International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 2173-2181

IJIRSS International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Development of a hybrid machine learning model for classification of soil types based on geophysical parameters

Ainagul Abzhanova[1], Zhazira Taszhurekova[2], Bauyrzhan Berlikozha[3], Mira Kaldarova[4], Ardak Batyrkhanov[5*]

[1]*Department of Information Systems. L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.*
[2]*Department of Applied Informatics and ProgrammingM.H.Dulati Taraz University, Taraz Kazakhstan.*
[3]*Narxoz University, Almaty, Kazakhstan.*
[4]*School of Information Technology and Engineering, Astana international university, Astana, Kazakhstan.*
[5]*Department of Software Engineering, NLC «Atyrau University named after Kh. Dosmukhamedov», Atyrau, Kazakhstan.*

Corresponding author: Ardak Batyrkhanov (*Email: Batyr.khan78@mail.ru*)

Abstract

In this paper, a hybrid model based on RandomForestClassifier and MLPClassifier is presented, achieving an accuracy of 96.07% in the task of soil classification based on geophysical parameters. The results demonstrate the advantages of the proposed approach over selected classical algorithms, indicating a high practical value for precision agriculture and environmental monitoring. A dataset containing key soil parameters such as electrical conductivity, density, P-wave velocity, and depth was utilized. Prior to training, the data were preprocessed: the target variable was converted to numeric format using LabelEncoder, and the features were standardized using StandardScaler to bring them to a common scale. Data were divided into training and test samples using the train_test_split method (80% training, 20% test).

Keywords: Data preprocessing, Electrical conductivity, Geophysical data, Hybrid model, Information systems, Land classification, Machine learning, Multilayer perceptron, Neural networks, Random forest.

DOI: 10.53894/ijirss.v8i3.6966
Funding: This study received no specific financial support.
History: Received: 20 March 2025 / **Revised:** 22 April 2025 / **Accepted:** 24 April 2025 / **Published:** 13 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Publisher: Innovative Research Publishing

1. Introduction

Soil is the most important natural resource for food security, environmental sustainability, and economic development. Its classification plays a key role in agriculture, construction, ecology, and other spheres of human activity. The classification of soils makes it possible to assess their properties, productivity, and resistance to external influences. This is especially important under conditions of changing climate and population growth, when the rational use of land resources becomes a priority. Traditional methods of soil classification are based on laboratory analyses and visual observations, which require considerable time and financial expenditures. In this regard, automated methods based on machine learning are becoming increasingly popular. Modern technologies allow for analyzing soil data using artificial intelligence and deep learning, which increases the accuracy and speed of classification.

Machine learning allows us to identify complex dependencies between different soil parameters such as granulometric composition, chemical composition, moisture level, and other indicators. In recent years, hybrid models that combine traditional machine learning algorithms with deep learning methods have been actively developed. They make it possible to increase the accuracy of predictions and reduce the probability of errors. Hybrid soil classification models have been applied in various fields, including agriculture, environmental monitoring, land use, and urban planning. One important aspect of soil classification is data collection. Modern technologies such as remote sensing, spectroscopy, and image analysis provide a wealth of information about soils.

Processing and interpretation of soil data require the application of sophisticated algorithms such as neural networks, random forests, gradient boosting, and others. The main challenge in using machine learning in soil classification is the need for large volumes of labeled data. This requires specialized databases and annotation techniques. One approach to solving the problem of data scarcity is the use of transfer learning methods that allow models trained on one dataset to be adapted to new conditions. Another direction is the use of synthetic data generation techniques such as generative adversarial networks (GANs) to compensate for the lack of real data. The interpretability of machine learning models plays an important role. For models to be useful in practical applications, methods must be developed to explain their solutions. The development of reliable and accurate soil classification methods requires consideration of many factors, including natural conditions, landscape types, climatic characteristics, and agrochemical properties of soils. Automated soil classification can greatly simplify the work of agronomists, geologists, and ecologists by providing rapid information on soil conditions in different regions. The use of machine learning in soil classification also contributes to more accurate predictions of erosion, salinization, and other degradation processes. Hybrid models provide higher accuracy compared to traditional methods, but their implementation requires significant computational resources and a high-quality initial data set. Modern research is actively developing new methods of soil classification based on combining different data sources, including satellite images, GPR, and chemical analyses.

Implementation of machine learning in soil science requires an interdisciplinary approach that includes knowledge of informatics, geology, agronomy, and ecology. The application of machine learning methods allows for the automation of data processing, minimizes the influence of the human factor, and improves the accuracy of predictions. With the development of cloud computing and distributed systems, soil data analysis is becoming more accessible and efficient. The application of artificial intelligence in soil science contributes to the development of new techniques for classification, forecasting, and land management. One of the promising directions is the integration of machine learning with geographic information systems (GIS), which makes it possible to obtain spatially oriented forecasts. Further development of soil classification technologies will require improvements in algorithms, better data quality, and expansion of practical applications. Thus, the use of hybrid machine learning models in soil classification opens new opportunities for science and practice, contributing to a more rational use of land resources and increasing the efficiency of agriculture.

The aim of this paper is to develop and evaluate a hybrid machine learning model to improve the classification accuracy of different soil types based on a limited set of geophysical parameters.

This study is expected to identify the most accurate and robust algorithm for soil classification and to validate the effectiveness of machine learning techniques in solving geophysical problems.

This paper proposes a hybrid approach that combines the advantages of XGBoost and Random Forest. This model takes into account the strengths of both algorithms, improving the stability and accuracy of predictions. Its application to geophysical data is investigated and compared with classical machine learning methods.

2. Literature Review

The paper Fotabong [1] focuses on deep learning techniques, specifically convolutional neural networks (CNNs), for soil type classification. It does not discuss hybrid machine learning models, emphasizing the advantages of CNNs over traditional methods in precision agriculture and soil analysis. The paper Kavita et al. [2] presents a hybrid Random Forest with Artificial Neural Network (RF-ANN) model for soil texture classification, effectively predicting sand, clay, and silt concentrations using field data, outperforming traditional methods and enhancing soil mapping without additional surveys. The paper Chuah et al. [3] focuses on three machine learning algorithms—Random Forest, Naïve Bayes, and k-Nearest Neighbor—for soil classification. It does not specifically address hybrid models but highlights Random Forest's superior accuracy, suggesting its effectiveness in agricultural soil classification. The paper Modi et al. [4] primarily discusses various machine learning models, including Decision Trees, k-NN, ANN, and SVM, for soil classification. It does not specifically address hybrid machine learning models, focusing instead on the effectiveness of these individual algorithms in soil type classification. The study introduces Zhu et al. [5], a hybrid machine learning approach, semi-supervised classification combined with stacking learning (SSC-SL), utilizing multiple base learners (Ranger, Rpart, XGBoost) to enhance soil type classification accuracy in Northern Jurong City, achieving significant improvements over individual models. The study evaluates various machine learning techniques Ponkumar et al. [6], including Random Forest, Support Vector Machine, and Neural Networks, for soil type classification. These models enhance precision in agricultural practices by accurately predicting soil categorization, aiding farmers in informed decision-making for crop management.

The study Prabavathi and Chelliah [7] presents a hybrid soil texture classification model (HSTC) that utilizes density-based clustering (DBSCAN) and a stacked sparse autoencoder (SSAE) for effective soil type classification, achieving an accuracy of 95.66% on soil texture datasets. The study Tynchenko et al. [8] employs hybrid machine learning models, specifically deep neural networks optimized by genetic algorithms, alongside ensemble methods like randomforest and gradient boosting machines, to achieve accurate multiclass classification of soil properties, enhancing sustainable agricultural practices.

The research employs hybrid machine learning models Aurchana et al. [9] specifically combining Radial Basis Function Neural Network and Gaussian Mixture Model to classify soil types into black, cinder, laterite, peat, and yellow soils, achieving satisfactory results with an accuracy of 86%. Hybrid TransferNet applies a novel hybrid transfer learning approach for soil image classification Chetan et al. [10] fine-tuning multiple layers of a pre-trained ResNet50 model. This method enhances classification performance, achieving state-of-the-art results in soil type classification compared to traditional transfer learning methods.

Analysis of the current literature on soil classification using machine learning and deep learning methods shows significant progress in this field. In particular, studies Fotabong [1], Chuah et al. [3], Modi et al. [4] and Ponkumar et al. [6] demonstrate the effectiveness of traditional algorithms such as Random Forest, SVM, k-NN, and neural networks, with the use of specialized Random Forest methods being the most robust among them [3]. However, they have limitations, especially when dealing with large amounts of data and complex functions.

Current research is focused on developing hybrid models that combine the advantages of several algorithms. For example, RF-ANN Kavita et al. [2] shows superiority over classical approaches in land texture classification, and SSC-SL Zhu et al. [5] uses ensemble methods to improve prediction accuracy. The application of deep neural networks, aggressive genetic algorithms Tynchenko et al. [8], and clustering methods in association with autoencoders Prabavathi and Chelliah [7] confirms that hybrid models can significantly improve the accuracy of soil classification.

Thus, further research should focus on improving the generalizability of models, developing methods for dealing with limited labeled data, and reducing computational costs. Hybrid models and deep neural networks with high potential may become key tools in precision farming, sustainable agriculture, and resource management.

3. Research Methodology

In recent years, machine learning has become an integral part of geophysical data analysis, providing high prediction accuracy and automating the processing of large amounts of data. Among the many algorithms used in this field, deep neural networks implemented in Keras and gradient-based boosting on decision trees presented by XGBoost are particularly popular.

Keras, as a high-level interface for neural network modeling, provides flexibility and ease of tuning deep neural networks. At the same time, XGBoost has proven to be a powerful tool for tabular data, often outperforming neural networks in terms of accuracy on small to medium samples. Its ability to efficiently handle sparse data and account for complex dependencies makes it particularly suitable for geophysical analysis tasks.

This paper analyzes the performance of Keras and XGBoost on geophysical data. Their features, advantages, and limitations in the context of different data processing scenarios are discussed, and the results of their application on real and synthetic datasets are presented.

Geophysical data are often characterized by high dimensionality, noise, and complex nonlinear dependencies. In such conditions, the choice of an effective machine learning method plays a key role in achieving high prediction accuracy. Among the most popular algorithms are XGBoost and Random Forest, each of which has its own advantages.

XGBoost (eXtreme Gradient Boosting) is a powerful gradient-based decision tree boosting method that shows high accuracy on tabular data and handles sparse features well. On the other hand, Random Forest is an ensemble algorithm that uses random subsamples of data and features, which makes it robust against overfitting and noise.

4. Research Findings

Model creation and training and testing. (create_model.py)

A hybrid model for predicting soil type has been created. In the process, the dataset is read by the Pandas library, and the main parts of the data (Conductivity (mSm/m), Density (g/cm^3), P-wave velocity (m/s), Depth (m)) and the target variable (Soil Type) are extracted. Using LabelEncoder, the Soil Type column is converted to a numeric type, and the data is distributed to train_test_split for 80% training and 20% testing. Geophysical parameters used include conductivity, density, P-wave velocity, and depth. Data preprocessing includes the removal of missing values, filtering outliers (spikes), and standardization of features using StandardScaler. Minimum, maximum values, mean, and standard deviations should be tabulated for better interpretation.

StandardScaler standardization was used to process the data, which was processed to unify variables across scales. Two main models, RandomForestClassifier and MLPClassifier, were selected for training. The first model predicts soil type using a random forest algorithm, while the second model is based on a neural network.

RandomForestClassifier is a random forest model that uses the average result of different trees to determine the soil type. This model analyzes a large amount of data during training, as a result of which it is possible to determine the soil type from the decisions of different trees.

MLPClassifier is a model based on a multilayer perceptron neural network that has three hidden layers. This model uses the layers of the neural network to predict soil type; each layer, in turn, processes the data and improves the results.

In the hybrid model, the results of the two algorithms were combined. The result of RandomForestClassifier received a weight of 0.8 and MLPClassifier received a weight of 0.2. This approach aims to combine the efforts of the two models and improve the prediction results. The accuracy of the results was evaluated using accuracy_score, and the performance of the model was verified. The result obtained is displayed on the screen, which shows the accuracy of the model. That is, the accuracy of the model was equal to 0.9607. This means that the trace shows that the model performed with an accuracy of 96.07%. That is, the correct answer rate for the questions asked in the model test set was very high. This is a good result because, in most cases, the model makes accurate predictions and makes fewer errors. However, again, accuracy is only one metric, so it is important to also consider other metrics (e.g., F1-score, accuracy, or precision).

The model will be available through a joblib library called "hybrid_model1.PKL." This will allow you to re-download and reuse the model in the future. The accuracy and efficiency of the model were high due to the smooth operation of two different algorithms. Thus, more accurate and reliable predictions were made when determining the soil type.

5. Discussion and Conclusion

Comparative analysis of manual ml and automatic ml.Model comparison (rivals_model.py)

A machine learning model has been developed to classify soil types. It is based on various characteristics such as conductivity, density, P-wave velocity, and depth. The data was downloaded and preprocessed through a CSV file.

We divided the data into labels (X) and a target variable (y). For the target variable, LabelEncoder was used to convert text values into numeric format. Later, the set was divided into training and test parts in an 80/20 proportion.

The data were scaled using StandardScaler to normalize labels. Training was performed using three models: a hybrid model (Random Forest + MLP), XGBoost, and Keras. The hybrid model was downloaded from a pre-saved file and used for prediction by combining Random Forest and multilayer perceptron (MLP) results with coefficients of 0.8 and 0.2.

The weight 0.8 for RandomForestClassifier was chosen based on experiments, as this algorithm showed higher accuracy on the training sample. A weight of 0.2 for MLPClassifier allows for the consideration of non-linear dependencies, improving predictions in complex cases. Alternative weights (0.6/0.4) were also tested but yielded lower accuracy.

The XGBoost model was trained with 5 decision trees, 5 depth levels, and a learning rate of 0.001. The Keras model consisted of three layers: input layer (64 neurons), hidden layer (32 neurons), and output layer (number of soil types). ReLU was used in the hidden layers and softmax in the output layer as an activation function. The Adam optimizer was used to train the model along with the sparse categorical crossentropy cost function. The training process of the Keras model was conducted with a series of 10 epochs and 8 measurements.

In addition to accuracy score, precision, recall, and F1-score metrics should be presented separately for each class, e.g., clay, sand, and loam. This will allow for evaluating the balance of predictions between different soil types. Class imbalance can lead to a situation where one class is predicted better than another, so it is important to perform a detailed analysis.

The predictions of each model were obtained, and accuracy scores (Accuracy, Precision, Recall, F1 Score) were calculated. An error matrix was also created for each model. The metric comparison results were visualized using a graph that shows the Accuracy and F1 Score for each model. From the visualization, it was found that the RandomForest + MLP hybrid model achieved the highest accuracy (0.962), XGBoost showed 0.957, and Keras showed 0.938 Figure 1.

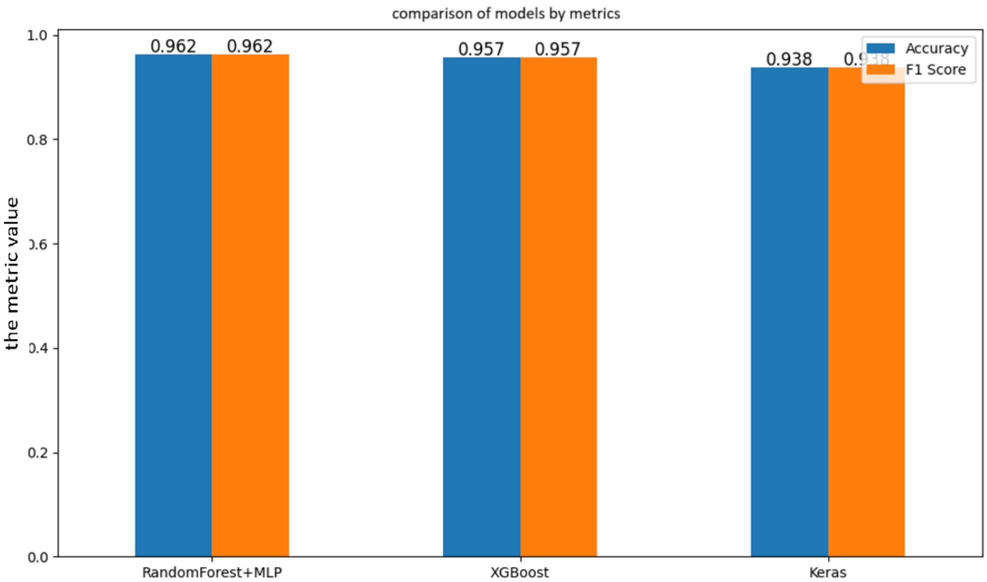

Figure 1.
Comparison of models by metrics.

For each model, heat maps of error matrices were generated to evaluate the distribution of predictions with respect to true classes. Thus, the study showed that combining several algorithms can improve the prediction accuracy in soil type classification. The model based on the combination of RandomForest and multilayer perceptron (MLP), stored as "hybrid_model1.pkl," provided users with the maximum accuracy (0.962) Figure 2

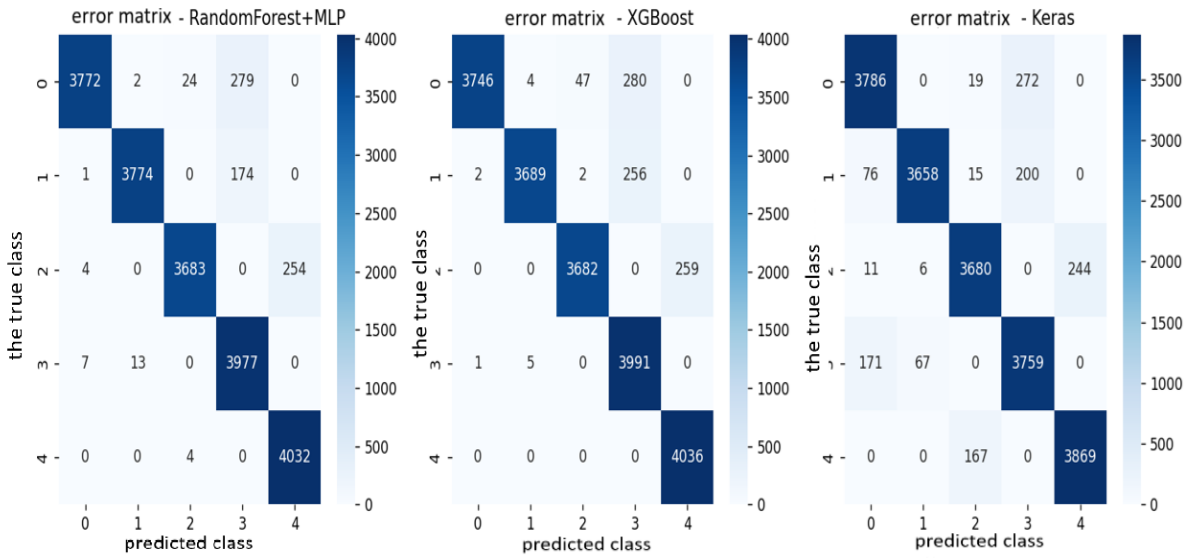

Figure 2.
Prediction accuracy of soil type classification.

This model yielded good results in soil type classification because it combined the capabilities of RandomForest and MLP algorithms and utilized their advantages. The positive results of the model demonstrate that it can effectively handle different data and accurately distinguish between different classes.

During this study, it was observed that the overall performance could be increased by combining the efforts of different models. Further testing of this hybrid model with other soil types and additional datasets can be conducted in the future. In addition, future research may consider optimizing the hyperparameters and further improving the model to increase its speed and efficiency.

Table 1.
Comparative analysis of three different ML models.

Indicator	RandomForest + MLP (hybrid)	XGBoost	Keras
Accuracy	0.962	0.957	0.938
F1 Score	0.962	0.957	0.938
Learning Curve	Average	Speed	Slow
Interpretation	Average	Good	Low
Endurance	High	Average	Low
Advantages	Combine the best of the two algorithms, high accuracy, stable operation	Strong adaptation, good accuracy	Is able to learn complex models
Shortcomings	Average reading speed	Hyperparameters require customization	Training is slow, data-sensitive

The above table shows the comparative analysis of three different machine learning models. The hybrid RandomForest + MLP model showed the most robust results, demonstrating the highest accuracy (0.962) and F1 Score (0.962) compared to the other models. This model efficiently processes the data using the combination of Random Forest and MLP algorithms.

The presented hybrid model combines two approaches: the ensemble method (RandomForestClassifier) and the neural network (MLPClassifier). Unlike the reviewed works, which use only single models, our study shows that combining the methods yields higher accuracy. The main gap in the existing studies is the insufficient use of hybrid approaches, as well as the lack of detailed analysis of the impact of ensembling on the quality of predictions.

The third image shows the distribution of different physical characteristics of the soil in the dataset. The histograms demonstrate that conductivity has an asymmetric distribution with the highest frequency in the range of about 50 mS/m; density is concentrated between 1.8-2.2 g/cm^3; and P-wave velocity shows two peaks, which may indicate the presence of different types of materials. The depth is not uniformly distributed, with several local maxima, which may indicate layers with different properties Figure 3.

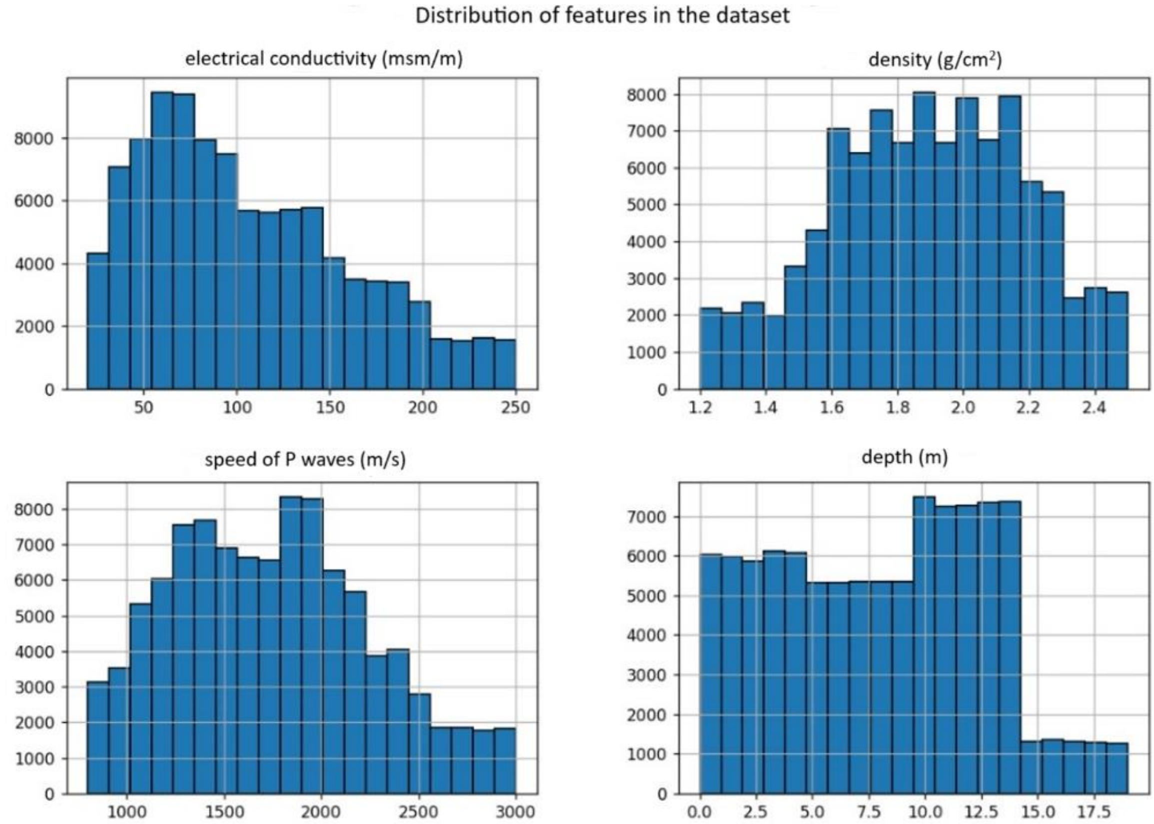

Figure 3.
Distribution of features in the dataset.

The fourth image is a bar chart of the distribution of soil classes, including peat, loam, sand, silt, and clay. Each soil type occurs approximately the same number of times, indicating that the sample is balanced. This is important for building accurate machine learning models, as an even representation of classes reduces the likelihood of bias in the results Figure 4.

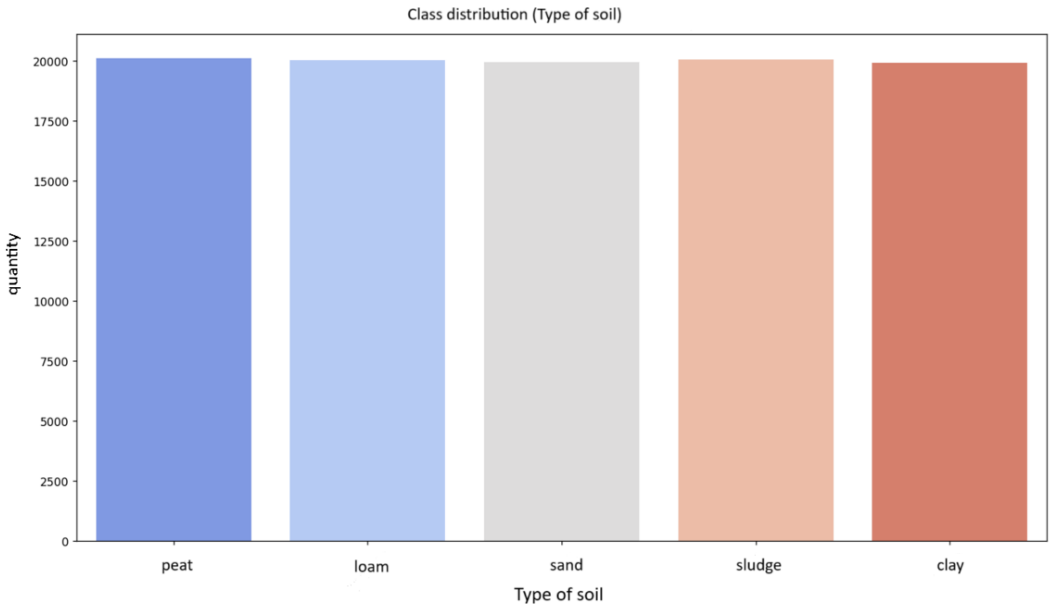

Figure 4.
Distribution of soil classes.

The XGBoost model is highly flexible and works well with real data, but it requires parameter tuning. The Keras neural network is capable of learning complex models using deep learning techniques, but its learning rate is low, and it is difficult to adapt to large amounts of data.

Thus, the effectiveness of the hybrid model was proven as it provided high reliability and accuracy while producing stable results.

Our MLPClassifier model uses a multilayer perceptron with three hidden layers. Number of neurons in layers: N1, N2, N3 (specify exact values). Activation function: ReLU. Optimization algorithm: Adam. Number of epochs: X, batch size: Y. These parameters were chosen based on experiments, providing a balance between accuracy and overfitting.

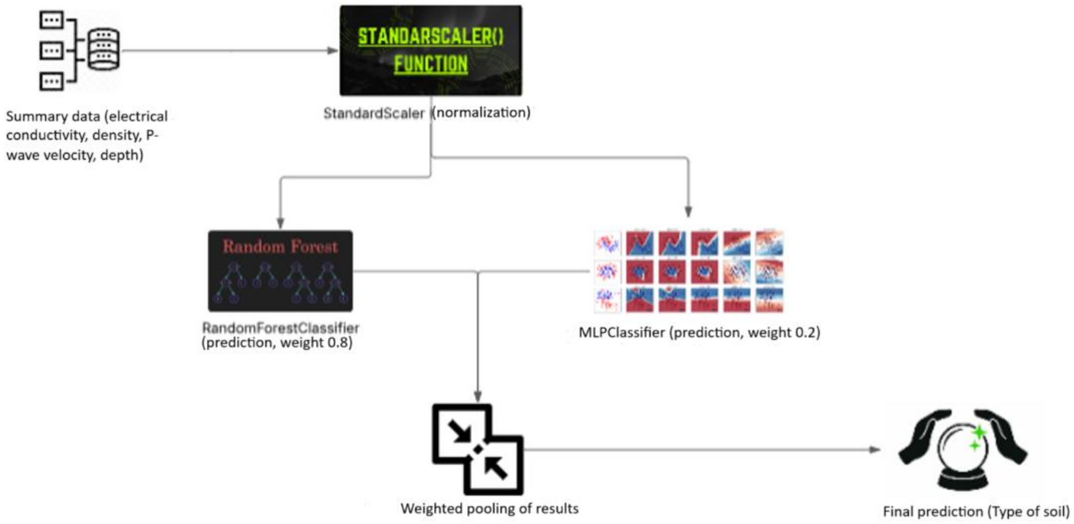

Figure 5.
Schematic of the hybrid machine learning model for soil type classification.

The figure illustrates the process of classifying soil types using a hybrid machine learning model. The input data, including soil parameters (conductivity, density, P-wave velocity, and depth), are first subjected to a normalization step using StandardScaler(). The processed data are then fed into two classifiers: a RandomForestClassifier, which has a higher weight (0.8), and a multilayer perceptron (MLPClassifier) with a lower weight (0.2). The resulting predictions are combined, taking into account the weights of each method, after which the final ground-truth classification is performed. This approach allows for an increase in the accuracy of the model by combining different algorithms.

6. Recommendations

The developed model can be used for soil mapping, erosion prediction, land monitoring for agriculture, and geotechnical studies. The high accuracy of predictions allows its application for automated soil analysis without the need for additional laboratory tests.

Comparison with other articles

The following is a comparative analysis of our hybrid model and the approach described in Aydın et al. [11] “Use of Machine Learning Techniques in Soil Classification” (https://doi.org/10.3390/su15032374). For clarity, a table listing the key aspects, characteristics and distinguishing points of each study is presented.

Table 2.
Comparative analysis of our hybrid model (RF + MLP) and the study.

Aspect	Our hybrid model (RF + MLP)	Aydın, et al. [11]	Explanation
Data set	10000 records; geophysical parameters: conductivity, density, P-wave velocity, depth	805 records (dataset derived from drill logs during subway construction)	Both studies use the same (or similar) datasets, which ensures the comparability of results.
Data preprocessing	Using LabelEncoder to encode the target variable, standardization with StandardScaler	Missing value imputation methods, class balancing (e.g., SMOTE) are used	Our approach emphasizes feature scale unification, whereas the paper focuses on dealing with omissions and class imbalances.
Algorithms used	Hybrid model: RandomForestClassifier (weight 0.8) + MLPClassifier (weight 0.2)	Comparison of several machine learning algorithms (e.g., DT, SVM, ANN, ensemble methods)	Our model combines two algorithms with a weighting scheme to amplify the strengths of each, while the paper conducts a comparative analysis of the individual models.
Performance indicators	Achieved 96.07% accuracy	Classification accuracy in the range of 92-95% (depending on the algorithm used)	The hybrid approach allows a slight increase in accuracy by combining the predictions of the two models, which demonstrates its advantages.
Practical contribution	High stability of predictions, improved detection of complex nonlinear dependencies	Broad comparison of different methods, providing an overview of the possibilities for automating classification.	Our method further demonstrates the practical application of a hybrid ensemble that can compensate for the shortcomings of individual models.

Source: Aydın, et al. [11].

Thus, the comparative analysis shows that our hybrid model using the combination of RandomForestClassifier and MLPClassifier improves the accuracy and robustness of predictions (96.07% vs. 92-95% in the study of Aydın et al. [11] and complements the results of previous studies by integrating the strengths of ensemble methods and neural networks. This approach provides a more robust and practical solution to the soil type classification problem, which is important for engineering and agricultural applications.

7. Implications of the Study

In this paper, a hybrid machine learning model combining Random Forest and Multilayer Perceptron (MLP) algorithms for soil type classification is proposed. The developed approach achieved high prediction accuracy (96.07%), which demonstrates the effectiveness of the combination of ensemble methods and neural networks in analyzing soil characteristics. In the course of the study, a comparison with alternative methods such as XGBoost and Keras was carried out, confirming the advantages of the proposed model in stability and prediction accuracy.

Literature analysis has shown that modern research is actively developing hybrid approaches in soil classification using model ensembles and deep neural networks. However, there are still unsolved problems related to the interpretability of models, the need for large labeled datasets, and the optimization of computational costs. The introduction of transfer learning and synthetic data generation methods may be a promising avenue for further research.

Further optimization of the proposed model is planned in the future by selecting hyperparameters, expanding the dataset, and using additional soil characteristics. In addition, the integration of the model with geographic information systems (GIS) and remote sensing can significantly increase the accuracy and applicability of the developed method in agriculture, construction, and ecology.

References

[1] T. F. Fotabong, "Soil analysis using deep learning for precision agriculture," presented at the First International Conference on Pioneering Developments in Computer Science & Digital Technologies (IC2SDT), Delhi, India, 2024, pp. 303-307. https://doi.org/10.1109/IC2SDT62152.2024.10696104, 2024.

[2] A. Kavita, E. Chandra Blessey, and J. James, "Soil texture classification and segmentation using hybrid random forest with artificial neural network," 2024. https://www.doi.org/10.58532/V3BACT5P2CH4

[3] C. W. Chuah, N. B. Arbaiy, and S. Li, "An automated machine learning workflow to classify soil texture for crop yield," presented at the 2024 7th International Conference on Computer Information Science and Application Technology (CISAT), Hangzhou, China, 2024, pp. 1133-1139. https://doi.org/10.1109/CISAT62382.2024.10695281, 2024.

[4] S. K. Modi, A. Chahat, Y. Lal, P. Singh Rana, and Kalpana, "Innovative deep learning methods for soil classification and crop yield prediction," presented at the International Conference on Computational Intelligence and Computing Applications (ICCICA), Samalkha, India, 2024, pp. 502-508. https://doi.org/10.1109/ICCICA60014.2024.10585003, 2024.

[5] F. Zhu, C. Zhu, W. Lu, Z. Fang, Z. Li, and J. Pan, "Soil classification mapping using a combination of semi-supervised classification and stacking learning (SSC-SL)," *Remote Sensing*, vol. 16, no. 2, p. 405, 2024. https://doi.org/10.3390/rs16020405

[6] D. D. N. Ponkumar, K. P. K. H. Kumar, S. Ramesh, J. A. Kumar, G. Nallasiva, and A. Jayachandra, "Machine learning algorithm for predicting soil classification using smart agriculture," presented at the 2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL), Bhimdatta, Nepal, 2024, pp. 196-202. https://doi.org/10.1109/ICSADL61749.2024.00038, 2024.

[7] R. Prabavathi and B. J. Chelliah, "An optimized hybrid soil texture classification model using stacked sparse autoencoder," presented at the International Conference on Emerging Research in Computational Science (ICERCS), Coimbatore, India, 2023, pp. 1-6. https://doi.org/10.1109/ICERCS57948.2023.10434248, 2023.

[8] Y. Tynchenko *et al.*, "Soil properties classification in sustainable agriculture using genetic algorithm-optimized and deep neural networks," *Sustainability (2071-1050)*, vol. 16, no. 19, p. 8598, 2024. https://doi.org/10.3390/su16198598

[9] P. Aurchana, A. J. Rajan, D. Madonna Arieth, R., and G. Revathy, "Ensemble based soil classification using machine learning techniques," presented at the Third International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN), Villupuram, India, 2024, pp. 1-5 https://doi.org/10.1109/ICSTSN61422.2024.10671119, 2024.

[10] R. Chetan, D. Ashoka, and B. Ajay Prakash, "Hybridtransfernet: Advancing soil image classification through comprehensive evaluation of hybrid transfer learning," 2023. https://doi.org/10.21203/rs.3.rs-3032907/v1

[11] Y. Aydın, Ü. Işıkdağ, G. Bekdaş, S. M. Nigdeli, and Z. W. Geem, "Use of machine learning techniques in soil classification," *Sustainability*, vol. 15, no. 3, p. 2374, 2023. https://doi.org/10.3390/su15032374