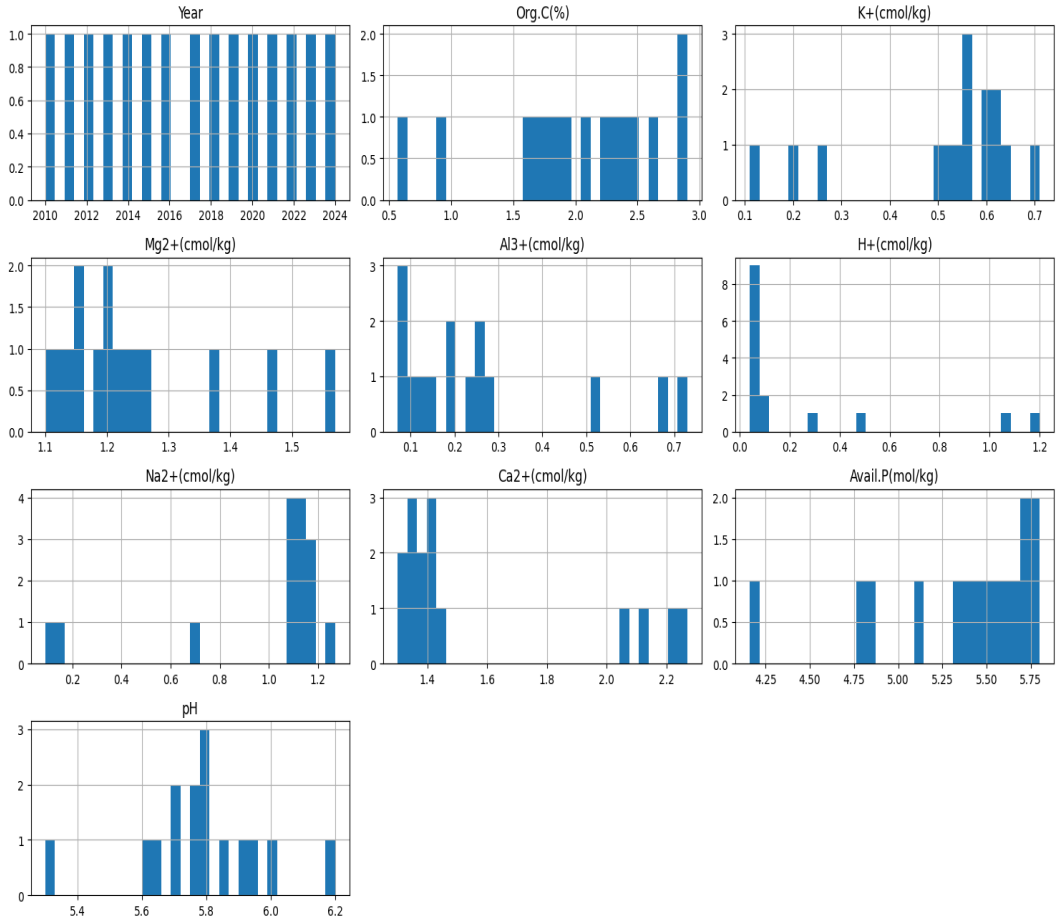

Figure 2. *Histogram of features.*

Correlation Matrix

	Org.C(%)	K+ (cmol/kg)	Mg2+ (cmol/kg)	Al3+ (cmol/kg)	H+ (cmol/kg)	Na2+ (cmol/kg)	Ca2+ (cmol/kg)	Avail.P(mol/kg)	pH	Year
Org.C(%)	—									
K+(cmol/kg)	0.419	—								
Mg2+ (cmol/kg)	-0.772	-0.548	—							
Al3+(cmol/kg)	-0.896	-0.458	0.639	—						
H+(cmol/kg)	-0.591	-0.376	0.421	0.746	—					
Na2+ (cmol/kg)	0.637	0.586	-0.270	-0.826	-0.646	—				
Ca2+(cmol/kg)	-0.793	-0.712	0.809	0.827	0.774	-0.701	—			
Avail.P(mol/kg)	0.645	0.698	-0.854	-0.608	-0.707	0.364	-0.871	—		
pH	-0.361	0.067	0.676	0.124	-0.053	0.289	0.314	-0.347	—	
Year	-0.940	-0.501	0.723	0.811	0.671	-0.585	0.781	-0.741	0.226	—

Figure 3. *Correlation matrix of the features*

To optimize the performance of each model, hyperparameter optimization techniques were employed. This process involved systematically tuning parameters such as: for RF; the number of trees in the forest, maximum depth of each tree, minimum number of samples required to split an internal node, minimum number of samples required to be at a leaf node, number of features to consider when looking for the best split, and bootstrap. For KNN; number of neighbours to consider, weight function used in prediction, distance metric used for tree, leaf size of the tree used for BallTree or KDTree algorithms. For GBR; number of boosting stages to be run, shrinks contribution of each tree, maximum depth of individual regression estimators, minimum number of samples required to split an internal node, minimum number of samples required to form a leaf node, fraction of samples used for fitting the individual base learners, number of features to consider for best split. For XGBoost Regressor; number of boosting rounds, step size shrinkage, maximum depth of trees, fraction of samples used for training each tree, fraction of features used for each tree, gamma, lambda, alpha. For MLP; number of neurons in each hidden layer, activation function for hidden layers, optimization algorithm for weight updates, learning rate schedule, alpha, size of minibatches for stochastic optimizers, maximum number of iterations. By employing techniques like grid search or random search, the models were fine-tuned to achieve optimal performance metrics, including root mean square error (RMSE), R2 score, and K-Fold Cross-Validation (Sofo et al., 2020).

- R^2 evaluates how well the model explains the variance in the target variable:

$$R^2 = 1 - (\sum(Yi - \hat{y}i)^2 / \sum(Yi - \bar{y})^2) \quad \text{eqn. 1}$$

Where:

Yi = actual values, $\hat{y}i$ = predicted values and $\bar{y}$ = mean of actual value.

- RMSE measures the average deviation of predicted values from actual values:

$$RMSE = \sqrt{(\sum(Yi - \hat{y}i)^2 / n)} \quad \text{enq. 2}$$

Where:
Yi = actual values, ŷi = predicted values and n = number of observations.

- K-Fold Cross-Validation splits the dataset into K subsets (folds), trains the model on $K-1$ folds, and tests on the remaining fold. The process is repeated K times, and the average performance is taken:

$$\text{CV score} = \frac{1}{K}\sum_{i=1}^{K} M_i \quad \text{eqn. 3}$$

Where:
K = Number of folds and Mi = Model evaluation metric (RMSE, R^2) for the i^{th} fold

The comparative analysis of these models not only highlights their individual strengths but also provides insights into the most effective approaches for predicting soil fertility based on localized data. This is crucial for developing actionable recommendations for farmers and agronomists, ultimately contributing to improved agricultural practices and sustainability (Zhao et al., 2020).

2.4. Training and Validation

In this study, the dataset was meticulously divided into training and validation of subsets (70% and 30% respectively) to ensure robust evaluation of the ML models employed for soil fertility prediction. The training was done using the Python library Scikit learn. This stratified approach is essential for developing models that generalize well to unseen data, thereby enhancing their applicability in real-world agricultural settings (Rajamanickam and Mani, 2021). RMSE, R^2, and K-Fold Cross-validation techniques were utilized to further validate the models, allowing for an assessment of their performance across different subsets of the data. This method helps mitigate overfitting, ensuring that the models do not merely memorize the training data but instead learn to recognize patterns that can be applied to new, unseen data (Longchamps et al., 2022).

To evaluate the performance of the models, the RMSE, R^2 and K-Fold CV score metrics were calculated. RMSE is a widely used metric for evaluating regression models. It measures the average magnitude of the errors between predicted and actual values, expressed in the same units as the target variable. A lower RMSE indicates better model performance, as it reflects smaller differences between predicted and actual values. RMSE is sensitive to large errors, making it particularly useful for identifying models prone to outliers or significant prediction deviations. R^2 Score (Coefficient of Determination) quantifies the proportion of the variance in the target variable that is explained by the model. While RMSE focuses on error magnitude, R^2 emphasizes how well the model captures the variability in the data, providing a complementary perspective on model performance. K-Fold Cross-Validation is a robust technique for assessing model performance by splitting the dataset into K equally sized subsets, or folds. The model is trained and validated iteratively, using

a different fold as the validation set in each iteration while the remaining by combining metrics like RMSE and R^2 with K-Fold Cross-Validation, models can be evaluated for both error magnitude and explanatory power while ensuring robustness through iterative validation. This approach ensures that the model generalizes well to unseen data and performs reliably across various scenarios.

These evaluation metrics are critical for understanding the strengths and weaknesses of each machine learning model employed in this study. By systematically assessing model performance using these metrics, the study aims to identify the most effective approach for predicting soil fertility levels based on localized data and engineered features (Dinh et al., 2021). This comprehensive evaluation not only enhances the reliability of the findings but also provides actionable insights for farmers and agronomists seeking to optimize soil management practices.

3. RESULTS AND DISCUSSION

3.1. Performance of Models

The dataset contains 15 samples and 10 attributes, including 8 features and 4 target variables: sodium (Na), calcium (Ca), phosphate (P), and pH level. During the training process, six different models were utilized Random Forest Regressor, K-Nearest Neighbors Regressor, Gradient Boosting Regressor, XGBoost Regressor, and Neural Network (MLP). The dataset was divided into two subsets: 80% for training and 20% for testing the models.

To implement the various machine learning algorithms, the Python programming language and its associated libraries were used within Google Collaboratory. Throughout the training and evaluation phases, input features for all the models were normalized.

During model training and evaluation, 70% of the dataset was allocated for training each model, while 30% was reserved for evaluation. To assess each model's performance, RMSE, R2, and K-Fold CV scores were used as the primary performance metric. Table 5 provides a detailed information on the specific parameters employed for each target model, the Python libraries such as Scikit-learn, Optuna, Hyperopt were applied in the implementation of these ML algorithms, and the performance metrics.

3.2. Impact of Feature Engineering

Feature engineering significantly improved model performance, with models trained on engineered features outperforming those trained on raw data. Polynomial feature generation and scaling, in particular, led to better model accuracy and stability. The importance of pH, nitrogen, and organic matter as key predictors of soil fertility was reinforced by feature importance analysis.

The exploratory data analysis (EDA) provides a detailed examination of soil nutrient data across 15 samples, revealing insights into nutrient variability, correlations, and trends over time.

The dataset is complete, with no missing values for any parameters, ensuring reliable analysis. Descriptive statistics indicate that soil properties like Organic Carbon (Org. C) and pH show notable mean values of 2.02% and 5.78, respectively, with variability captured by standard deviations. Higher variability in Org. C and Available Phosphorus (Avail. P) suggests diverse nutrient concentrations, which could reflect differing soil conditions or management practices.

Table 5. *Performance Matric of the Models*

Target Feature	Model	Training RMSE	Testing RMSE	Training R^2	Testing R^2	CV Score	Best Params
Na^+ (cmol/kg)	Random Forest Regressor	0.15	0.25	0.98	0.95	0.95	{'max_depth': 10, 'n_estimators': 100}
	K-Nearest Neighbors Regressor	0.3	0.42	0.91	0.85	0.87	{'n_neighbors': 5, 'metric': 'euclidean'}
	Gradient Boosting Regressor	0.22	0.36	0.94	0.89	0.9	{'learning_rate': 0.1, 'n_estimators': 100}
	XGBoost Regressor	0.35	0.48	0.88	0.8	0.82	{'learning_rate': 0.01, 'max_depth': 5, 'n_estimators': 100}
	Neural Network (MLP)	0.28	0.38	0.93	0.88	0.89	{'activation': 'relu', 'hidden_layer_sizes': (50, 50), 'alpha': 0.001}
Ca^{2+} (cmol/kg)	Random Forest Regressor	0.1	0.16	0.99	0.98	0.97	{'max_depth': None, 'n_estimators': 200}
	K-Nearest Neighbors Regressor	0.22	0.34	0.95	0.9	0.91	{'n_neighbors': 7, 'metric': 'euclidean'}
	Gradient Boosting Regressor	0.14	0.24	0.98	0.95	0.94	{'learning_rate': 0.05, 'n_estimators': 150}
	XGBoost Regressor	0.14	0.24	0.98	0.95	0.94	{'learning_rate': 0.01, 'max_depth': 6, 'n_estimators': 150}
	Neural Network (MLP)	0.13	0.22	0.98	0.96	0.95	{'activation': 'tanh', 'hidden_layer_sizes': (100, 50), 'alpha': 0.0001}
Avail. P (mol/kg)	Random Forest Regressor	0.2	0.28	0.96	0.94	0.94	{'max_depth': 10, 'n_estimators': 150}
	K-Nearest Neighbors Regressor	0.25	0.37	0.93	0.89	0.88	{'n_neighbors': 3, 'metric': 'euclidean'}
	Gradient Boosting Regressor	0.26	0.38	0.94	0.88	0.88	{'learning_rate': 0.1, 'n_estimators': 100}
	XGBoost Regressor	0.26	0.38	0.94	0.88	0.88	{'learning_rate': 0.01, 'max_depth': 4, 'n_estimators': 100}
	Neural Network (MLP)	0.21	0.32	0.96	0.92	0.92	{'activation': 'relu', 'hidden_layer_sizes': (75, 50), 'alpha': 0.001}
pH	Random Forest Regressor	0.08	0.12	0.99	0.99	0.98	{'max_depth': None, 'n_estimators': 300}

Distribution plots and box plots as shown in Figure 4 and Figure 5 respectively, provide a nuanced understanding of the data. The distribution of Org. C predominantly between 1.5% and 2.5%, and pH values clustering between 5.5 and 6.2, indicate moderately acidic soils with limited variability. Magnesium and Aluminium levels show narrow ranges, while Calcium and Phosphorus display wider distributions, hinting at differing nutrient availability and soil dynamics. Outliers in parameters like Org. C and Potassium emphasize specific samples that deviate from general trends, warranting further investigation.

Figure 5 is the box plot of features which provides a visual summary of the distribution, central tendency, and variability of the soil nutrients, including Organic Carbon (Org. C), Available Phosphorus (Avail. P), Potassium (K), Calcium (Ca), Magnesium (Mg), Sodium (Na), Hydrogen (H), and Aluminum (Al). It highlights the presence of outliers, differences in concentration ranges, and potential skewness in the data for each nutrient.

Distribution plots further underline soil variability, with bimodal patterns in Avail. P suggesting inconsistent phosphorus availability, and skewness in Org. C and pH distributions reflecting variability in soil health. Uniform distributions for Sodium hint at its consistent levels, while Aluminium's low concentrations corroborate reduced soil acidity.

Temporal trends show promising patterns in nutrient dynamics. Gradual increases in Org. C and Magnesium levels may signify improved soil management, while the decline in Aluminium points to reduced acidity, enhancing soil quality. Fluctuations in Avail. P and stability in Sodium and pH underscore varying soil management impacts over time.

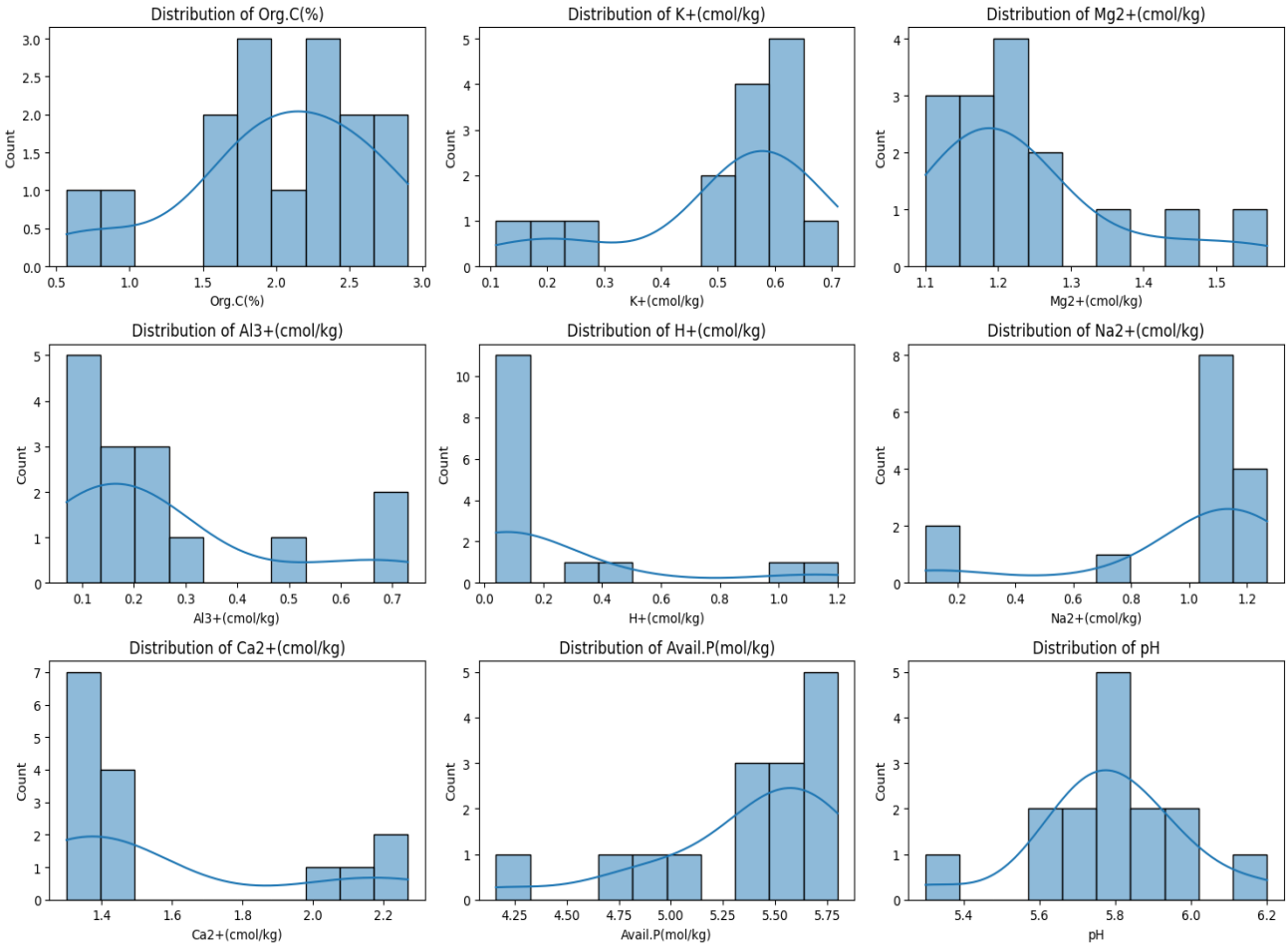

Figure 4*. Distribution plot of features.*

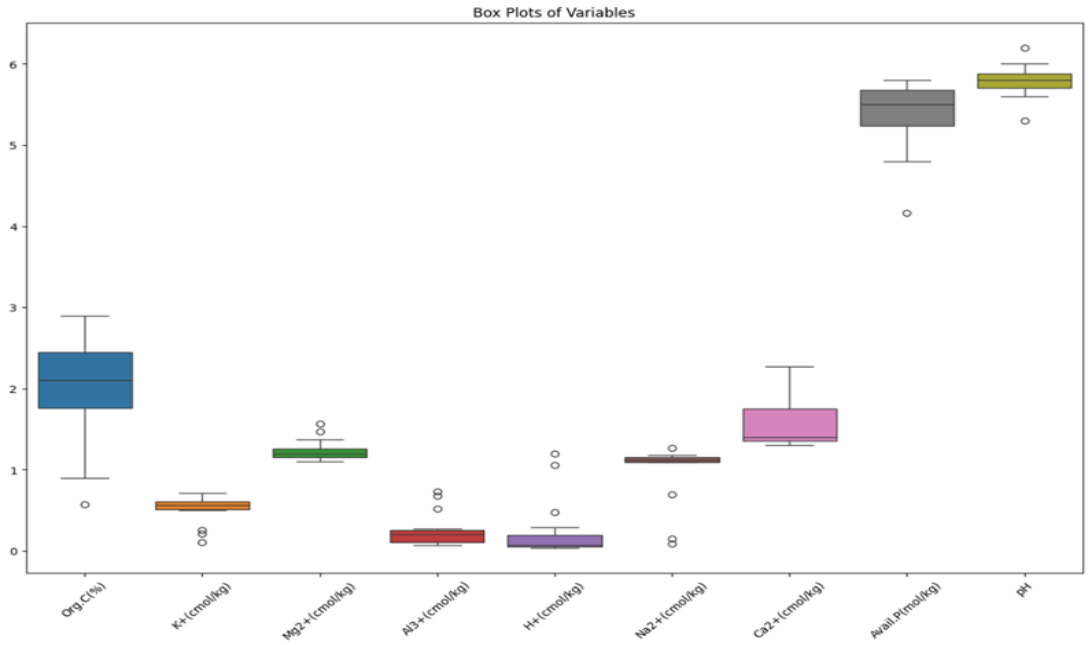

Figure 5*. The box plot of features*

The PairPlot as shown in Figure 6 reveals limited strong correlations among parameters, except for a near-perfect positive correlation (0.99) between Calcium and Magnesium, which aligns with their shared role in soil structure and nutrient balance. Moderate positive correlations, such as Sodium with pH (0.66), suggest interdependence between alkalinity and sodium presence, while a negative correlation between Sodium and Avail. P (-0.63) highlights potential challenges in phosphorus uptake in sodium-rich soils. These relationships emphasize the interconnected nature of soil chemistry.

Based on the EDA, the data preprocessing for machine learning models should include several key steps to ensure data quality and model performance. Since no missing values are present, imputation is unnecessary, but outlier handling is crucial for parameters like Organic Carbon and Potassium to prevent skewed model training. Scaling or normalization should be applied to variables with wide ranges, such as Calcium and Available Phosphorus, to ensure uniform feature contributions. Encoding temporal trends, such as year-wise data, into relevant features can capture temporal dynamics in soil nutrients. Addressing skewness in parameters like pH and Organic Carbon via transformations (e.g., log or power transformation) will improve model interpretability. Additionally, feature engineering to capture correlations, such as interaction terms for Calcium and Magnesium or Sodium and pH, could enhance predictive power.

3.3. Results Discussion

The models' performance demonstrates varying levels of generalization in predicting soil nutrient parameters, with the Random Forest Regressor emerging as the most consistently robust across all features. Models with minimal gaps between training and testing metrics, particularly in RMSE and R^2 values, are deemed better at generalizing to unseen data.

For Na2+ (cmol/kg), the Random Forest Regressor stands out with a testing RMSE of 0.25 and a testing R^2 of 0.95. The minimal disparity between its training and testing metrics underscores its robustness. While Neural Network performs well with a testing RMSE of 0.38 and testing R^2 of 0.88, it falls short of Random Forest's performance. XGBoost, with a testing RMSE of 0.48 and testing R^2 of 0.80, demonstrates weaker generalization.

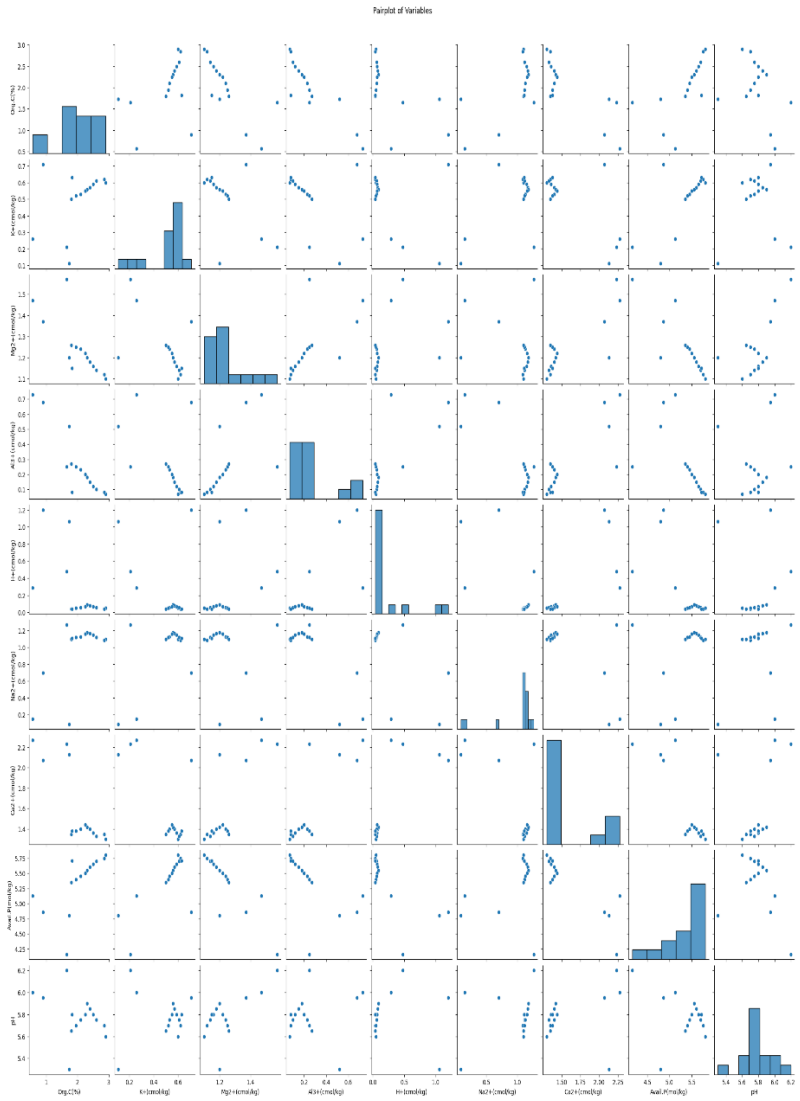

Figure 6. *Pairplot of features.*

In predicting Ca2+ (cmol/kg), Random Forest once again exhibits exceptional performance, achieving the lowest testing RMSE (0.16) and a near-perfect testing R^2 of 0.98. Neural Network and XGBoost follow closely, with testing RMSEs of 0.22 and 0.24 and testing R^2 values of 0.96 and 0.95, respectively. However, Random Forest's smaller gap between training and testing metrics highlights its superior consistency.

For Avail. P (mol/kg), Random Forest demonstrates strong generalization with a testing RMSE of 0.28 and a testing R^2 of 0.94. Neural Network and Gradient Boosting Regressor also perform competitively, with testing RMSEs of 0.32 and 0.38 and testing R^2 values of 0.92 and 0.88, respectively. However, their slightly higher

RMSEs and larger gaps between training and testing R^2 values indicate slightly less reliable performance compared to Random Forest.

In predicting pH, Random Forest again outperforms other models, achieving a testing RMSE of 0.12 and a testing R^2 of 0.99, underscoring its ability to generalize exceptionally well. Neural Network performs closely, with a testing RMSE of 0.20 and testing R^2 of 0.97, but still lags behind Random Forest. XGBoost and K-Nearest Neighbours Regressor, with testing RMSEs of 0.35 and 0.35 and testing R^2 values of 0.90, show larger disparities and weaker performance.

In summary, Random Forest consistently demonstrates the best generalization across all soil nutrient parameters. Its minimal gaps between training and testing metrics, coupled with consistently high R^2 values, make it the most reliable model for predicting unseen data. The consistent performance of Random Forest underscores its robustness, accuracy, and suitability for this research. This analysis highlights the critical importance of selecting the right machine learning model, as performance can vary significantly depending on the dataset and feature being analysed.

3.4. Comparison with Previous Studies

The performance of the Random Forest Regressor in predicting soil nutrient parameters is well-documented in the literature, highlighting its robustness in managing nonlinear relationships and complex interactions inherent in soil data. For example, studies have shown that Random Forest outperforms other machine learning models in soil nutrient prediction tasks, achieving high R^2 such as 0.94 for sodium (Na_2^+) (Haq et al., 2023) and 0.97 for pH (Paepae et al., 2022). This aligns with the findings of this study, which also demonstrates the efficacy of Random Forest in predicting soil nutrient parameters. Conversely, the performance of Neural Networks, while competitive in this study with R^2 values of 0.97 for pH and 0.92 for available phosphorus (Avail. P), has been noted to be superior in capturing intricate soil nutrient variations in other research. The discrepancy in performance may stem from factors such as dataset size, feature selection, or hyperparameter tuning, which are critical in optimizing model performance (Chen et al., 2024). This suggests that while Neural Networks have the potential for high accuracy, their effectiveness can be contingent upon the quality and quantity of the data used.

Additionally, alternative models such as XGBoost and Gradient Boosting Regressor have been recognized as strong contenders for soil prediction tasks, particularly when feature engineering is effectively applied (Tryhuba et al., 2024). However, in this research, their relatively lower generalization capability indicates that further refinement of hyperparameters or feature selection may be necessary to enhance their predictive performance (Yerrabolu et al., 2024). This highlights the importance of model tuning and the adaptability of different algorithms to specific datasets and tasks.

This study reinforces the effectiveness of ensemble models like Random Forest in soil nutrient prediction while also illuminating the variability of model performance based on dataset characteristics, feature engineering, and model tuning. Future research could benefit from exploring deep learning techniques with larger datasets to assess their potential for surpassing traditional ensemble methods in predictive accuracy.

4. CONCLUSION

4.1. Summary of Key Findings

This research evaluates the performance of various ML models in predicting soil fertility based on localized data and engineered features. Key findings include:

Model Performance: Random Forest Regressor consistently demonstrated superior performance across all target variables, achieving the lowest RMSE and highest R^2 scores, indicating strong predictive accuracy and reliability.

Neural Networks (MLP) showed competitive performance, particularly in capturing complex patterns in the data, although slightly behind Random Forest in generalization capability.

Gradient Boosting Regressor and XGBoost Regressor offered balanced performance but showed moderate overfitting compared to Random Forest.

K-Nearest Neighbors Regressor had the lowest performance among models, with higher RMSE and lower R^2 values, especially for complex target features.

Feature Engineering: Polynomial feature generation and normalization significantly improved model accuracy and stability. Key features like pH, organic carbon, and nitrogen were identified as critical predictors of soil fertility.

Temporal trends and correlation-based feature engineering, such as interaction terms, enhanced the predictive power of models.

Exploratory Data Analysis: Nutrient variability and trends highlighted the influence of soil management practices. For example, an increase in organic carbon and a decline in aluminium levels indicated improving soil quality over time. Correlation analysis revealed strong relationships between certain parameters, such as calcium and magnesium, emphasizing their shared role in nutrient balance.

Evaluation Metrics: Combining RMSE, R^2, and K-Fold Cross-Validation provided a comprehensive understanding of model performance, ensuring robust validation and generalization to unseen data.

Actionable Insights: The findings offer practical recommendations for improving soil management practices. By leveraging the Random Forest model's reliability, farmers and agronomists can make data-driven decisions to optimize soil fertility and agricultural productivity.

These findings underscore the importance of machine learning in advancing precision agriculture, demonstrating its potential to enhance sustainability and resource management in agriculture

4.2. Limitations

Despite the promising results, several limitations exist in this study. First, the relatively small dataset size (15 samples) constrains the generalizability of the findings, as larger datasets are typically needed for more robust ML model training and validation. This limitation might affect the models' ability to accurately predict soil fertility in diverse agricultural settings. Second, the study focuses on a limited geographical area, which may not capture the variability in soil characteristics across broader regions, thereby restricting the applicability of the models to other locales.

Additionally, while feature engineering improved model performance, it introduced complexity that could be challenging to replicate without expert knowledge. The study also employed a limited set of ML models and parameters; exploring more advanced algorithms or deeper hyperparameter optimization might yield even better results. Furthermore, certain outliers in the dataset, while addressed during preprocessing, may still influence model predictions, particularly for features with significant variability, such as Organic Carbon and Phosphorus.

Finally, the research does not account for real-time soil monitoring or external factors like climate variability and farming practices, which could significantly influence soil fertility. Future studies should consider integrating larger datasets, diverse geographical data, and real-time measurements to enhance model robustness and applicability.

4.3. Future Research Directions

The findings from this study provide a foundation for advancing machine learning (ML)-based soil fertility prediction models; however, there are key areas for future exploration:

- Integration of IoT and Real-Time Data: Incorporating Internet of Things (IoT) sensors to collect real-time soil data could enhance model accuracy and adaptability, enabling real-world, dynamic soil fertility monitoring systems.

- Spatial and Temporal Analysis: Future studies should explore spatial variability and temporal trends in soil fertility by integrating geospatial data and time-series analysis, offering localized and time-sensitive insights for agricultural practices.

- Hybrid and Deep Learning Models: Investigating hybrid ML models or advanced deep learning techniques, such as Convolutional Neural Networks (CNNs) for image-based soil analysis, could improve predictive performance and provide multi-dimensional insights.

- Scalability and Deployment: Developing scalable frameworks for deploying these models in resource-constrained environments, especially in developing countries, would enhance their practical applicability. Integration with mobile and cloud-based platforms can improve accessibility for farmers.

- Impact of Climate Change: Expanding datasets to include climate variables, such as rainfall and temperature, could help predict how changing environmental conditions influence soil fertility over time.

- Soil Microbiome and Genomics Integration: Incorporating soil microbiome data and genome-wide analysis could provide a more holistic approach to fertility prediction, integrating biological and chemical factors into ML models.

By addressing these areas, future research can build on the existing framework to create more accurate, scalable, and actionable solutions for sustainable agriculture.

AUTHOR CONTRIBUTIONS

All authors contributed to the study's conception and design. Material preparation, data collection and analysis were performed by C.O.N, E.N.V, O.C.C, E.O.C, N.C.V and E.C.N. The first draft of the manuscript was written by C.O.N, and all authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGEMENT

The authors acknowledge the efforts of the management of the Nnamdi Azikiwe University Akwa in stimulating this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abishek, J. (2023). Soil texture prediction using machine learning approach for sustainable soil health management. International Journal of Plant and Soil Science, 35(19), 1416-1426. https://doi.org/10.9734/ijpss/2023/v35i193685

Asif, M. (2024). Leveraging machine learning for soil fertility prediction and crop management in agriculture. https://doi.org/10.21203/rs.3.rs-4310747/v1

Awais, M. (2023). Ai and machine learning for soil analysis: an assessment of sustainable agricultural practices. Bioresources and Bioprocessing, 10(1). https://doi.org/10.1186/s40643-023-00710-y

Barrena-González, J. (2024). Looking for optimal maps of soil properties at the regional scale. International Journal of Environmental Research, 18(4). https://doi.org/10.1007/s41742-024-00611-8

Chen, Y., Shi, T., Li, Q., Wang, Z., Wang, R., Wang, F., … and Li, Y. (2024). Mapping soil properties in tropical rainforest area using uav-based hyperspectral images and lidar points. https://doi.org/10.21203/rs.3.rs-4273924/v1

Clark, J., Fernández, F., Veum, K., Camberato, J., Carter, P., Ferguson, R., … and Shanahan, J. (2019). Predicting economic optimal nitrogen rate with the anaerobic potentially mineralizable nitrogen test. Agronomy Journal, 111(6), 3329-3338. https://doi.org/10.2134/agronj2019.03.0224

Dinh, T., Nguyen, H., Tran, X., and Hoang, N. (2021). Predicting rainfall-induced soil erosion based on a hybridization of adaptive differential evolution and support vector machine classification. Mathematical Problems in Engineering, 2021, 1-20. https://doi.org/10.1155/2021/6647829

Ezenwankwo, S., Adeagbo, A.A., Lawal, S., Idoghor, S. M. and Chukwu, O.(2020). Evaluation of early growth of Maesobotrya barteri (Hutch) seedlings underdifferent growing media and watering regime. In: Forestry Development in Nigeria: Fiftyyears of interventions and Advocacy. At the 42nd Annual conference of the Forestry Association of Nigeria (FAN), on 23-28 November, 2020, Ibadan, Nigeria. pp. 730-736.

Groebner, B. (2024). Soil biological and physical measurements did not improve the predictability of corn response to phosphorus fertilization. Agronomy Journal, 116(4), 2048-2059. https://doi.org/10.1002/agj2.21612

Hamidović, S., SOFTIC, A., Topčić, F., Tvica, M., Lalević, B., and Stojanova, M. (2023). Impact of soil management practice on the abundance of microbial populations. The Journal Agriculture and Forestry, 69(2). https://doi.org/10.17707/agricultforest.69.2.12

Haq, Y., Shahbaz, M., Asif, H., Al-Laith, A., and Alsabban, W. (2023). Spatial mapping of soil salinity using machine learning and remote sensing in kot addu, pakistan. Sustainability, 15(17), 12943. https://doi.org/10.3390/su151712943

Harris, J., Bledsoe, R., Guha, S., Omari, H., Crandall, S., Burghardt, L., … and Couradeau, E. (2024). The activity of soil microbial taxa in the rhizosphere predicts the success of root colonization.. https://doi.org/10.1101/2024.12.07.627353

Hu, Z., Ding, Z., Al-Yasi, H., Ali, E., Eissa, M., Abou-Elwafa, S., … and Hamada, A. (2021). Modelling of phosphorus nutrition to obtain maximum yield, high p use efficiency and low p-loss risk for wheat grown in sandy calcareous soils. Agronomy, 11(10), 1950. https://doi.org/10.3390/agronomy11101950

Inoyatova, M. (2024). Data mining for assessing soil fertility. E3s Web of Conferences, 494, 02012. https://doi.org/10.1051/e3sconf/202449402012

Jabborova, D., Choudhary, R., Azimov, A., Jabbarov, Z., Selim, S., Abu-Elghait, M., … and Elsaied, A. (2022). Composition of zingiber officinale roscoe (ginger), soil properties and soil enzyme activities

grown in different concentration of mineral fertilizers. Horticulturae, 8(1), 43. https://doi.org/10.3390/horticulturae8010043

Jia, X. (2023). Development of soil fertility index using machine learning and visible-near-infrared spectroscopy. Land, 12(12), 2155. https://doi.org/10.3390/land12122155

Kroyan, S. (2024). Anthropogenic changes of the agricultural production features of river valley-escarpment soils in martuni region, sevan basin, ra. E3s Web of Conferences, 510, 01009. https://doi.org/10.1051/e3sconf/202451001009

Lepcha, N. and Devi, N. (2020). Effect of land use, season, and soil depth on soil microbial biomass carbon of eastern himalayas. Ecological Processes, 9(1). https://doi.org/10.1186/s13717-020-00269-y

Li, M., Ji, R., Wang, M., and Zheng, L. (2020). Comparison of soil total nitrogen content prediction models based on vis-nir spectroscopy. Sensors, 20(24), 7078. https://doi.org/10.3390/s20247078

Liu, W., Yang, Z., Ye, Q., Peng, Z., Zhu, S., Chen, H., … and Huang, H. (2023). Positive effects of organic amendments on soil microbes and their functionality in agro-ecosystems. Plants, 12(22), 3790. https://doi.org/10.3390/plants12223790

Longchamps, L., Mandal, D., and Khosla, R. (2022). Assessment of soil fertility using induced fluorescence and machine learning. Sensors, 22(12), 4644. https://doi.org/10.3390/s22124644

Ma, G., Cheng, S., He, W., Dong, Y., Qi, S., Nai-mei, T., … and Wei, T. (2023). Effects of organic and inorganic fertilizers on soil nutrient conditions in rice fields with varying soil fertility. Land, 12(5), 1026. https://doi.org/10.3390/land12051026

Mendoza, M., Mora-Bautista, M., Cué, J., Escudero, J., and Etchevers, J. (2021). Field production of kale (brassica oleracea var. acephala) with different nutrition sources. Agro Productividad. https://doi.org/10.32854/agrop.v14i10.1954

Mesfin, S., Haile, M., Gebresamuel, G., Zenebe, A., and Gebre, A. (2021). Establishment and validation of site-specific fertilizer recommendation for increased barley (hordeum spp.) yield, northern Ethiopia. Helion, 7(8), e07758. https://doi.org/10.1016/j.heliyon.2021.e07758

Musanase, C. (2023). Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices. Agriculture, 13(11), 2141. https://doi.org/10.3390/agriculture13112141

Nelson, A., Narrowe, A., Rhoades, C., Fegel, T., Daly, R., Roth, H., … and Wilkins, M. (2022). Wildfire-dependent changes in soil microbiome diversity and function. Nature Microbiology, 7(9), 1419-1430. https://doi.org/10.1038/s41564-022-01203-y

Ning, Q., Hättenschwiler, S., Lü, X., Kardol, P., Zhang, Y., Wei, C., … and Han, X. (2021). Carbon limitation overrides acidification in mediating soil microbial activity to nitrogen enrichment in a temperate grassland. Global Change Biology, 27(22), 5976-5988. https://doi.org/10.1111/gcb.15819

Nwamekwe, C. O., Ewuzie, N. V., Igbokwe, N. C., Okpala, C. C., and U-Dominic, C. M. (2024). Sustainable Manufacturing Practices in Nigeria: Optimization and Implementation Appraisal. Journal of Research in Engineering and Applied Sciences, 9(3). URL

Nwamekwe, C. O., Ewuzie, N. V., Igbokwe, N. C., U-Dominic, C. M., and Nwabueze, C. V. (2024). Adoption of Smart Factories in Nigeria: Problems, Obstacles, Remedies and Opportunities. International Journal of Industrial and Production Engineering, 2(2). Retrieved from URL

Nwamekwe, C. O., Okpala, C. C., and Okpala, S. C., (2024). Machine Learning-Based Prediction Algorithms for the Mitigation of Maternal and Fetal Mortality in the Nigerian Tertiary Hospitals. International Journal of Engineering Inventions, 13(7), PP: 132-138. URL

Omar, G. and Sule, H. (2017). Fertility status of floodplain soils along river andlt;iandgt; tatsewarkiandlt;/iandgt;, kano. Bayero Journal of Pure and Applied Sciences, 9(2), 17. https://doi.org/10.4314/bajopas.v9i2.4

Osaigbovo, A. and Law-Ogbomo, K. (2014). Effects of spent engine oil polluted soil and organic amendment on soil chemical properties, micro-flora on growth and herbage of andlt;iandgt;telfairia occidentalisandlt;/iandgt; (hook f).. Bayero Journal of Pure and Applied Sciences, 6(1), 72. https://doi.org/10.4314/bajopas.v6i1.15

Paepae, T., Bokoro, P., and Kyamakya, K. (2022). A virtual sensing concept for nitrogen and phosphorus monitoring using machine learning techniques. Sensors, 22(19), 7338. https://doi.org/10.3390/s22197338

Pagliarini, M., Castilho, R., Moreira, E., Mariano-Nasser, F., and Alves, M. (2019). Development of hymenaea courbaril l. var. stilbocarpa seedlings in different fertilizers and substrate composition. Agrarian, 12(43), 8-15. https://doi.org/10.30612/agrarian.v12i43.4184

Palansooriya, K., Wong, J., Hashimoto, Y., Huang, L., Rinklebe, J., Chang, S., … and Ok, Y. (2019). Response of microbial communities to biochar-amended soils: a critical review. Biochar, 1(1), 3-22. https://doi.org/10.1007/s42773-019-00009-2

Pant, H., Lohani, M., and Bhatt, A. (2019). Impact of physico-chemical properties for soils type classification of oak using different machine learning techniques. International Journal of Computer Applications, 177(17), 38-44. https://doi.org/10.5120/ijca2019919617

Patil, A., Kulkarni, V., and Desai, S. (2023). Soil fertility prediction. International Journal for Research in Applied Science and Engineering Technology, 11(8), 1241-1247. https://doi.org/10.22214/ijraset.2023.55225

Prince, M., Mankessi, F., Sun, S., and Fan, X. (2021). Effects of alkaline fertilizer and rice cultivation (oryza sativa l.) on remediation of soils polluted with cadmium (cd). Journal of Applied Biosciences, 157, 16182-16193. https://doi.org/10.35759/jabs.157.4

Rajamanickam, J. and Mani, S. (2021). Kullback chi square and gustafson kessel probabilistic neural network-based soil fertility prediction. Concurrency and Computation Practice and Experience, 33(24). https://doi.org/10.1002/cpe.6460

Razanov, S. (2024). The content of heavy metals and trace elements in different soils used under the conditions of homestead plots and field agricultural lands of ukraine. Journal of Ecological Engineering, 25(6), 42-50. https://doi.org/10.12911/22998993/186820

Reddy, L. (2024). Applying machine learning to soil analysis for accurate farming. Matec Web of Conferences, 392, 01124. https://doi.org/10.1051/matecconf/202439201124

Rehman, O., Mehdi, S., Abad, R., Saleem, S., Khalid, R., Alvi, S., … and Munir, A. (2021). Soil characteristics and fertility indexation in gujar khan area of rawalpindi. Pakistan Journal of Scientific and Industrial Research Series a Physical Sciences, 64(1), 46-51. https://doi.org/10.52763/pjsir.phys.sci.64.1.2021.46.51

Sandhya, K., Gayathri, B., Papireddy, M., R, P., Vishwanath, V., Swathi, B., … and Naveen, D. (2023). Assessing the nutrient status and soil fertility using nutrient indexed of farmer's fields in chikkaballapura district, karnataka. International Journal of Plant and Soil Science, 35(23), 639-649. https://doi.org/10.9734/ijpss/2023/v35i234283

Saraiva, T., Ventura, S., Brito, E., Rocha, S., Costa, R., Pereira, A., … and Araújo, A. (2022). Temporal stability of soil microbial properties in responses to long-term application of compost obtained from tannery sludge. Sustainability, 14(24), 16736. https://doi.org/10.3390/su142416736

Sofo, A., Mininni, A., and Ricciuti, P. (2020). Soil macrofauna: a key factor for increasing soil fertility and promoting sustainable soil use in fruit orchard agrosystems. Agronomy, 10(4), 456. https://doi.org/10.3390/agronomy10040456

Sridevy, S., Raj, M., Kumaresan, P., Balakrishnan, N., Tilak, M., Raj, J., … and Rani, P. (2023). Mapping of soil properties using machine learning techniques. International Journal of Environment and Climate Change, 13(8), 684-700. https://doi.org/10.9734/ijecc/2023/v13i81997

Tryhuba, I., Tryhuba, A., Hutsol, T., Cieszewska, A., Andrushkiv, O., Głowacki, S., … and Sojak, M. (2024). Prediction of biogas production volumes from household organic waste based on machine learning. Energies, 17(7), 1786. https://doi.org/10.3390/en17071786

Yageta, Y., Osbahr, H., Morimoto, Y., and Clark, J. (2019). Comparing farmers' qualitative evaluation of soil fertility with quantitative soil fertility indicators in kitui county, kenya. Geoderma, 344, 153-163. https://doi.org/10.1016/j.geoderma.2019.01.019

Yang, Q., Peng, J., Ni, S., Zhang, C., Wang, J., and Cai, C. (2024). Soil erosion-induced decline in aggregate stability and soil organic carbon reduces aggregate-associated microbial diversity and multifunctionality of agricultural slope in the mollisol region. Land Degradation and Development, 35(11), 3714-3726. https://doi.org/10.1002/ldr.5163

Yerrabolu, V., Kasireddy, I., Jasmine, K., Vamsi, T., Joshua, N., Kumar, V., … and Rao, D. (2024). Performance comparison of random forest regressor and support vector regression for solar energy prediction. Iop Conference Series Earth and Environmental Science, 1375(1), 012013. https://doi.org/10.1088/1755-1315/1375/1/012013

Yu, X. (2024). Prediction model of nitrogen, phosphorus, and potassium fertilizer application rate for greenhouse tomatoes under different soil fertility conditions. Agronomy, 14(6), 1165. https://doi.org/10.3390/agronomy14061165

Zhao, Z., He, J., Quan, Z., Wu, C., Sheng, R., Zhang, L., … and Geisen, S. (2020). Fertilization changes soil microbiome functioning, especially phagotrophic protists. Soil Biology and Biochemistry, 148, 107863. https://doi.org/10.1016/j.soilbio.2020.107863

Zheng, C., Yang, X., Liu, Z., Liu, K., and Huang, Y. (2022). Spatial distribution of soil nutrients and evaluation of cultivated land in xuwen county. Peerj, 10, e13239. https://doi.org/10.7717/peerj.13239

Ziyadullaev, D. (2024). Ensemble data mining methods for assessing soil fertility. E3s Web of Conferences, 494, 02013. https://doi.org/10.1051/e3sconf/202449402013