

Article

Estimation of Rock Shear Strength Parameters from Geophysical Indicators Using an Equilibrium-Optimized Multilayer Perceptron

Baohua Liu ^{1,2,3}, Ze Xiang ^{1,2,*} and Hang Lin ^{3,*} 

¹ School of Civil Architectural Engineering, Shaoyang University, Shaoyang 422000, China; liubaohuasyxy@163.com

² Hunan Engineering Research Center of Bamboo Fiber Building Materials, Shaoyang 422000, China

³ School of Resources and Safety Engineering, Central South University, Changsha 410083, China

* Correspondence: xiangz@hnu.edu.cn (Z.X.); linhangabc@126.com or hanglin@csu.edu.cn (H.L.)

Abstract

Accurate and scalable estimation of rock shear strength parameters is essential for remote-sensing-supported geological hazard assessment, slope stability evaluation, and engineering geological mapping. However, determining cohesion and internal friction angle requires multiple triaxial tests under different confining pressures, which are time-consuming, costly, and difficult to apply widely. To support remote-sensing-oriented geoscience and civil engineering applications, this study develops a hybrid machine learning framework for estimating cohesion and internal friction angle from geophysical and mechanical indicators. A cross-source database was compiled from published rock records collected from the Jinchuan mining area in China and the Luhri area in India. After completeness screening and unit harmonization, 213 mixed-lithology cases were retained for modeling, with P-wave velocity, density, uniaxial compressive strength, and tensile strength used as input variables. An Equilibrium Optimizer was coupled with a multilayer perceptron to optimize the network weights and biases, and model performance was evaluated using five-fold cross-validation, independent testing, repeated runs, and comparisons with conventional MLP and several typical machine learning models. The proposed EO-MLP model achieved high prediction accuracy, with test-set coefficient of determination values of 0.946 for internal friction angle and 0.983 for cohesion and corresponding RMSE values of 1.072 and 0.671, respectively. Robust scaler normalization produced the best performance among the three tested normalization strategies. SHapley Additive exPlanations analysis indicated that density was the dominant predictor of cohesion, whereas uniaxial compressive strength and P-wave velocity made the largest contributions to internal friction angle prediction. The proposed framework provides an indirect data-driven tool for estimating shear strength parameters and can complement engineering-geological investigation and rock engineering design.



Received: 17 June 2026

Revised: 14 July 2026

Accepted: 16 July 2026

Published: 30 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: rock shear strength; equilibrium optimizer; geophysical parameters; machine learning; data-driven prediction

1. Introduction

Rock materials often serve as critical load-bearing components in major infrastructure systems, and their mechanical responses play a pivotal part in maintaining the operational

safety and long-term stability of structures [1,2]. Typically, rock masses are not only inherently flawed with internal defects such as microcracks and voids but are also frequently subjected to various external influences, including multi-source disturbances and environmental factors [3,4]. As a result, they exhibit pronounced anisotropy and heterogeneity. Under extreme loading conditions or in high-risk environments, including deep tunnels, steep rock slopes, or geological repositories for nuclear waste, even localized instability within the rock mass can rapidly evolve into through-going failure under stress, potentially triggering large-scale structural collapse [5]. Cohesion (c) and internal friction angle (φ) are two fundamental indices used to characterize the shear strength parameters [6]. These parameters are widely employed in slope stability analysis, underground excavation design, and geohazard forecasting [7]. φ reflects resistance mobilized along the shear plane, and c represents the bonding strength between rock particles. Inaccurate estimation of these two parameters may lead to erroneous stability assessments and increased risks [8]. Therefore, developing predictive models that reliably estimate the φ and c is essential for improving the reliability of geotechnical assessments and supporting structural optimization in rock engineering applications [9,10].

Over the past few decades, the shear characteristics of rock materials have been extensively investigated [11]. As shown in Figure 1, φ and c are commonly determined from triaxial compression tests using the Mohr–Coulomb failure envelope. c corresponds to the intercept on the envelope, while φ is derived from the slope of the envelope. Although this method is widely regarded as a standard approach, it faces significant practical limitations. In many engineering scenarios, difficulties in specimen preparation, limitations in testing precision, and challenges in reproducing in-situ stress conditions lead to high costs and limited data representativeness [12]. To overcome these limitations, numerous empirical approaches have been proposed to estimate φ and c from readily available rock properties. For example, Edimann et al. [13] assessed the feasibility of estimating φ and c using uniaxial compressive strength (UCS), porosity, and other properties in the absence of triaxial test results. Hajdarwish [14] investigated mudrocks and employed both bivariate and multivariate regression analyses to quantify the relationships between φ/c and key index properties, including slake durability index, specific gravity, absorption, void ratio, and liquid limit. Recently, Karakul and Ulusay [15] developed a predictive model for the φ based on theoretical tensile strength and UCS, demonstrating improved predictive performance over traditional approaches. However, despite their simplicity and computational efficiency, such empirical models are often limited to specific lithologies and fail to capture the nonlinear behavior induced by material heterogeneity under complex loading conditions. Therefore, developing alternative predictive approaches that are more robust, generalizable, and capable of capturing complex mechanical behavior remains an important research direction.

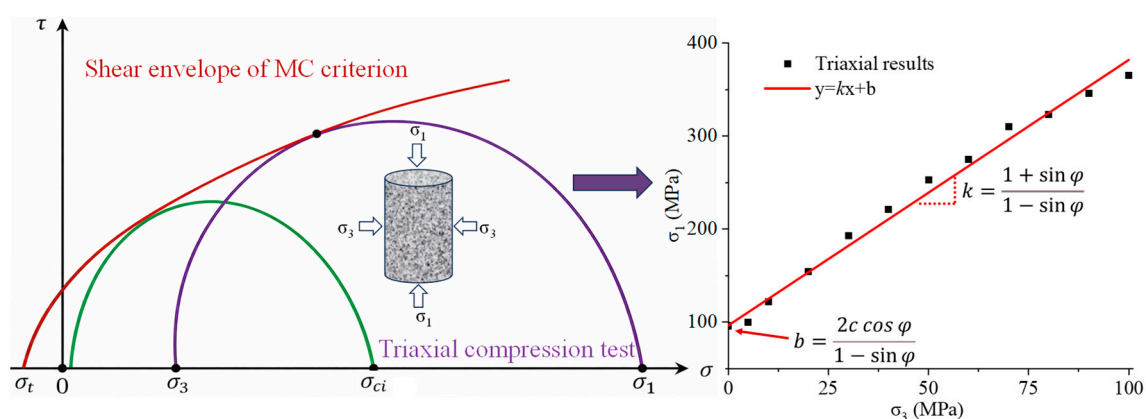


Figure 1. Regression-based determination of φ and c from triaxial test data.

Machine learning (ML) methods offer strong adaptability and generalization capability for engineering prediction tasks involving high-dimensional, nonlinear, and strongly coupled geotechnical parameters [16,17]. Unlike traditional techniques, machine learning methods can efficiently capture complex dependencies between mechanical characteristics and multiple factors, enabling fast prediction of key parameters [18,19]. For example, Shen and Jimenez [20] proposed a genetic programming-based model for estimating φ and c from UCS and tensile strength in the absence of triaxial test data. The model exhibited high prediction accuracy when compared with both traditional regression models and triaxial test results, offering an efficient and cost-effective solution for early-stage projects or cases with limited experimental conditions. In another study, Shahani et al. [21] applied various advanced ML models including lasso regression, ridge regression, and decision trees to develop predictive models for c and φ . Their study demonstrated the feasibility of the ML approach; however, the chosen base learners were not subjected to systematic hyperparameter optimization. More recently, Li et al. [22] constructed ML models with five ensemble-tree algorithms to predict c and φ , and employed Bayesian optimization to tune model hyperparameters, thereby improving predictive accuracy. Despite these advances, further improvements in optimization efficiency and robustness remain desirable. In this context, hybrid predictive frameworks that combine metaheuristic algorithms with machine learning models have attracted increasing attention because of their strong capability to capture complex nonlinear relationships and to improve predictive accuracy and generalization across different datasets.

Most existing studies rely on single-lithology or relatively homogeneous datasets with random splits, which may overestimate model generalization in engineering scenarios. To address this limitation, a mixed-lithology, cross-source dataset covering diverse lithologies and multiple experimental protocols was assembled in this study. Based on this database, an efficient and robust predictive framework was developed by integrating the Equilibrium Optimizer (EO) with a Multilayer Perceptron (MLP). The proposed framework was evaluated to assess its predictive accuracy and generalization capability. The results provide a reliable data-driven tool for parameter estimation in civil engineering and offer methodological support for the further development of intelligent prediction techniques.

2. Dataset and Preprocessing

2.1. Data Gathering

In practice, the use of open and representative datasets improves the transparency and scientific value of data-driven studies [23,24]. Accordingly, this study compiled a comprehensive database consisting of 611 representative cases from the Jinchuan mining area in China and the Luhri area in India, based on previously published studies [22,25]. As illustrated in Figure 2, the database encompasses a variety of rock types, including quartzite, limestone, quartz mica schist, and slate. The inclusion of samples from two geographically distinct regions increases the variability of both the input variables and the target parameters. This provides a broader basis for examining the relationships among rock physical properties, mechanical characteristics, φ , and c under different lithological conditions. The studies [22,25] reported extensive laboratory experiments such as uniaxial compression tests, Brazilian tensile tests, triaxial compression tests, and velocity measurements used to determine the corresponding mechanical characteristics. To prepare this database, records were screened for completeness: among the 597 cases reported in Ref. [25], 199 met the six-variable (P-wave velocity, density, UCS, tensile strength, φ and c) requirement and were retained; an additional 14 complete cases were incorporated from Ref. [22]. After merging and unit harmonization, the final modeling database comprised 213 cases.

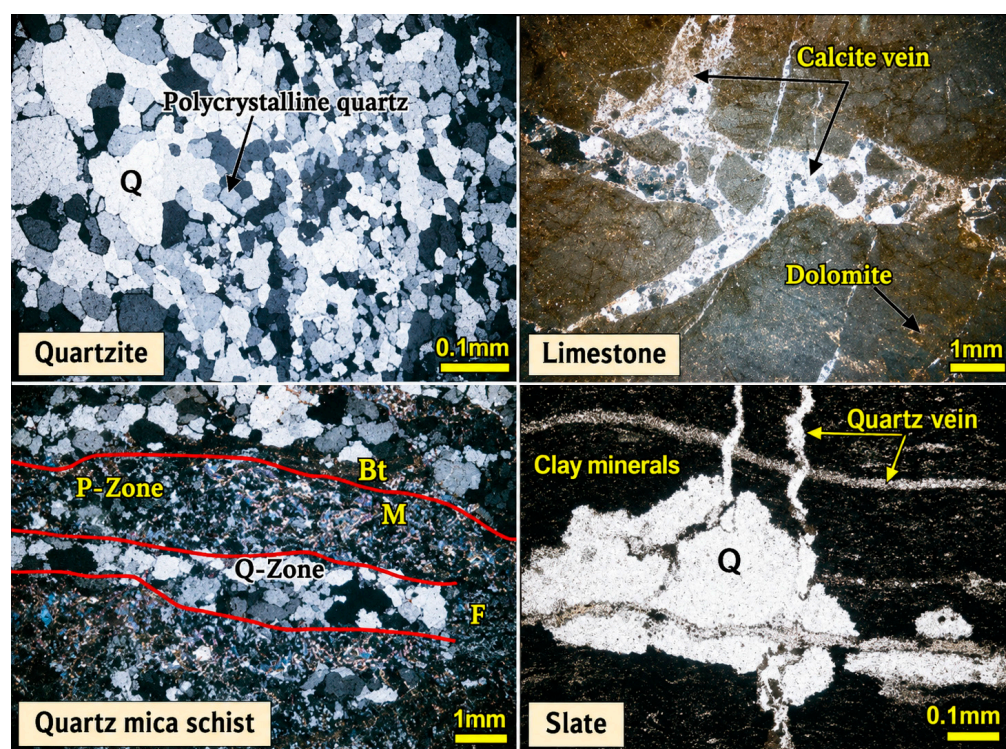


Figure 2. Representative microphotographs of specimens used for modeling [25].

2.2. Data Analysis and Normalization

In ML modeling, the representativeness and quality of the database play a critical role in model performance [26,27]. Prior to model development, a statistical analysis was performed on the compiled database to examine the distribution characteristics of each variable and to support subsequent data preprocessing. Table 1 summarizes the key statistical descriptors of all variables, including the coefficient of variation (CV), Median (Med) and mean absolute deviation (Mad). They help identify potential differences in scale and distribution among the variables before model training. As shown in Table 1, both the input and output variables exhibit nonzero skewness, indicating asymmetric data distributions. The negative kurtosis values further suggest that the distributions are relatively flat and dispersed.

Table 1. Detailed statistics of the database.

	Number	Standard Deviation	Skewness	Kurtosis	CV	Mad	Med
P-wave velocity	213	1095.590	−0.073	−0.908	0.250	921.597	4259.38
Density	213	0.101	−0.130	−0.163	0.038	0.081	2.67
UCS	213	56.584	0.346	−0.884	0.461	46.920	116.6
Tensile strength	213	7.316	0.240	−0.820	0.481	6.000	14.9
c	213	5.286	0.485	−0.262	0.281	4.263	18.39
φ	213	4.348	−0.045	−0.868	0.128	3.673	34.15

Appropriate data visualization facilitates a deeper understanding of data characteristics and distributions, thereby helping identify potential issues and improve the performance of ML models [28,29]. Compared with conventional box plots, violin plots are more suitable for comparative analysis across multiple variables, as they allow intuitive identification of differences in distribution patterns. Figure 3 illustrates the distribution of all variables in the database using violin plots. Both the input and output variables cover relatively wide value ranges and exhibit generally balanced and approximately symmetric

distributions, with no obvious outliers. These observations suggest that this database is of high quality and provides a reliable basis for subsequent model training and generalization.

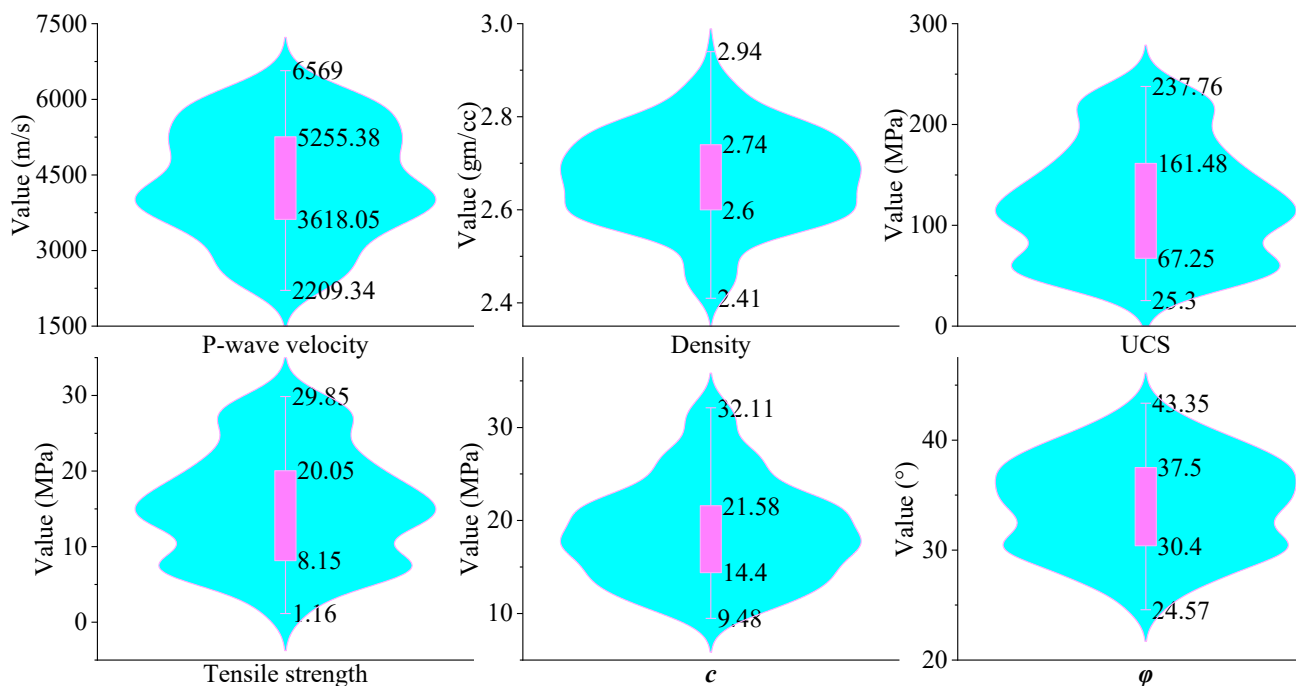


Figure 3. Visualization of the input and output features.

To improve generalization and curb overfitting, the full database was not used for fitting. Instead, following widely accepted practices in the literatures [30,31], this database was first randomly divided into a training set and an independent test set at a ratio of 70:30. This strategy helps assess the predictive capability on unseen samples and ensures its robustness and reliability in applications. To further enhance the stability of model training and ensure comparability among input variables, data normalization was performed prior to model development. Given the differences in dimensions and wide variability of feature values, directly feeding the original data into the model may adversely affect training convergence and overall performance. Accordingly, this study employed the robust scaler normalization technique [32]. The normalization process is given by Equation (1). For each variable, the median and interquartile range, defined as the difference between the 75th and 25th percentiles, were calculated using the training set only. The fitted scaling parameters were subsequently applied to both the training and test sets. This method can eliminate the magnitude difference of the original data, avoid accidental phenomena and speed up the calculation speed of the ML model.

$$\bar{Y} = \frac{Y - Med}{IQR} \quad (1)$$

where Y and $\bar{Y}$ denote the original and scaled data. IQR quantifies data dispersion as the difference between the 75th and 25th percentiles. Med denotes the median value.

2.3. Cross-Validation and Performance Evaluation

To enhance the accuracy and robustness, k-fold cross-validation (k-f CV) method was adopted as the primary technique for assessing model performance. As a widely adopted resampling method in machine learning, k-f CV method partitions the original database into k approximately equal and mutually exclusive subsets [33], as shown in Figure 4. At each iteration, one subset is reserved for testing, with the others used for training. The

process cycled through all folds so that every subset served once as the testing subset. Performance measures from the k runs were then averaged to obtain a stable estimate of generalization. In comparison with a single train-test split, k -f CV method offers lower evaluation bias and variance, effectively mitigating performance fluctuations caused by different data splits. It is worth noting that for data with relatively small sample sizes or imbalanced class distributions, this method maximizes data utilization and improves the reliability of performance assessment by stabilizing model predictions [34]. In this study, 5-f CV method was selected to measure the predictive capability on different data subsets.

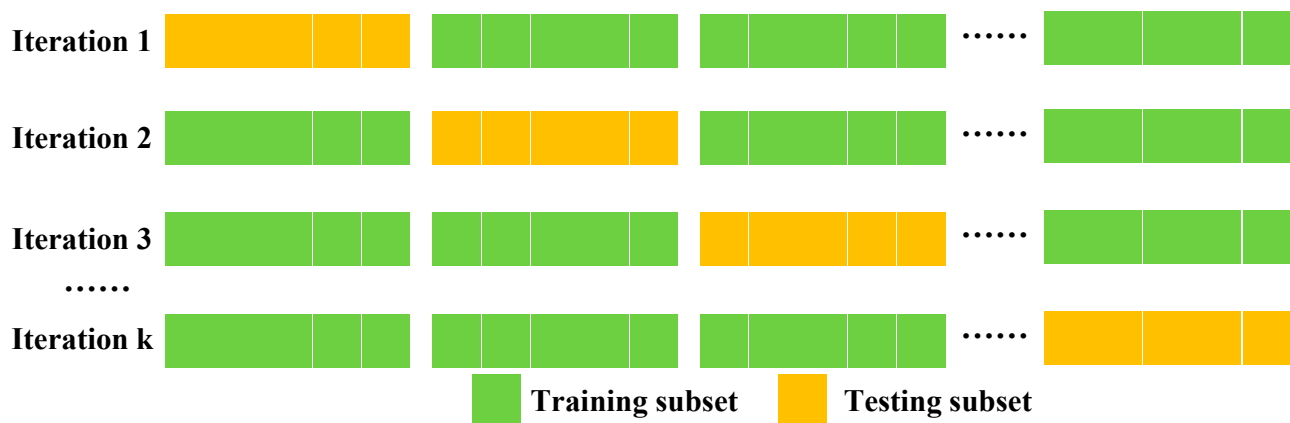


Figure 4. Diagram of k -f CV method.

Prior to evaluating the predictive capability of ML models, it is essential to define the performance metrics explicitly [35]. The choice of appropriate performance metrics directly influences the reliability of model assessment. To comprehensively evaluate model performance, this study adopts three widely used performance metrics: R^2 , MAE, and RMSE, as defined in Equations (2)–(4) [36]. These metrics characterize model performance from different perspectives and are commonly employed in ML models for their complementary ability to reflect prediction accuracy and error behavior. Since different models may yield varying prediction outcomes on the same data, these metrics provide an objective basis for identifying the most accurate model. In general, a higher R^2 value (approaching 1) and lower MAE and RMSE values (approaching 0) indicate superior predictive performance [37,38].

$$R^2 = 1 - \frac{\sum_{i=1}^M (y_i^{mea} - y_i^{pre})^2}{\sum_{i=1}^M (y_i^{mea} - \bar{y})^2} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^M |y_i^{mea} - y_i^{pre}|}{M} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^M (y_i^{mea} - y_i^{pre})^2}{M}} \quad (4)$$

where y_i^{mea} and y_i^{pre} refer to the observed and predicted values, respectively, M indicates the sample size, and $\bar{y}$ is the average of the observations.

3. Methodology

3.1. Multilayer Perceptron

MLP adopts a feedforward neural architecture that mimics the connectivity and information transmission mechanisms of biological neurons in the human brain [39]. As one of the fundamental structures in artificial neural networks, the MLP has garnered widespread attention in engineering applications due to its powerful capability to approximate complex nonlinear functions. In geotechnical engineering, MLP models have been successfully employed in a variety of tasks, such as mechanical parameter prediction and slope stability evaluation. A comprehensive overview of the fundamental principles and practical implementations of MLP models can be found in existing reviews [40,41].

As shown in Figure 5, the MLP comprises input, hidden, and output layers arranged in a feedforward architecture, with the input and output neuron numbers defined by the feature and target dimensions, respectively. By contrast, the network depth, the number of neurons, and the activation functions are tunable hyperparameters that should be selected according to the complexity of the prediction task. Common activation functions such as ReLU, tanh, and sigmoid are defined in Equations (5)–(7). These architectural and functional settings jointly determine the learning capacity and generalization performance of the MLP. In this study, separate MLP models were developed for cohesion and internal friction angle. Both models adopted a fixed 4–16–8–1 architecture, with ReLU activation functions in the two hidden layers and a linear output layer. These architectural settings were specified before EO optimization and were not included in the EO search space.

$$f(x) = \max(0, x) \quad (5)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (6)$$

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (7)$$

where x denotes the pre-activation input, and $f(x)$ represents the output after applying the activation function.

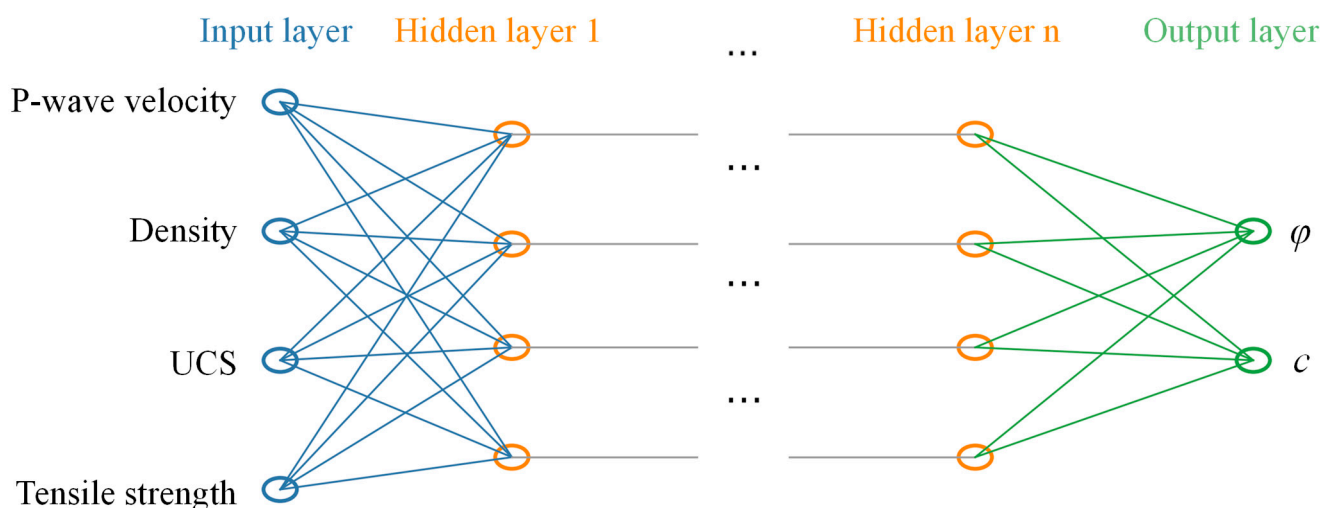


Figure 5. Schematic diagram of MLP.

3.2. Hybrid EO-MLP Model

Although the MLP demonstrates strong capabilities in approximating nonlinear functions and modeling complex dynamic systems, it still suffers from several inherent limitations. These limitations include initialization-dependent performance, premature conver-

gence to local optima, and difficulty in selecting an appropriate learning rate [42]. Such drawbacks are primarily attributed to the high dependence and uncertainty associated with hyperparameter configuration. Therefore, efficiently and reliably determining the hyperparameters of MLP models has become one of the key issues in improving their generalization ability and predictive performance.

In recent years, the equilibrium optimizer (EO) has shown promising performance in solving complex optimization problems [43]. Unlike conventional stochastic methods, EO mimics the dynamic mass balance behavior of substances in a control volume, where particles iteratively adjust their concentrations toward an equilibrium state. In the context of EO, the optimal solution is conceptualized as the desired equilibrium condition. In EO, each candidate solution is assigned a concentration value that is iteratively updated according to mass transfer and dynamic equilibrium principles. A distinctive feature of EO is the incorporation of an equilibrium pool, which stores several elite solutions to guide the search process and enhance convergence stability. EO effectively balances global exploration and local exploitation through an adaptive search mechanism. By dynamically adjusting the equilibrium factors and perturbation rates, the algorithm maintains population diversity while accelerating convergence toward high-quality solutions. Owing to its structural simplicity and robust global search capability, EO has demonstrated its applicability across diverse fields, including numerical optimization, parameter estimation, and the weight optimization of hybrid ML models [44].

As illustrated in Figure 6, EO updates the concentration of each search agent under the guidance of candidate solutions selected from the equilibrium pool. In the figure, C denotes the concentration vector of the current candidate solution, while $C_{eq(1)}$, $C_{eq(2)}$, $C_{eq(3)}$, and $C_{eq(4)}$ represent the four best-so-far equilibrium candidates. $C_{eq(ave)}$ is the arithmetic mean of these four candidates, and the five vectors together constitute the equilibrium pool. During each iteration, one member of the equilibrium pool is selected to guide the update of the current candidate. The arrows in Figure 6 indicate the corresponding update directions, whereas the dashed circles represent possible updated positions of the candidate solutions. In the early stages of optimization, the equilibrium candidates are widely dispersed throughout the search space, and the exponential term produces relatively large fluctuations. This facilitates global exploration by enabling the search agents to traverse different regions of the domain. As the algorithm progresses, the equilibrium candidates tend to cluster around the global optimum, and the exponential term generates smaller perturbations, thereby promoting local exploitation through finer step sizes. This dynamic updating mechanism balances global exploration and local exploitation and contributes to the convergence of the algorithm. The same updating principle can be extended to high-dimensional search spaces, where the concentration vector is adjusted within an n -dimensional hyperspace. Further details of EO can be found in Refs. [45,46].

To reduce the sensitivity of the conventional MLP to its initial parameters, EO was used to optimize the connection weights and biases of the fixed network architecture. For each prediction target, the 225 trainable weights and biases were encoded as a 225-dimensional candidate vector. The network architecture and activation functions remained unchanged throughout the EO optimization process. Given the MLP's powerful nonlinear approximation ability and EO's superior convergence and balance in global optimization, this paper introduces a hybrid EO-MLP framework to improve the predictive accuracy and generalization capability for φ and c . The existing literature appears to lack a systematic evaluation of combining the equilibrium optimizer with MLP for the prediction of φ and c . This research gap provides both theoretical motivation and practical potential for the integration of EO into this context, while also offering a new direction for the application of intelligent optimization algorithms in geotechnical modeling. The workflow of the pro-

posed hybrid EO-MLP model is visually summarized in Figure 7. The workflow integrates data preprocessing, parameter optimization, model training, and performance evaluation within a unified procedure.

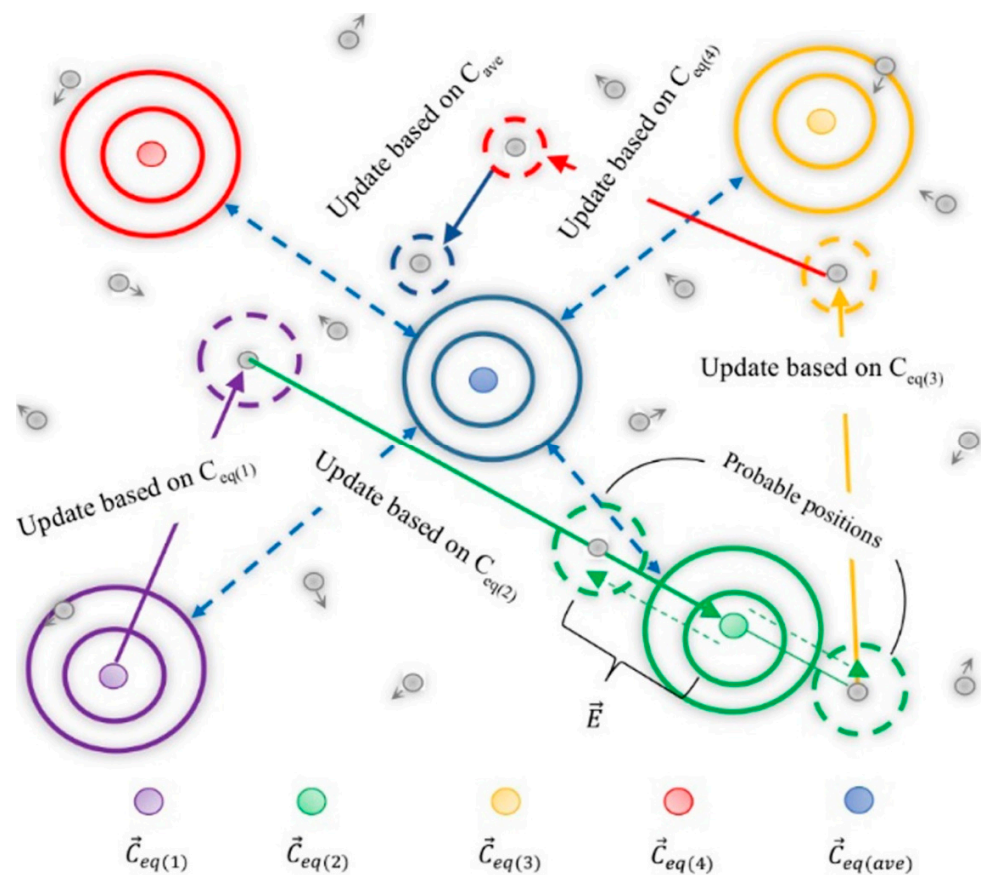


Figure 6. Schematic diagram of EO [43].

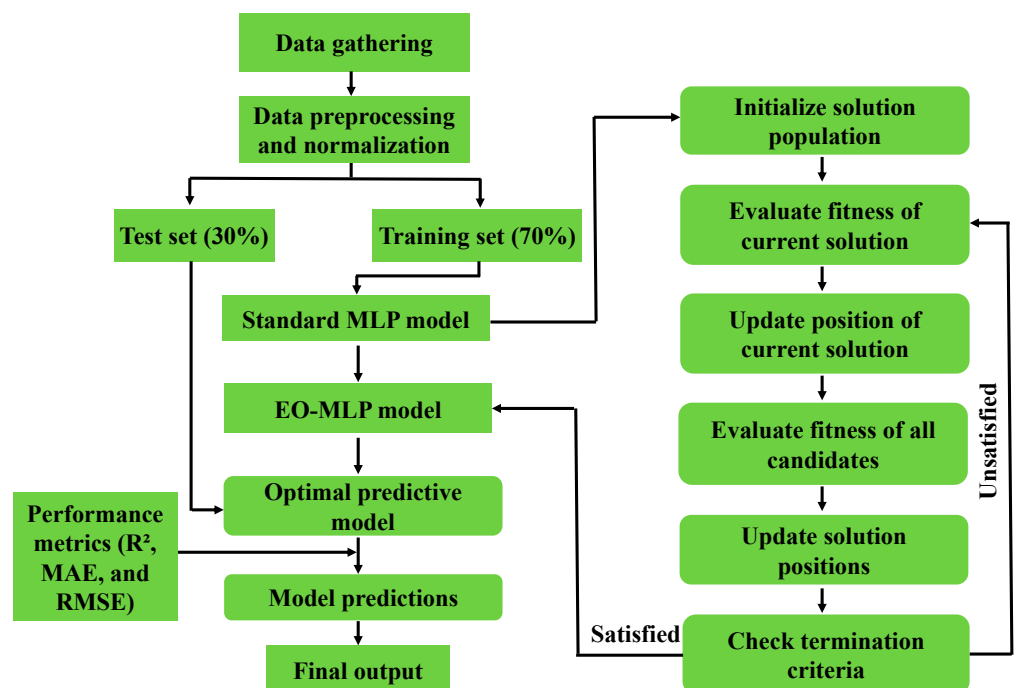


Figure 7. Schematic illustration of the EO-MLP model.

4. Results and Analysis

4.1. Model Results

The predictive performance for φ and c was comprehensively evaluated by comparing the EO-MLP model with a conventional MLP model. Figure 8 compares predicted and measured results during model calibration and validation. The conventional MLP model was trained solely using the standard backpropagation algorithm without hyperparameter optimization. In contrast, the EO-MLP model employed EO to optimize the connection weights and biases of the fixed MLP architecture. The number of hidden layers, neurons, and activation functions remained unchanged during optimization. Both models were independently trained and tested over 40 independent runs, with the best-performing result selected as the final output. Therefore, the reported metrics represent the best observed performance among the 40 runs rather than their average performance.

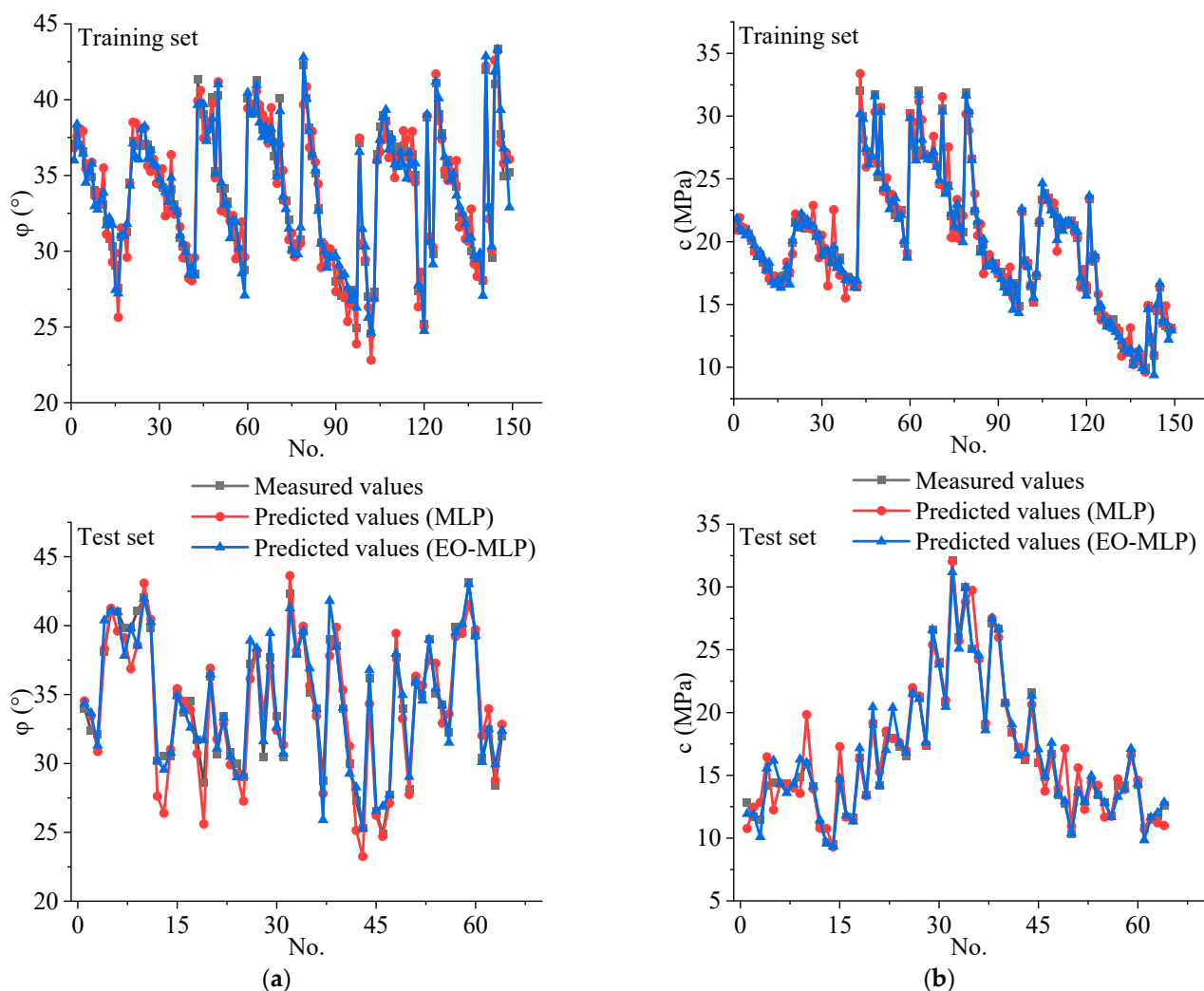


Figure 8. Comparisons between the measured and predicted results, (a) φ , (b) c . No. denotes the sequential sample number within the training or test set.

The EO-MLP model exhibits superior performance across the entire sample range (see Figure 8). Notably, in regions characterized by significant fluctuations or local extremes in φ and c , the EO-MLP model achieves closer agreement between predicted and experimental results than the conventional MLP model. To quantitatively assess model performance, Table 2 reports performance metrics including R^2 , MAE, and RMSE. The results show that the EO-MLP model achieved a test R^2 of 0.946 and RMSE of 1.072 in predicting φ , which

is significantly better than the corresponding values of 0.909 and 1.391 obtained by the MLP model. For c , EO-MLP improved R^2 from 0.941 to 0.983 and reduced RMSE from 1.249 to 0.671 relative to the MLP model. The same improvement trend was observed for both target variables, indicating that the benefit of EO optimization was not limited to a single prediction task. The reduction in MAE and RMSE also shows that the improvement was reflected not only in the goodness of fit but also in the magnitude of the prediction errors. These results suggest that the EO-based optimization enabled the MLP to capture the nonlinear relationships between the input variables and the two shear-strength parameters more effectively. The training-set metrics were consistently better than the corresponding test-set metrics for both models. This difference is commonly observed because model parameters are directly calibrated using the training samples, whereas the test set is used to evaluate predictive performance on previously unseen data [47].

Table 2. Comparative evaluation of predictive models.

		φ			c		
		R^2	MAE	RMSE	R^2	MAE	RMSE
EO-MLP	Training set	0.966	0.588	0.770	0.991	0.363	0.468
	Test set	0.946	0.692	1.072	0.983	0.434	0.671
MLP	Training set	0.937	0.850	1.056	0.969	0.542	0.876
	Test set	0.909	1.129	1.391	0.941	0.726	1.249

Overall, the EO-MLP model not only exhibits significantly improved accuracy compared to the MLP model but also demonstrates enhanced robustness against overfitting, underscoring its effectiveness and applicability in predicting shear strength parameters. By leveraging its capability for automatic parameter optimization, the model offers an efficient alternative for rapid acquisition of shear strength parameters in rock structure design. In contrast to traditional approaches that rely heavily on extensive physical testing and expert judgment, this model substantially reduces experimental costs and technical barriers, providing a promising tool for rapid prediction and intelligent decision-making in practice.

4.2. Comparison with Different Data Preprocessing Methods

In machine-learning modeling, data scaling is an important part of the preprocessing procedure, exerting a direct influence on model performance and convergence efficiency [48,49]. Raw datasets often exhibit inconsistencies in units, uneven distributions, and large numerical ranges. Without appropriate normalization, the model may become insufficiently sensitive to certain features, thereby impairing its overall predictive accuracy [50]. Normalization transforms input features into comparable or standardized numerical ranges, effectively eliminating scale discrepancies among physical quantities and enhancing inter-feature comparability [51]. Therefore, applying an appropriate normalization scheme before training improves learning from heterogeneous inputs and strengthens numerical stability and generalization. Three normalization strategies were considered: robust scaling, minimum-maximum normalization method, and Z-score normalization. The first reduces the influence of outliers using the median and interquartile range, the second maps variables into a predefined interval through linear transformation, and the third rescales variables to zero mean and unit variance. These methods have been shown in previous studies to exhibit good robustness and adaptability [52]. This section provides a systematic comparison and analysis of these three normalization techniques. The formulation for robust scaling is provided in Equation (1), while the corresponding expressions

for minimum-maximum normalization method and Z-score normalization are given in Equation (8) and (9), respectively.

$$\bar{Y} = \frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \quad (8)$$

$$\bar{Y} = \frac{Y - \mu}{\sigma} \quad (9)$$

where σ denotes the standard deviation, while μ represents the arithmetic mean.

Except for the use of different normalization techniques, all other modeling and training procedures were kept consistent. As illustrated in Figure 9, the performance of three normalization methods was compared in predicting φ and c . Among them, robust scaler consistently yielded superior performance, achieving the lowest MAE and RMSE values across both targets (e.g., 0.434 and 0.671 for c ; 0.692 and 1.072 for φ). Moreover, the robust scaler yielded the highest R^2 values for both prediction tasks, reaching 0.946 and 0.983 for φ and c , respectively, which indicates stronger generalization and predictive capability. Therefore, it can be concluded that applying robust scaler in the EO-MLP model leads to the most accurate and reliable prediction results.

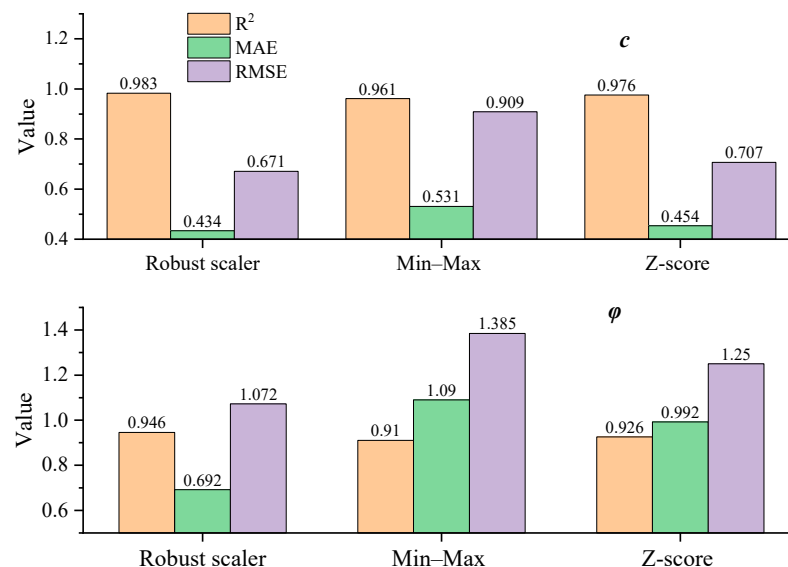


Figure 9. Comparison of performance metrics under different data preprocessing methods.

4.3. Comparison with Previous Studies

For different ML model outputs of prediction results, evaluation indices in Equations (2)–(4) changes accordingly. Comparing these indices allows for determining the most accurate predictive model [53]. To validate the generalization and predictive performance, it was systematically compared with four typical ML models, namely support vector machine (SVM), ridge regression, lasso regression, and decision tree (DT). These models employed the same data constructed in Section 2, ensuring that the comparison was based on a consistent and comprehensive database [21]. Moreover, all models employed the same data partitioning strategy as the proposed model, with the same 70:30 training-to-test ratio, and adopted five-fold cross-validation to ensure fairness and consistency in evaluation. Given that different ML models yield varying prediction results, the corresponding statistical indicators defined in Equations (2)–(4) also vary accordingly. Therefore, comparative analysis of these metrics (RMSE, MAE, and R^2) can effectively determine the optimal predictive model. The performance of each model in predicting φ and c is summarized in Figure 10. The proposed EO-MLP model delivers the best overall performance, with

superior metric values and the highest R^2 among all evaluated models. For φ , the EO-MLP model outperforms SVM and DT, showcasing its superior capability to capture complex nonlinear relationships and interactions. The larger differences in φ prediction, mainly reflect the different nonlinear representation capacities of the models. Linear models tend to underfit the coupled feature–target relationships, whereas EO-MLP model combines nonlinear mapping with EO-based initialization.

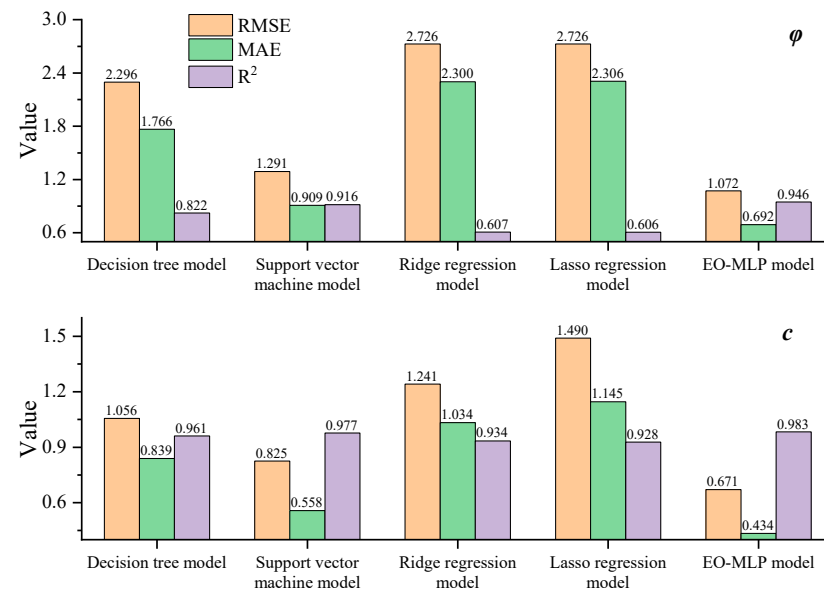


Figure 10. Performance evaluation of different ML models.

In addition to the comparison presented in Figure 10, a further benchmarking analysis was performed to evaluate the EO-MLP model with several representative recent models for predicting φ and c , including those proposed by Lei et al. [54], Li et al. [22], and Cheng et al. [55]. The corresponding test-set performance metrics, namely R^2 and MAE, are summarized in Table 3. For these previous studies, the reported values were taken directly from the original publications. As shown in Table 3, the EO-MLP model exhibits slightly better average performance for both φ and c . However, this advantage should be interpreted with caution, as the reported differences may be affected by inconsistencies in test sets and modeling settings, such as training ratios, input variables, and data scaling strategies. Since the original codes of these models were not publicly available, they could not be re-evaluated on the same database. Therefore, the present comparison should be regarded as indicative rather than conclusive.

Table 3. Quantitative assessment of model prediction accuracy.

	φ		c	
	R^2	MAE	R^2	MAE
EO-MLP model	0.946	0.692	0.983	0.434
Adaboost model [22]	0.892	1.284	0.979	0.612
PSO-ANN model [55]	0.942	0.691	0.976	0.522
ISSA-XGBoost model [54]	0.932	0.695	0.982	0.450

4.4. SHapley Additive exPlanations Analysis

Widely used in machine learning method, feature importance analysis quantifies how each input variable affects the predicted response [56]. To further clarify the influence of individual input variables on φ and c estimation, SHapley Additive exPlanations (SHAP)

can be incorporated in future interpretable ML models to quantify feature contributions [57]. SHAP treats each feature as a marginal contributor to the model output: by evaluating the incremental effect of a feature when added to the model and aggregating over all possible feature permutations, a global importance measure is obtained. This analysis facilitates understanding of the model's internal behavior and enhances interpretability [58]. Equation (10) provides the computational formulation of SHAP value. Further details on SHAP can be found in Refs. [59,60].

$$\phi_i = \sum_{S \subseteq N/\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \quad (10)$$

where $v(S)$ is the prediction produced with features limited to subset, N represents the full set of input features.

SHAP analysis was conducted to quantify the contribution of each input variable to the EO-MLP predictions. DeepExplainer was selected because it is suitable for neural-network-based predictive models. For each data partition, the corresponding training samples were used as the background reference database, while SHAP values were evaluated on the held-out test samples. The analysis was performed using the scaled input variables consistent with model training. Figure 11 presents the mean absolute SHAP value of four input variables for φ and c . Density shows the highest importance (0.4186) for predicting c , followed by UCS (0.3759), P-wave velocity (0.2306), and tensile strength (0.1881). UCS and P-wave velocity rank as the top contributors to φ (0.2898 and 0.2108), while tensile strength also contributes significantly (0.1901). In contrast, the importance of density is relatively low (0.0886), indicating a weaker effect on φ under the current modeling conditions. It should be noted that although all input variables affect the model output to varying degrees, their relative importance may vary considerably depending on data composition, modeling assumptions, and geological conditions. Therefore, the results presented here only reflect relative trends under the specific modeling scenario and should not be generalized. In modeling, the physical relevance and potential interactions of all input features should be thoroughly considered, regardless of their individual importance scores.

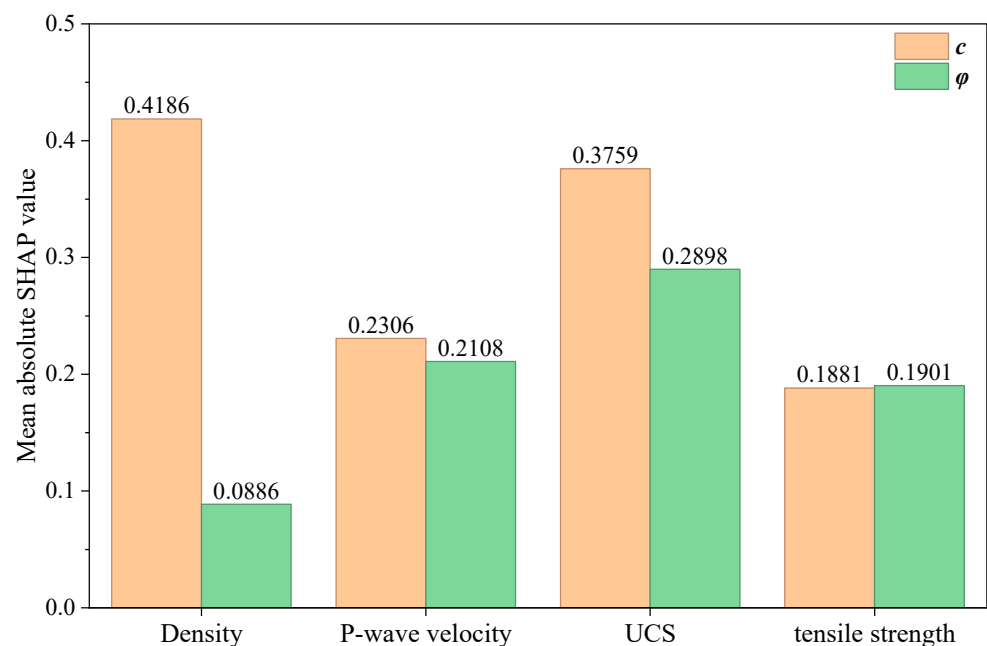


Figure 11. SHAP results of inputs.

5. Discussion

5.1. Comparison with Different Train Ratio

The training-to-test ratio was varied to assess its effect on predictive performance, while all other factors were kept constant. The training-to-test split ratios (i.e., train ratio) were set at 70%, 75%, 80%, 85%, and 90%, with each ratio corresponding to a different partition of the database. The results are presented in Figure 12, which shows the Mean R^2 obtained from several hundred model runs on the test set.

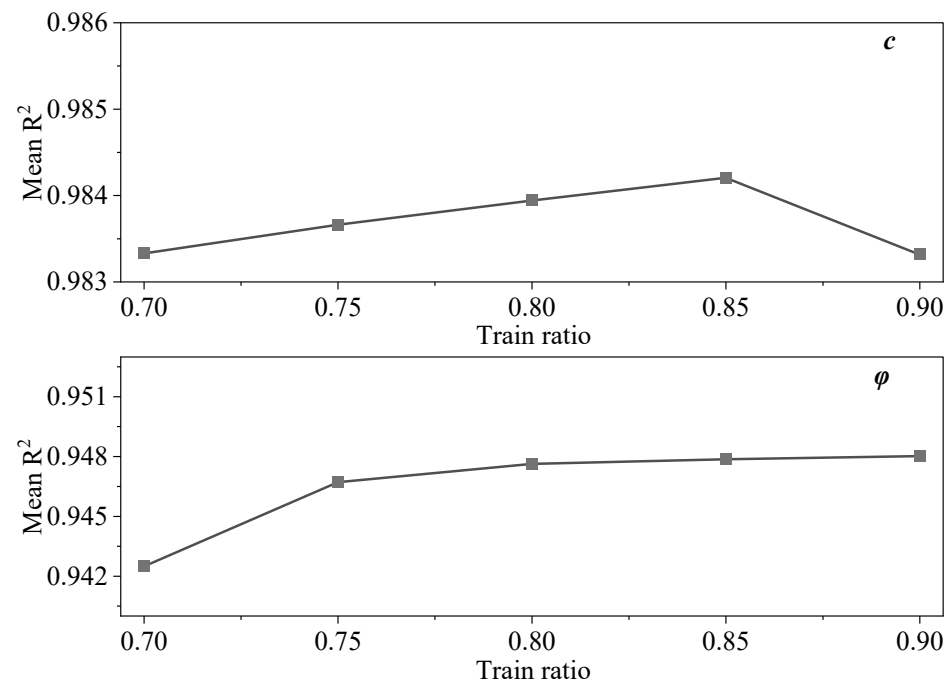


Figure 12. Comparison of performance metrics of different train ratio.

The mean R^2 generally increases with the proportion of training data, suggesting that additional calibration samples improve the model's capacity to learn representative relationships and enhance its predictive robustness. This trend indicates that enlarging the training subset can provide more stable parameter calibration and improve prediction reliability for unseen samples. However, the improvement in Mean R^2 is relatively marginal beyond the 80% train ratio, suggesting that after a certain threshold, increasing the size of the training set may not result in a significant boost in performance.

Interestingly, although the general trend suggests better performance with more training data, the differences in Mean R^2 between the various ratios are not drastic. This observation suggests that, within the chosen range of training ratios, the model achieves a sufficiently high performance even with relatively smaller training datasets. This could mean that, for some applications, a smaller training set may be sufficient to achieve an optimal model without needing to allocate an excessive amount of data for training. While a larger training set generally improves the predictive power of the model, the gains diminish as the training data grows, highlighting the potential for efficient model development with smaller datasets in certain contexts. Future studies may explore the trade-off between training time, data requirements, and model accuracy to optimize the use of available data for predicting ϕ and c .

5.2. Contribution and Limitations

The proposed method enables rapid and accurate evaluation of ϕ and c without resorting to complex analytical derivations or costly physical tests. Unlike most published

evaluations conducted on single-lithology, relatively homogeneous datasets with random splits, the present assessment adopts a mixed-rock setting that spans multiple lithologies and multi-source experimental records. A multi-lithology, cross-source dataset together with an external validation set is employed to examine applicability and robustness under realistic cross-scenario (out-of-distribution) conditions. Under a common protocol, established base-lines (AdaBoost, PSO-ANN, ISSA-XGBoost) reproduced on the same corpus exhibit marked performance degradation on cross-source splits, whereas the EO-MLP model maintains lower error and more stable calibration, indicating superior robustness to source shift.

Although the ML framework demonstrates high accuracy and robustness in predicting shear strength parameters, several limitations remain. First, ML model performance is highly dependent on the size and distribution of the training samples [61]. Despite incorporating multi-source data to improve generalization, the current database comprises only 213 samples, which limits its coverage of diverse geological conditions. Second, the input features mainly rely on macroscopic parameters, while microstructural information such as joint distribution, mineral composition, and pore network characteristics has not yet been fully integrated. These factors are likely to affect shear strength parameters and merit further investigation [62]. In addition, the current modeling effort prioritizes predictive accuracy, while offering limited insight into the decision-making process. In ensemble models or deep learning architectures, nonlinear coupling among variables often reduces interpretability, which may hinder trust in the model for high-risk geotechnical decision-making [63]. Another potential issue lies in multicollinearity among some input features, which can introduce redundancy and increase model uncertainty. In practice, although the training efficiency of the model has been optimized in this study, its response speed and stability under resource-constrained conditions still require further validation. The present work does not yet include an analysis of predictive uncertainty.

Future research should focus on building a larger, geologically diverse training dataset, particularly incorporating in-situ triaxial test results from actual engineering projects. The integration of microstructural analysis and digital rock physics techniques could provide a more direct connection between experimental parameters and real geological structures. In addition, uncertainty analysis can be used to assess reliability, evaluate whether estimates are adequate, and identify potential systematic bias. Finally, efforts should be made to improve field applicability by adopting model compression techniques and edge computing frameworks, thereby enhancing the practical utility of ML-driven predictions for geotechnical parameters.

6. Conclusions

1. The Equilibrium Optimizer was applied to optimize the weights and biases of a fixed MLP architecture, which improved the prediction accuracy for φ and c . The proposed EO-MLP model exhibited superior performance for both φ and c . On the test set it attains $R^2 = 0.946$ /RMSE = 1.072 for φ and $R^2 = 0.983$ /RMSE = 0.671 for c , outperforming the conventional MLP model. An external cross-study benchmark further indicates that EO-MLP model is superior to three recent ML models.
2. SHapley Additive exPlanations analysis revealed that, for c , density had the highest contribution (0.4186), followed by uniaxial compressive strength (0.3759), P-wave velocity (0.2306), and tensile strength (0.1881). In contrast, for φ , the dominant variables were UCS (0.2898), P-wave velocity (0.2108), and tensile strength (0.1901), whereas density contributed the least (0.0886).
3. Compared to four classical machine learning models, the EO-MLP model consistently outperformed all alternatives across all evaluation metrics. For c , the EO-MLP model achieved an R^2 of 0.983, notably higher than the support vector machine model (0.977).

and decision tree model (0.948). For φ , its R^2 reached 0.946, surpassing the support vector machine model (0.916) and decision tree model (0.822). Furthermore, among the tested data preprocessing methods applied to this database for predicting φ and c , the robust scaler normalization delivered the most accurate results, characterized by the lowest MAE and RMSE together with the highest R^2 .

Author Contributions: Conceptualization, B.L., Z.X. and H.L.; Software, Z.X.; Validation, B.L.; Formal analysis, B.L. and H.L.; Investigation, B.L. and Z.X.; Data curation, Z.X.; Writing—original draft, B.L. and H.L.; Writing—review & editing, B.L. and H.L.; Visualization, H.L.; Project administration, Z.X. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Boulanger-Martel, V.; Bussière, B.; Côté, J. Resistance of a soapstone waste rock to freeze-thaw and wet-dry cycles: Implications for use in a reclamation cover in the Canadian Arctic. *Bull. Eng. Geol. Environ.* **2021**, *80*, 41–54. [\[CrossRef\]](#)
2. Han, Y.; Shu, R.H.; Chen, Y.F.; Feng, F. A new method to determine the crack closure stress based on stress difference. *Theor. Appl. Fract. Mech.* **2022**, *119*, 103337. [\[CrossRef\]](#)
3. Grindheim, B.; Li, C.C.; Høien, A.H.; Lia, L. Behavior of a rock mass in uplift field tests of rock anchors. *Rock Mech. Rock Eng.* **2024**, *57*, 2339–2364. [\[CrossRef\]](#)
4. Xie, S.; Li, J.; Wang, S.; Li, X. Experimental study on dynamic shear properties of infilled joints under dynamic shearing: Insights from normal stress and shear rate. *Rock Mech. Rock Eng.* **2026**. [\[CrossRef\]](#)
5. Xie, S.J.; Ma, T.X.; Huang, Y.Q.; Lin, H.; Shi, Z.M. New shear constitutive model for rock discontinuities based on maximum entropy theory: Performance analysis and engineering application. *Bull. Eng. Geol. Environ.* **2026**, *85*, 159. [\[CrossRef\]](#)
6. Kalantari, S.; Baghbanan, A.; Hashemalhosseini, H. An analytical model for estimating rock strength parameters from small-scale drilling data. *J. Rock Mech. Geotech. Eng.* **2019**, *11*, 135–145. [\[CrossRef\]](#)
7. Renani, H.R.; Martin, C.D. Factor of safety of strain-softening slopes. *J. Rock Mech. Geotech. Eng.* **2020**, *12*, 473–483. [\[CrossRef\]](#)
8. Li, D.; Chen, Y.; Lu, W.; Zhou, C. Stochastic response surface method for reliability analysis of rock slopes involving correlated non-normal variables. *Comput. Geotech.* **2011**, *38*, 58–68. [\[CrossRef\]](#)
9. Rullière, A.; Rivard, P.; Peyras, L.; Breul, P. Influence of roughness on the apparent cohesion of rock joints at low normal stresses. *J. Geotech. Geoenvironmental Eng.* **2020**, *146*, 04020003. [\[CrossRef\]](#)
10. Alidaryan, M.; Khosravi, M.H.; Bahaaddini, M.; Moosavi, M.; Roshan, H. Mobilization of cohesion and friction angle of intact rocks in the shearing process. *Rock Mech. Rock Eng.* **2023**, *56*, 8221–8233. [\[CrossRef\]](#)
11. Xie, S.; Li, J.; Wang, S.; Li, X. An experimental investigation on the dynamic shear characteristics of wet joints. *Int. J. Rock Mech. Min. Sci.* **2025**, *194*, 106193. [\[CrossRef\]](#)
12. Aghababaei, M.; Behnia, M.; Moradian, O. Experimental investigation on strength and failure behavior of carbonate rocks under multistage triaxial compression. *Int. J. Rock Mech. Min. Sci.* **2019**, *123*, 104099. [\[CrossRef\]](#)
13. Edimann, K.; Somerville, J.M.; Smart, B.G.D.; Hamilton, S.A.; Crawford, B.R. Predicting rock mechanical properties from wireline porosities. In Proceedings of the SPE/ISRM Rock Mechanics in Petroleum Engineering, Trondheim, Norway, 8–10 July 1998.
14. Hajdarwish, A.; Shakoor, A.; Wells, N.A. Investigating statistical relationships among clay mineralogy, index engineering properties, and shear strength parameters of mudrocks. *Eng. Geol.* **2013**, *159*, 45–58. [\[CrossRef\]](#)
15. Karakul, H.; Ulusay, R. An alternative method for predicting internal friction angle of rock materials. *Environ. Earth Sci.* **2024**, *83*, 306. [\[CrossRef\]](#)
16. Patil, U.S.; Kene, S.D. Sustainable Concrete Incorporating Agro-Industrial Wastes (WHA-CHA-ASW): Experimental Investigation and Machine Learning-Based Modeling. *J. Mater. Eng. Perform.* **2026**. [\[CrossRef\]](#)

17. Zhang, H.; Xu, H.Q.; Xie, Y.Q.; Zhang, M.D.; Li, R.; Li, G. High-Temperature Oxidation Behaviors of AlCoCrFeNi and CoCrFeNiNb Eutectic High-Entropy Alloys: A Machine Learning-Assisted Design and Mechanism Study. *J. Mater. Eng. Perform.* **2026**. [\[CrossRef\]](#)
18. Liu, S.L.; Wang, L.Q.; Zhang, W.A.; Sun, W.X.; Wang, Y.H.; Liu, J.P. Physics-informed optimization for a data-driven approach in landslide susceptibility evaluation. *J. Rock Mech. Geotech. Eng.* **2024**, *16*, 3192–3205. [\[CrossRef\]](#)
19. Xie, S.J.; Jiang, Z.Y.; Lin, H.; Ma, T.X.; Peng, K.; Liu, H.W.; Liu, B.H. A new integrated intelligent computing paradigm for predicting joints shear strength. *Geosci. Front.* **2024**, *15*, 101884. [\[CrossRef\]](#)
20. Shen, J.; Jimenez, R. Predicting the shear strength parameters of sandstone using genetic programming. *Bull. Eng. Geol. Environ.* **2018**, *77*, 1647–1662. [\[CrossRef\]](#)
21. Shahani, N.M.; Ullah, B.; Shah, K.S.; Hassan, F.U.; Ali, R.; Elkotb, M.A.; Ghoneim, M.E.; Tag-Eldin, E.M. Predicting angle of internal friction and cohesion of rocks based on machine learning algorithms. *Mathematics* **2022**, *10*, 3875. [\[CrossRef\]](#)
22. Li, D.; Yang, B.; Liu, Z.; Liu, Y.; Zhao, J. Prediction method of rock cohesion and internal friction angle based on ensemble tree algorithm. *Gold Sci. Technol.* **2024**, *32*, 847–859.
23. Rathor, A.P.S.; Tabaroei, A.; Dey, A.; Sharma, J.K. A Physics-Informed Machine Learning Framework for Strength Prediction, Uncertainty Quantification, and Optimization of Chemically Stabilized Soils. *J. Mater. Eng. Perform.* **2026**, 1–24. [\[CrossRef\]](#)
24. Reddy, S.P.; Ammisetti, D.K.; Juvvala, S.R.; Reddy, A.V.N. Experimental Investigation and Machine Learning Modeling on the Tribological Characteristics of Heat Treated AA7075/B4C/BN/SiC Hybrid Composites. *J. Mater. Eng. Perform.* **2026**, 1–21. [\[CrossRef\]](#)
25. Kainthola, A.; Singh, P.K.; Verma, D.; Singh, R.; Sarkar, K.; Singh, T.N. Prediction of strength parameters of himalayan rocks: A statistical and anfis approach. *Geotech. Geol. Eng.* **2015**, *33*, 1255–1278. [\[CrossRef\]](#)
26. Kurnaz, T.F.; Erden, C.; Kökçam, A.H.; Dagdeviren, U.; Demir, A.S. A hyper parameterized artificial neural network approach for prediction of the factor of safety against liquefaction. *Eng. Geol.* **2023**, *319*, 107109. [\[CrossRef\]](#)
27. Schmüdderich, C.; Machacek, J.; Prada-Sarmiento, L.F.; Staubach, P.; Wichtmann, T. Strain-dependent slope stability for earthquake loading. *Comput. Geotech.* **2022**, *152*, 105048. [\[CrossRef\]](#)
28. Endert, A.; Ribarsky, W.; Turkay, C.; Wong, B.W.; Nabney, I.; Blanco, I.D.; Rossi, F. The state of the art in integrating machine learning into visual analytics. *Comput. Graph. Forum* **2017**, *36*, 458–486. [\[CrossRef\]](#)
29. Liu, B.; Wen, H.J.; Di, M.R.; Huang, J.H.; Liao, M.Y.; Yu, J.Y.; Xiang, Y.T. Mapping and interpretability of aftershock hazards using hybrid machine learning algorithms. *J. Rock Mech. Geotech. Eng.* **2025**, *17*, 4908–4932. [\[CrossRef\]](#)
30. Tuan, N.S.; To, Q.D.; Vu, M.N.; Nguyen, T.D.; Nguyen, T.T. Study the elastic properties and the anisotropy of rocks using different machine learning methods. *Geophys. Prospect.* **2020**, *68*, 2557–2578. [\[CrossRef\]](#)
31. Koohmishi, M.; Guo, Y. Machine learning approach to railway ballast degradation prognosis considering crumb rubber modification and parent rock strength. *Constr. Build. Mater.* **2023**, *409*, 133985. [\[CrossRef\]](#)
32. Ahsan, M.M.; Mahmud, M.A.P.; Saha, P.K.; Gupta, K.D.; Siddique, Z. Effect of data scaling methods on machine learning algorithms and model performance. *Technologies* **2021**, *9*, 52. [\[CrossRef\]](#)
33. Wang, J.; Zuo, Y.J.; Dong, L.J.; Yan, X.H. Identifying different classes of geoacoustic events using machine learning. *Int. J. Rock Mech. Min. Sci.* **2025**, *192*, 106144. [\[CrossRef\]](#)
34. Wu, L.Y.; Li, J.H.; Zhang, J.W.; Wang, Z.F.; Tong, J.B.; Ding, F.; Li, M.; Feng, Y.; Li, H. Prediction model for the compressive strength of rock based on stacking ensemble learning and shapley additive explanations. *Bull. Eng. Geol. Environ.* **2024**, *83*, 1–23. [\[CrossRef\]](#)
35. Lett, E.; La Cava, W.G. Translating intersectionality to fair machine learning in health sciences. *Nat. Mach. Intell.* **2023**, *5*, 476–479. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Xie, S.J.; Lin, H.; Duan, H.Y. A novel criterion for yield shear displacement of rock discontinuities based on renormalization group theory. *Eng. Geol.* **2023**, *314*, 107008. [\[CrossRef\]](#)
37. Liu, C.-Y.; Ku, C.-Y.; Wu, T.-Y. Advancing rock mass classification using machine learning approach. *Bull. Eng. Geol. Environ.* **2025**, *84*, 612. [\[CrossRef\]](#)
38. Xie, S.; Lin, H.; Duan, H.; Chen, Y. Modeling description of interface shear deformation: A theoretical study on damage statistical distributions. *Constr. Build. Mater.* **2023**, *394*, 132052. [\[CrossRef\]](#)
39. Ebrahimabadi, A.; Azimipour, M.; Bahreini, A. Prediction of roadheaders' performance using artificial neural network approaches (MLP and KOSFM). *J. Rock Mech. Geotech. Eng.* **2015**, *7*, 573–583. [\[CrossRef\]](#)
40. Zhao, S.; Tan, D.Y.; Lin, S.Q.; Yin, Z.Y.; Yin, J.H. A deep learning-based approach with anti-noise ability for identification of rock microcracks using distributed fibre optic sensing data. *Int. J. Rock Mech. Min. Sci.* **2023**, *170*, 1–23. [\[CrossRef\]](#)
41. Gurgenc, E.; Altay, O.; Altay, E.V. AOSMA-MLP: A novel method for hybrid metaheuristics Artificial Neural Networks and a new approach for prediction of geothermal reservoir temperature. *Appl. Sci.* **2024**, *14*, 3534. [\[CrossRef\]](#)
42. Golabchi, Y.; Perras, M.A.; Khan, U.T. Comparison of different classification and regression machine learning algorithms for the prediction of excavation damage zone depths. *Rock Mech. Rock Eng.* **2025**, *58*, 10439–10461. [\[CrossRef\]](#)

43. Faramarzi, A.; Heidarinejad, M.; Stephens, B.; Mirjalili, S. Equilibrium optimizer: A novel optimization algorithm. *Knowl.-Based Syst.* **2020**, *191*, 105190. [[CrossRef](#)]
44. Gao, Y.Y.; Zhou, Y.Q.; Luo, Q.F. An efficient binary equilibrium optimizer algorithm for feature selection. *IEEE Access* **2020**, *8*, 140936–140963. [[CrossRef](#)]
45. Yu, Z.; Shi, X.Z.; Miao, X.H.; Zhou, J.; Khandelwal, M.; Chen, X.; Qiu, Y.G. Intelligent modeling of blast-induced rock movement prediction using dimensional analysis and optimized artificial neural network technique. *Int. J. Rock Mech. Min. Sci.* **2021**, *143*, 104794. [[CrossRef](#)]
46. Sun, J.L.; Wu, S.C.; Wang, H.; Wang, T.; Geng, X.J.; Zhang, Y.J. Inversion of surrounding rock mechanical parameters in a soft rock tunnel based on a hybrid model EO-LightGBM. *Rock Mech. Rock Eng.* **2023**, *56*, 6691–6707. [[CrossRef](#)]
47. Liu, H.Y.; Ma, T.X.; Lin, Y.; Peng, K.; Hu, X.Q. Deep learning in rockburst intensity level prediction: Performance evaluation and comparison of the ngo-cnn-bigru-attention model. *Appl. Sci.* **2024**, *14*, 5719. [[CrossRef](#)]
48. Li, Q.; Li, J.P.; Duan, L.C.; Tan, S.C. Prediction of rock abrasivity and hardness from mineral composition. *Int. J. Rock Mech. Min. Sci.* **2021**, *140*, 104658. [[CrossRef](#)]
49. Sun, Z.; Li, Y.; Su, L.; Niu, D.; Luo, D.; He, W. Investigation of electrical resistivity for fiber-reinforced coral aggregate concrete. *Constr. Build. Mater.* **2024**, *414*, 135011. [[CrossRef](#)]
50. Ma, T.; Chen, C.; Shen, L.; Luo, K.; Jiang, Z.; Yang, Z.; Chen, H.; Peng, K. Physics-informed neural networks for capturing the true relationships between parameters to predict the dynamic triaxial strength of rocks in cold environments. *Meas. J. Int. Meas. Confed.* **2026**, *257*, 118900. [[CrossRef](#)]
51. Lin, H.; Chen, Y.; Yao, R.; Sun, Z.; Zhou, X. Hybrid machine learning models to predict the shear strength of discontinuities with different joint wall compressive strength. *Nondestruct. Test. Eval.* **2024**, *40*, 2439–2459. [[CrossRef](#)]
52. Pirunjinda, S.; Jongpradist, P.; Ryu, D.W.; Inoue, J.; Yazdani, M.; Kim, H.M. Ground subsidence hazard evaluation in large reclaimed area using unsupervised machine learning with parameter estimation and spatial interpolation. *Eng. Geol.* **2026**, *360*, 108450. [[CrossRef](#)]
53. Lin, H.; Chen, Y.; Wang, Y.; Cao, R.; Li, J. Statistical damage shear constitutive model of rock joints under seepage pressure. *Front. Earth Sci.* **2020**, *8*, 232. [[CrossRef](#)]
54. Lei, D.; Zhang, Y.; Lu, Z.; Wang, G.; Lai, Z.; Lin, M.; Chen, Y. A machine learning framework for predicting shear strength properties of rock materials. *Sci. Rep.* **2025**, *15*, 8748. [[CrossRef](#)] [[PubMed](#)]
55. Cheng, Y.; He, D.; Liu, H.; Wang, G. Shear strength parameters prediction of rock materials using hybrid machine learning model. *Nondestruct. Test. Eval.* **2025**, *41*, 1–24. [[CrossRef](#)]
56. Sun, Z.; Li, Y.L.; Yang, Y.X.; Su, L. Splitting tensile strength of basalt fiber reinforced coral aggregate concrete: Optimized XGBoost models and experimental validation. *Constr. Build. Mater.* **2024**, *416*, 135133. [[CrossRef](#)]
57. Jas, K.; Mangalathu, S.; Dodagoudar, G.R. Evaluation and analysis of liquefaction potential of gravelly soils using explainable probabilistic machine learning model. *Comput. Geotech.* **2024**, *167*, 106051. [[CrossRef](#)]
58. Ma, T.; Hu, X.; Liu, H.; Peng, K.; Lin, Y.; Chen, Y.; Luo, K. Elastic modulus prediction for high-temperature treated rock using multi-step hybrid ensemble model combined with coronavirus herd immunity optimizer. *Meas. J. Int. Meas. Confed.* **2025**, *240*, 115596. [[CrossRef](#)]
59. Pradhan, B.; Lee, S.; Dikshit, A.; Kim, H. Spatial flood susceptibility mapping using an explainable artificial intelligence (XAI) model. *Geosci. Front.* **2023**, *14*, 101625. [[CrossRef](#)]
60. Gharavi, A.; O’Sullivan, A.; Haddad, M.; Hassan, M.G.; Alasmar, R.; Yousefi, P.; Al-Saegh, S. Explainable machine learning and deep learning for productive zone identification in tight sandstone reservoirs: Integrating PROMETHEE-II and class imbalance handling. *Geoenergy Sci. Eng.* **2026**, *262*, 214436. [[CrossRef](#)]
61. Sun, Z.; Li, Y.; Han, T.; Su, L.; Zhu, X.; He, J. Performance evaluation of hybrid fiber-reinforced concrete based on electrical resistivity: Experimental and data-driven method. *Constr. Build. Mater.* **2024**, *446*, 137992. [[CrossRef](#)]
62. Xie, S.; Lin, H.; Duan, H.; Yao, R.; Ma, T. A novel triaxial strength criterion for rocks based on the ultimate strength and its application. *Geoenergy Sci. Eng.* **2025**, *246*, 213590. [[CrossRef](#)]
63. Takaew, P.; Xia Jianhong, C.; Doucet Luc, S. Machine learning and tectonic setting determination: Bridging the gap between Earth scientists and data scientists. *Geosci. Front.* **2024**, *15*, 101726. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.