Vietnam Journal of Earth Sciences, 48(2), 128−144, https://doi.org/10.15625/2615-9783/24022

Vietnam Journal of Earth Sciences
https://vjs.ac.vn/index.php/jse

Optimizing XGBoost, CatBoost, and Bagging models for predicting the maximum dry density of compacted soil using grid search hyperparameter tuning

Binh Thai Pham[*1], Indra Prakash[2]

[1]*Geotechnical and Artificial Intelligence research group, University of Transport Technology, Hanoi 100000, Vietnam*
[2]*Formerly Dy. Director General, Geological Survey of India, Gujarat, India*

Received 07 November 2025; Received in revised form 12 December 2025; Accepted 29 December 2025

ABSTRACT

In geotechnical engineering, an accurate estimation of maximum dry density (MDD) is essential to ensure the stability of geotechnical structures such as roads, embankments, and foundations. While traditional laboratory methods, such as the Proctor compaction test, are reliable, they are often labor-intensive and time-consuming. Therefore, the main aim of this study is to develop efficient data-driven models, including XGBoost (XGB), CatBoost (CAB), and Bagging (BAG), for rapid and reliable estimation of MDD using easily measurable soil properties. A dataset of 214 soil samples collected from the Van Don-Mong Cai expressway construction project (Vietnam) comprising eight key input variables was used: gravel content, coarse and fine sand contents, silt and clay content, optimum moisture content, liquid limit, plastic limit, and plasticity index. Model performance was evaluated using R^2, RMSE, MAE, and a Taylor diagram. Results indicate that the Grid Search-optimized BAG model achieved the best performance, with R^2 values of 0.94 and 0.81 for the training and test datasets, respectively, and the lowest RMSE and MAE. Optimized CAB showed comparable performance, while XGB exhibited relatively lower generalization capability. Optimized CAB yielded similar results, whereas optimized XGB performed worse. The significance of this study lies in demonstrating that ensemble learning models, particularly Bagging, can provide accurate, physically interpretable predictions of MDD, thereby reducing reliance on extensive laboratory testing. The novelty of this work lies in the systematic comparison of optimized ensemble models using a real construction dataset, combined with interpretability analysis via partial dependence plots consistent with established soil mechanics principles. These findings highlight the potential of optimized machine learning models as practical tools for modern geotechnical engineering applications.

Keywords: Maximum dry density, soil compaction, ensemble learning, XGBoost, CatBoost, Bagging, grid search optimization.

1. Introduction

The maximum dry density (MDD) of soil is a fundamental parameter in geotechnical engineering, directly influencing the design, stability, and longevity of infrastructure such as roads, embankments, and foundations (Zhao et al., 2024a). Accurate estimation of MDD ensures proper compaction, optimizes

*Corresponding author, Email: binhpt@utt.edu.vn

bearing capacity, and minimizes settlement, which are essential for safe and durable construction (Zhao et al., 2024a). Traditionally, laboratory-based methods, most notably the standard and modified Proctor compaction tests, have been the benchmark for determining MDD (Demir, 2024). Although reliable, these tests are labor-intensive, time-consuming, and often impractical for large-scale or time-sensitive projects. Moreover, they require significant sample preparation, controlled conditions, and skilled personnel, making them less feasible in remote or resource-constrained settings. To overcome these limitations, researchers have explored predictive approaches based on easily measurable soil properties, such as Atterberg limits, grain-size distribution, and moisture content. Early studies primarily relied on empirical correlations and statistical regression models. While these approaches are interpretable and straightforward, their ability to capture the complex, non-linear behavior of soil compaction is limited, resulting in reduced prediction accuracy under heterogeneous soil conditions.

Recently, machine learning (ML) methods have emerged as powerful alternatives for modeling and solving complex geotechnical problems, including geotechnical prediction (Nhat et al., 2025; Prakash et al., 2024). ML models can learn non-linear relationships directly from data and have been successfully applied to predict various soil and foundation parameters, including MDD. Several studies have demonstrated the potential of individual ML models and hybrid approaches for estimating compaction characteristics. However, many existing studies rely on default model parameters or limited tuning, which can significantly constrain model performance and robustness. Moreover, systematic comparisons of optimized ensemble learning models using real construction datasets remain limited in the literature.

ML is fundamentally based on the principle of learning patterns from data to make predictions about unseen or future data. In predictive tasks, ML models are trained on historical datasets where the relationship between input features and output targets is known. Through this process, the model learns how input variables influence the outcome. ML models have been successfully applied to predict various soil compaction properties in geotechnical engineering. More specifically, Zhao and Guan (2024) estimated the MDD of stabilized soils using both individual and hybrid machine learning techniques such as Support Vector Regression (SVR) enhanced with three optimization algorithms: Manta Ray Foraging Optimization (MRFO), Artificial Rabbits Optimization (ARO), and Improved Manta-Ray Foraging Optimizer (IMRFO), and stated that among the models tested, the hybrid SVAR model (SVR combined with ARO) achieved the best performance. Omer (2025) investigated the prediction of modified Proctor compaction parameters, specifically optimum moisture content (OMC) and MDD, using a combination of multivariable mathematical models and ML techniques such as linear regression (LR), multiple linear regression (MLR), nonlinear regression (NLR), and M5P decision tree and artificial neural networks (ANN). He showed that, among these models, the ANN outperformed the others. Verma et al. (2024) tested five ML algorithms: K-Nearest Neighbor (KNN), Gaussian Process Regression (GPR), Random Forest (RF), and Extreme Gradient Boosting (XGB), each enhanced through Particle Swarm Optimization (PSO), were used for estimating modified Proctor compaction parameters such as MDD and OMC, and showed that among the models, XGB showed the highest predictive capability for MDD, followed by KNN, GPR, and RF. At the same time, OMC prediction was best performed by XGB, KNN, RF, and GPR, respectively.

In general, many studies on MDD prediction have relied on default model parameters or limited hyperparameter tuning, which can significantly constrain model performance (Ali et al., 2024; Ewusi-Wilson et al., 2025). Furthermore, comprehensive comparisons of ensemble models, particularly under systematically optimized conditions, are rare in the literature. This study addresses these gaps by developing and rigorously optimizing XGBoost (XGB), CatBoost (CAB), and Bagging (BAG) models for MDD prediction using Grid Search hyperparameter tuning. A diverse dataset encompassing a wide range of soil types and index properties, collected from the Van Don-Mong Cai expressway construction project (Vietnam), was used to train and validate the models. The performance of the models was assessed using popular evaluation metrics, including root mean square error (RMSE), coefficient of determination (R^2), and mean absolute error (MAE). The novelty of this work lies in its systematic comparison of state-of-the-art ensemble methods under optimized conditions, offering not only improved predictive performance but also insights into the physical interpretability of model outputs. This contributes to the development of efficient, reliable, and scalable tools for modern geotechnical engineering applications.

Based on these considerations, the objectives of the present study are as follows: (i) to develop and optimize XGBoost, CatBoost, and Bagging ensemble models for predicting the maximum dry density of compacted soils using basic soil properties; (ii) to systematically evaluate and compare the performance of these models using multiple statistical metrics and graphical validation tools. (iii) to interpret model behavior using partial dependence plots and assess consistency with fundamental soil mechanics principles; and (iv) to demonstrate the practical applicability of optimized ensemble models as reliable alternatives to time-consuming laboratory compaction tests. By addressing these objectives, this study advances the use of explainable, optimized machine learning techniques in geotechnical engineering, providing both methodological insights and practical guidance for soil compaction assessment in real-world construction projects.

2. Materials and Methods

2.1. Data Used

In this study, the dataset for predicting the MDD of soil was compiled from the Van Don - Mong Cai expressway construction project (Vietnam). This dataset has been reported in previous studies, and appropriate references have now been added to acknowledge earlier use of similar data. The dataset comprises 214 soil samples spanning a wide range of soil types commonly encountered in highway construction projects.

The selection of these eight parameters was based on their widespread use in geotechnical practice and their direct physical relevance to soil compaction behavior. Grain size fractions (G, CS, FS, and SC) control particle packing, interlocking, and void ratio. Optimum moisture content governs lubrication and particle rearrangement during compaction. Atterberg limits (LL, PL, and PI) represent soil plasticity and water affinity, which strongly influence achievable dry density. Other potentially relevant parameters, such as compaction energy or mineralogical composition, were not included due to data unavailability and are acknowledged as limitations and future research opportunities.

All soil properties were measured following standard geotechnical testing procedures, ensuring data reliability and consistency. Table 1 presents the statistical characteristics of the dataset, including the mean, standard deviation, minimum, maximum, and quartiles, providing insight into data variability and representativeness.

The wide range of values confirms that the dataset captures diverse soil conditions, which is essential for robust model training and validation.

Table 1. Statistical analysis of the data used

No	Variables	Mean	Std	Min	25%	50%	75%	Max
1	G (%)	22.057056	13.295476	0.000	9.0750	24.750	31.7000	51.40
2	CS (%)	24.101028	7.017101	3.000	20.7000	23.700	27.7750	46.30
3	FS (%)	9.034766	6.467583	2.500	4.6000	7.250	11.0000	41.50
4	SC (%)	44.807103	10.447274	17.870	37.7500	44.550	49.2000	88.70
5	O	1.508879	0.372560	0.120	1.2525	1.510	1.7700	2.94
6	LL (%)	39.514533	6.173263	2.080	36.6375	39.990	43.5075	48.45
7	PL (%)	20.318411	3.067936	1.170	19.2925	20.835	21.8875	28.49
8	PI (%)	19.197991	4.077597	0.910	16.8300	18.435	22.3200	27.48
9	MDD (g cm^{-3})	1.880640	0.117635	1.672	1.8220	1.871	1.9610	2.14

Figure 1 illustrates the distribution of each variable, highlighting their diversity and implications for predictive modeling. G distribution is bimodal, with peaks at low and high values, indicating soils with minimal and substantial gravel content. (Fig. 1a) CS is slightly right-skewed, with most samples between 15% and 30% (Fig. 1b). FS shows a strong right skew, with most samples having low FS content (Fig. 1c). SC is near-normally distributed around 45%, reflecting balanced representation (Fig. 1d). O is right-skewed, clustering around 0.015 (Fig. 1e). Atterberg limits (LL, PL, PI) are slightly right-skewed, indicating moderate to high plasticity (Fig. 1f–h). MDD shows a bimodal distribution with peaks around 1.8 g cm^{-3} and 2.0 g cm^{-3}, ensuring robust model training across varied compaction scenarios (Fig. 1i).

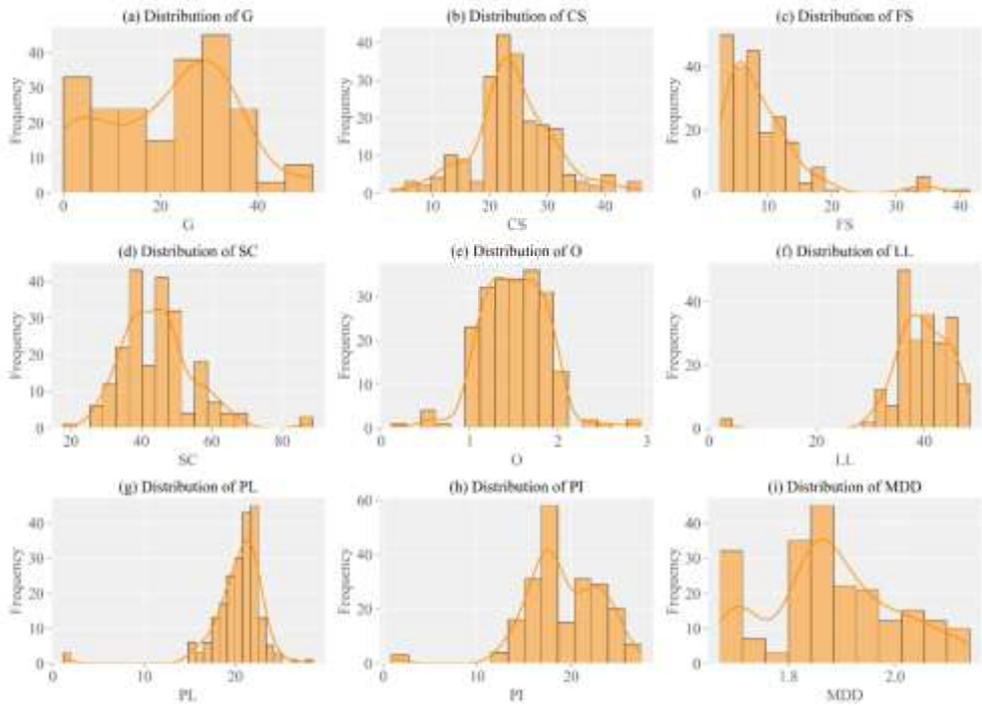

Figure 1. Distribution of variables used in the modeling

Strong correlations among some Atterberg limit parameters indicate potential multicollinearity; however, these variables were retained due to their physical significance in soil mechanics. The limitations of Pearson's correlation in detecting multicollinearity are acknowledged, and advanced diagnostics, such as the Variance Inflation Factor (VIF), are suggested for future studies.

Figure 2 presents the correlation matrix of the variables, highlighting linear relationships with MDD. G has a strong positive correlation with MDD (0.62), while SC shows a strong negative correlation (-0.59). FS has a moderate negative correlation (-0.29), and CS has a negligible correlation (-0.03). The Atterberg limits (LL, PL, PI) and O show weak negative correlations with MDD (-0.19 to -0.23). Strong intercorrelations exist among LL, PL, and PI (e.g., LL and PI: 0.90), and a strong negative correlation exists between G and SC (-0.74). These insights guide feature selection and highlight potential multicollinearity.

Figure 2. Correlation Matrix of variables used in the modeling

To improve model performance and ensure fair comparison among machine learning algorithms, standard normalization was applied to all input variables. This transformation rescales features to have zero mean and unit variance, preventing variables with larger numerical ranges from dominating the learning process. Normalization enhances model convergence and stability but does not eliminate outliers, which is a recognized limitation.

2.2. Methods used

2.2.1. XGBoost (XGB)

XGB, developed by Chen and Guestrin (2016), is an optimized gradient boosting algorithm renowned for its high performance and scalability in regression and classification tasks. Its ability to model complex nonlinear relationships and handle structured data makes it ideal for predicting soil MDD in geotechnical engineering (Li et al., 2024). Its regularization mechanism helps control model complexity and reduce overfitting, particularly on relatively small datasets. Grid Search was selected because the dataset is moderate in size (214 samples), the hyperparameter space is manageable, and the method is transparent and reproducible, which is essential for engineering applications. Each combination of hyperparameters was evaluated using cross-validation, and the optimal set was selected based on prediction error.

XGB employs gradient boosting, where an ensemble of weak learners (typically decision trees) is built sequentially, with each tree correcting the errors of the previous ensemble (Chen and Guestrin, 2016; Qiu et al., 2022). The model minimizes a loss function by adding trees that predict residuals. For a sample x_i, the prediction is given by (Chen and Guestrin, 2016):

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), f_k \in \mathrm{F} \qquad (1)$$

where F is the space of regression trees, and f_k represents the k-th tree.

The objective function to be minimized at each iteration combines a differentiable convex loss function l (measuring the difference between the predicted and actual values) and a regularization term Ω (penalizing model complexity):

$$\mathrm{L}(\phi)=\sum_{i=1}^{n} l(y_i,\hat{y}_i)+\sum_{k=1}^{K}\Omega(f_k) \qquad (2)$$

where y_i is defined as the true value, $\hat{y}_i$ is defined as the predicted value, and n is the number of samples.

The regularization term is typically defined as:

$$\Omega(f)=\gamma T+\frac{1}{2}\lambda\,|\,w|^2 \qquad (3)$$

Where T n is the number of leaves in the tree, w is the vector of leaf weights, γ and λ are regularization parameters controlling model complexity and overfitting.

2.2.2. CatBoost (CAB)

CatBoost, developed by Yandex in 2017 (Cai et al., 2024), is a gradient boosting algorithm designed to handle categorical features efficiently and reduce overfitting. Its strong predictive performance and minimal tuning requirements make it suitable for regression tasks, such as predicting soil MDD. CatBoost builds an ensemble of decision trees sequentially, using ordered boosting to prevent target leakage and efficient encoding for categorical features (Hancock and Khoshgoftaar, 2020). The prediction for a sample x_i after K iterations is (Prokhorenkova et al., 2018):

$$\hat{y}_i^{(t)}=\hat{y}_i^{(t-1)}+f_t(x_i) \qquad (4)$$

Where f_t is the decision tree added at the iteration t?

With ordered boosting, for a permutation σ of the dataset, the prediction for the i-th sample is:

$$\hat{y}_{\sigma(i)}^{(t)}=\hat{y}_{\sigma(i)}^{(t-1)}+f_t\big(x_{\sigma(i)};\{x_{\sigma(j)},y_{\sigma(j)}\}_{j<i}\big) \qquad (5)$$

For categorical features, the target statistic for a category c in the sample i is calculated as:

$$TS_i=\frac{\sum j<i[x_{\sigma(j)}=c]\cdot y_{\sigma(j)}+a\cdot p}{\sum_{j<i}[x_{\sigma(j)}=c]+a} \qquad (6)$$

where a is a regularization parameter and p is the prior (often the mean target value). The overall objective function minimized by CatBoost is:

$$\mathrm{L}=\sum_{i=1}^{n} L(y_i,\hat{y}_i^{(T)})+\sum_{t=1}^{T}\Omega(f_t) \qquad (7)$$

where L is the loss function (such as mean squared error for regression), and Ω is a regularization term.

2.2.3. Bagging (BAG)

BAG, introduced by Breiman (1996), is an ensemble method that reduces model variance by training multiple base learners on bootstrap samples and averaging their predictions. Its simplicity and effectiveness make it suitable for predicting soil MDD. The main principle of BAG is to generate multiple bootstrap samples from the training data (Breiman, 1996). Each bootstrap sample is created by randomly sampling, with replacement, from the original dataset, resulting in several datasets of the same size as the original (Breiman, 1996). A separate base estimator (often a decision tree regressor) is trained on each bootstrap sample. For regression tasks, the final prediction for a given input is obtained by averaging the predictions from all base estimators (Nguyen et al., 2022). In this study, BAG is applied to model the relationship between soil properties and MDD, leveraging its ability to capture diverse patterns in the data and to reduce the risk of overfitting (Ewusi-Wilson et al., 2025).

BAG generates B bootstrap samples, Db (for b = 1, 2,…, B) by sampling with replacement from the dataset D. A base regressor hb (e.g., a decision tree) is trained on each sample. The final prediction for input x is (Breiman 1996):

$$\hat{y}_{BAG}(x) = \frac{1}{B}\sum_{b=1}^{B} f^{(b)}(x) \quad (8)$$

where B is the number of base estimators, and $f^{(b)}(x)$ is the prediction of the b-th estimator for input x.

2.2.4. Grid search optimization

Grid search optimization systematically tunes hyperparameters to maximize model performance (Zhao et al., 2024b). The main principle of grid search optimization is to define a grid of possible values for each model hyperparameter (Ashwini et al., 2024). The algorithm then trains and evaluates the model for every possible combination of these values, typically using cross-validation to ensure robust performance estimation (Ayalew et al., 2024). It is applied in this study to optimize XGB, CAB, and BAG models for predicting soil MDD. Grid search evaluates all combinations in a predefined hyperparameter grid Θ, using cross-validation to assess performance via a metric S (e.g., mean squared error). The optimal hyperparameters θ^* are:

$$\theta^* = \arg\min_{\theta\in G} S(\theta) \quad (9)$$

Where $S(\theta)$ is the evaluation metric (e.g., validation error) for the model trained with hyperparameters θ.

Although Grid Search can be computationally expensive, its use in this study is justified by the moderate dataset size (214 samples) and the limited hyperparameter space, which allows exhaustive, transparent, and reproducible optimization. More advanced techniques, such as Random Search or Bayesian optimization, are better suited for larger datasets and are therefore suggested for future research.

2.2.5. Training and Testing Strategy

The dataset was divided into training and testing subsets using a standard 70:30 split to evaluate model generalization performance. Given the limited dataset size (214 samples), this ratio provides sufficient data for model learning while retaining an independent validation subset, thereby achieving a balanced trade-off between model stability and generalization and remaining consistent with standard practice in geotechnical ML studies.

2.2.6. Validation metrics

Validation metrics are critical for assessing the predictive performance and generalization ability of ML models (Nguyen et al., 2023; Thai et al., 2022). In this study, we employ three widely used metrics: the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE) to evaluate the effectiveness of XGB, CAB, and BAG regression models in predicting the MDD of compacted soil. These metrics provide complementary insights into model accuracy, error magnitude, and the ability to capture variance in the observed MDD data.

R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variables (Ngo et al., 2022a). It is calculated as (Nguyen et al., 2023; Prakash et al., 2022):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \quad (10)$$

where y_i is the observed value, $\hat{y}_i$ is the predicted value, $\overline{y}$ is the mean of observed values, and n is the number of samples. An R^2 value closer to 1 indicates better model performance (Prakash and Phan, 2023; Le et al., 2022).

RMSE quantifies the average magnitude of the prediction errors, giving higher weight to larger errors. It is defined as (Pal et al., 2024; Pham et al., 2021):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (11)$$

MAE measures the average absolute difference between the observed and predicted values, providing a straightforward interpretation of prediction accuracy (Ngo et al., 2022b; Duc et al., 2025):

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (12)$$

Each metric has its limitations. R^2 can be misleading for non-linear models or in datasets with low variance. RMSE may disproportionately penalize large errors due to squaring, potentially overstating the impact of outliers. MAE, while more stable in the presence of outliers, does not penalize large errors as heavily as RMSE. By combining all three metrics, this study provides a balanced, detailed assessment of the regression models' predictive performance, accounting for both average error and variability in predictive quality.

In the current study, the Taylor Diagram (Taylor, 2001) was also used to validate and compare the models. It is a specialized graphical tool designed to concisely summarize how well model predictions match observations (Taylor, 2001). The strength of the Taylor Diagram lies in its ability to simultaneously present three important statistical measures like R^2, RMSE, and MAE within a single, intuitive polar plot. This integrated approach allows for clear visual comparison of models in terms of their ability to capture observed variability, maintain strong correlation, and minimize prediction error (Taylor, 2001).

2.2.7. Partial Dependence Plots (PDP)

PDP, introduced by Friedman (2001), are a powerful tool for interpreting machine learning models by visualizing the marginal effect of one or two input features on the predicted outcome. PDP illustrates how predictions change as a function of a specific feature, averaging out the effects of all other features (Friedman, 2001). This is particularly useful for complex, non-linear models like XGB, CAB, and BAG, where direct interpretation is challenging. In this study, PDPs are employed to analyze how soil properties affect the predicted MDD of compacted soil, ensuring model predictions align with geotechnical domain knowledge.

Mathematically, the partial dependence function for a feature x_s is defined as (Friedman, 2001):

$$PD(x_s) = \frac{1}{n}\sum_{i=1}^{n} f(x_s, x_C^{(i)}) \quad (16)$$

where f is the trained model, x_s is the value of the feature of interest, $x_C^{(i)}$ represents the values of all other features for the i-th instance in the dataset, and n is the total number of samples. The function (PD(x_s)) thus represents the average predicted outcome as x_s varies, holding all other features at their observed values.

3. Results and analysis

3.1. Comparison of model performance using Grid search optimization

Table 2 compares the performance of XGB, CAB, and BAG models for predicting soil MDD, with and without Grid search optimization. Without Grid search, all models achieve high training R^2 scores (1.00 for XGB and CAB, 0.94 for BAG), but their testing R^2 scores are lower (0.82 for XGB, 0.87 for CAB, 0.79 for BAG), indicating overfitting. This is confirmed by near-zero training RMSE and MAE values for XGB and CAB, in contrast to higher testing errors. With Grid search optimization, the models show more balanced performance: training R^2 values decrease slightly (0.89 for XGB, 0.91 for CAB, 0.94 for BAG), while testing R^2 values improve or remain stable (0.85 for XGB, 0.89 for CAB, 0.81 for BAG). Testing RMSE and MAE also decrease, e.g., XGB's testing RMSE improves from 0.07 to 0.05, demonstrating better generalization. In general, Grid search optimization effectively tunes hyperparameters to balance model complexity and predictive accuracy, reducing overfitting and enhancing robustness.

In general, Grid Search improved all applied BAG, CAB, and XGB models by methodically identifying the most suitable hyperparameter settings. Optimized models are superior because they strike an optimal balance between learning and generalization (Zhao et al., 2024b). In predicting soil compaction parameters such as MDD, this approach yields more accurate, practical, and trustworthy predictions, which are essential for geotechnical and construction applications where design reliability is critical.

Table 2. Comparison of model performance using Grid search optimization

Metric	Dataset	Without Grid search optimization			With Grid search optimization		
		XGB	CAB	BAG	XGB	CAB	BAG
R^2	Training	1	1	0.94	0.82	0.91	0.94
	Testing	0.82	0.87	0.79	0.66	0.79	0.81
RMSE	Training	0	0	0.03	0.05	0.03	0.03
	Testing	0.05	0.04	0.06	0.07	0.06	0.05
MAE	Training	0	0	0.02	0.04	0.03	0.02
	Testing	0.03	0.03	0.04	0.06	0.05	0.04

3.2. Optimized model validation and comparison

Figure 3 presents scatter plots comparing actual and predicted MDD values for the optimized XGB, CAB, and BAG models. In the training dataset, all models show strong agreement between predicted and observed values. However, differences in generalization become apparent in the testing dataset. Figure 3 presents scatter plots comparing actual and predicted MDD values for XGB, CAB, and BAG models, optimized using Grid search. Regarding R^2, the XGB model achieves an R^2 of 0.82 on the training set and 0.66 on the test set (Fig. 3a, b). The wider spread of points around the diagonal in the testing plot indicates a reasonable fit to training data but reduced accuracy on unseen data, suggesting limited generalization. The CAB model performs better, with R^2 values of 0.91 (training) and 0.79 (testing) (Fig. 3c, d). Points are more closely aligned with the diagonal, particularly in the testing set, demonstrating improved capture of data relationships and better generalization than XGB. The BAG model shows the strongest performance, with R^2 values of 0.94 for training and 0.81 for testing (Fig. 3e, f). The tight clustering of points along the 1:1 line confirms its ability to capture the underlying relationships between soil properties and MDD.

Figure 4 presents line plots comparing actual and predicted MDD values for optimized XGB, CAB, and BAG models on the training (Fig. 4a) and testing (Fig. 4b) datasets. These plots assess each model's ability to capture MDD variations after Grid search optimization. In the training dataset (Fig. 4a), all models closely track the actual MDD values, with predicted lines nearly overlapping the dashed line representing true values. BAG and CAB show slightly better alignment at extreme values (peaks and troughs), indicating a marginally better fit compared to XGB. In the testing dataset (Fig. 4b), BAG and CAB maintain closer alignment with actual values, especially in regions with rapid MDD fluctuations. XGB, while effective, shows larger deviations at specific points, indicating slightly weaker generalization.

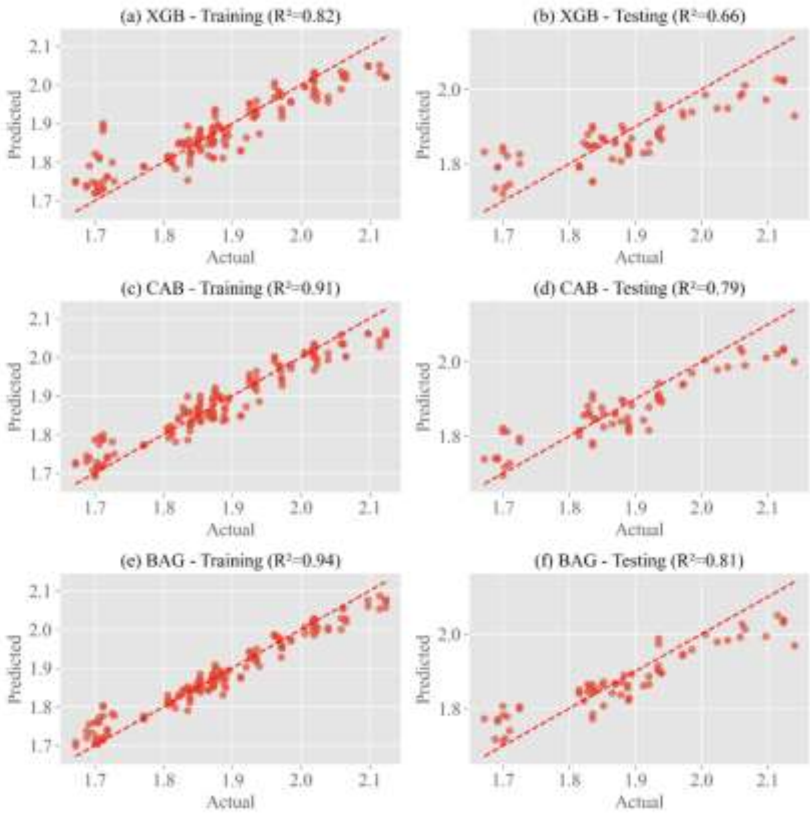

Figure 3. Scatter plots of actual vs. predicted MDD for optimized models: (a) XGB training, (b) XGB testing, (c) CAB training, (d) CAB testing, (e) BAG training, and (f) BAG testing

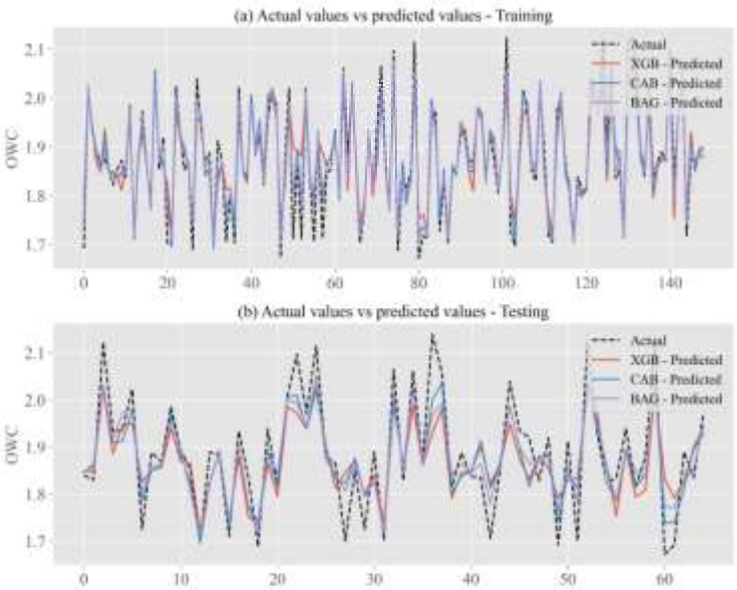

Figure 4. Line plots of actual vs predicted values of the models: (a) training and (b) testing

Figure 5 summarizes the performance metrics R², RMSE, and MAE for optimized XGB, CAB, and BAG models on training and testing datasets, using bar charts to highlight predictive accuracy and generalization after Grid search optimization. Figure 5a shows R² values, with higher values indicating a better fit. The BAG model achieves the highest R² (0.94 training, 0.81 testing), followed by CAB (0.91 training, 0.79 testing) and XGB (0.82 training, 0.66 testing). BAG's superior R² reflects its ability to explain data variance and generalize effectively. Figure 5b shows RMSE, with lower values indicating better performance. BAG records the lowest RMSE (0.03 training, 0.05 testing), followed by CAB (0.03 training, 0.06 testing) and XGB (0.04 training, 0.07 testing). BAG's low RMSE underscores its precision in MDD predictions. Figure 5c shows MAE, with BAG again outperforming others (0.02 training, 0.04 testing), followed by CAB (0.03 training, 0.05 testing) and XGB (0.04 training, 0.06 testing). These results confirm BAG's robust accuracy, supporting its use in geotechnical engineering for reliable soil compaction analysis.

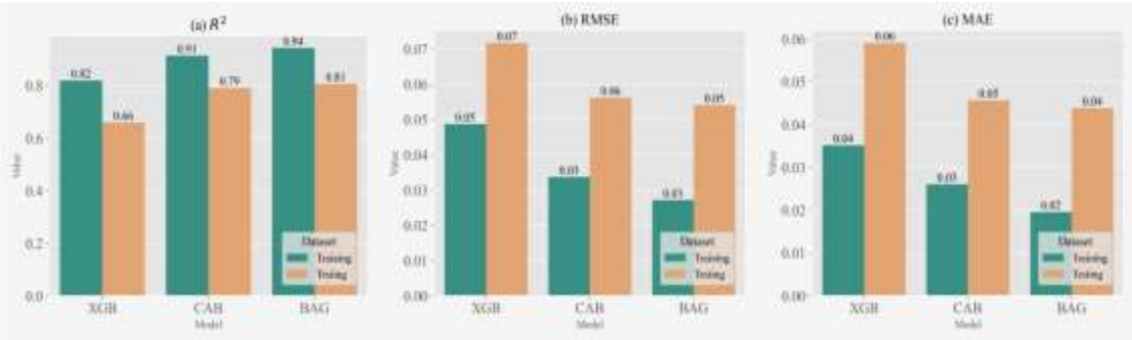

Figure 5. Bar charts of validation metrics: (a) R^2, (b) RMSE, and (c) MAE

Figure 6 uses Taylor diagrams to compare the performance of optimized XGB, CAB, and BAG models based on RMSE, MAE, and R² for training and testing datasets. Taylor diagrams visualize how closely predictions match actual MDD values, with points nearer the reference point indicating better performance. Figures 6a, b show RMSE for training and testing. BAG is closest to the reference (0.03 training, 0.05 testing), followed by CAB (0.03 training, 0.06 testing) and XGB (0.04 training, 0.07 testing), confirming BAG's superior precision. Figs 6c,d display MAE, with BAG nearest the reference (0.02 training, 0.04 testing), followed by CAB (0.03 training, 0.05 testing) and XGB (0.04 training, 0.06 testing), reinforcing BAG's ability to minimize prediction errors. Figures 6e, f illustrate R², with BAG closest to the reference (0.94 training, 0.81 testing), followed by CAB (0.91 training, 0.79 testing) and XGB (0.82 training, 0.66 testing). BAG's high R² values highlight its excellent fit and generalization. Overall, Fig. 6 confirms that BAG, followed by CAB, delivers the most accurate and reliable MDD predictions, with XGB showing reduced performance on unseen data. These findings emphasize the value of ensemble methods, particularly BAG, for optimizing soil compaction predictions in geotechnical engineering.

Taylor diagrams in Fig. 6 further validate model performance by jointly representing correlation, error magnitude, and variability. The BAG model appears closest to the reference point for all metrics, confirming its superior predictive accuracy and reliability. CAB also shows strong performance, while XGB remains comparatively less stable on the testing dataset.

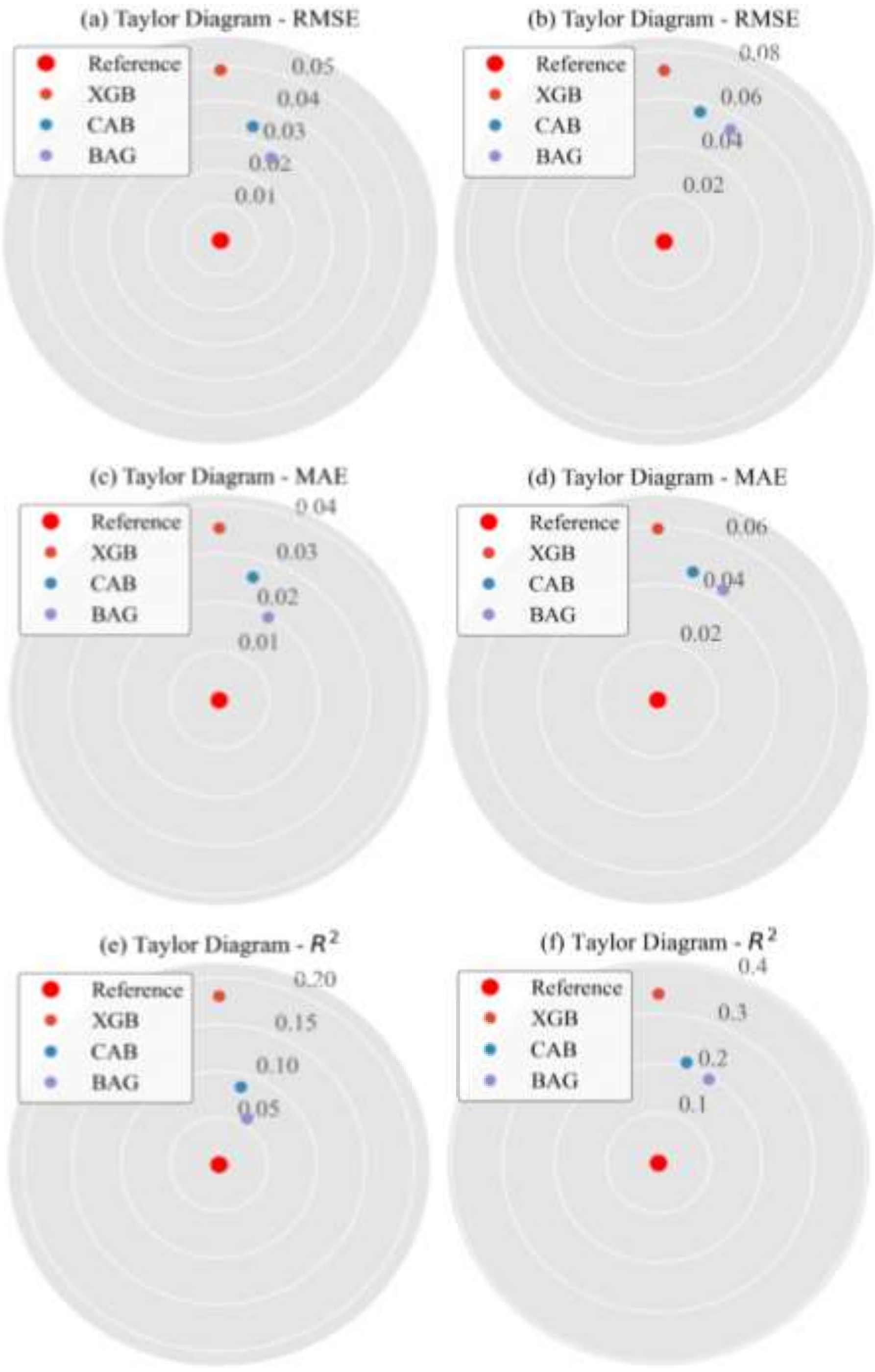

Figure 6. Taylor diagram analysis of model performance: (a) RMSE training, (b) RMSE testing, (c) MAE training, (d) MAE testing, (e) R^2 training, and (f) R^2 testing

In short, the validation and comparison of the models clearly demonstrate the effectiveness of ensemble learning, particularly BAG, when optimized via Grid

Search for predicting MDD in compacted soils. Among the three models evaluated, BAG consistently achieved the best performance on both the training and test datasets, compared with XGB and CAB. This can be attributed to its ability to reduce variance and its robustness to the natural heterogeneity and noise commonly present in geotechnical datasets (Breiman, 1996). By aggregating multiple base learners, BAG improves generalization and mitigates overfitting, further enhanced by systematic hyperparameter tuning via Grid Search (Ayalew et al., 2024).

In comparison with earlier studies (Tiwari et al., 2024; Günaydın, 2009), the optimized BAG model offers notable improvements. In this study, the BAG model achieved a test R^2 of 0.81, along with lower RMSE and MAE values, indicating superior predictive capability (Table 3). Earlier studies primarily relied on artificial neural networks or single machine learning models, often without systematic hyperparameter optimization. In comparison, the optimized BAG model in this study achieves a higher testing R^2 (0.81) and lower error metrics, demonstrating improved predictive capability. The improved performance can be attributed to three main factors: (i) the use of ensemble learning, which reduces variance and enhances generalization; (ii) systematic Grid Search optimization, which identifies optimal hyperparameter combinations; and (iii) the use of a real construction dataset representing practical soil conditions. These aspects collectively explain why the BAG and CAB models outperform many previously reported models.

Table 3. Comparison of model performance metrics for MDD prediction across studies

Works	Models	R^2 (Testing)	RMSE (Testing)	MAE (Testing)
Günaydın (2009)	ANN	0.75 (average)	-	-
Tiwari et al. (2024)	ANN	-	0.0522	-
	KNN	-	0.1366	-
This study	Optimized BAG	0.81	0.05	0.04
	Optimized CAB	0.79	0.06	0.05
	Optimized XGB	0.66	0.07	0.06

3.2. Partial Dependence Plot (PDP) and Physical Interpretation

Figure 7 presents a PDP illustrating the marginal effect of individual soil properties on the predicted MDD, while averaging out the effects of other variables. The best BAG model was selected for its robustness in capturing feature interactions and importance. Figure 7a shows that G has a strong positive effect on MDD, with MDD increasing by approximately 0.02 g/cm³ per 10% increase in G above 20%. This reflects gravel's role in enhancing soil packing and reducing voids, consistent with geotechnical principles. Figure 7b indicates that CS has a positive but less pronounced effect, with MDD rising steadily up to 30% CS, then plateauing. This suggests that coarse sand improves grain interlocking, but its impact diminishes at higher concentrations due to interactions with other fractions. Figure 7c reveals a non-linear trend for FS: MDD increases slightly up to 15% FS, then decreases by approximately 0.01 g/cm³ per 10% increase thereafter. Small amounts of fine sand fill voids effectively, but excessive amounts may disrupt particle rearrangement. Figure 7d for SC, 7f for LL, 7g for PL, and 7h for PI all show negative relationships with MDD. As SC, LL, PL, or PI increases, MDD decreases by approximately 0.03–0.05 g/cm³ per 10-unit increase. These properties, associated with higher soil plasticity (defined by Atterberg limits, which measure the water content boundaries of soil plasticity), increase cohesion and water retention, reducing compaction efficiency and achievable dry density. Figure 7e displays a non-linear relationship between O and MDD, with MDD peaking at approximately 10–12% moisture, then decreasing by 0.02 g/cm³ per 5% increase in O. This reflects the balance between sufficient moisture for particle lubrication and excessive moisture that reduces density in finer soils.

In summary, the PDP in Fig. 7 demonstrates that coarser fractions (G and CS) enhance MDD, while finer fractions and plasticity-related properties (SC, LL, PL, PI) reduce it. The non-linear effects of FS and O highlight the importance of optimizing soil composition. These findings align with soil mechanics principles and provide actionable insights for designing soil mixtures and compaction strategies to achieve optimal dry density in geotechnical engineering.

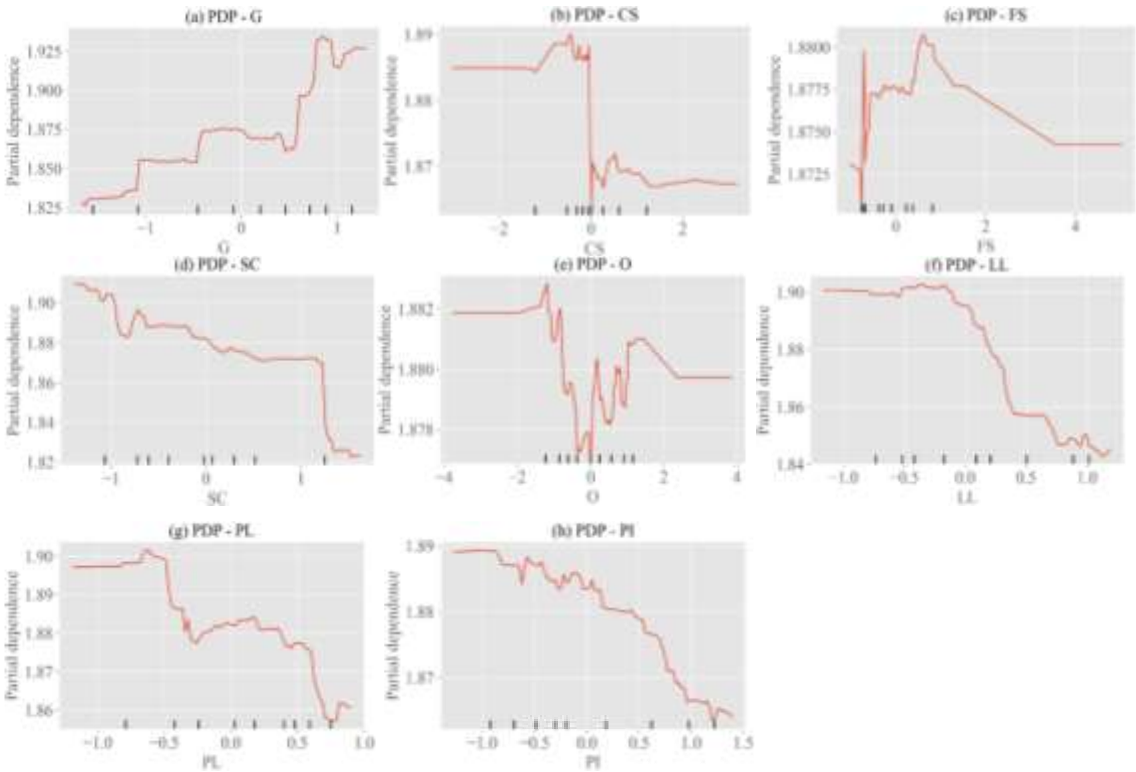

Figure 7. PDPs for input variables for the RF model:
(a) G, (b) CS, (c) FS, (d) SC, (e) O, (f) LL, (g) PL, and (h) PI

4. Conclusions

This study addressed the challenge of accurately predicting the MDD of compacted soils by applying advanced ensemble machine learning algorithms, XGB, CAB, and BAG, optimized through Grid Search hyperparameter tuning. These models were developed using a comprehensive dataset of key soil properties and evaluated using standard performance metrics, including R^2, RMSE, and MAE. Among the three models, the Grid Search-optimized BAG algorithm demonstrated superior predictive performance, outperforming XGB and CAB in both accuracy and generalization. Its predictions aligned closely with established soil mechanics principles, highlighting, for instance, the positive influence of coarse materials and the negative impact of fines and plasticity, thereby supporting the model's physical interpretability. These findings underscore the potential of ensemble learning, particularly BAG, as a practical and reliable alternative to traditional laboratory testing methods such as the Proctor compaction test.

While the results are promising, this study has certain limitations that open avenues for future research. The dataset, though diverse, could be expanded to include a wider range of soil types, compaction conditions, and geographic regions to enhance the robustness and generalizability of the models. Additionally, exploring more advanced feature engineering techniques and incorporating emerging algorithms may further improve predictive performance. Nevertheless, the systematic model development, rigorous optimization, and validation undertaken in this study establish a strong foundation for integrating machine learning into geotechnical engineering practice.

References

Ali H.F.H., Omer B., Mohammed A.S., Faraj R.H., 2024. Predicting the maximum dry density and optimum moisture content from soil index properties using efficient soft computing techniques. Neural Computing and Applications, 36(19), 11339–11369. https://doi.org/10.1007/s00521-024-09734-7.

Ashwini S., Arunkumar J., Prabu R.T., Singh N.H., Singh N.P., 2024. Diagnosis and multi-classification of lung diseases in CXR images using optimized deep convolutional neural network. Soft Computing, 28(7), 6219–6233. https://doi.org/10.1007/s00500-023-09480-3.

Ayalew A.A., Bitew G., Salau A.O., Tegegne M.L., Gupta S.K., Israr M., Tegegne T., Assegie T.A., 2024. Grid Search Hyperparameters Tuning with Supervised Machine Learning for Awngi Language Named Entity Recognition. In: 2024 Second International Conference Computational and Characterization Techniques in Engineering & Sciences (IC3TES). IEEE, 1–6. Doi: 10.1109/IC3TES62412.2024.10877504.

Breiman L., 1996. Bagging predictors. Machine learning, 24, 123–140. https://doi.org/10.1007/BF00058655.

Cai Y., Yuan Y., Zhou A., 2024. Predictive slope stability early warning model based on CatBoost. Scientific Reports, 14(1), 25727. https://doi.org/10.1038/s41598-024-77058-6.

Chen T., Guestrin C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 785–794. https://doi.org/10.1145/2939672.2939785.

Demir Z., 2024. Effects of vermicompost and salinity on proctor optimum water content, maximum dry bulk density and consistency of a sandy clay loam soil. Communications in Soil Science and Plant Analysis, 55(12), 1747–1767. https://doi.org/10.1080/00103624.2024.2328622.

Duc N.D., Nguyen M.D., Prakash I., Van H.N., Van Le H., Thai P.B., 2025. Prediction of safety factor for slope stability using machine learning models. Vietnam Journal of Earth Sciences, 47(2), 182–200. https://doi.org/10.15625/2615-9783/22196.

Ewusi-Wilson R., Yendaw J.A., Sebbeh-Newton S., Ike E., Ayeh F.J.F., 2025. Explainable Artificial Intelligence Estimation of Maximum Dry Density in Soil Compaction Based on Basic Soil Properties and Compaction Energy. Transportation Infrastructure Geotechnology, 12(2), 1–30. https://doi.org/10.1007/s40515-025-00552-5.

Friedman J.H., 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics, 29(5), 1189–1232. Doi: 10.1214/aos/1013203451.

Günaydın O., 2009. Estimation of soil compaction parameters by using statistical analyses and artificial neural networks. Environmental Geology, 57(1), 203–215. Doi: 10.1007/s00254-008-1300-6.

Hancock J.T., Khoshgoftaar T.M., 2020. CatBoost for big data: an interdisciplinary review. Journal of big data, 7(1), 94. https://doi.org/10.1186/s40537-020-00369-8.

Le T.-H., Nguyen H.-L., Pham C.-T., 2022. Artificial intelligence approach to predict the dynamic modulus of asphalt concrete mixtures. Journal of Science and Transport Technology, 22–30. https://doi.org/10.58845/jstt.utt.2022.en.2.2.22-30.

Li B., You Z., Ni K., Wang Y., 2024. Prediction of Soil Compaction parameters using machine learning models. Applied Sciences, 14(7), 2716. https://doi.org/10.3390/app14072716.

Ngo T.Q., Nguyen L.Q., Tran V.Q., 2022b. Predicting tensile strength of cemented paste backfill with aid of second order polynomial regression. Journal of Science and Transport Technology, 43–51. https://doi.org/10.58845/jstt.utt.2022.en.2.4.43-51.

Ngo T.-T.T., Pham H.T., Acosta J., Derrible S., 2022a. Predicting bike-sharing demand using random forest. Journal of Science and Transport Technology, 2(2), 13–21. https://doi.org/10.58845/jstt.utt.2022.en.2.2.13-21.

Nguyen D.D., Nguyen H.P., Vu D.Q., Prakash I., Pham B.T., 2023. Using GA-ANFIS machine learning model for forecasting the load bearing capacity of driven piles. Journal of Science and Transport Technology, 3(2), 26–33. https://doi.org/10.58845/jstt.utt.2023.en.3.2.26-33.

Nguyen D.D., Roussis P.C., Pham B.T., Ferentinou M., Mamou A., Vu D.Q., Bui Q.-A.T., Trong D.K., Asteris P.G., 2022. Bagging and multilayer perceptron hybrid intelligence models predicting the swelling potential of soil. Transportation Geotechnics, 36, 100797. https://doi.org/10.1016/j.trgeo.2022.100797.

Nhat V.H., Trinh P.T., Cam L.V., Dieu B.T., Van H.L., Prakash I., Anh N.N., Van H.N., Thanh N.D., Thao N.P., 2025. Mapping Cadmium Contamination Potential in Surface Soil for Civil Engineering Applications: A Comparative Study of Machine Learning and Deep Learning Models in the Gianh River Basin, Vietnam. Journal of Science and Transport Technology, 48–70. https://doi.org/10.58845/jstt.utt.2025.en.5.2.48-70.

Omer B., 2025. Machine learning techniques and multivariable mathematical models for predicting modified soil compaction parameters based on particle size and consistency limits. Modeling Earth Systems and Environment, 11(1), 1–46. https://doi.org/10.1007/s40808-024-02247-1.

Pal S., Hieu V.T., Nguyen D.D., Vu D.Q., Prakash I., 2024. Investigation of Support Vector Machines with Different Kernel Functions for Prediction of Compressive Strength of Concrete. Journal of Science and Transport Technology, 55–68. https://doi.org/10.58845/jstt.utt.2024.en.4.2.55-68.

Pham B.T., Amiri M., Nguyen M.D., Ngo T.Q., Nguyen K.T., Tran H.T., Vu H., Anh B.T.Q., Van L.H., Prakash I., 2021. Estimation of shear strength parameters of soil using Optimized Inference Intelligence System. Vietnam Journal of Earth Sciences, 43(2), 189–198. https://doi.org/10.15625/2615-9783/15926.

Prakash I., Kumar R., Nguyen T.-A., Vu P.T., 2022. Development of effective XGB model to predict the Axial Load Capacity of circular CFST columns. Journal of Science and Transport Technology, 26–42. https://doi.org/10.58845/jstt.utt.2022.en.2.4.26-42.

Prakash I., Nguyen D.D., Tuan N.T., Van P.T., Van H.L., 2024. Landslide susceptibility zoning: integrating multiple Intelligent models with SHAP Analysis. Journal of Science and Transport Technology, 23–41. https://doi.org/10.58845/jstt.utt.2024.en.4.1.23-41.

Prakash I., Phan T.-N., 2023. Estimating the compressive strength of self-compacting concrete with fiber using an extreme gradient boosting model. Journal of Science and Transport Technology, 3(1), 12–25. https://doi.org/10.58845/jstt.utt.2023.en.3.1.12-25.

Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A., 2018. CatBoost: unbiased boosting with categorical features. Advances in Neural Information Processing Systems, 31.

Qiu Y., Zhou J., Khandelwal M., Yang H., Yang P., Li C., 2022. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Engineering with Computers, 38(5), 4145–4162. https://doi.org/10.1007/s00366-021-01393-9.

Taylor K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of geophysical research: atmospheres, 106(D7), 7183–7192. https://doi.org/10.1029/2000JD900719.

Thai P.B., Nguyen D.D., Thi Q.-A.B., Nguyen M.D., Vu T.T., Prakash I., 2022. Estimation of load-bearing capacity of bored piles using machine learning models. Vietnam Journal of Earth Sciences, 44(4), 470–480. https://doi.org/10.15625/2615-9783/17177.

Tiwari L.B., Burman A., Samui P., 2024. A comparative study of soft computing paradigms for modelling soil compaction parameters. Transportation Infrastructure Geotechnology, 1–19. https://doi.org/10.1007/s40515-024-00436-0.

Verma G., Kumar B., Ransinchung R.N.G., 2024. Particle swarm optimization-based machine learning algorithms for developing the modified proctor compaction parameter prediction software. Transportation Infrastructure Geotechnology, 11(4), 1492–1519. https://doi.org/10.1007/s40515-023-00326-x.

Zhao L., Guan G.D., 2024. Maximum dry density estimation of stabilized soil via machine learning techniques in individual and hybrid approaches. Journal of Ambient Intelligence and Humanized Computing, 15(11), 3831–3846. https://doi.org/10.1007/s12652-024-04860-5.

Zhao Q., Liu K., Xiong C., Deng X., Yang S., 2024a. Estimating the maximum dry density of soil via least square support vector regression individual and hybrid forms. Indian Geotechnical Journal, 1–13. https://doi.org/10.1007/s40098-024-00952-3.

Zhao Y., Zhang W., Liu X., 2024b. Grid search with a weighted error function: Hyper-parameter optimization for financial time series forecasting. Applied Soft Computing, 154, 111362. https://doi.org/10.1016/j.asoc.2024.111362.