American Journal of Robotics and Intelligent Systems
2026, Vol. 1, No. 3, pp. 73–83
https://doi.org/10.11648/j.ajris.20260103.11

SciencePG
Science Publishing Group

Research Article

Analysis and Prediction of Soil Fertility Using Machine Learning Techniques

Tigist Mintesnot Tewabe*

Department of Information Technology, University of Gondar, Gondar, Ethiopia

Abstract

Soil fertility is a crucial aspect of agricultural productivity and sustainability, determines the soil's capacity to provide essential nutrients necessary for plant growth and development. This study focuses on the analysis and prediction of soil fertility using ensemble learning techniques in the South Gondar Zone. By examining various soil parameters, including macro and micronutrients, soil structure, pH, and organic matter content, the research aims to develop predictive models that accurately assess soil fertility levels. The objective of this study is to analyze and predict soil fertility using machine learning techniques, specifically targeting the south Gondar Zone, in the Amhara region of Ethiopia. The dataset for this study comprises 20,168 instances, including both fertile and non-fertile samples, with 17 selected attributes. Several machine learning models were evaluated on both the original and SMOTE-balanced datasets. The models included Random Forest, AdaBoost, and XGBoost classifiers. I applied Random Forest classifier consistently demonstrated the highest performance, with testing accuracies of 94.55% on the original dataset and 94.37% on the SMOTE-balanced dataset. AdaBoost also showed strong performance, achieving testing accuracies of 94.6% and 94.31% on the original and SMOTE-balanced datasets, respectively. XGBoost performed well but was slightly less accurate compared to the ensemble methods. However, XGBoost showed robust performance on both datasets, with testing accuracies of 94.04% and 94.27%. Feature importance analysis using the Random Forest classifier has shown that factors such as Clay, CEC, $CaCO_3$, Sand, and Mn significantly impact soil fertility prediction. as a conclusion Random Forest classifier, an ensemble-based learning technique, was the most reliable and accurate model for predicting soil fertility. These results highlight the importance of specific soil properties in determining fertility and can guide targeted soil management practices to improve agricultural productivity.

Keywords

Random Forest, AdaBoost, Random Forest Classifier, SMOTE

1. Background and Problem Statement

The increasing global demand for food and energy, coupled with limited access to essential agricultural resources such as land and water, poses a significant challenge to sustainable agricultural production, particularly in developing regions like sub-Saharan Africa. Ethiopia, with a rapidly growing population exceeding 110 million, faces mounting pressure to enhance agricultural productivity while preserving soil health and natural resources. Declining soil fertility, driven by continuous cultivation, reduced fallowing, and

*Correspondence: Tigist Mintesnot Tewabe (tigist.mintesnot@uog.edu.et), Tigist Mintesnot Tewabe (tigistmintesnot@gmail.com)

Received: 6 March 2026; **Accepted:** 28 July 2026; **Published:** 17 August 2026

competing uses of organic inputs such as dung and crop residues for energy, has further intensified the need for efficient nutrient management [1, 2, 14, 15]. The judicious use of fertilizers, combined with appropriate agronomic practices, has been shown to improve crop yields and nutrient use efficiency while minimizing environmental impacts. However, effective nutrient management requires accurate assessment of soil properties and site-specific recommendations. Advances in smart farming, particularly through machine learning and data science, provide new opportunities to analyze complex soil and environmental data for improved soil fertility prediction [3, 4]. By leveraging historical and real-time data such as temperature, rainfall, humidity, and soil pH, machine learning models can estimate key soil nutrients like nitrogen, phosphorus, and potassium, enabling data-driven decision-making. Therefore, this study aims to conduct a comparative analysis of soil fertility prediction using various machine learning algorithms, evaluating their performance in assessing soil fertility and supporting sustainable agricultural productivity.

Ethiopia is rich in agricultural products that significantly influence the country's economy, yet traditional farming practices on the same plots over time have led to soil nutrient depletion and imbalance, reducing farmers' yields and negatively impacting national economic performance. To address this, scholars have recommended the use of chemical fertilizers, especially given the scarcity of organic alternatives, and awareness campaigns have helped many farmers adopt them [5, 7]. However, the effectiveness of chemical fertilizers varies by location due to differences in initial soil fertility, and improper application can lead to lodging, high production costs, and reduced benefits. To optimize fertilizer use, agricultural researchers have conducted field experiments, collecting soil samples from representative farms across various woredas to assess initial soil fertility [8]. Numerous studies have explored machine learning techniques for soil fertility prediction: for instance, Random Forest algorithms have been applied to predict soil grades based on macro- and micro-nutrients, though linear models often lacked accuracy for predicting nutrient quantities from environmental factors. Other research compared Decision Tree, Random Forest, Gradient Boosting, and K-Nearest Neighbors, finding showed that Random Forest generally superior in AUC accuracy, though Phosphorous prediction remained challenging due to dataset limitations. Some studies also used Decision Tree, KNN, and Naive Bayes to predict crop yields from factors like soil pH, moisture, sunlight, and temperature, yet these datasets were often imbalanced and provided limited insight into the influence of these factors on fertility. Importantly, few studies in Ethiopia's agricultural sector have incorporated feature engineering to create new predictive features, relying instead on pre-existing datasets from the Ministry of Agriculture or other sources. Existing research also suffers from limitations in study area, data size, predictor variables, feature selection methods, and handling of imbalanced data. To address these gaps, this study aims to employ advanced machine learning techniques to identify key characteristics influencing soil fertility in Ethiopia, focusing on the Amhara Region (South Gondar zone), and will explore which attributes are most important for prediction, which classification algorithms are most suitable, and the predictive performance of the proposed models [6].

2. Objectives of the Study

2.1. General Objective

The main objective of this research was to analyze soil properties and predict soil fertility using machine learning techniques, particularly ensemble learning methods.

2.2. Specific Objectives

1) To identify important soil attributes that affect fertility.
2) To compare different machine learning classification models.
3) To develop a predictive model that accurately predicts soil fertility.
4) To improve prediction accuracy using data preprocessing and sampling techniques.

3. Research Methodology

3.1. Dataset Preparation

Collecting and structuring the data that is used for analysis is one of the tasks that needs close attention in the process of data mining. The data for this specific research was collected from National Soil Testing Center in south Gondar Zone (SGZNSTC). NSTC is established to develop guideline for fertilizer recommendations based on initial soil fertility status requirement as one of its major objectives. This in return achieves one of the goals of the government which is ensuring food security. In this connection, NSTC work hard with different laboratories in different parts of Ethiopia. These laboratories keep data in different soil parameters after field and laboratory experiments. In the soil, there are different nutrients in different proportion. Soil pH, nitrogen, phosphorous and organic matter are some of the soil properties (nutrients). These properties can contribute for effective growth of crops if they are in sufficient proportion, otherwise, they may cause damage on the growing crop. Moreover, the most important crops in South Gondar zone (woreta zuria wereda) such as wheat, rice, teff and other crops in terms of production and area coverage. So far, there are a number of records collected from experiments that have been made in different years. Among these, 2003 to 2013 data which is recent were chosen for this particular research since many of the soil parameters are there and they are also helpful to know the current situation. The record includes different crops. The original soil

prosperity data contain 20168 both fertile and not fertile instances and 16 features.

In general the dataset used in this study contained:

1) 20,168 soil samples.
2) 17 soil features such as clay, sand, pH, calcium carbonate, and other nutrients.
3) Two classes: fertile soil and non-fertile soil.

3.2. Data Preprocessing

Several preprocessing steps were performed to prepare the dataset for model training, including handling missing and duplicate values to ensure data quality, selecting relevant features using correlation analysis, removing highly correlated features to reduce redundancy and improve model efficiency, and balancing the dataset using SMOTE (Synthetic Minority Over-Sampling Technique) to address class imbalance and enhance model performance; as shown in the following figure, the dataset contains sixteen different attributes, and a detailed explanation of each attribute is provided in the the following Figure 1.

```
df = pd.read_csv('SGSFPDataset2016.csv')
df.head()
```

	pH	EC	OC	OM	N	P	K	Zn	Fe	Cu	Mn	Sand	Silt	Clay	CaCO3	CEC	Fertility
0	7.74	0.40	0.01	0.01	75	20.0	279.0	0.48	6.4	0.21	4.7	84.3	6.8	8.9	6.72	7.81	Fertile
1	9.02	0.31	0.02	0.03	85	15.7	247.0	0.27	6.4	0.16	5.6	90.4	3.9	5.7	4.61	7.19	Fertile
2	7.80	0.17	0.02	0.03	77	35.6	265.0	0.46	6.2	0.51	6.1	84.5	6.9	8.6	1.53	12.32	Fertile
3	8.36	0.02	0.03	0.05	106	6.4	127.0	0.50	3.1	0.28	2.3	93.9	1.7	4.4	0.00	1.60	Non Fertile
4	8.36	1.08	0.03	0.05	96	10.5	96.0	0.31	3.2	0.23	4.1	91.5	4.1	4.4	9.08	7.21	Non Fertile

Figure 1. *Dataset Preparation.*

The following Table 1 shows the attributes with their description and data type.

Table 1. *Missing Values Handling Method.*

S. No	Attribute name	Data type	Description
1	PH	Number	Soil pH Value
2	EC	Number	Electric Conductivity
3	OC	Number	Organic Carbon
4	OM	Number	Organic Matter
5	N	Number	Nitrogen Content
6	P	Number	Phosphorous Content
7	K	Number	Potassium Content
8	Ze	Number	Zinc Content
9	Fe	Number	Iron Content
10	Cu	Number	Copper Content
11	Mn	Number	Manganese Content
12	Slit	Number	Soil Composition
13	Clay	Number	Clay content of the soil
14	Sand	Number	Sand content of the soil
15	$CaCO_3$	Number	Sodium Bi-Carbonate Content

S. No	Attribute name	Data type	Description
16	CEC	Number	Cation Exchange Capacity
17	Fertility	Number	Fertile (1) or not fertile (0)

It refers to fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies. Fill in missing attribute or class values can be performed by using the attribute mean (or majority nominal value) or by using the attribute mean (majority nominal value) for all samples belonging to the same class.

```
In [70]: 1 df.isnull().sum()

Out[70]: pH          0
         EC          0
         OC          0
         OM          0
         N           0
         P           0
         K           1
         Zn          0
         Fe          0
         Cu          0
         Mn          0
         Sand      180
         Silt      360
         Clay        0
         CaCO3       0
         CEC         0
         Fertility   0
         dtype: int64
```

Figure 2. *Checking Dataset for null values.*

From the above figure, we observed the attribute sand and silt have a Missing value 180 and 360 respectively. Describe the above dataset, which shows the minimum value, maximum value, mean value, count, standard deviation, etc.

Finally, we filled the missing values in our features using the mean value of each feature, which means we filled the mean value to handle missing data. Check once more to see if there are any null values.

```
In [71]: 1 df.describe()
```

Out[71]:

	pH	EC	OC	OM	N	P	K	Zn	Fe	Cu	
count	20168.000000	20168.000000	20168.000000	20168.000000	20168.000000	20168.000000	20167.000000	20168.000000	20168.000000	20168.000000	201
mean	8.293199	0.296299	0.199299	0.338448	173.494942	13.603501	215.498636	0.496523	4.598453	0.312531	
std	0.541592	0.265754	0.172215	0.296920	61.378056	8.514737	86.512579	0.263062	1.876528	0.178082	
min	6.520000	0.020000	0.010000	0.010000	75.000000	1.800000	70.000000	0.040000	1.000000	0.010000	
25%	8.060000	0.130000	0.080000	0.130000	124.000000	7.900000	145.000000	0.310000	3.200000	0.160000	
50%	8.340000	0.190000	0.140000	0.240000	164.000000	10.500000	187.000000	0.460000	4.200000	0.280000	
75%	8.620000	0.360000	0.240000	0.410000	218.000000	18.700000	268.000000	0.640000	6.100000	0.440000	
max	9.480000	1.580000	0.740000	1.270000	278.000000	35.600000	480.000000	1.300000	9.100000	0.730000	

Figure 3. *Describe the dataset which shows the minimum value, maximum value, mean value.*

```
In [74]:  1 df.isnull().sum()

Out[74]: pH           0
         EC           0
         OC           0
         OM           0
         N            0
         P            0
         K            0
         Zn           0
         Fe           0
         Cu           0
         Mn           0
         Sand         0
         Silt         0
         Clay         0
         CaCO3        0
         CEC          0
         Fertility    0
         dtype: int64
```

Figure 4. *The existence there are any null values.*

As shown Figure 3. we observe that there are no null values in any column indicating that the dataset is complete and does not require any sort of treatment for missing values.

3.3. Identify Outliers

Cleaning data can also be done by correcting inconsistent data.

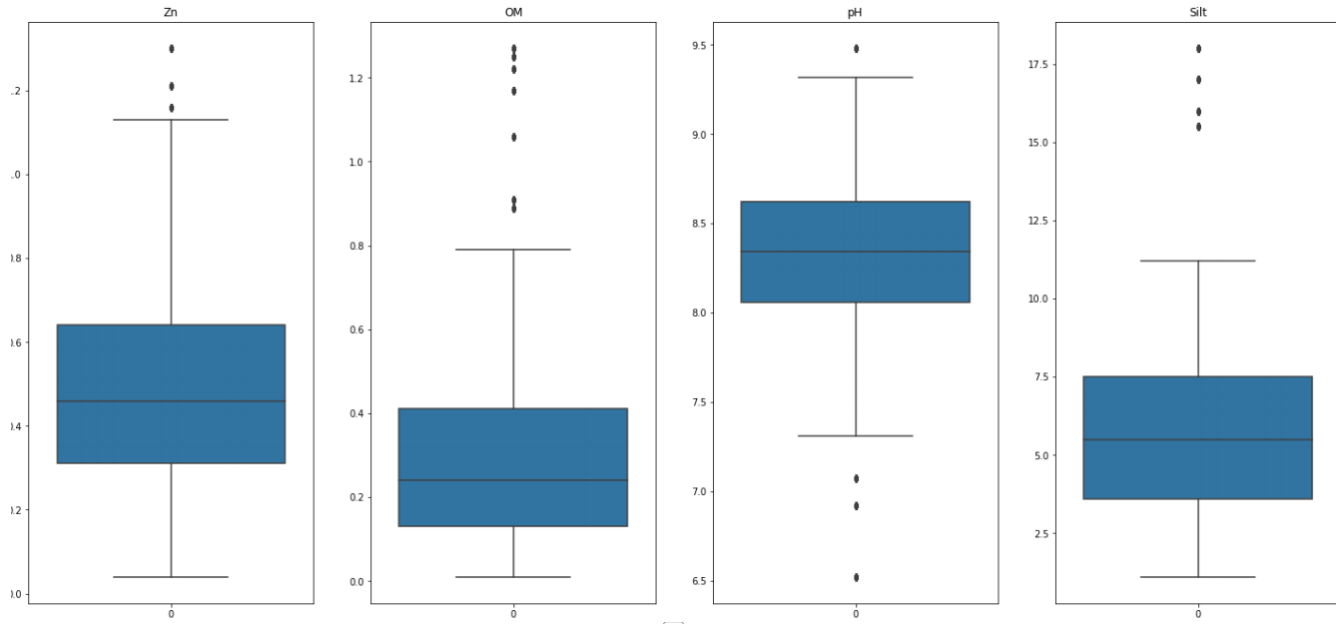

Figure 5. *The data cleaning to identify possible outliers.*

The data of this research share these problems. Hence it needs cleaning; without cleaning the data, analysis become under question and one may arrive at defective output which is a big loss in the research work. In this specific research, data cleaning was performed by consulting the domain expert as well as based on the researcher's own observation.

There were a number of records whose values of many attributes were missed. In this case, it is unadvisable to fill all these values of attributes.

3.4. Exploratory Data Analysis and Data Visualization

1) Fertility vs PH

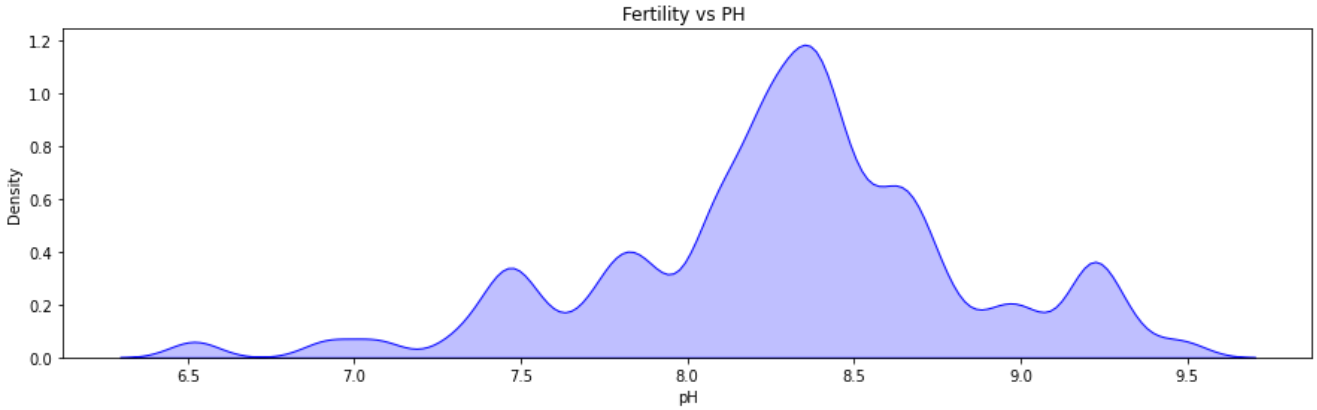

Figure 6. *The value difference between Fertility vs PH.*

From the above figure we observed between pH factor 6 and 8.5 has high forcibility indicators.

2) To understand the dataset's size and as shown Figure 6 below Dataset comprises of 20168 rows and 17 columns and there were 11168 fertile and 9000 not soil fertile.

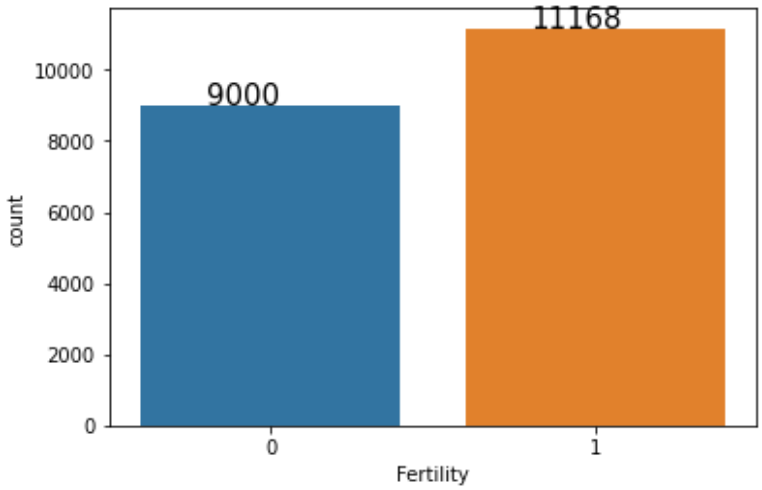

Figure 7. *Understand the dataset's size.*

3.5. Handling Class Imbalance

Sampling imbalance can be addressed using under-sampling, which removes instances from the majority class, over-sampling, which increases the minority class, or a combination of both techniques. In this study, the Synthetic Minority Over-sampling Technique (SMOTE) was used to correct class imbalance in the dataset. SMOTE works by generating synthetic data for the minority class using a nearest neighbor algorithm while maintaining the majority class. After applying SMOTE, the minority class increased from 9,000 to 11,168 samples, while the majority class remained unchanged.

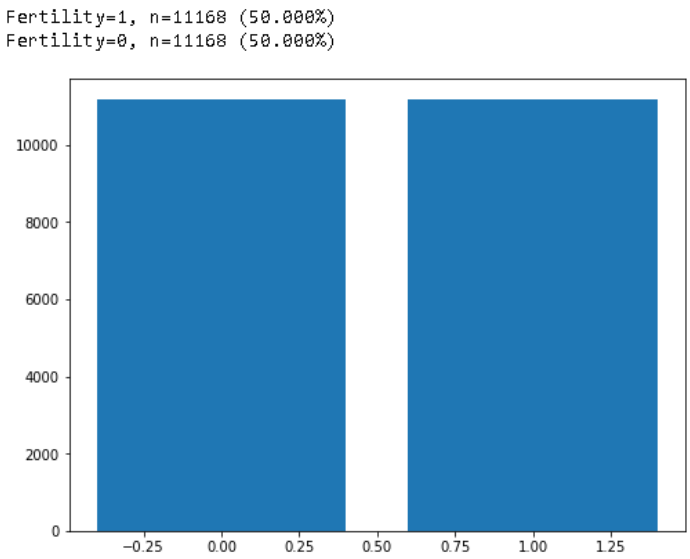

Figure 8. *SMOTE resembling Technique.*

3.6. Feature Correlation

Visualize the correlation of all the features using the heat map function of Seaborn. Using filtering metrics (Pearson correlation), the model evaluates the significance of each feature at this stage. We saw in the Figure 9 bellow, the Heatmap and Scatterplot; we can easily observe that N, OC and OM are highly correlated with 88% and also OC and OM are highly correlated 100.

Figure 9. *This figure illustrates feature selection processes.*

3.7. Tools for Predictive Analysis

Python is a free, open-source, generic language that can be easily integrated and improved with additional code, compared to tool suites. In addition, it has good visualization, plotting capabilities, and libraries with complete statistical algorithms. Since Python met all the needs and was easier to use for the author of the thesis, it was selected as the tool for the project. The Anaconda environment and Python programming language were used by the researcher, with label encoding via the Scikit-Learn package and Python code written within Jupiter Notebook.

The researchers attempt to evaluate the three categorization models in this part using various evaluation indicators. Using the classification techniques, we developed prediction models and trained them with training data, then ran test data for each model and validated them using the confusion matrix, recall, precision, F1-measure, and Roc curve. Following the evaluation, the results of the categorization models are discussed, and the research questions are attempted to be answered.

3.8. Building the Predictive Model

The main objective of this study is to use pre-processed labeled data to train supervised machine learning classification models that learn the relationship between input features and soil fertility status, identify the most relevant attributes, evaluate suitable algorithms, and accurately classify soil samples as fertile (1) or not fertile (0) to determine the best-performing model for soil fertility prediction.

3.8.1. Dividing the Data Set into Two Parts: Training and Testing

The researcher will separate the data used to train a model into two categories: "Training data" and "Testing data." The 'training data set' is used to train the classifier, while the 'test data set' is used to test it. The researcher wants to divide the dataset into a training set and a testing set using the function train test split (). Features, target, and test-size are the three parameters that must be passed. For consistency and fair comparison, we randomly split the dataset into a training set (80%) and a test set (20%). We used the same test set for all of the studies to ensure uniformity and fairness in the comparison. The dataset is separated into two halves in 80:20 ratios. According to this, 20% of the data is utilized for model testing, and 80% is used for model training.

3.8.2. Proposed Model Development

In this study, three classification algorithms—Random Forest, XGBoost, and AdaBoost—were trained and evaluated using cross-validation on both the original dataset and SMOTE up-sampled data, optimized through grid search hyperparameter tuning, assessed with a confusion matrix, and the most accurate model was selected to answer the soil fertility prediction research questions.

(i). Random Forest Classifier

According to [9] to the researchers, the ensemble machine learning method known as Random Forests (RF) can aid in classification and regression. It expands the fundamental idea of a single classification tree by increasing the quantity of classification trees used in the training phase. To classify an instance, each tree in the forest generates its own answer (vote for a class), and the model selects the class with the most votes from the forest's trees. One of the most significant advantages of RF over typical decision trees is its protection against over fitting, which allows the model to perform well.

(ii). AdaBoost

AdaBoost is a popular ensemble machine learning algorithm that improves classification accuracy by sequentially combining multiple weak learners into a strong learner, reducing bias and variance while iteratively correcting misclassification errors, making it an efficient and effective solution for predictive modeling tasks [10, 13].

(iii). XGBoost Classifier

XGBoost is an efficient and scalable ensemble algorithm that iteratively combines weak learners using gradient boosting with regularization to reduce overfitting, improve prediction accuracy, and minimize residual errors, while in this study the classification models were trained and evaluated on both the original and SMOTE-resampled datasets using resampling techniques to assess their performance [11, 12].

3.9. Evaluation Model Performance

In this study, model performance was evaluated using accuracy, precision, recall, and F1-score, where accuracy represents the proportion of correctly classified samples, and the confusion matrix is used as a performance evaluation tool to summarize and assess the classification results by comparing the predicted and actual values of the test data^73ç.

The following Table 2 shows Confusion Matrix for Evaluating Classifier Accuracy.

Table 2. The classifier accuracy for the confusion matrix of the model.

Predict	1	0
1	True positives (TP): Cases in which prediction is "1," soil fertility will be fertile, and it is fertile	False negatives (FN): Cases in which prediction is " not fertile," but actually it is fertile (type II error)
0	False positives (FP): Cases in which the prediction is "1," or fertile but it is not fertile (type I error)	True negatives (TN): Cases in which prediction is "not fertile," and it is not fertile

The confusion matrix evaluates classification performance by categorizing predictions into true positive (TP), true negative (TN), false positive (FP), and false negative (FN), which are then used to compute key performance metrics such as accuracy, precision, recall, F1-score, and area under the precision–recall curve to assess the effectiveness of the trained machine learning model [14, 15].

$$\text{Accuracy} = (TP + T\,N) / (TP + FP + F\,N + T\,N)$$

Precision:-

Precision is a performance metric derived from the confusion matrix that measures the proportion of correctly predicted positive cases out of all cases predicted as positive by the model. It indicates how accurately the model identifies positive samples and reflects its ability to minimize false positive (FP) errors. A low precision value means the model incorrectly classifies many negative (0) samples as positive (1), showing poor reliability in positive predictions.

$$\text{Precision} = (TP) / (TP + FP)$$

In terms of mathematics, the sole difference between precision and recall is that recall considers false negatives rather than false positives.

Recall:-

The recall is the benchmark that can be obtained from the confusion matrix. This statistic represents the percentage of positive samples (1) that are properly identified out of all positive (1) samples fed into the model. To put it another way, this statistic illustrates how well the model recognizes the positive cases in the dataset. Equation 3.11 illustrates this metric's mathematical representation.

$$\text{Recall} = (TP) / (TP + F\,N)$$

Precision, on the other hand, refers to the proportion of relevant results that are correctly classified by the algorithm, whilst recall refers to the total number of relevant results successfully categorized by the algorithm. The F1 Score is the harmonic mean of precision and recall. The precision and recall impacts are combined to produce the value shown by the F1 score.

$$\text{F1 Score} = 2*\text{Precision}*\text{Recall} / (\text{Precision} + \text{Recall})$$

Classification error/misclassification: is defined as the ratio of the incorrectly recognized (either positive or negative) in relation to the whole sample size. In our study this shows if the classifiers classify the soil fertility as 'not fertile (0) and fertile (1)'

$$\text{Error Rate (\%)} = ((FP + FN) / (TP + TN + FP + FN))*$$

3.10. Implementation Tools

The study was implemented using the Python programming language within the Jupyter Notebook environment, supported by the Scikit-learn machine learning library for model development and evaluation, and managed through the Anaconda environment for package management and reproducibility. These tools were used for data preprocessing, visualization, and model development.

4. Key Findings

Important Soil Factors Affecting Fertility

The most important predictors of soil fertility were:

1) Clay content
2) CEC (Cation Exchange Capacity)
3) Calcium Carbonate ($CaCO_3$)
4) Sand content
5) Manganese (Mn)

These variables had the strongest influence on soil fertility prediction

Model Performance

The performance evaluation of the tested machine learning models showed that the Random Forest classifier performed the best among all models used in the study. It achieved an accuracy of approximately 94.37%, demonstrating its strong capability in predicting soil fertility accurately. In addition to Random Forest, the AdaBoost and XGBoost classifiers also showed high performance, with accuracy levels above 94%, indicating that these ensemble learning techniques are highly effective for classification tasks involving soil data. The high accuracy achieved by all three models confirms that machine learning approaches are reliable and efficient tools for predicting soil fertility. These results highlight the potential of machine learning to analyze complex soil characteristics and pro-

vide accurate predictions, which can support better agricultural planning and soil management decisions.

5. Conclusion

The study successfully demonstrated that machine learning techniques, particularly ensemble models such as Random Forest, are highly effective for predicting soil fertility. The results showed that machine learning models can predict soil fertility with high accuracy by analyzing complex relationships among soil properties. Ensemble models, which combine multiple learning algorithms, provided better performance and more reliable predictions compared to single models because they reduce errors and improve generalization. Additionally, the study confirmed that proper data preprocessing, including feature selection and handling class imbalance using sampling techniques such as SMOTE, significantly improves model performance and prediction accuracy. These findings indicate that machine learning can serve as a powerful tool to support farmers and agricultural experts in making informed decisions about fertilizer application, soil management, and crop planning, ultimately improving agricultural productivity and sustainability.

6. Importance of the Study

This research is important because it contributes significantly to improving agricultural productivity by providing accurate and data-driven methods for predicting soil fertility. By using machine learning techniques, the study enables more efficient and precise fertilizer application, ensuring that farmers apply the right type and amount of fertilizer based on the actual condition of the soil. This helps reduce unnecessary fertilizer use, which in turn lowers farming costs and minimizes environmental impact. Additionally, the predictive model developed in this research serves as a valuable decision-support tool for farmers, agricultural experts, and policymakers, helping them make informed decisions regarding soil management, crop planning, and resource allocation. Ultimately, the study supports sustainable agriculture in Ethiopia by promoting efficient soil management practices, improving crop yields, conserving resources, and enhancing long-term agricultural productivity.

7. Recommendation

The study recommends the use of machine learning models as a decision-support tool in agricultural practices to improve soil fertility assessment and fertilizer management. By integrating machine learning into agricultural decision-making, farmers and agricultural experts can make more accurate and data-driven decisions, which can lead to increased productivity and efficient use of resources. The study also suggests expanding the dataset by including additional soil properties, environmental factors, and climatic variables to further improve the accuracy and reliability of the prediction models. Incorporating more diverse and comprehensive data will enhance the model's ability to generalize across different soil conditions. Furthermore, the study recommends applying the developed model to other geographic regions to evaluate its effectiveness in different agricultural contexts and ensure its broader applicability. Finally, it emphasizes the importance of developing real-world applications, such as software systems or mobile-based decision-support tools, that can help farmers easily access soil fertility predictions and receive recommendations for improving soil management and crop production.

Abbreviations

AdaBoost	Adaptive Boosting
$CaCO_3$	Calcium Carbonatefp False Positives
CEC	Cation Exchange Capacity
FN	False Negative
KNN	K-Nearest Neighbors
Mn	Sand Content, and Manganese
TN	True Negative
TP	True Positives
SMOTE	Synthetic Minority Over-sampling Technique
XGBoost	Extreme Gradient Boosting

Author Contributions

Tigist Mintesnot Tewabe: Conceptualization, Data curation, Formal Analysis, Methodology, Writing – original draft, Writing – review & editing

Conflicts of Interest

The author declares that there is no conflict of interest.

References

[1] L. D. Tamene, T. Amede, J. Kihara, D. Tibebe, and S. Schulz, "A review of soil fertility management and crop response to fertilizer application in Ethiopia: towards development of site- and context-specific fertilizer recommendation," CIAT Publ., 2017.

[2] P. A. Sanchez, G. L. Denning, and G. Nziguheba, "The African green revolution moves forward," Food Secur., vol. 1, pp. 37–44, 2009.

[3] C. Yirga, T. Erkossa, and G. Agegnehu, "Soil acidity management." Ethiopian Institute of Agricultural Research, 2019.

[4] A. Deressa, M. Yli-Halla, M. Mohamed, and L. Wogi, "Phosphorus sorption characteristics of lime amended Ultisols and Alfisols in humid tropical Western Ethiopia," Open Access J. Environ. Soil Sci., 2020.

[5] B. Iticha and C. Takele, "Digital soil mapping for site-specific management of soils," Geoderma, vol. 351, pp. 85–91, 2019.

[6] J. Havlin and R. Heiniger, “Soil fertility management for better crop production,” Agronomy, vol. 10, no. 9. MDPI, p. 1349, 2020.

[7] T. Erkossa, A. Wudneh, B. Desalegn, and G. Taye, “Linking soil erosion to on-site financial cost: lessons from<? xmltex\backslashnewline?> watersheds in the Blue Nile basin,” Solid Earth, vol. 6, no. 2, pp. 765–774, 2015.

[8] G. Zelleke, G. Agegnehu, D. Abera, and S. Rashid, “Fertilizer and soil fertility potential in Ethiopia,” Gates Open Res, vol. 3, no. 482, p. 482, 2019.

[9] B. Nigussa, “A Trilingual Android Application with Automatic Malaria Detection from Microscopic Images of Red Blood Cells”, PhD thesis, Addis Ababa University, Ethiopia.

[10] y Y. Ding, H. Zhu, R. Chen, R. Li, “An Efficient AdaBoost Algorithm with the Multiple Thresholds Classification”, Appl. Sci. 2022, 12(12), 5872; https://doi.org/10.3390/app12125872

[11] C. D. Popa, R. Dan, L. Haidar, C. Popescu, R. Dan, T. Popa, and L. Petrescu,”Quantification of Cardiovascular Disease Risk Among Hypertensive Subjects in Active Romanian Population Using New Echocardiographic, Biological and Atherogenic Markers”, 2026, 62(1), 32; https://doi.org/10.3390/medicina62010032

[12] L. Hanif, Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction,

[13] S. Adane, P. Thomas, “Analysis and Prediction of Soil Fertility Using Machine Learning Techniques in North Wollo Zone, Amhara Region, in Ethiopia”, 1st Edition, 2025.

[14] N. Janvier, N. Arcade, N. Eric, and N. Jean, “Machine learning based soil fertility prediction,” Int. J. Innov. Sci. Eng. Technol., vol. 8, no. 7, pp. 141–146, 2021.

[15] A. R. RV, S. R. U, “ Enhancing Soil Fertility Prediction Through Advanced Modelling: Agrimind Intelligent Fertility Predictor”.