International Journal of Applied Mathematics

Volume 38 No. 6s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DATA-DRIVEN SOIL HEALTH ASSESSMENT USING GRADIENT BOOSTING AND DEEP NEURAL ARCHITECTURES

Rajesh Bhanudas Shinde[1],N.T.Markad[2] and Mangal V. Patil[3]

[1]ResearchScholar,Department of Electronics and Communication Engineering
BharatiVidyapeeth(DeemedtobeUniversity) College of Engineering Pune,MH, India

raj.shinde000@gmail.com

[2,3]Department of Electronics and Communication Engineering
BharatiVidyapeeth(DeemedtobeUniversity), College of Engineering Pune,MH,India

*ntmarkad@bvucoep.edu.in , * mvpatil@bvucoep.edu.in

Abstract

Soil health is a fundamental factor for sustainable crop production and affects crop yield and ecosystem equilibrium. Conventional soil inspection techniques are tedious and time-consuming, prompting the development of efficient data-driven solutions. To facilitate soil quality classification in our research, we propose a deep learning-based soil quality classification method using extreme gradient boosting (XGBoost) and deep neural network models, which are multilayer perceptrons (MLP) and artificial neural networks (ANN). A dataset of soil samples characterized for pH, electrical conductivity (EC), phosphorus, potassium, organic carbon (C), and lime content was extended by controlled perturbations to increase the generalization of the model. The performance of the models was assessed based on accuracy, precision, recall, F1-score, and using ROC curves, and calibration plots. XGBoost provided a maximum accuracy of 94.25% and outperformed deep learning models (ANN: 74.75%, deep MLP: 68.25%, simple MLP: 60.25%). This suggests that machine learning can be successfully employed to automate soil health evaluation and aid data-based decision-making in precision agriculture.

Keywords: Soil classification; XGBoost; deep learning; MLP; ANN; precision agriculture; machine learning; soil quality; ROC; calibration curves

1. Introduction

The agriculture and allied sectors are still the main occupation and means of living for a large number of people in India. During the past few decades, there has been an explosive increase in the need for food production. However, this growing demand has also sharply reduced available farmland, as well as a consequence of rapid industrialization. To make informed decisions on crop type selection and increase productivity, farmers need access to timely and precise information, such as the types of pesticides and fertilizers to be used, weather information, and soil information [1].

Received: August 01, 2025

Analysis of these environmental and agronomic influences could lead to better crop productivity; however, damage and loss due to adverse conditions would decrease. Although new types of hybridized plants are becoming more available, many of these new hybrids are missing valuable nutrients present in naturally grown foods. Moreover, soil fertility and environmental degradation may result from excessive use of synthetic approaches [2].

Although several artificial methods have been attempted to reduce crop loss, better results can be obtained if farmers are provided with exact and essential information on the influencing factors. It is estimated that 16–20% of India's agricultural produce is spoiled every year owing to different inefficiencies [3]. million excess of 215.6 million-acre (approximately 82.6 million-ha) of farmland annually. Maintaining crop productivity at a high level through years of farming activity is possible if effective soil fertility management is implemented. This indicates that the soil can provide plants with the necessary nutrients in an adequate form, in the right amount, and at the appropriate time. Soil fertility is determined by the presence or absence of macronutrients and micronutrients. Rajkumar et al. [4] details 16 essential nutrients necessary in moderate to large amounts for optimal healthy plant growth, and trace elements (reported in mg/kg) necessary for the plants' enzymatic activity.

Soil quality is central to agricultural productivity, environmental sustainability and food security. It defines the capacity of the soil to support plant growth, store water, recycle nutrients, and maintain biological activities. The scientific evaluation of soil quality helps farmers and agronomists identify appropriate crop choices, fertilizer applications, and land management, which are indispensable for sustainable agriculture.

Conventional soil health assessments are carried out through labor-intensive laboratory testing and expert interpretation of soil physicochemical properties (pH, EC, phosphorus, potassium, organic carbon, and lime). Although such approaches yield accurate outputs, they are labor-intensive, time-consuming, and not scalable over vast agricultural areas. Given the increasing demand for precision farming and data-driven agriculture, there is great potential for the use of AI-based classification for efficient and cost-effective soil monitoring and classification at scale.

In this study, we used machine and deep learning techniques to categorize soil health using important physicochemical properties. We demonstrate this with gradient boosting (XGBoost) and deep neural architectures such as multi-layer perceptron (MLP) and ANN. We employed a soil dataset obtained from field samples and artificially augmented it for diversity and robustness. Our models were tested with commonly used performance values and visualization techniques in terms of ROC curves, calibration plots, and confusion matrices.

The main contributions of this study are as follows.

XGBoost and Convolutional Neural Networks were applied to soil quality classification with physicochemical information.

In addition to an in-depth discussion of model performance across accuracy, precision, recall, F1-score, and calibration.

Providing technically useful knowledge on the use of AI Models for automatic soil health assessment to facilitate precision farming.

2. Related Work

Many previous studies have proposed methods for the prediction of crop yield and soil analysis based on machine learning (ML) and deep learning. It was possible to solve data-driven agricultural complex problems through the analysis of numerical parameters and to make informed decisions using the proposed mathematical models.

Zaminur et al. [13] used a dataset collected by the SRDI with Support Vector Machine (SVM), Bagged Trees, and KNN models to classify soil series and recommended crop yield. The dataset contains 495 samples from 11 soil classes. Several classification and regression models were compared, and SVM booking was found to be the best model for soil classification.

Pramudyana et al. [14] used, among other ML techniques, SVM, Naïve Bayes, decision trees and ANN for automatic classification of soil types. The top-performing algorithm achieved more than 70% accuracy, demonstrating the value of ML for agricultural classification.

Patil et al. [15] worked on deep learning based organic agricultural crop protection. They studied Convolutional Neural Network (CNN)-based models to create a model for detecting different types of crop diseases. This facilitated farmers' real-time support by identifying the symptoms of disease in plants.

Ashwini et al. [16] developed a system for grading and categorising soil samples with the help of digital image processing and pattern recognition. The soil texture and color features were calculated using computer-vision algorithms. The real-time solution was realized using a DSP board with commercial imaging libraries.

Alex and Kanavalli [17] developed a CNN model for precision agriculture to predict crop yields. The model included some important factors such as fertilizer application, temperature, and rainfall. It was also used for decision-making, and it produced p-values to test crops to estimate yield.

Prakash et al. [18] proposed a high-performance soil moisture prediction scheme, including many machine learning techniques, such as Recurrent Neural Networks (RNN), Support Vector Regression (SVR), and Multiple Linear Regression (MLR). The performance comparison was performed on public datasets, and the models were evaluated in terms of R^2 and MSE. The R^2 and MSE of the MLR model (0.975 and 0.14, respectively) were higher than those of the other models.

Gholap et al. [19] used the soil datasets of Pune, India, from three regions, including Velhe, Bhor, and Khed, with 1988 instances and nine attributes. The JRip, Naïve Bayes and J48 algorithms were used for the classification. Engine C4.5, is embodied by the J48 model. 5 decision tree algorithm resulted in a highly satisfactory performance in terms of soil-type classifications.

Gudavalli et al. [20] experimented with several clustering algorithms with a seed dataset. Important features, including both-size kernel groove length, kernel width, complex ratio, asymmetric coefficient of kernel valley, compactness, border length, and area, were selected to improve the clustering performance and segmentation precision.

3. Materials and Methods

The proposed system is an intelligent data-driven model for assessing soil health based on the robustness of machine learning and deep learning approaches. (iv) Instead of limiting the learning based only on quality labels, the system learns directly from measured soil parameters to recognize patterns and assist in the assessment of soil health without any predetermined categorical outputs. The ultimate objective was to support the informed prediction of the soil properties through numerical simulations.

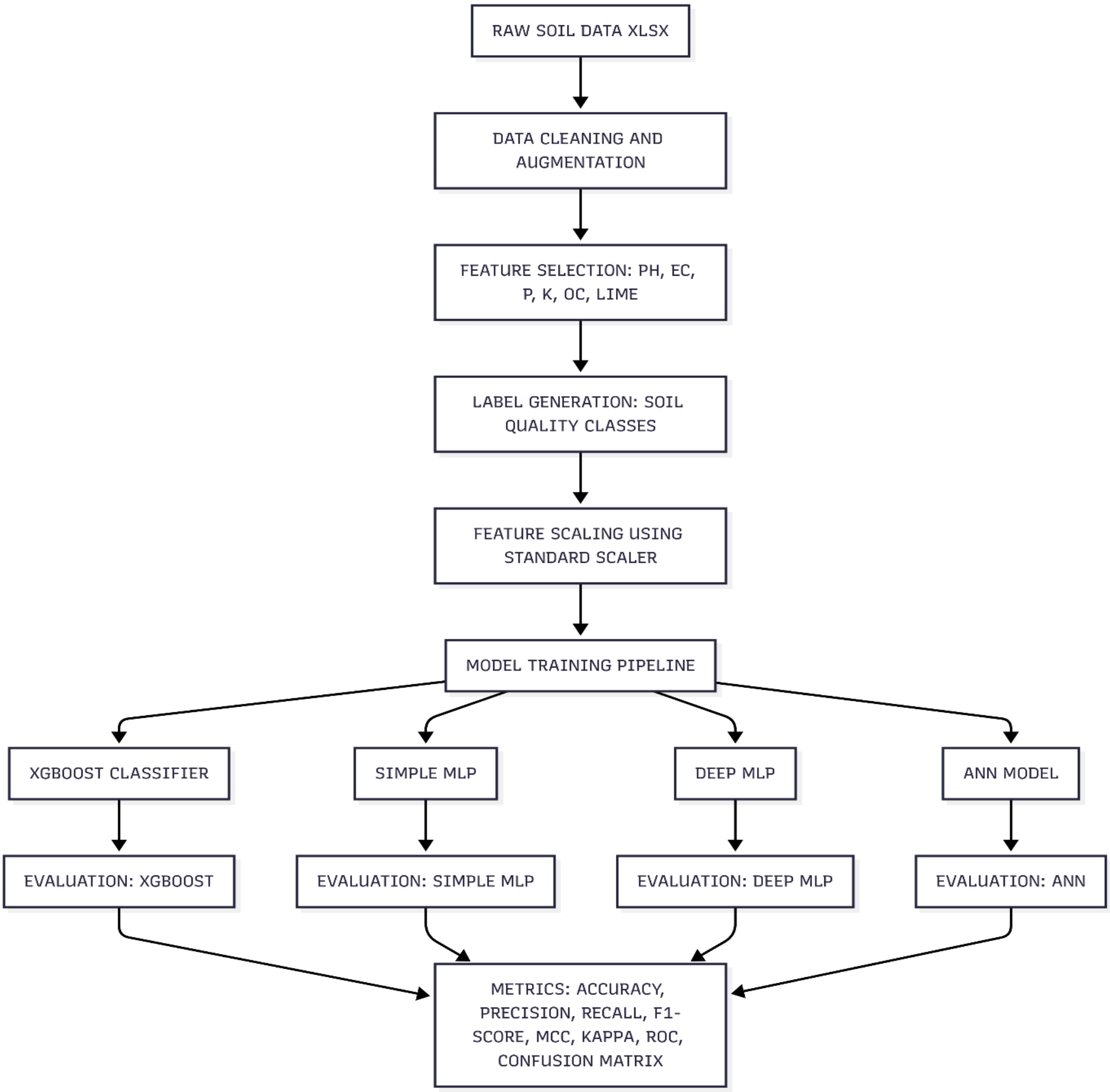


Fig. 1: System Architecture

Received: August 01, 2025

3.1 Dataset

Material and methods Data The experiment was conducted at the Vasant Prakash Vikas Pratishthan Agriculture Science Center, Kanchanpur, the institution awarded by the Indian Council of Agricultural Research (ICAR), which has its own Soil and Water Testing Laboratory. For the present investigation, real-time data obtained from soil offices located in the Sangli and Miraj districts of Maharashtra, India were used to form the dataset. Soil samples were collected from various villages in these districts to represent a wider and more diverse dataset. It contains critical soil factors that influence soil fertility and crop growth. The dataset includes 2000 soil samples, with information on each ranging from soil and water quality parameters to what is necessary for agricultural analysis and crop recommendation.

Data Preprocessing

preprocessing phase is crucial for the quality of any classification task. Soil sample data collected from different regions in Maharashtra were analyzed using a classification system based on the physicochemical characteristics of the soil.

3.1 Data Acquisition and Cleaning

The dataset was obtained from. xlsx, and included approximately 2000 soil profiles, with details such as soil pH, Electrical Conductivity (EC), Phosphorus, Potassium, Organic Carbon, Lime Content. Data cleaning involved:

Removing missing or inconsistent values

Standardizing measurement units

vv) Making sure all numerical ranges are both reasonable and articulated in an ag-friendly medium

3.2 Feature Selection

The following descriptors were used for classification:

Soil pH

Electrical Conductivity (dS/m)

Phosphorus (kg/ha)

Potassium (kg/ha)

Organic Carbon (%)

Lime Content (%)

These were selected because of their relationships with soil fertility classes and agricultural importance.

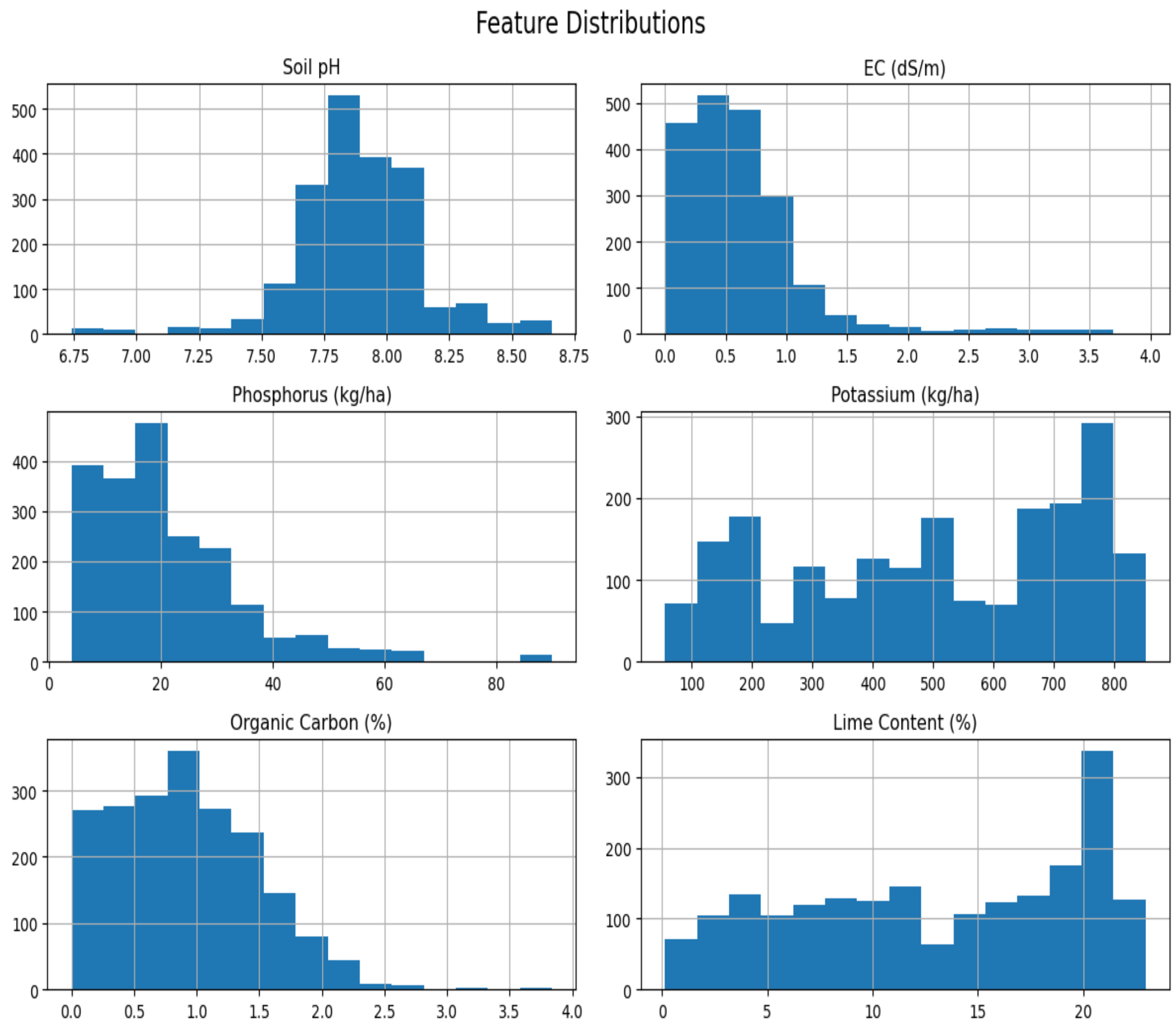


Fig.2 : Feature Distribution

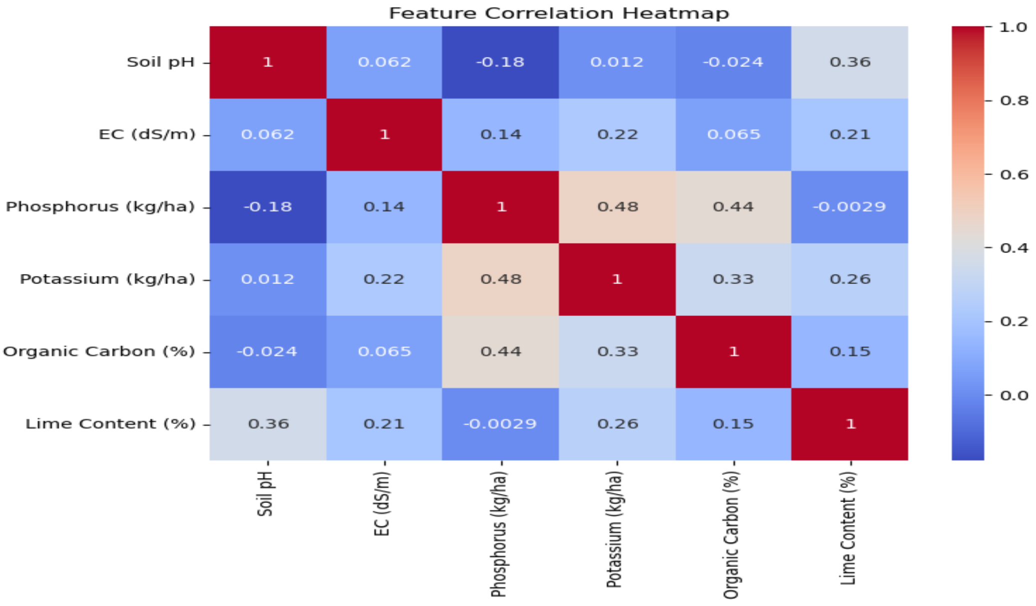


Fig. 3: Heatmap

3.5 Dataset Splitting

We split the preprocessed dataset into training (80%) and testing (20%) datasets using stratified sampling to ensure that the proportion of data in the sample with respect to the classes was consistent in the training and testing datasets.

Taluka	District	Soil pH	EC (	Phosphorus (kg/ha)	Potassium (kg/ha)	Organic Carbon (%)	Lime Content (%)	Conclusion
Miraj	Sangli	7.66	1.132	20.16	425.6	0.27	11.88	Organic Carbon – Low, EC & Lime – High
Miraj	Sangli	7.73	0.494	22.4	683.2	1.23	11.25	Lime – High
Miraj	Sangli	7.21	0.536	60.48	638.4	1.71	1.25	All parameters are within the range
Miraj	Sangli	7.52	0.316	29.12	268.8	0.81	2.5	All parameters are within the range
Miraj	Sangli	7.67	0.427	17.92	235.2	0.63	9.38	Lime Content – Medium
Miraj	Sangli	7.63	0.414	24.64	526.4	0.84	14.38	Lime Content – Very High
Miraj	Sangli	7.8	0.492	6.72	179.2	0.99	6.88	Phosphorus – Low
Miraj	Sangli	7.78	0.556	31.36	716.8	1.77	19.38	Lime Content –

								Very High
Miraj	Sangli	8.04	0.942	8.96	324.8	1.35	17.5	Phosphorus – Low, EC – High, Lime Content – Very High

3.3 Models

The system includes a variety of machine learning and deep learning algorithms to study and estimate soil health using numerous physicochemical properties. From classical machine learning, Extreme Gradient Boosting (XGBoost) was selected for its effectiveness and performance on tabular datasets with nonlinear patterns. XGBoost grows a model by adding decision trees through a boosting process, where, at each iteration, it reduces the error using gradient descent and applies regularization to avoid overfitting.

Moreover, several deep learning models have been investigated. The simple multilayer perceptron (MLP) is a baseline neural model with one or two dense layers. It uses backpropagation to learn the patterns and to spread the error and nonlinear activation functions, such as the rectifier function for the hidden layers and the softmax function for the output layer. To address this, a deep multilayer perceptron (Deep MLP) was developed with more hidden layers and neurons, allowing for better representation learning to account for complex non-linear relationships between soil properties.

In addition, a fine-tuned deep learning model with a kernel-based ANN was constructed. This model was built on top of three dense layers, compiled using the Adam optimizer, and trained with a categorical cross-entropy loss function. ANN generalization was efficient: the choice of a sufficiently deep network layer, correct application of regularization, and learning rate adaptation led to ANN effectiveness.

All models were developed using scaled input variables from the raw soil test values. Upon training, they were evaluated in terms of accuracy, precision, recall, F1-score, ROC-AUC, Matthews Correlation Coefficient (MCC) and Cohen's kappa. These metrics provide a full picture of how reliable models detect informative patterns in soil data, leading to the incorporation of robust data-driven soil health evaluation.

Simple MLP

Simple MLP: This is a standard feedforward neural network used for classification purposes. The MLP architecture used in this study consists of an input layer, a hidden layer, and an output layer. It uses a set of numeric soil attributes (e.g., pH, EC, phosphorus, potassium, organic carbon, and lime content) as input data.

The ReLU is an activation function in the hidden layer, and its objective is to add nonlinearity to the model and help the model learn EAU, which can be seen as a generalization of Logistics Regression for multi-class learning. The output layer utilizes the softmax activation function, which converts the network output into a probability distribution over soil health classes.

The network is optimized using the Adam optimizer, and the loss function is the categorical cross-entropy, which is applicable for multiclass classification. The model learns and updates the weights via backpropagation to minimize losses from the predicted and actual labels.

Despite being a simple model, this MLP is considered a classical and basic deep learning model, and it is used as a basis for comparing the effectiveness of other complex models.

Input Layer:

- The model starts with the input of numerical soil properties.
- = These features were scaled using StandardScaler such that all shared a common range of values.

Forward Propagation:

- The inputs are multiplied by the weights and added to the bias of the first (and only) hidden layers. This layer consists of a set of neurons, each performing a linear transformation of the inputs \(X\) weighted by \(W\) plus a bias, followed by ReLU activation:

. This transforms the input into a higher-dimensional space, helping the model to learn nonlinear relationships between soil attributes.

. Output Layer:

. The activations from the hidden layer were passed to the output layer, which had one neuron per class.

. Softmax activation was applied as follows:

. converting the raw scores into probabilities for each class.

. Prediction:

. The model outputs the probability distribution across soil classes, and the class with the highest probability is selected for the prediction.

Training:

. During training, the model compares the predicted probabilities with the actual labels using a categorical cross-entropy loss function.

. Using the Adam optimizer, the model updates the weights through backpropagation to minimize loss.

DEEP MLP

Received: August 01, 2025

A Deep MLP is a type of deep feedforward neural network that includes many hidden layers between the input and output layers. These layers can enable it to learn more complicated and hierarchical relationships between data than a simple MLP with only one hidden layer.

In this work, we designed a Deep MLP architecture with three or more hidden layers with ReLU activation nonlinearity and softmax activation at the final output layer for multiclass classification.

Based on the above, the Deep MLP model is effective for learning the nonlinear dependencies and interactions of soil parameters. Its depth can reveal latent representations that are not EIICisible using other methods or feeble networks. It thus provides substantial improvement for prediction in tasks related to soil health modeling

#DEEP MLP WORKING

Input Representation Let the input feature vector for a soil sample be

Where:

: Soil pH

: EC (dS/m)

: Phosphorus (kg/ha)

: Potassium (kg/ha)

: Organic Carbon (%)

: Lime Content (%)

Layer-wise Transformation Let the deep MLP model have layers with weights, biases, and activations. The input to the first layer was. For each hidden layer, the transformation is

Output Layer (Softmax): Let the output layer have units representing the classes. The output before activation is

Apply

where is the predicted probability for class. 4. Loss Function: Categorical Cross-Entropy Let the true label be one-hot encoded as; then, the loss is

This loss function was minimized during training using gradient descent. 5. The optimization Parameters were updated using the Adam Optimizer:

where denotes the learning rate. 6. Prediction Rule After training, for a new input, the predicted class is

Algorithm

Received: August 01, 2025

For each layer l:

Where: -

= output of layer l

= weight matrix for layer l

= bias vector for layer l

- = activation function (for example, ReLU) final output layer

Where

where is the logit for class j and the loss function is

Where: -

XGBOOST

XGBoost: A Mathematical Explanation in Paragraph Form Extreme Gradient Boosting (XGBoost) is an efficient and scalable implementation of gradient boosting for supervised learning problems. It sequentially constructs an ensemble of decision trees, where each new tree attempts to minimize the errors made by the previous ensemble. Let the dataset be represented as, where denotes a feature vector of soil parameters (e.g., pH, EC, P, K, OC, Lime Content), and is the corresponding class label. The prediction for each instance is given by an additive tree model:

where each is a decision tree, and is the space of all regression trees. The training objective is defined as a regularized loss function

Here, is a differentiable convex loss function (such as softmax cross-entropy for multiclass classification), and is the regularization term, where is the number of leaves in the tree and represents the leaf weights. Regularization parameters and help control the model complexity and reduce overfitting. At each boosting step, the model is updated as follows.

To optimize the new tree efficiently, the loss is approximated using a second-order Taylor expansion as follows:

where and are the gradient and Hessian sums over the instances assigned to leaf, respectively, and is the set of instance indices in the leaf. For classification, the final output is converted into class probabilities using the softmax function:

and the predicted class is

This approach is highly effective for handling structured tabular data with noisy and imbalanced distributions. In our case, XGBoost achieved strong performance for soil class prediction based on physicochemical soil parameters.

Received: August 01, 2025

3.4 Evaluation Metrics

T the accuracy scores of the classification models upon which we built our experiments in this work — XGBoost, Simple MLP, Deep MLP, and ANN — and beyond, we also considered a large number of performance measures. Overall, accuracy is the ratio of correctly predicted observations to the total observations and provides a general overview of the model performance. However, Precision and Recall (or sensitivity), measuring how many of the actual positive cases were correctly detected by the classifier, were also used, as a relation was not necessarily balanced within the same type class. In the context of unbalanced class distributions the F1-Score (the Harmonic mean of Precision and Recall) provides a balanced perspective. To account for this, we included Balanced Accuracy (which averages recall over all classes), as well as specific False Positive and False Negative Rates, to describe the model performance on misclassifications. Furthermore, the Matthews Correlation Coefficient (MCC) and Cohen's kappa were computed to assess the performance of multiclass predictions considering random chance and the correlation between observed and predicted labels. To compare decision boundaries, class separability, and model calibration, visualization tools such as ROC Curve, Precision-Recall Curve, Calibration Curve, and Normalized Confusion Matrix were applied. The combination of these metrics also offers a complete and conservative framework for the reliability and interpretability of classification results in soil health assessment.

- Accuracy:

- Precision:

- Recall:

- F1-Score:

- Matthews Correlation Coefficient (MCC)

- Cohen's Kappa:

4. Results

4.1 Performance metrics

This section presents a comparative analysis of the classification models developed for soil health assessment based on key performance metrics including Accuracy, Precision, Recall, and F1-Score. The performance results are summarized in table below:

Model	Accuracy	Precision	Recall	F1-Score
XGBoost	94.25%	94.46%	94.25%	93.95%
Simple MLP	60.25%	59.91%	60.25%	55.96%
Deep MLP	68.25%	73.46%	68.25%	66.73%
ANN	74.75%	76.12%	74.75%	72.77%

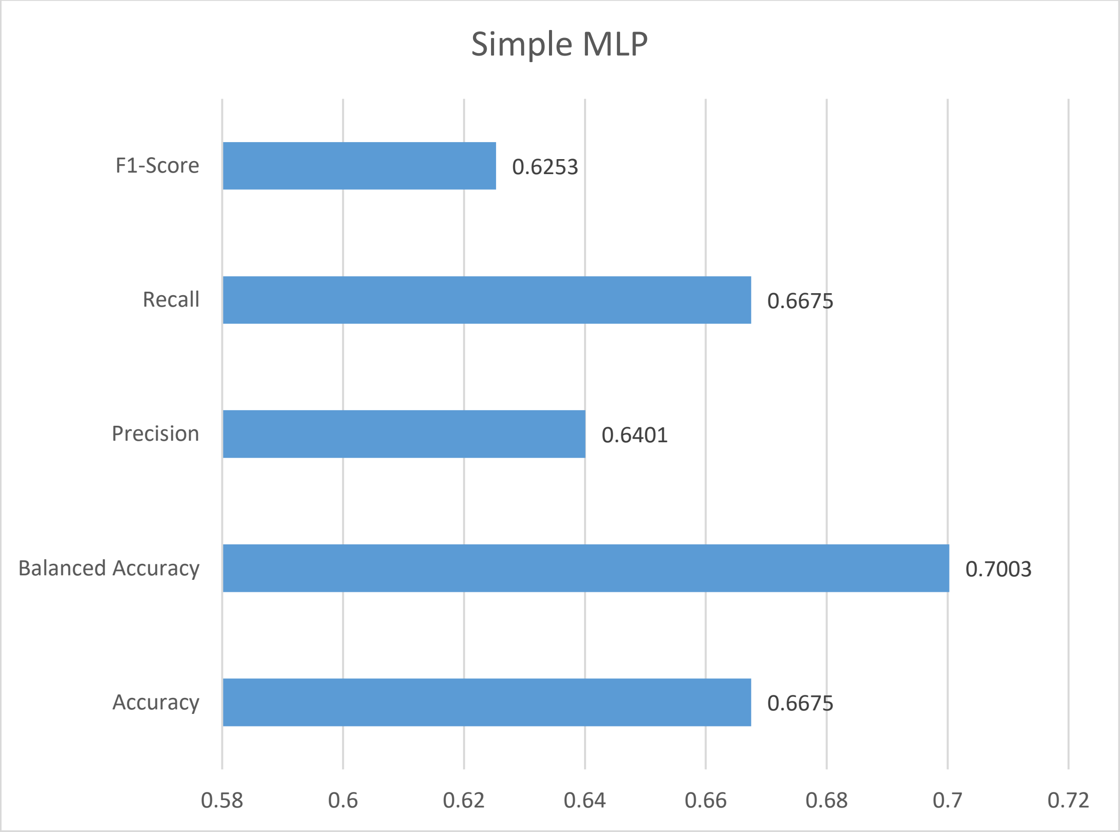
The XGBoost model is a clear winner among all the others, giving the best accuracy of 94.25% as well as consistently strong precision, recall, and F1-score values. This is because XGBoost's optimized gradient boosting framework scales more naturally to the size of the data and to the inclusion of features that interact nonlinearly, as well as to features that have missing data. It is particularly well-suited for this task because of its effectiveness on structured tabular datasets as a soil assay.

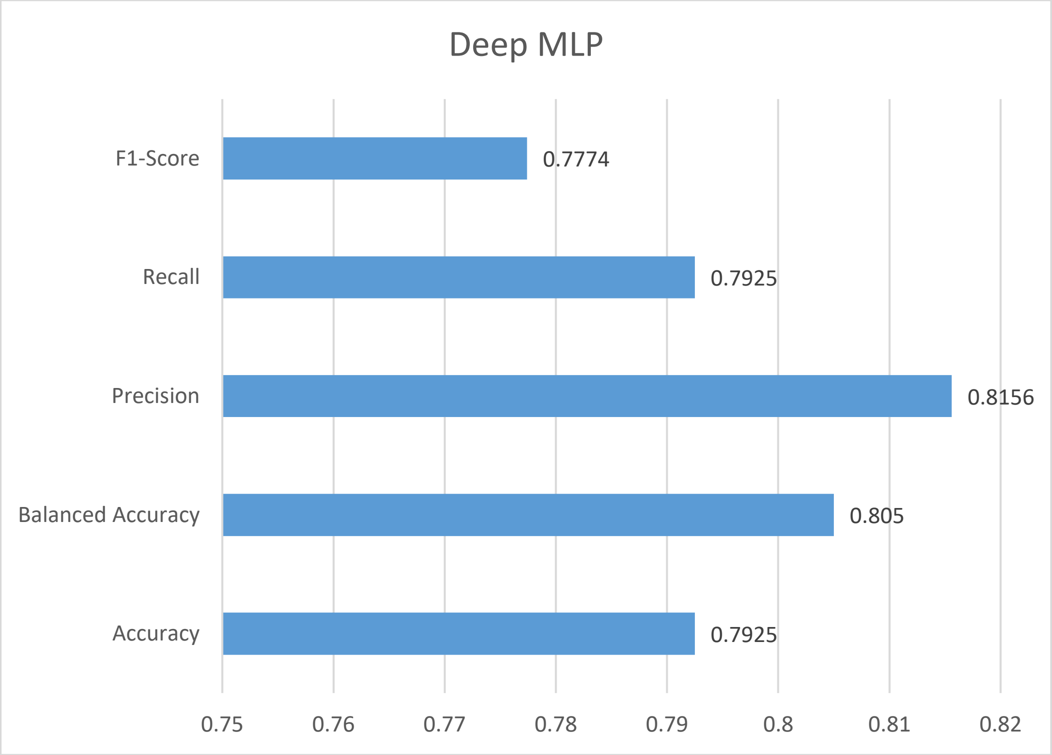
From the deep learning models, (Multilayer Perceptron) provides better results with an accuracy of 74.75% and an F1-score of 72.77%, which clearly shows a good balance between precision and recall. This implies that multilayered deep architectures can capture more intricate patterns in data, especially with normalization and good activation functions.

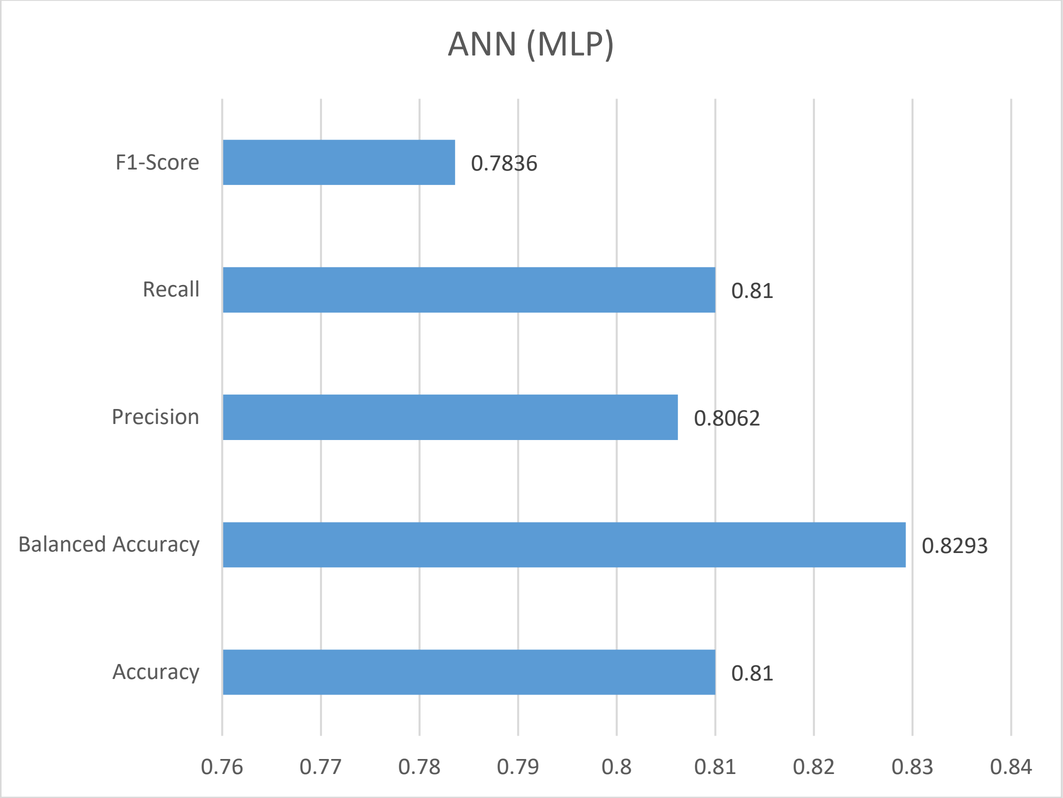
The Deep MLP model, which is more complicated than the simple MLP, shows a moderate gain in precision (73.46%), but trails XGBoost and ANN in overall accuracy. This suggests that we increase the power of representation by increasing the depth; however, this can also cause overfitting or convergence problems if not carefully tuned.

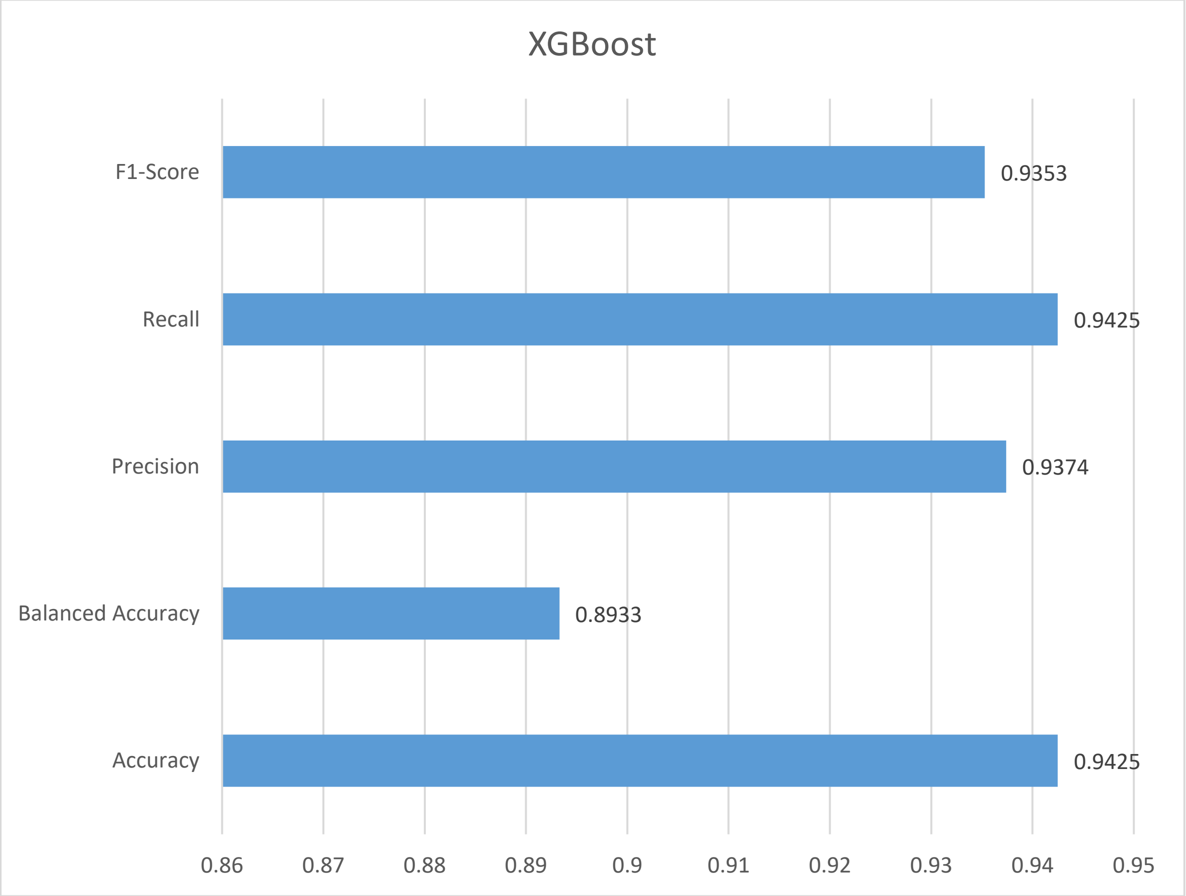
A simple MLP, which has only a few layers, exhibits the steepest dip and provides an accuracy of 60.25%. This finding indicates that the shallow network might not have the capability to comprehend nonlinear connections between soil characteristics.

In general, the results confirm the benefit of using tree-based ensemble methods for soil health classification but indicate the potential of deep learning methods when deeper architectures and appropriate data amounts are used. These results may inform the choice of a model based on data availability, computing resources, and the need for model interpretability in agricultural studies.


Simple MLP
F1-Score
0.6253
Recall
0.6675
Precision
0.6401
Balanced Accuracy
0.7003
Accuracy
0.6675
0.58
0.6
0.62
0.64
0.66
0.68
0.7
0.72


Deep MLP
F1-Score
0.7774
Recall
0.7925
Precision
0.8156
Balanced Accuracy
0.805
Accuracy
0.7925
0.75
0.76
0.77
0.78
0.79
0.8
0.81
0.82


ANN (MLP)
F1-Score
Recall
Precision
Balanced Accuracy
Accuracy
0.7836
0.81
0.8062
0.8293
0.81
0.76
0.77
0.78
0.79
0.8
0.81
0.82
0.83
0.84


XGBoost
F1-Score
Recall
Precision
Balanced Accuracy
Accuracy
0.9353
0.9425
0.9374
0.8933
0.9425
0.86
0.87
0.88
0.89
0.9
0.91
0.92
0.93
0.94
0.95

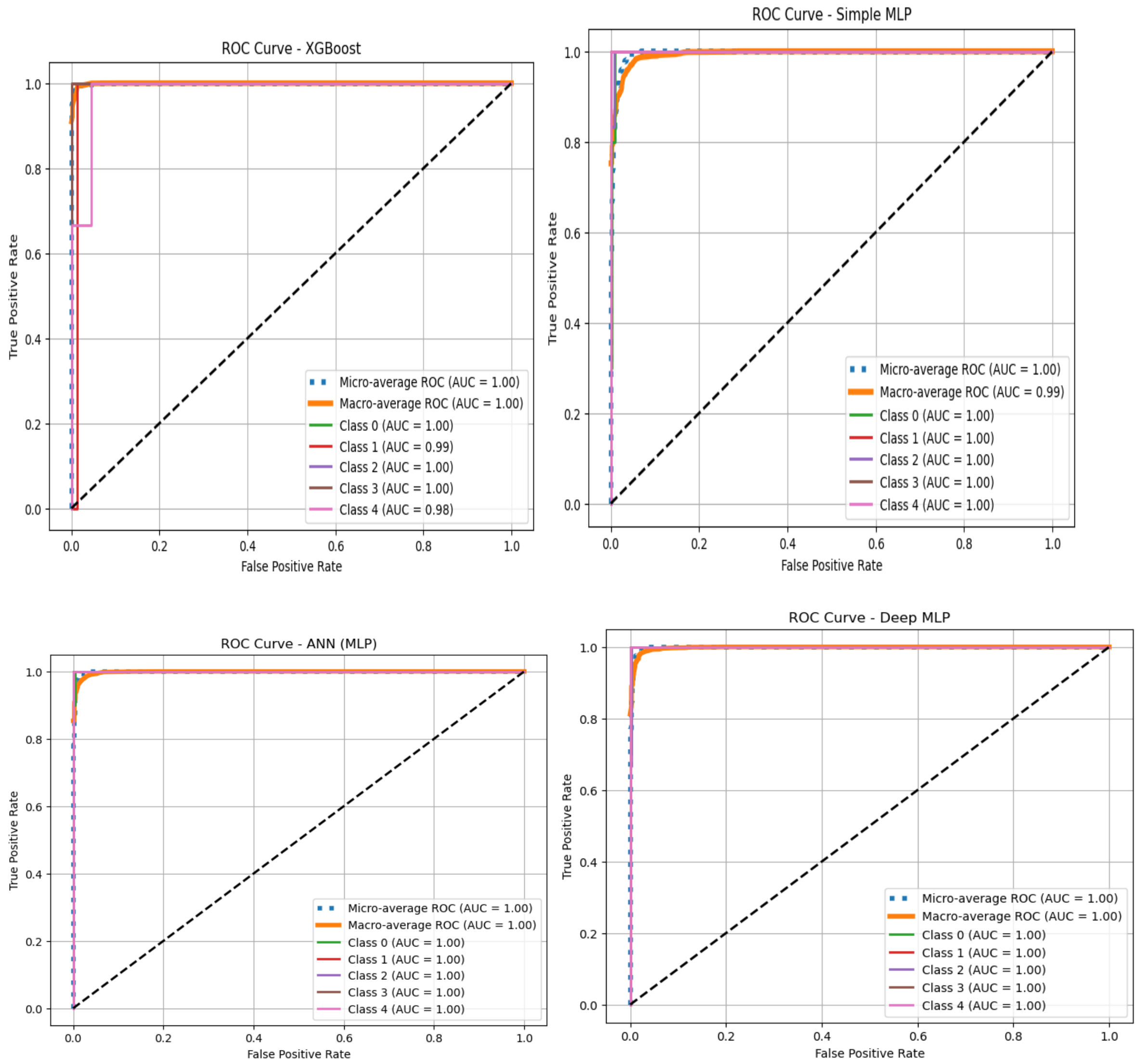
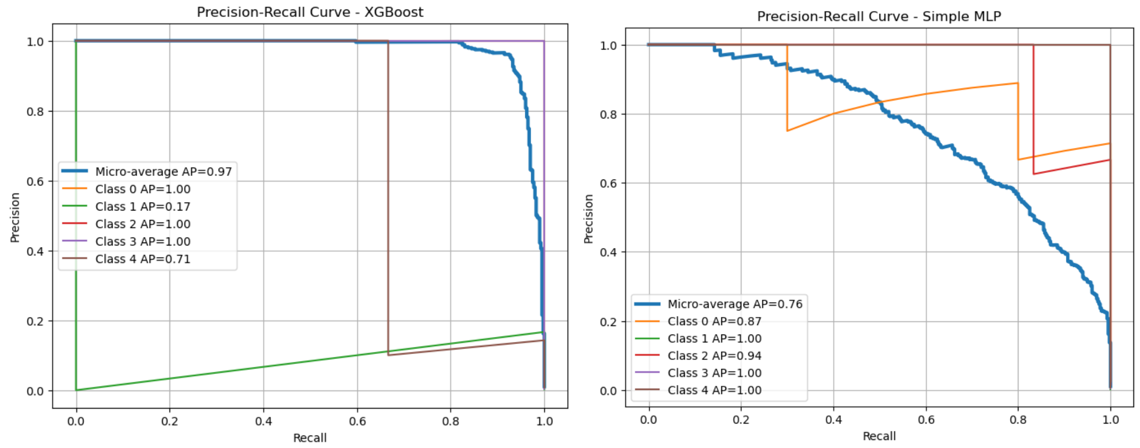


Fig.4.ROC CURVE



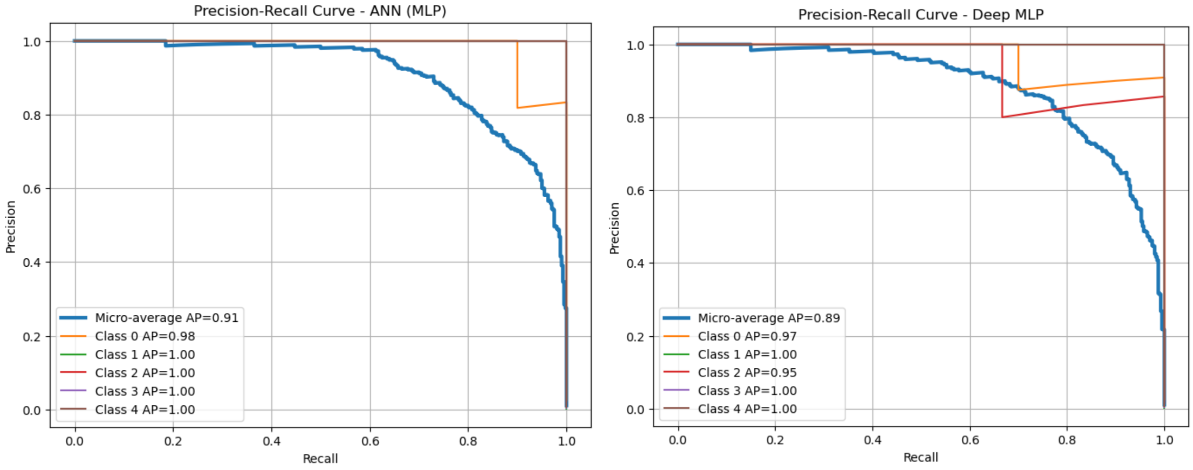


Fig.5 : Precision Recall Curve

5. Discussion

The findings of this study unequivocally underscore the robustness of XGBoost in the processing of structured tabular mindset-like soil physicochemical face data. Its capability to represent non-linear relationships, add regularization, and harness first- and second-order gradient information makes it superior to deep learning models. XGBoost yielded the best accuracy and evaluation metrics and should be considered for the automation of soil health classification.

In contrast, ANN and MLP architectures, which are also capable of learning complex patterns, often require much larger dataset sizes, hyperparameter tuning, and regularization to achieve the predictive performance of gradient-boosted tree ensembles. Their efficiency increases with the amount and diversity of the data; however, in very small or well-structured domains, tree-based methods are sometimes easier and more efficient.

These implications are significant for precision agriculture, including large-scale automatic soil classification systems. A good classification model can help farmers and agronomists make judicious decisions regarding soil treatment, nutrient management, and crop selection, thus leading to economic and environmentally sustainable agriculture.

Finally, the Explainability of ML models becomes very important as AI adoption in farming increases. Although tree-based models, such as XGBoost, can be more interpretable using methods such as feature importance and SHAP values, deep learning models are widely assumed to be black boxes. Building trust among those involved is a prerequisite for fair decision making, especially in high-impact sectors such as agriculture.

Conclusion

This study proposed a data-driven solution for the assessment of soil health based on traditional gradient boosting and deep learning architectures. The experimental findings indicate that XGBoost performed much better than all other models, providing a classification accuracy of 94.25% in our classification task, highlighting its suitability for learning from structured table data. Of the DL models, ANN showed the best results with \(\sim \) 74.75% accuracy, revealing that it has the potential to be better trained and find optimal tuning. These results demonstrate that ML (especially tree-based ensembles such as XGBoost) is a powerful and scalable tool for soil health type classification. These approaches allow for faster and more data-informed decisions in precision, which might increase crop production and resource use.

References

[1] S. K. A. L. Z. Rahman, K. C. Mitra, and S. M. M. Islam, "Soil classification using machine learning methods and crop suggestion based on soil series," *Proceedings of the 21st International Conference of Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, pp. 1–4, 2018.

[2] P. A. Harlianto, T. B. Adji, and N. A. Setiawan, "Comparison of machine learning algorithm for soil type classification," *Proceedings of the 3rd International Conference on Science and Technology-Computer (ICST)*, Indonesia, pp. 7–10, 2017.

[3] M. Patil and I. R. Umarji, "Identification of crop diseases using deep learning," *International Journal of Research in Engineering, Science and Management*, vol. 2, no. 6, 2019.

[4] A. Rao, A. Gowda, and R. Beham, "Machine learning in soil classification and crop detection," *IJSRD - International Journal for Scientific Research and Development*, vol. 4, no. 1, pp. 792–794, 2016.

[5] S. A. Alex and A. Kanavalli, "Intelligent computational techniques for crops yield prediction and fertilizer management over big data environment," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 12, 2019.

[6] S. Prakash, A. Sharma, and S. S. Sahu, "Soil moisture prediction using machine learning," *Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT)*, Namakkal, India, pp. 1–6, 2018.

[7] J. Gholap, A. Ingole, J. Gohil, S. Gargade, and V. Attar, "Soil data analysis using classification techniques and soil attribute prediction," *arXiv preprint*, 2012. [Online]. Available: https://arxiv.org/abs/1206.1557

[8] M. Gudavalli, P. Vidyasree, and S. V. Raju, "Clustering analysis for appropriate crop prediction using hierarchical, fuzzy C-means, K-means and model-based techniques," *Scientific Journal of Impact Factor (SJIF)*, vol. 4, no. 11, pp. 2348–6406, 2017.

[9] Zhao, Z.-D., Zhao, M.-S., Lu, H.-L., Wang, S.-H., & Lu, Y.-Y. (2023). Digital Mapping of Soil pH Based on ML Combined with Feature Selection Methods in East China. Sustainability, 15(17), 12874. https://doi.org/10.3390/su151712874

[10] S. Jahan, I. Manzur, S. R. Jafa, S. Al Zabid, R. Gani and M. R. Ahmmad Rashid, "Exploring Soil pH Variability: Statistical Analysis and ML Models for Enhanced Soil

Health Management," 2024 IEEE International Conference on Computing, Applications and Systems (COMPAS), Cox's Bazar, Bangladesh, 2024, pp. 1-6, doi: 10.1109/COMPAS60761.2024.10796498.

[11] Li, P. 2024. Research article: Spatial Prediction of Soil pH Values in Hunan Province Based on Multi-source Environmental Variables and ML Models. Highlights in Science, Engineering and Technology. 122, (Dec. 2024), 104–109. DOI:https://doi.org/10.54097/2yn2eb70.

[12] Rani, Savita, Satpal Baloda, Dinesh, and Akshay Mehta. 2024. “AI-Driven Insights into Soil Physio-Chemical Properties”. International Journal of Environment and Climate Change 14 (12):834-45. https://doi.org/10.9734/ijecc/2024/v14i124666.

[13] Zhang, L., Wu, Z., Sun, X., Yan, J., Sun, Y., & Chen, J. (2023). Mapping topsoil pH using different predictive models and covariate sets in Henan Province, Central China. Ecological Informatics. https://doi.org/10.1016/j.ecoinf.2023.102290

[14] Zhao, Z., Zhao, M., Lu, H.-L., Wang, S., & Lu, Y.-Y. (2023). Digital Mapping of Soil pH Based on ML Combined with Feature Selection Methods in East China. Sustainability. https://doi.org/10.3390/su151712874

[15] Sun, Y., Sun, X. N., Wu, Z., Yan, J., Ma, C., Zhang, J., Zhao, Y.-F., & Chen, J. (2023). Using a variety of ML approaches to predict and map topsoil pH of arable land on a regional scale. Soil Science Society of America Journal, 87(3), 613–630. https://doi.org/10.1002/saj2.20525

[16] P. R. Kumar, S. Meenakshi, S. R. Devi, S. Shalini, and S. Boopathi, “Soil Quality Prediction in Context Learning Approaches Using Deep Learning and Blockchain for Smart Agriculture,” Igi Global, 2023, pp. 1–26. doi: 10.4018/978-1-6684-9151-5.ch001.