



Explainable metaheuristic-optimized gradient boosting decision trees for slope stability prediction

Ning Wang · Manuel Pastor ·
Saeid Moussavi Tayyebi · Taraka Pattela ·
Guoyan Zhao · Meng Wang

Received: 22 May 2025 / Accepted: 31 October 2025
© The Author(s) 2025

Abstract Slope stability prediction remains challenging due to multi-factor interactions and the limited capacity of traditional methods to capture nonlinear behavior. Existing AI-driven approaches can show dataset-dependent performance and limited interpretability, which hinders adoption. This study proposes three novel hybrid models integrating metaheuristic optimization and explainable artificial intelligence (XAI) to address this challenge. Three metaheuristic algorithms, Alpha evolutionary algorithm (AEA), Sine cosine algorithm (SCA), and Arithmetic optimization algorithm (AOA), are employed to optimize the hyperparameters of gradient boosting decision trees (GBDT), forming hybrid models (AEA-GBDT, SCA-GBDT, AOA-GBDT). A comprehensive database of 436 slope cases with six inputs (H , β , γ , c ,

φ , and r_u) and slope state was compiled. Five-fold cross-validation (CV) was used for all models to ensure fair evaluation and limit overfitting. The Friedman–Nemenyi test analysis identified AEA-GBDT as the top model. AEA-GBDT achieved superior performance with the highest accuracy (0.920), precision (0.934), recall (0.914), $F_{1\text{-score}}$ (0.924), Cohen’s kappa (0.840), and AUC (0.960). To contextualize performance, XGBoost, AdaBoost, Random Forest, and SVM were also evaluated. Shapley Additive explanations (SHAP) indicate that φ and c dominate the predictions, while r_u exhibits nonlinear interactions with c . This method narrows the gap between black-box AI models and engineering transparency and provides interpretable, data-driven support for slope design and risk management.

N. Wang · G. Zhao · M. Wang (✉)
School of Resources and Safety Engineering, Central
South University, Changsha 410083, China
e-mail: mwanglh@csu.edu.cn

N. Wang · M. Pastor · S. M. Tayyebi
Department of Applied Mathematics, ETS Ingenieros de
Camino, Universidad Politécnica de Madrid, Calle del
Profesor Aranguren, 3, 28040 Madrid, Spain

T. Pattela
Department of Earth, Environment and Physical Sciences,
University of Siena, Via Laterano 8, 53100 Siena, Italy

M. Wang
Department of Mining and Materials Engineering, McGill
University, Montreal H3A 0E8, Canada

Highlights

- Developed three hybrid models (AEA-GBDT, SCA-GBDT, AOA-GBDT) integrating metaheuristic optimization with gradient boosting decision trees for slope stability prediction.
- Compiled a comprehensive database of 436 soil slopes combining real and simulated engineering cases.
- The AEA-GBDT model achieved superior performance with the highest accuracy (0.920), precision (0.934), recall (0.914), $F_{1\text{-score}}$ (0.924), Cohen’s kappa (0.840), and AUC (0.960).

- SHAP analysis enhanced interpretability, identifying cohesion (c) and internal friction angle (ϕ) as the dominant controlling factors.
- Independent validation on 16 loess slope cases confirmed excellent generalization and engineering applicability.

Keywords Slope stability · Hybrid ensemble models · Metaheuristic optimization · Shapley Additive explanations

1 Introduction

Slope stability is a crucial topic in geotechnical and mining engineering. It directly affects safety and environmental protection (Liu et al. 2014; Deng et al. 2020; Pham et al. 2021; Aminpour et al. 2023). It involves the ability of natural or artificial slopes to maintain their geometric form and structural integrity under various external factors (Feng and Hudson 2004; Griffiths and Marquez 2007). Slope failures are commonly observed in mountain roads, railways, water conservancy and hydropower projects, mining operations, and other engineering projects. Common failures include landslides, collapses, and debris flows, which threaten lives and property and lead to severe economic losses and ecological losses (Ray et al. 2020; Gao et al. 2023; Pastor et al. 2024). Slope stability is influenced by a variety of factors, including the material properties, structural conditions, groundwater, topographic and geomorphological features, climatic, and human activity (Matsui and San 1992; Ahour et al. 2020; Huang et al. 2023). These interactions make slope stability analysis a multi-dimensional and nonlinear problem. In recent years, the advancement of large-scale engineering projects has posed greater challenges to the accuracy and reliability of slope stability predictions (Wang et al. 2005; Ahmed 2015; Qi and Tang 2018a).

Slope stability research methods have evolved from qualitative to quantitative approaches and from simple to complex models. Early methods relied on engineering judgement and simplified mechanical models, such as the limit equilibrium method (Acevedo et al. 2021), the ordinary method of slices (Duncan and Wright 1980), and the Bishop method (Ullah et al. 2020). These methods, while simplifying

problems, provided reliable analytical tools for engineering practice. With the advance of computing, numerical methods such as the finite element method (Matsui and San 1992; Griffiths and Marquez 2007; Liu et al. 2015) and the discrete element methods (Tayyebi et al. 2022) became widely used. These accommodate complex geometries, material nonlinearity, and dynamic loading, improving accuracy and applicability of slope stability analyses (Salmasi and Jafari 2019). Recognizing the inherent uncertainties in geological parameters, researchers began to incorporate probabilistic statistical methods into slope stability analysis (Zhang et al. 2022b). For example, Li and Lumb (1987) proposed a slope stability assessment method based on reliability theory, marking a shift from deterministic analysis to risk assessment. Despite these advances, traditional methods remain limited for high-dimensional and nonlinear problems.

With the rapid development of artificial intelligence (AI) technology, data-driven methods have been increasingly used for slope stability analysis. Pham et al. (2021) introduced the XBB-CM method and achieved an accuracy of 90.3% on 153 cases. Zhou et al. (2019) expanded the database to 221 cases and obtained 86% accuracy. Lin et al. (2022) employed principal component analysis combined with ensemble learning, though concerns regarding duplicate cases in the appendix dataset were raised. Qi and Tang (2018b) compared six single models and an ensemble method, and showed that algorithm choice leads to different results on the same data. More advanced strategies have also been explored in recent years. Kardani et al. (2021) proposed a stacking approach combined with finite element analysis to improve the predictive capacity of slope stability models. Zhang et al. (2022a) proposed selective integration by edge distance minimization to improve generalization. Niu et al. (2020) coupled PCA with a GA-tuned SVM for highway slopes and reported strong accuracy, while Anuragi and Kishan (2025) used an improved PSO to optimize SVM, achieving superior classification performance. Das et al. (2011) applied neural networks optimized by differential evolution to slope stability prediction. Tiwari et al. (2025) evaluated 633 slope cases using ensemble and hybrid models with evolutionary optimization, demonstrating dataset-dependent performance.

AI methods have been widely applied in slope stability analysis (Gordan et al. 2016; Lin et al. 2022;

Liu et al. 2023). Large databases can be integrated, and complex nonlinear patterns can be captured, which are often difficult for traditional approaches (Tang et al. 2019; Bardhan and Samui 2022; Yang et al. 2023b). However, two major limitations have been recognized (Cheng and Hoang 2015; Zhou et al. 2022). First, the performance of AI models is highly sensitive to hyperparameters. Different parameter settings within the same model can produce inconsistent prediction results (Qiu et al. 2024). Second, the “black-box” nature of AI models remains a challenge, as the internal mechanisms of predictions are difficult to interpret (Mangalathu et al. 2020). To overcome these issues, hybrid ensemble learning has been developed in recent years. By combining ensemble learners with metaheuristic optimization, prediction accuracy and robustness can be improved (Fatty et al. 2023). In this context, gradient boosting decision tree (GBDT) is well suited to tabular and nonlinear slope data. Recent metaheuristics, such as the alpha evolution algorithm (AEA), the sine cosine algorithm (SCA), and the arithmetic optimization algorithm (AOA), build on traditional methods. They were selected in this study because they represent three distinct and recently developed paradigms: evolutionary adaptation, trigonometric oscillatory updates, and operator-driven exploration–exploitation balancing. They often achieve more stable convergence under comparable computational budgets. In parallel, the development of explainable artificial intelligence (XAI) has provided new solutions. Transparent and credible decision-making bases can be offered, while high predictive performance is maintained (Chen et al. 2023; Gao et al. 2024; Tiwari and Das 2025).

To bridge the gap in predictive performance and interpretability, three hybrid models were developed. Initially, a database of 436 slope cases was assembled and statistically profiled. Subsequently, GBDT was employed as the base learner, and its hyperparameters were optimized with three complementary metaheuristics: AEA, SCA, and AOA. These complementary strategies support efficient exploration of high-dimensional hyperparameters spaces and the identification of near-optimal hyperparameter combinations. Model performance was assessed using accuracy, precision, recall, $F_{1\text{-score}}$, and Cohen’s kappa. To determine the optimal model, the Friedman-Nemenyi test was conducted. Furthermore, Shapley Additive Explanations (SHAP) were applied to the optimal

hybrid ensemble model to ensure transparency and interpretability of predictions. This study deepens the understanding of slope stability prediction and promotes interdisciplinary innovation between artificial intelligence and slope engineering.

2 Slope stability datasets

In this study, a comprehensive slope database was developed, primarily derived from the survey work conducted by Chakraborty et al. (Zhu 2005; Chakraborty and Goswami 2017; Zhou et al. 2019; Kardani et al. 2021). The dataset consists exclusively of soil slopes, and all unstable cases correspond to circular failure mechanisms, which are the most common mode of failure in soil slopes. A total of 436 slope cases were included, comprising 226 stable and 210 failure samples. Among them, 286 were collected from documented case histories in Australia, Canada, China, and the United States, ensuring practical engineering applicability. The remaining 150 cases were generated through finite element simulations reported in the literature, which provided systematic parameter exploration and broad combinations of slope conditions. This composition ensures that the database captures both real-world slope and representative simulated conditions, thereby providing a robust basis for model training and validation.

To accurately predict slope stability, it is essential to consider a wide range of factors (Shen et al. 2024). Slope stability is influenced by a combination of internal and external factors, such as material properties, topographic features, groundwater conditions, and human activity (Wang et al. 2020; Acevedo et al. 2021). Local slope instability occurs under the action of gravity and pore pressure when the maximum shear force on the potential sliding surface exceeds the shear strength (Moses et al. 2020). The schematic of the slope structure is illustrated in Fig. 1.

The main influencing factors can be summarized into physico-mechanical parameters, geometrical structural parameters, and hydrological conditions of the slope (Lin et al. 2018, 2022). In this study, six key factors were selected as inputs to the prediction model: slope height (H), slope angle (β), unit weight (γ), cohesion (c), internal friction angle (φ), and pore pressure ratio (r_u). The descriptive statistics of the slope database are presented in Table 1. Other factors such as precipitation,

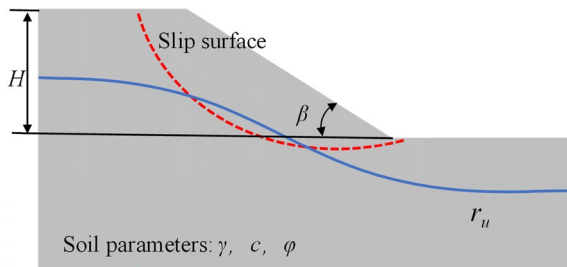


Fig. 1 Slope instability and slope structural parameters

stratigraphy, and seismic activity are also known to influence slope stability (Yang et al. 2023a). However, consistent and quantitative data for these factors are difficult to obtain from slope engineering case records (Segoni et al. 2025). Previous studies have demonstrated that reliable predictions can still be achieved based on the six selected factors (Manouchehrian et al. 2014; Kang et al. 2017; Wang et al. 2023).

The distribution of sample features and the correlation between them were quantitatively analyzed, with results shown in the matrix scatter plot in Fig. 2. Each cell in the figure represents the relationship between two variables, with the horizontal and vertical axes indicating their respective scales. The data is divided into two groups based on slope state, represented by different colors (blue for stable and red for failure). The diagonal of the matrix contains histograms showing the distribution of each variable, which are similar for both groups. Each cell also displays the linear correlation coefficient between the two features, ranging from -1 to 1 . The upper triangular part of Fig. 2 intuitively illustrates the correlation coefficients and p -values of each parameter under the two slope stability states. The significance levels are indicated by asterisks ('***', '**', '*', '.', ''), corresponding to p -values of (0, 0.001, 0.01, 0.05, 0.1, 1), respectively. Among them, H and γ show

a strong correlation, with a coefficient of 0.382. Values close to 1 or -1 indicate a strong correlation, whereas values close to 0 indicate no linear correlation. Each scatter plot includes a trend line representing the linear regression relationship between the two variables. The generally small slopes of these trend lines suggest that the correlations between most variables are low, specifically within the range of 0–0.3. This implies that most variables are either uncorrelated or weakly correlated, indicating their relative independence. Given this independence, it is reasonable and necessary to combine these selected features for a comprehensive prediction of slope stability. In addition, the statistical description in Table 1 showed that the slope cases span a wide range of parameter values (e.g., cohesion up to 300 kPa and slope height up to 565 m), which enhances the representativeness of the dataset. These characteristics imply that the prediction problem is high-dimensional and nonlinear, making it unsuitable for simple linear or empirical models. Consequently, advanced ensemble learning methods such as GBDT, which are capable of capturing nonlinear interactions and handling heterogeneous data distributions, are well suited for this dataset. All input variables were carefully checked, and no abnormal or unrealistic values were detected, ensuring that the dataset remained within plausible geotechnical ranges. To improve the model's prediction accuracy and efficiency, the sample attributes were standardized to dimensionless values, normalizing the dataset to a range of $[0,1]$ before model training.

3 Methodology

3.1 Gradient boosting decision tree (GBDT)

The Gradient boosting decision tree (GBDT) is an ensemble learning method proposed by Professor

Table 1 Descriptive statistics of slope database

Indicator	Count	Mean	Minimum	Median	Maximum	Standard deviation
Unit weight γ (kN/m ³)	436	21.45	10.06	21.20	31.30	3.72
Cohesion c (kPa)	436	37.22	0.00	28.95	300.00	39.49
Internal friction angle φ (°)	436	26.85	0.00	27.07	57.36	10.02
Slope angle β (°)	436	34.76	4.24	35.00	65.00	11.04
Slope height H (m)	436	80.97	3.45	52.00	565.00	88.08
Pore pressure ratio (r_u)	436	0.23	0.00	0.20	1.00	0.28



Fig. 2 Scatter plot of feature correlation matrix of database

Friedman (2001), which uses decision trees as base estimators combined with gradient boosting. GBDT has become a highly influential algorithm in machine learning, particularly effective on structured data. In this study, classification trees are used as the base estimators. A classification tree consists of nodes and edges, where leaf nodes represent the slope stability state, and internal nodes represent input metrics, including H , β , γ , c , ϕ , and r_u . At the core of GBDT is an iterative process where decision trees fit the residuals to influence the structure of subsequent trees. The residuals from the previous round are fitted using the gradient information of the loss function, resulting in a decrease in residuals with each iteration. The prediction from each round of decision trees gradually approaches the true value, eventually forming the GBDT model through iterative learning. The GBDT model training process is

illustrated in Fig. 3. Hyperparameter selection is crucial for the accuracy and applicability of the GBDT model (Hoang and Pham 2016; Mahmoodzadeh et al. 2022). Its main hyperparameters include $N_{\text{estimators}}$, Learning_rate , Max_depth , and Min_samples_split , which are optimized in this study.

3.2 Alpha evolution algorithm (AEA)

The Alpha evolutionary algorithm (AEA) is a recently proposed population-based metaheuristic (Gao and Zhang 2024). It was designed to improve optimization in high-dimensional and nonlinear search spaces. AEA integrates three main strategies: alpha operators, evolutionary path adaptation, and matrix generation. These components allow AEA to balance global exploration and local exploitation. Conventional algorithms

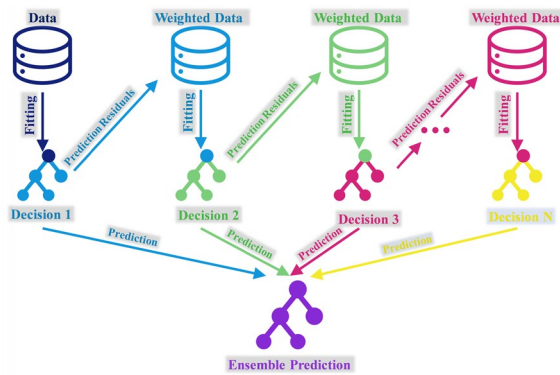


Fig. 3 GBDT Model Structure Diagram

such as PSO often face limitations. They may converge prematurely or require extensive parameter tuning. AEA avoids these issues by introducing adaptive update rules. The alpha operator diversifies the population by generating new basis vectors and adaptive step sizes. The evolutionary path mechanism accumulates historical information to guide the search. The matrix generation mechanism recombines candidate solutions to cover the feasible space more effectively. AEA is therefore well suited for complex engineering problems with multimodal objectives and irregular search spaces.

The search process starts with initialization of candidate solutions:

$$X_i^0 = lb + (ub - lb) \cdot rand(0,1) \quad (1)$$

where X_i^0 is the i -th solution, and lb and ub are the lower and upper bounds of the domain.

The **alpha operator** then updates each solution adaptively:

$$X_i^{t+1} = X_i^t + \alpha \cdot M \cdot (X_r^t - X_s^t) \quad (2)$$

where α is a control factor, M is an evolution matrix, X_r^t and X_s^t are randomly selected solutions. This step maintains diversity and introduces directed search.

The **evolutionary path mechanism** accumulates information across generations:

$$p^{t+1} = \lambda p^t + \delta (X_i^t - \bar{X}^t) \quad (3)$$

where p^t is the path vector, λ is a damping factor, δ is a learning coefficient, and $\bar{X}^t$ is the mean solution at iteration t . This strategy improves adaptability and reduces local stagnation.

By combining these mechanisms, AEA achieves both exploration and refinement. It also requires fewer parameters than GA or DE. It shows strong robustness in high-dimensional, nonlinear, and multimodal problems. In this study, AEA was applied to optimize the hyperparameters of the GBDT model. For further mathematical details, readers are referred to Gao and Zhang (2024).

3.3 Sine cosine algorithm (SCA)

Sine cosine algorithm (SCA), proposed by Mirjalili, is a metaheuristic global optimization algorithm based on trigonometric (sine and cosine) optimization mechanisms (Mirjalili 2016). The method is based on mathematical properties of sine and cosine functions. It guides search agents along oscillatory trajectories. This design enables a flexible switch between global exploration and local exploitation. The algorithm initializes a set of candidate solutions in the feasible domain. Each solution is updated according to sine and cosine operators:

$$X_i^{t+1} = X_i^t + r_1 \times \sin(r_2) \times |r_3 P_i^t - X_i^t| \quad (4)$$

$$X_i^{t+1} = X_i^t + r_1 \times \cos(r_2) \times |r_3 P_i^t - X_i^t| \quad (5)$$

where X_i^t is the current solution, and P^t is the best solution found so far. r_1, r_2, r_3 are random control parameters. These parameters adjust the search step and direction. When $\sin$ or $\cos$ values are positive, agents move toward the best solution. When values are negative, they move away to enhance exploration. This oscillatory mechanism helps the algorithm escape local optima. The control parameters gradually decrease during iterations, which shifts the focus from exploration to exploitation.

Compared with other metaheuristic algorithms, SCA is simpler and requires fewer parameters. It has strong ability to cover the search space in early iterations. It also shows competitive performance in multimodal and engineering optimization problems. In this study, SCA was applied to optimize the hyperparameters of the GBDT model. Its oscillatory update rules provide a complementary search strategy to AEA and AOA (Mirjalili 2016).

3.4 Arithmetic optimization algorithm (AOA)

The Arithmetic optimization algorithm (AOA) is a novel metaheuristic optimization algorithm founded on quadratic mixed operations in arithmetic. It was introduced by Abualigah et al. (2021). The method is inspired by the basic arithmetic operators: addition, subtraction, multiplication, and division. These operators are used to guide the search agents toward global or local regions. The search process begins with a set of candidate solutions. Each solution is updated using arithmetic operators and adaptive control parameters. In the exploration phase, multiplication and division expand the search space:

$$X_i^{t+1} = X_i^t \times (r_1 \cdot \frac{U_b - L_b}{2}) \quad (6)$$

while in the exploitation phase, addition and subtraction refine the solutions near promising regions:

$$X_i^{t+1} = X_i^t \pm r_2 \cdot (P^t - X_i^t) \quad (7)$$

where X_i^t is the current solution, P^t is the best solution so far, r_1 and r_2 are adaptive random parameters. L_b and U_b denote the lower and upper bounds of the search space.

The algorithm employs a Mathematical Optimizer Accelerated function to control the transition between exploration and exploitation. In early iterations, exploration dominates through multiplication and division. In later iterations, exploitation dominates through addition and subtraction. This dynamic adjustment allows AOA to maintain diversity while improving convergence speed. Compared with other metaheuristics, AOA is simple to implement and requires few parameters. In this study, AOA was used to optimize the GBDT model. Its arithmetic-based strategy complements the evolutionary adaptation of AEA and the oscillatory trajectories of SCA. Together, these algorithms provide diverse and robust optimization approaches. For detailed formulations, readers are referred to Abualigah et al. (2021).

3.5 GBDT-based hybrid models

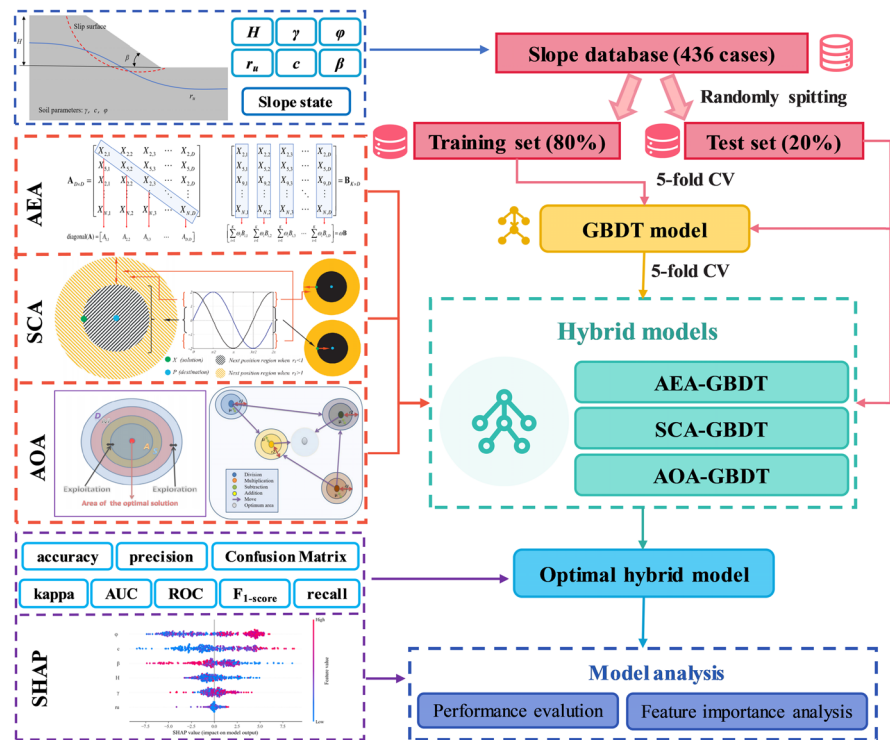
The study aims to optimize the GBDT model for slope stability prediction. For this purpose, three metaheuristic optimization algorithms, AEA, SCA, and AOA are

employed to optimize combination of hyperparameters for the GBDT model, yielding AEA-GBDT, SCA-GBDT, and AOA-GBDT. The modeling process of AEA-GBDT, SCA-GBDT and AOA-GBDT is shown in Fig. 4. The hybrid models development process is as follows: (i) Dataset division: The dataset was split 80% for training and 20% for testing, and the same split was used for every hybrid model to ensure comparability; (ii) Population initialization: When using the three metaheuristic optimization algorithms, the hyperparameters of the GBDT were first randomly generated as the initial solution. Considering the number of populations, each hybrid model (AEA-GBDT, SCA-GBDT, and AOA-GBDT) was developed for different population sizes; (iii) Iterative loop: This step iteratively searches for the optimal combination of hyperparameters for the GBDT model by calculating the fitness function and performing cross-validation (CV). This process continues until the maximum number of iterations was reached. (iv) Obtaining the optimal hyperparameter combination: The hyperparameter combination of each hybrid GBDT model was determined by iteration of the optimization algorithms. The optimal hybrid GBDT model was identified by combining multiple evaluation metrics through the Friedman-Nemenyi test. (v) Model interpretability analysis: The interpretability of the optimal GBDT was examined with SHAP to quantify global importance and case-level contributions. Overall, this study proposed three metaheuristic algorithms to optimize the GBDT model. The results demonstrated the superior performance of the hybrid model in slope stability prediction.

3.6 Models evaluation criteria

Model evaluation is a crucial step following model development. In this study, accuracy, precision, recall, F_1 -score, and Cohen's kappa were selected as evaluation metrics (Qiu and Zhou 2023). These five metrics provide complementary perspectives. Accuracy reflects overall correctness. Precision and recall capture class specific reliability. F_1 measures the balance between precision and recall. Cohen's kappa accounts for agreement beyond chance. Together, they form a standard and widely accepted set of indicators for evaluating classification models. All of them are positive measures, meaning that higher values indicate better model performance. Each metric is calculated using the confusion matrix's four fundamental quantities: true positives (TP), true negatives (TN), false positives (FP), and false

Fig. 4 The establishment process of the hybrid models



negatives (FN). TN and TP denote the number of correctly classified negative and positive samples, respectively, FN and FP refer to the number of mis-classified negative and positive samples, respectively.

Accuracy measures the proportion of samples correctly predicted out of the total samples, expressed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

However, when the dataset is unbalanced, accuracy may not adequately reflect the model's performance, necessitating the consideration of additional metrics. Precision is the proportion of samples predicted to be in the positive category that are actually in the positive category:

$$Precision = \frac{TP}{TP + FP} \quad (9)$$

Recall is the proportion of samples that are actually positive classes that are correctly predicted as positive:

$$Recall = \frac{TP}{TP + FN} \quad (10)$$

Thus, precision and recall are two complementary metrics. The $F_{1-score}$, which is the harmonic mean of precision and recall, is used to measure the model's balance between precision and recall:

$$F_{1-Score} = \frac{2 \times (Precision * Recall)}{Precision + Recall} \quad (11)$$

The Cohen's kappa coefficient is used to measure the agreement between the predictions and the actual results:

$$Kappa = \frac{p_o - p_e}{1 - p_e} = \frac{N \sum_{i=1}^n X_{ii} - \sum_{i=1}^n (X_{i+} \times X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} \times X_{+i})} \quad (12)$$

where, n denotes the category, N represents the sum of the number of categories, X_{ii} denotes the diagonal element of the error matrix, X_{i+} denotes the sum of columns of the category, and X_{+i} denotes the sum of rows of the category.

In addition to the above five metrics, the receiver operating characteristic (ROC) curve and its **area under the curve (AUC)** assess discrimination independent of the decision threshold (Zhou et al. 2022). ROC plots the true-positive rate (TPR) against the false-positive rate (FPR) across thresholds, and larger AUC indicates better

discrimination. The horizontal axis of the ROC curve is the FPR, and the vertical axis is the TPR. FPR and TPR are represented by Eqs. (24) and (25):

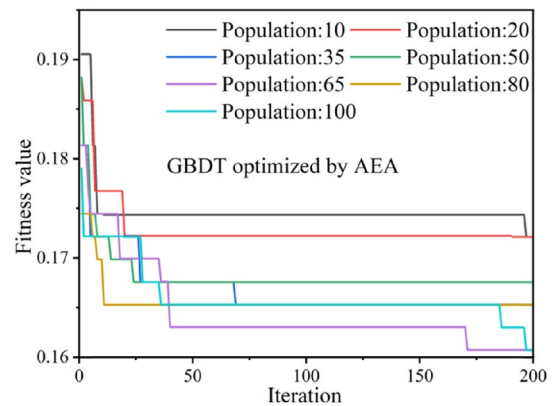
$$TPR = \frac{TP}{TP + FN} \quad (13)$$

$$FPR = \frac{FP}{TP + TN} \quad (14)$$

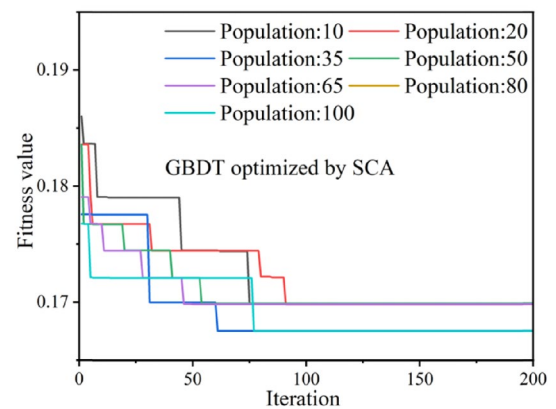
4 Results and discussion

4.1 Hyperparameters tuning

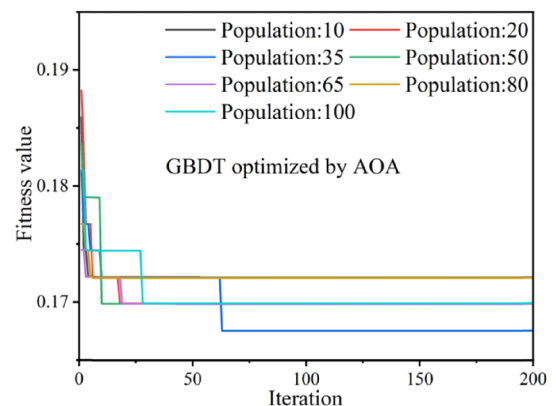
Due to its reliable performance, GBDT was used as the predictive model for slope stability in this study. To enhance model accuracy and robustness in identifying the nonlinear patterns of the data, the GBDT model was optimized using three different algorithms: AEA, SCA, and AOA. The hyperparameters of GBDT that were optimized included *N_estimators*, *Learning_rate*, *Max_depth*, *Min_samples_split*. This section discusses the optimization process and results for AEA-GBDT, SCA-GBDT, and AOA-GBDT models across seven different population sizes (10, 20, 35, 50, 65, 80, 100). Each training setup consisted of 200 iterations, and five-fold CV was employed to ensure the generalization capability of the models. Seven fitness curves were obtained for each hybrid model, as shown in Fig. 5. The similarity of the convergence curves of the three hybrid models verifies the effectiveness and consistency of this study in determining the optimal population size. Larger population sizes accelerated convergence rates but increased computational costs. Larger population sizes also result in longer computation times, necessitating a balance between model convergence speed and computation time in practical applications. The three hybrid models converge rapidly within the initial 20 iterations. While AEA and SCA exhibit faster iteration convergence speeds, this does not directly translate to computational speed, as SCA and AOA often requires more computational time per iteration. Furthermore, the AEA-GBDT hybrid model consistently finds the lowest fitness value. It should be highlighted that determining the optimal population size based solely on the convergence curve is insufficient.



(a) GBDT optimized by AEA



(b) GBDT optimized by SCA



(c) GBDT optimized by AOA

Fig. 5 Fitness curves of GBDT optimized by three optimization algorithms

A more integrated and accurate assessment method is required to make a well-informed decision.

In order to compare the performance of GBDT models under different optimization strategies, traditional evaluation metrics were used in this study. The specific performance of AEA-GBDT, SCA-GBDT, and AOA-GBDT hybrid models under different optimization strategies was shown in Table 2. In Table 2, the bold values indicate the best performance of each model across all population sizes for the corresponding evaluation metrics. From the results in Table 2, it can be seen that the AEA-GBDT for a population size of 65 produces better test set results than other population sizes, with evaluation metrics of accuracy=0.920, precision=0.934, recall=0.914, $F_{1\text{-score}}$ =0.924, and Cohen's kappa=0.840. The base GBDT model shows the lowest performance according to the results in Table 2, with evaluation metrics of accuracy=0.829, precision=0.847, recall=0.829, $F_{1\text{-score}}$ =0.838, and Cohen's kappa=0.658. This is evident from the results of the other hybrid models, which demonstrate that the different optimization algorithms have significantly improved the

performance of the GBDT model. As shown in Table 2, the training times of all models are reported. Larger population sizes improved model performance but also increased training time. Among the three optimizers, AEA-GBDT required the least time at comparable settings. It should be noted that although training is time-consuming, the prediction process runs at the second level, ensuring practical applicability.

Our initial evaluation metrics suggested that AEA(65)-GBDT outperformed other hybrid models in most aspects. However, to rigorously assess the statistical significance of these performance differences, we employed the methodology recommended by Demšar (2006). This approach involves two key statistical analyses: the non-parametric Friedman test and the Nemenyi post-hoc test. The Friedman test, analogous to a repeated-measures non-parametric ANOVA, was used to evaluate the overall performance of multiple hybrid models in predicting slope stability. This test is particularly suitable for our study as it does not assume normality of the data and can handle multiple comparisons effectively. In the Friedman test, models with better

Table 2 Classification performance of GBDT model after optimization with different strategies

Model	Population	Training time	Accuracy	Precision	Recall	$F_{1\text{-score}}$	Kappa
GBDT	—	4s	0.829	0.847	0.829	0.838	0.658
AEA-GBDT	10	830s	0.875	0.891	0.872	0.881	0.749
	20	1520s	0.875	0.891	0.872	0.881	0.749
	35	2250 s	0.909	0.933	0.893	0.913	0.817
	50	3470s	0.886	0.893	0.893	0.893	0.771
	65	4130 s	0.920	0.934	0.914	0.924	0.840
	80	5650s	0.897	0.931	0.872	0.901	0.795
	100	7120s	0.909	0.933	0.893	0.913	0.817
SCA-GBDT	10	1600 s	0.897	0.886	0.936	0.907	0.790
	20	3020 s	0.863	0.906	0.829	0.866	0.727
	35	4240 s	0.886	0.877	0.914	0.895	0.771
	50	6880 s	0.909	0.933	0.893	0.913	0.817
	65	8230 s	0.886	0.877	0.914	0.895	0.771
	80	11240 s	0.875	0.891	0.872	0.881	0.749
	100	14390 s	0.875	0.891	0.872	0.881	0.749
AOA-GBDT	10	1180s	0.852	0.886	0.829	0.857	0.704
	20	1970s	0.886	0.893	0.893	0.893	0.771
	35	3380s	0.886	0.893	0.893	0.893	0.771
	50	4730s	0.897	0.913	0.893	0.903	0.794
	65	6080s	0.886	0.893	0.893	0.893	0.771
	80	7450s	0.863	0.857	0.893	0.875	0.725
	100	9600 s	0.909	0.913	0.914	0.913	0.817

performance receive higher rankings. Following the Friedman test, we applied the Nemenyi test to identify specific significant differences between the constructed slope stability prediction models. The Nemenyi test also allows for the calculation of the critical difference (CD) between model performances, providing a threshold for statistical significance.

The Friedman-Nemenyi test statistics are calculated as follows:

$$\chi^2_F = \frac{12N}{\lambda(\lambda+1)} \left[\sum_{j=1}^{\lambda} R_{0j}^2 - \frac{\lambda(\lambda+1)^2}{4} \right] \quad (15)$$

$$CD = q_{\alpha} \times \sqrt{\frac{\lambda(\lambda+1)}{6N}} \quad (16)$$

where λ represents the number of models, N represents the amount of data, R_{0j}^2 represents the rank of the j th model, and q_{α} is the critical value in the Nemenyi test, which depends on the number of slope cases and the significance level α . The Friedman-Nemenyi test complement traditional evaluation indicators such as accuracy and precision, offering a robust statistical method for model comparison. Figure 6 provides a visual representation of the hybrid

models' performance rankings under different measurement optimizations, derived from these statistical tests. Analysis of Fig. 6 reveals that AEA(65)-GBDT demonstrated superior overall performance at $\alpha=0.05$ ($CD=1.41$) on the test set. This finding indicated that AEA(65)-GBDT not only outperforms other AEA-GBDT models but also surpasses the SCA-GBDT and AOA-GBDT models in comprehensive model performance. Table 3 presents the optimal population size and hyperparameter combinations for the GBDT model under the three optimization methods (AEA, SCA, and AOA). The superiority of AEA-GBDT can be attributed to the ability of AEA to balance exploration and exploitation in the search space, and its effectiveness in avoiding local optima.

4.2 Analysis on model performance

The confusion matrix is a key tool for evaluating the accuracy of classification models, providing an intuitive representation of the disparity between predicted and actual values. In the test set of 88 samples, there are 41 stable and 47 unstable samples, as illustrated in Fig. 7. The more consistent the predicted labels from the hybrid models are with the true labels, the better the model's performance. The values along

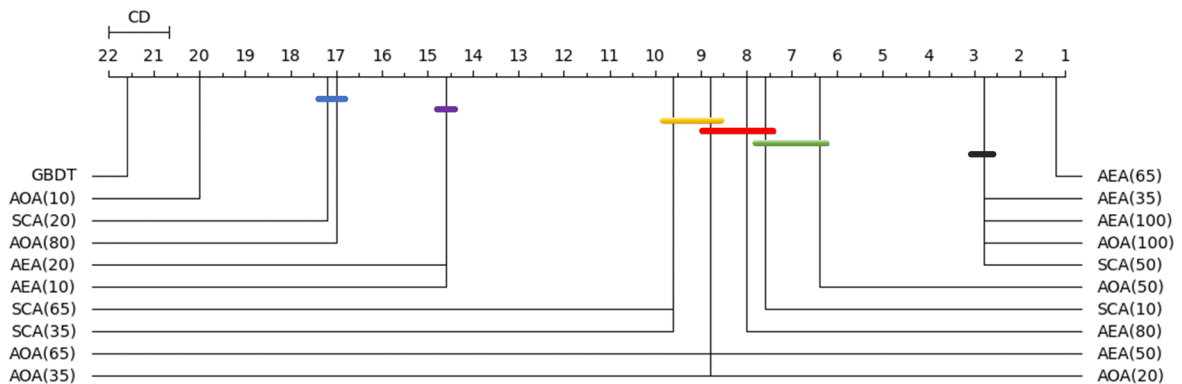


Fig. 6 Comprehensive comparison of different hybrid models

Table 3 Optimal parameter combination settings for three hybrid models

Parameters	Population size	N_estimators	learning_rate	max_depth	min_samples_split
Range	[10–100]	[1–100]	[0.001–0.5]	[0–100]	[2–50]
AEA-GBDT	65	100	0.102	19	13
SCA-GBDT	50	69	0.1779	19	10
AOA-GBDT	100	49	0.094	16	20

the diagonal of the matrix represent the number of correct predictions for each class (stable and unstable). The basic GBDT model correctly predicted 73 samples. For the hybrid models AEA-GBDT, SCA-GBDT, and AOA-GBDT, the predictions of 81, 80, and 80 samples, respectively, were consistent with the true labels. Further analysis found that the prediction results of the three hybrid models were similar, but much better than basic GBDT, XGBoost, RF, etc. Among them, the AEA-GBDT hybrid model had the best overall performance.

In addition to the confusion matrix, the receiver operating characteristic (ROC) curve and the area under curve (AUC) are also valuable tools for evaluating model performance, as illustrated in Fig. 8. From the various ROC curves in Fig. 8, it can be observed that the performance differences between the three hybrid models are minimal. The ROC curve of the AEA-GBDT model is the closest to the upper-left corner, indicating its superior performance. The AUC of the AEA-GBDT model is the highest among all models, with a value of 0.960. The SCA-GBDT and AOA-GBDT models also show strong performance. This further supports the significance of the three optimization strategies in improving slope stability assessment.

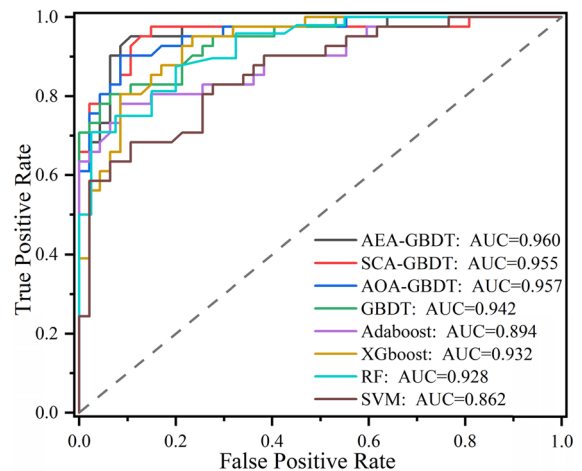


Fig. 8 Comparison of ROC curves and AUC of different models

4.3 Shapley additive explanations (SHAP)

In slope stability prediction, understanding the model's decision-making process is essential for enhancing its credibility and reliability. To gain deeper insights into the internal mechanisms and feature contributions of the model, this study applies SHAP (Delgado-Panadero et al. 2022). Rooted in the Shapley value theory of cooperative game theory, SHAP provides a consistent and equitable quantification of

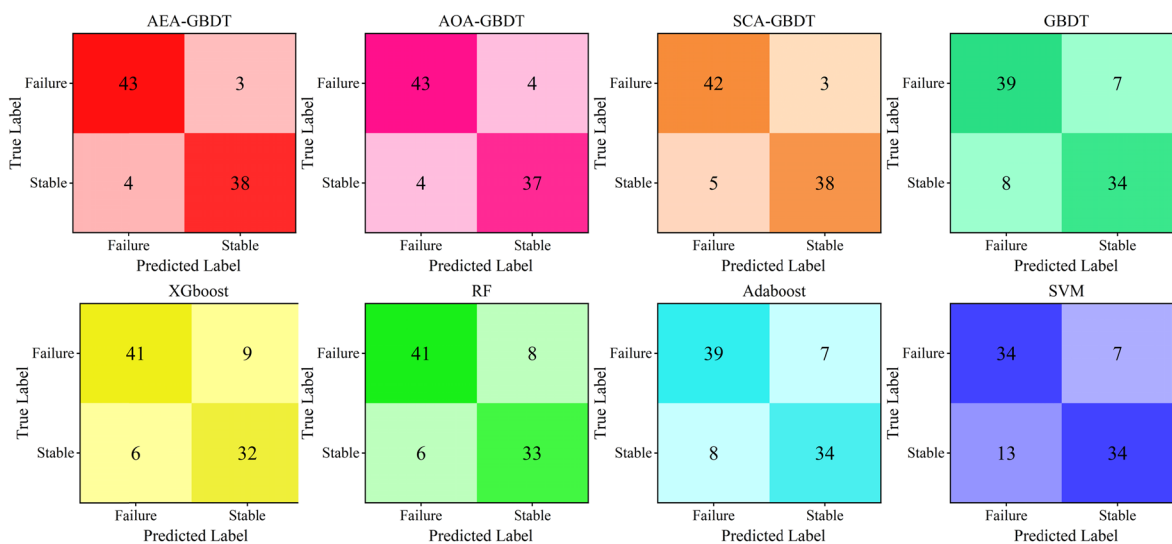


Fig. 7 Confusion matrix of different models on test data

each input feature's influence on the model's output (Parsa et al. 2020). Beyond identifying the most significant features for slope stability prediction, SHAP also elucidates how each feature affects the prediction results (Mangalathu et al. 2020). In this study, SHAP was utilized to analyze the optimal AEA-GBDT hybrid model.

As shown in Fig. 9, the SHAP force diagram provides explanatory insights into four sample cases of the model. Each feature in each sample is assigned a SHAP value, which quantifies its contribution to the model's prediction in a specific instance. The base value in the figure represents the model's initial prediction, while $f(x)$ denotes the final output after adjustments based on individual feature contributions. In the diagram, pink bars indicate features that positively influence the predicted value, whereas blue bars represent features that negatively impact it. The length of each segment reflects the magnitude of the feature's influence. A positive SHAP contribution suggests that the model is more inclined to classify the slope as stable. Among the analyzed cases, c has the most significant impact on the prediction outcome in Case A, while φ exerts the greatest influence in Cases B, C, and D. Conversely, r_u has the least impact in Cases A, B, and D. These results are consistent with classical

geotechnical principles, where cohesion and friction angle dominate shear strength and slope stability, while pore pressure reduces effective stress and weakens resistance.

Figure 10 presents the SHAP summary plot, which consolidates feature contribution rank across all samples. The horizontal axis represents the SHAP value magnitude, with the width of the plot indicating the density of data points. The color gradient, transitioning from blue to red, corresponds to increasing feature values. Specifically, larger φ and higher c correspond to larger SHAP values, indicating positive contribution to slope stability. In contrast, steeper and taller slopes (H and β) tend to reduce stability. Although r_u appears least influential in the standalone ranking, this does not imply insignificance in practice. According to Terzaghi's effective stress principle ($\sigma' = \sigma - u$), increasing r_u reduces the effective stress σ' , which reduces the mobilized shear strength commonly expressed via c' and φ' in the Mohr–Coulomb strength criterion ($\tau = c + \sigma' \tan \varphi$) (Terzaghi et al. 1996; Labuz and Zang 2014). Consequently, part of r_u 's impact is captured indirectly through its interactions with c and φ , leading to a smaller isolated SHAP value (Karir et al. 2022; Ahmad et al. 2023). The relatively low importance of γ likely

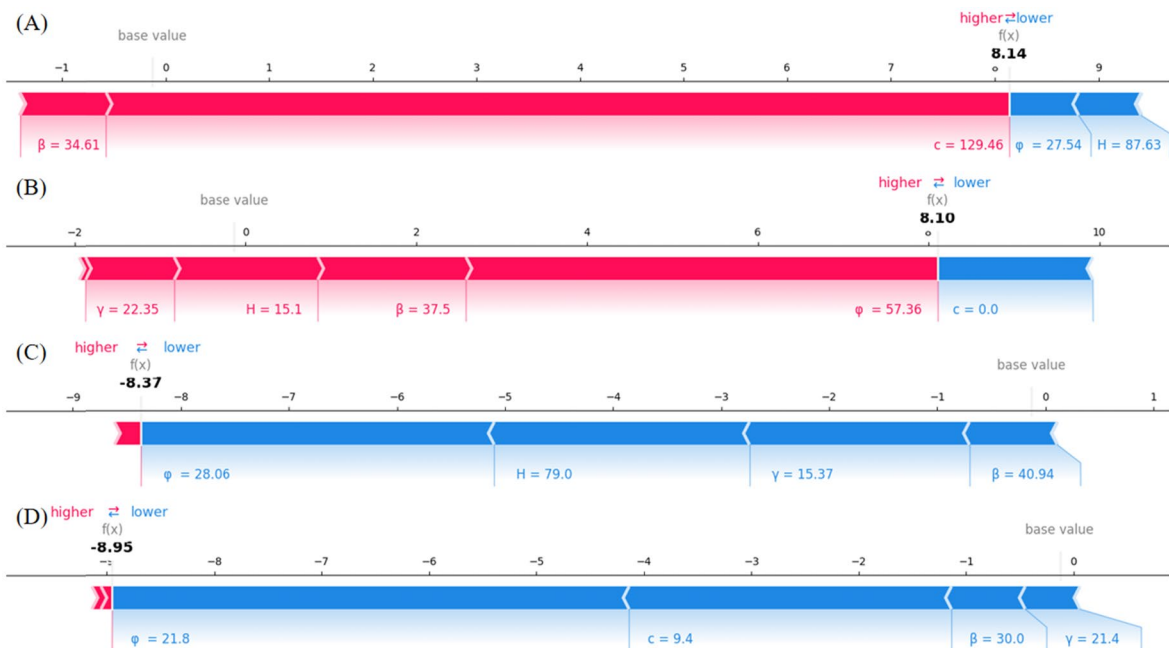
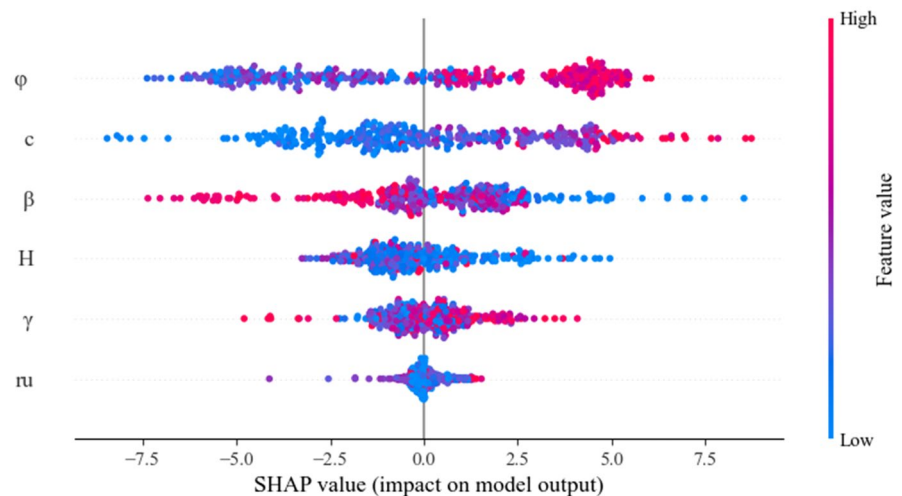


Fig. 9 SHAP analysis of four samples of the optimal mixed model

Fig. 10 The SHAP summary plot of AEA-GBDT model

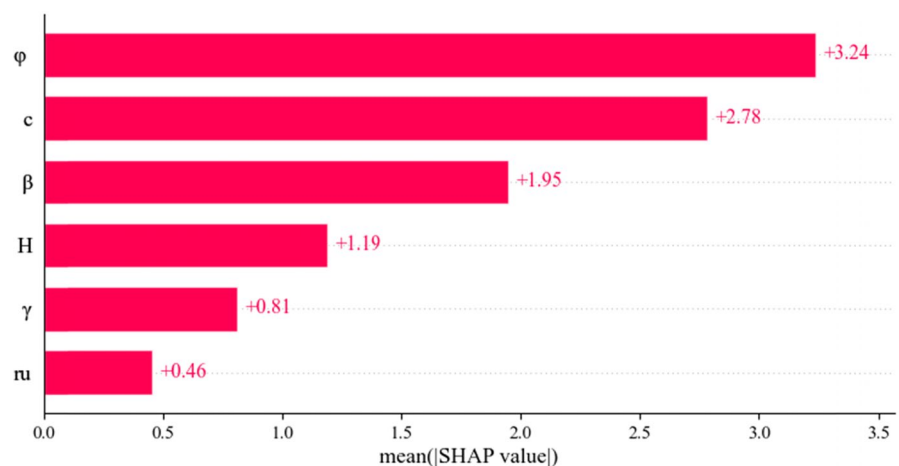


reflects its narrow range in this dataset, not its lack of physical relevance. γ governs the self-weight of the slope and interacts with shear strength, and previous studies have shown it to be an important factor under different conditions. The SHAP values of all samples of each feature are absolutely averaged and ranked according to the feature importance as shown in Fig. 11. It shows that ϕ , c and β are the dominant factors with average SHAP values of 3.24, 2.78 and 1.95, respectively, while H , γ and r_u have smaller contributions of 1.19, 0.81 and 0.46. These results highlight the consistency between model interpretation and geotechnical principles, while also revealing dataset-driven limitations.

The SHAP dependency plot illustrated in Fig. 12. The horizontal axis represents the feature value,

while the vertical axis denotes the corresponding SHAP value. Similar to a partial dependency plot, the SHAP dependency plot accounts for interaction effects among features and is defined only within the input space supported by the data. Several key observations emerge from the analysis. When c is low, the vast majority of samples exhibit SHAP values below 0, indicating an unstable state. Conversely, when c exceeds 50 kPa, most cases have SHAP values greater than 0, signifying a stable state. For β , when it exceeds 40° , a substantial number of samples display low SHAP values, reflecting a sharp increase in the proportion of unstable slopes. Additionally, a strong interaction is observed between r_u and c . According to Terzaghi's effective stress principle and Mohr–Coulomb strength criterion, higher r_u reduces

Fig. 11 SHAP absolute value average ranking of features of AEA-GBDT



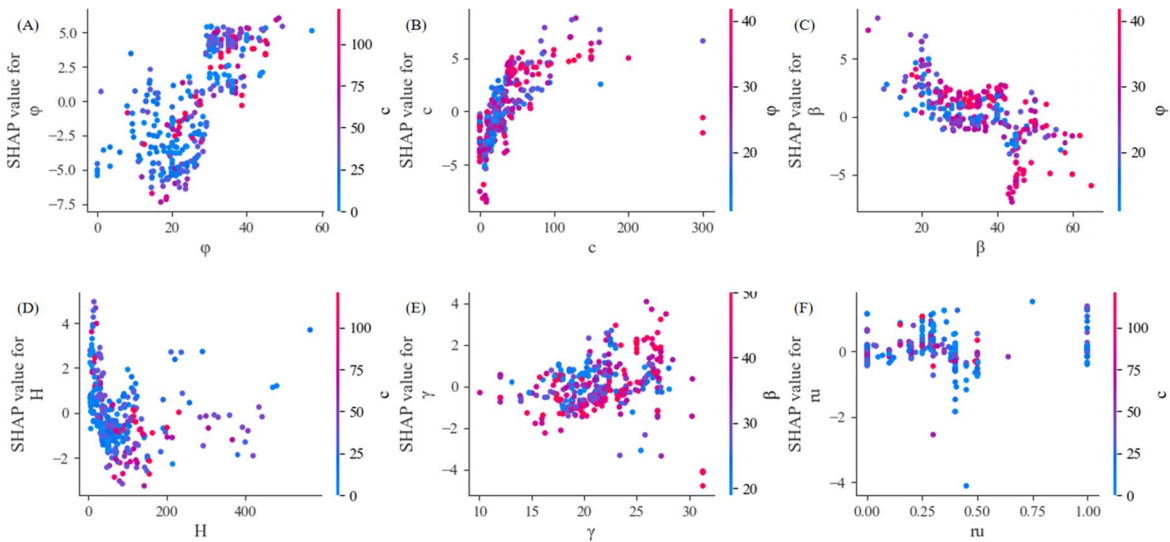


Fig. 12 SHAP dependency plot of the AEA-GBDT model

the effective normal stress σ' , which lowers the mobilized shear strength (often represented via the effective parameters c' and ϕ') (Terzaghi et al. 1996). Part of r_u 's influence therefore manifests through its interactions with c and ϕ . The link between strength reduction and instability is not absolute or linear, highlighting the nonlinear nature of these relationships (Lawal et al. 2024).

4.4 Comparison with previous studies

Comparing the proposed models with peer approaches is an effective way to demonstrate advancement. Table 4 compiles representative studies on soil slopes with circular slip failure under comparable problem settings. Because the database underpins model learning, studies with very small case bases were excluded. All included works adopt the same input indicators; only the study by Khanh Pham omitted r_u . Under these harmonized conditions, the metaheuristic optimized models AEA-GBDT, SCA-GBDT, and AOA-GBDT all achieve accuracies above 0.90, with AEA-GBDT reaching 0.92. Across most remaining metrics, the proposed models show

competitive performance, highlighting the effectiveness of metaheuristic optimizers such as AEA, SCA, and AOA for this task. Future research can integrate additional state-of-the-art optimizers with richer ensemble strategies to further enhance intelligent prediction and early warning of slope stability.

4.5 Model application

To further test the model, an out-of-sample verification was performed. 16 loess slope cases were compiled from several slope engineering projects (Gao 2011). These 16 loess slopes are independent of the 436-case database and were compiled from separate engineering reports. The AEA-GBDT model was then applied to these 16 cases. The comparison between the model prediction results and the literature results is shown in Table 5. The results show that the AEA-GBDT correctly predicted the stability outcomes for all 16 slope cases. The parameter range of the cases is large, which well verifies the external validation of the model. Together with the model results reported earlier, this case-based verification highlights the competitiveness and engineering practicality of AEA-GBDT for slope stability prediction.

Table 4 Results of relevant previous studies

Literature	Model	Data size	Accuracy	Precision	Recall	F ₁ -score	Kappa	AUC
Qi	SVM	148	0.860	–	–	–	–	0.967
	GBM	148	0.930	–	–	–	–	0.962
Zhou	GBM	221	0.865	–	–	–	0.832	0.900
	RF	221	0.801	–	–	–	0.624	0.897
Lin	GSA	132	0.889	–	–	–	0.778	0.889
	RF	132	0.833	–	–	–	0.667	0.833
Nhat-Duc Hoang	MO-LSVC	168	0.863	–	–	–	–	–
Khanh Pham	XGB-CM	153	0.903	–	–	0.914	–	0.950
Navid Kardani	Stacking model	150	0.910	–	–	0.905	–	0.904
Lin	Stacking model	444	0.847	–	–	–	0.691	0.945
Wang	SSA-RF	444	0.910	0.950	0.864	0.905	–	0.940
	HHO-RF	444	0.910	0.950	0.864	0.905	–	0.945
Satyam Tiwari	Adaboost	229	0.970	0.990	0.948	0.960	–	–
	XGboost	229	0.970	0.970	0.946	0.970	–	–
Huajin Zhang	MDMSE	422	0.881	–	–	0.904	–	–
this study	AEA-GBDT	436	0.920	0.934	0.914	0.924	0.840	0.960
	SCA-GBDT	436	0.909	0.933	0.893	0.913	0.817	0.955
	AOA-GBDT	436	0.909	0.913	0.914	0.913	0.817	0.957

Table 5 Slope stability prediction results for slope projects

Case	γ	c	φ	β	H	r_u	Actual status	AEA-GBDT
1	17.33	105.80	27.05	17.46	81.32	0.00	Stable	Stable
2	17.90	72.11	24.11	18.38	93.16	0.00	Stable	Stable
3	18.09	63.68	27.47	21.96	41.84	0.00	Stable	Stable
4	18.68	42.63	22.84	19.46	49.74	0.00	Stable	Stable
5	18.41	51.05	23.26	38.22	33.95	0.00	stable	stable
6	17.28	110.00	22.84	25.05	85.26	0.00	Stable	Stable
7	18.09	63.68	24.95	27.03	81.32	0.00	Stable	Stable
8	17.52	93.16	27.47	21.96	77.37	0.00	Stable	Stable
9	17.74	80.53	24.53	21.96	105.00	0.25	Stable	Stable
10	17.66	84.74	24.53	27.03	77.37	0.25	Stable	Stable
11	18.30	55.26	24.53	31.69	73.42	0.25	Failure	Failure
12	17.99	67.89	24.11	31.69	93.16	0.25	Failure	Failure
13	17.39	101.60	29.16	31.69	65.53	0.50	Stable	Stable
14	18.41	51.05	24.53	27.15	33.95	0.50	Stable	Stable
15	18.19	59.47	22.84	23.41	49.74	0.50	Failure	Failure
16	18.09	63.68	22.42	21.96	73.42	0.50	Failure	Failure

5 Limitations and future work

This study developed three hybrid models for slope stability prediction, achieving high accuracy with a dataset of 436 samples. Nevertheless, some limitations remain. First, despite using a comprehensive

database, several important factors such as seismic activity, rainfall, and subsurface stratigraphy were not included due to data unavailability; incorporating these factors in the future would enhance generalizability across diverse conditions. Second, slope instability is inherently a time-dependent process, and

future research should integrate time-series analysis and IoT-based monitoring. Deploying distributed sensor networks to collect data on rainfall, pore pressure, and seismic activity across multiple zones of a slope project will enable the model to adapt to real-time environmental changes, while robust pre-processing will be required to address sensor noise and missing records. Third, while the present framework focused on binary classification, extending it to multi-class labels could better align with engineering decision thresholds. Fourth, although the dataset was curated to avoid abnormal records, future work should further examine the robustness of the model under extreme-value conditions, which are common in geotechnical practice. Finally, incorporating physical knowledge into the model's loss function would further ensure that the framework captures both data-driven patterns and the established geomechanical principles governing slope instability, thereby improving its robustness and engineering relevance.

6 Conclusion

This study proposes three hybrid models for **slope stability prediction**, integrating metaheuristic algorithms (AEA, SCA and AOA) with GBDT. A comprehensive database of 436 slope **cases** was constructed, which was used to train and test the models. And all models underwent five-fold CV. The models were evaluated using accuracy, precision, recall, $F_{1\text{-score}}$, and Cohen's kappa, with additional analysis via ROC curves and AUC value. Additionally, SHAP were employed to interpret the model's predictions. The following conclusions can be drawn from the results obtained from this study:

1. The AEA-GBDT model performed excellently, achieving an accuracy of 0.920, a precision of 0.934, a recall of 0.914, an $F_{1\text{-score}}$ of 0.924, and a Cohen's kappa of 0.840. SCA-GBDT also showed strong performance with an accuracy of 0.909, a precision of 0.933, a recall of 0.893, an $F_{1\text{-score}}$ of 0.913, and a Cohen's kappa of 0.817. AOA-GBDT achieved an accuracy of 0.909, a precision of 0.913, a recall of 0.914, an $F_{1\text{-score}}$ of 0.913, and a Cohen's kappa of 0.817. The results indicated that AEA, SCA, and AOA improved

the performance of the base GBDT model by 11%, 9%, and 9%, respectively.

2. The SHAP analysis revealed the importance ranking of the features in AEA-GBDT model. SHAP ranked c and φ as dominant factors followed by H , β and γ . Although r_u showed lower standalone importance, increases in r_u reduce the effective normal stress σ' and thus the mobilized shear strength, so part of its impact appears through interactions with c and φ .
3. The Friedman-Nemenyi test was applied for a more objective evaluation of the model's overall performance. This statistical analysis confirmed that AEA-GBDT significantly outperforms SCA-GBDT, while SCA-GBDT and AOA-GBDT performed similarly. The AEA-GBDT model not only outperforms the baseline GBDT, RF, AdaBoost, and SVM models, but also demonstrates superior performance when compared with existing methods, further validating its advantages. Finally, the AEA-GBDT model was tested on 16 real-world loess slope cases. The model correctly predicted the stability outcomes for all 16 cases, demonstrating its strong external applicability. The broad parameter range of these cases underscores the model's ability to generalize across diverse conditions.

In conclusion, this study established a new benchmark for intelligent slope stability assessment, combining cutting-edge AI techniques with engineering-driven interpretability. The proposed framework not only improves predictive accuracy but also fosters trust in AI-driven solutions, paving the way for safer and more sustainable infrastructure development in slope stability.

Author contributions Conceptualization, MW and NW; methodology, GZ and NW; software, NW; validation, NW, MP, TP and ST; formal analysis, MW; investigation, MW; resources, MW, GZ, MW; data curation, NW; writing—original draft preparation, NW; writing—review and editing, NW, MP, MW, TP and ST; visualization, TP; supervision, MP, GZ; project administration, GZ; funding acquisition, NW, GZ. All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by the Education Department Foundation of Hunan Province (CX20230167).

Data availability Data will be made available on request.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

- Abualigah L, Diabat A, Mirjalili S et al (2021) The arithmetic optimization algorithm. *Comput Methods Appl Mech Eng* 376:113609. <https://doi.org/10.1016/j.cma.2020.113609>
- Acevedo AMG, Passini LDB, Talamini AA et al (2021) Assessing limit equilibrium method approach and mapping critical areas for slope stability analysis in Serra do Mar Paranaense—Brazil. *Environ Earth Sci* 80:572. <https://doi.org/10.1007/s12665-021-09863-5>
- Ahmad F, Tang X, Hu J et al (2023) Improved prediction of slope stability under static and dynamic conditions using tree-based models. *Comput Model Eng Sci* 137:455–487. <https://doi.org/10.32604/cmes.2023.025993>
- Ahmed B (2015) Landslide susceptibility modelling applying user-defined weighting and data-driven statistical techniques in Cox's Bazar Municipality, Bangladesh. *Nat Hazards* 79:1707–1737. <https://doi.org/10.1007/s11069-015-1922-4>
- Ahour M, Hataf N, Azar E (2020) A mathematical model based on artificial neural networks to predict the stability of rock slopes using the generalized Hoek-Brown failure criterion. *Geotech Geol Eng* 38:587–604. <https://doi.org/10.1007/s10706-019-01049-y>
- Aminpour M, Alaie R, Khosravi S et al (2023) Slope stability machine learning predictions on spatially variable random fields with and without factor of safety calculations. *Comput Geotech* 153:105094. <https://doi.org/10.1016/j.compgeo.2022.105094>
- Anuragi SK, Kishan D (2025) Analysis of slope stability using innovative hybrid BPSO-SVM machine learning techniques for enhance environmental sustainability. *Metall Mater Eng* 31:718–730
- Bardhan A, Samui P (2022) Probabilistic slope stability analysis of heavy-haul freight corridor using a hybrid machine learning paradigm. *Transp Geotech* 37:100815. <https://doi.org/10.1016/j.trgeo.2022.100815>
- Chakraborty A, Goswami D (2017) Prediction of slope stability using multiple linear regression (MLR) and artificial neural network (ANN). *Arab J Geosci* 10:385. <https://doi.org/10.1007/s12517-017-3167-x>
- Chen G, Deng W, Lin M, Lv J (2023) Slope stability analysis based on convolutional neural network and digital twin. *Nat Hazards* 118:1427–1443. <https://doi.org/10.1007/s11069-023-06055-1>
- Cheng M-Y, Hoang N-D (2015) Typhoon-induced slope collapse assessment using a novel bee colony optimized support vector classifier. *Nat Hazards* 78:1961–1978. <https://doi.org/10.1007/s11069-015-1813-8>
- Das SK, Biswal RK, Sivakugan N, Das B (2011) Classification of slopes and prediction of factor of safety using differential evolution neural networks. *Environ Earth Sci* 64:201–210
- Delgado-Panadero Á, Hernández-Lorca B, García-Ordás MT, Benítez-Andrades JA (2022) Implementing local-explainability in gradient boosting trees: feature contribution. *Inf Sci* 589:199–212. <https://doi.org/10.1016/j.ins.2021.12.111>
- Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 7:1–30
- Deng C, Hu H, Zhang T, Chen J (2020) Rock slope stability analysis and charts based on hybrid online sequential extreme learning machine model. *Earth Sci Inform* 13:729–746. <https://doi.org/10.1007/s12145-020-00458-5>
- Duncan JM, Wright SG (1980) The accuracy of equilibrium methods of slope stability analysis. *Eng Geol* 16:5–17. [https://doi.org/10.1016/0013-7952\(80\)90003-4](https://doi.org/10.1016/0013-7952(80)90003-4)
- Fatty A, Li A-J, Qian Z-G (2023) An interpretable evolutionary extreme gradient boosting algorithm for rock slope stability assessment. *Multimed Tools Appl* 83:46851–46874. <https://doi.org/10.1007/s11042-023-17445-9>
- Feng X-T, Hudson JA (2004) The ways ahead for rock engineering design methodologies. *Int J Rock Mech Min Sci* 41:255–273. [https://doi.org/10.1016/S1365-1609\(03\)00112-6](https://doi.org/10.1016/S1365-1609(03)00112-6)
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat*. <https://doi.org/10.1214/aos/1013203451>
- Gao J (2011) Prediction model for stability of high loess slopes. *Chin J Geo Eng* 33:170–176
- Gao H, Zhang Q (2024) Alpha evolution: an efficient evolutionary algorithm with evolution path adaptation and matrix generation. *Eng Appl Artif Intell* 137:109202. <https://doi.org/10.1016/j.engappai.2024.109202>
- Gao L, Pastor M, Li T et al (2023) A framework coupled neural networks and SPH depth integrated model for landslide propagation warning. *Acta Geotech* 18:3863–3888. <https://doi.org/10.1007/s11440-022-01774-4>
- Gao W, Zang M, Mei G (2024) Exploring influence of groundwater and lithology on data-driven stability prediction of soil slopes using explainable machine learning: a case

- study. *Bull Eng Geol Environ* 83:2. <https://doi.org/10.1007/s10064-023-03466-z>
- Gordan B, Jahed Armaghani D, Hajihassani M, Monjezi M (2016) Prediction of seismic slope stability through combination of particle swarm optimization and neural network. *Eng Comput* 32:85–97. <https://doi.org/10.1007/s00366-015-0400-7>
- Griffiths DV, Marquez RM (2007) Three-dimensional slope stability analysis by elasto-plastic finite elements. *Géotechnique* 57:537–546. <https://doi.org/10.1680/geot.2007.57.6.537>
- Hoang N-D, Pham A-D (2016) Hybrid artificial intelligence approach based on metaheuristic and machine learning for slope stability assessment: a multinational data analysis. *Expert Syst Appl* 46:60–68. <https://doi.org/10.1016/j.eswa.2015.10.020>
- Huang F, Xiong H, Chen S et al (2023) Slope stability prediction based on a long short-term memory neural network: comparisons with convolutional neural networks, support vector machines and random forest models. *Int J Coal Sci Technol* 10:18. <https://doi.org/10.1007/s40789-023-00579-4>
- Kang F, Xu B, Li J, Zhao S (2017) Slope stability evaluation using Gaussian processes with various covariance functions. *Appl Soft Comput* 60:387–396. <https://doi.org/10.1016/j.asoc.2017.07.011>
- Kardani N, Zhou A, Nazem M, Shen S-L (2021) Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. *J Rock Mech Geotech Eng* 13:188–201. <https://doi.org/10.1016/j.jrmge.2020.05.011>
- Karir D, Ray A, Kumar Bharati A et al (2022) Stability prediction of a natural and man-made slope using various machine learning algorithms. *Transp Geotech* 34:100745. <https://doi.org/10.1016/j.trgeo.2022.100745>
- Labuz JF, Zang A (2014) Mohr–Coulomb failure criterion. In: *The ISRM suggested methods for rock characterization, testing and monitoring: 2007–2014*. Springer, pp. 227–231
- Lawal AI, Hosseini S, Kim M et al (2024) Prediction of factor of safety of slopes using stochastically modified ANN and classical methods: a rigorous statistical model selection approach. *Nat Hazards* 120:2035–2056. <https://doi.org/10.1007/s11069-023-06275-5>
- Li KS, Lumb P (1987) Probabilistic design of slopes. *Can Geotech J*. <https://doi.org/10.1139/t87-068>
- Lin Y, Zhou K, Li J (2018) Prediction of slope stability using four supervised learning methods. *IEEE Access* 6:31169–31179. <https://doi.org/10.1109/ACCESS.2018.2843787>
- Lin S, Zheng H, Han B et al (2022) Comparative performance of eight ensemble learning approaches for the development of models of slope stability prediction. *Acta Geotech* 17:1477–1502. <https://doi.org/10.1007/s11440-021-01440-1>
- Liu Z, Shao J, Xu W et al (2014) An extreme learning machine approach for slope stability evaluation and prediction. *Nat Hazards* 73:787–804. <https://doi.org/10.1007/s11069-014-1106-7>
- Liu SY, Shao LT, Li HJ (2015) Slope stability analysis using the limit equilibrium method and two finite element methods. *Comput Geotech* 63:291–298. <https://doi.org/10.1016/j.compgeo.2014.10.008>
- Liu L, Zhao G, Liang W (2023) Slope stability prediction using k-NN-based optimum-path forest approach. *Mathematics* 11:3071. <https://doi.org/10.3390/math11143071>
- Mahmoodzadeh A, Mohammadi M, Farid Hama Ali H et al (2022) Prediction of safety factors for slope stability: comparison of machine learning techniques. *Nat Hazards* 111:1771–1799. <https://doi.org/10.1007/s11069-021-05115-8>
- Mangalathu S, Hwang S-H, Jeon J-S (2020) Failure mode and effects analysis of RC members based on machine-learning-based SHapley additive exPlanations (SHAP) approach. *Eng Struct* 219:110927. <https://doi.org/10.1016/j.engstruct.2020.110927>
- Manouchehrian A, Gholamnejad J, Sharifzadeh M (2014) Development of a model for analysis of slope stability for circular mode failure using genetic algorithm. *Environ Earth Sci* 71:1267–1277. <https://doi.org/10.1007/s12665-013-2531-8>
- Matsui T, San K-C (1992) Finite element slope stability analysis by shear strength reduction technique. *Soils Found* 32:59–70. <https://doi.org/10.3208/sandf1972.32.59>
- Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. *Knowl-Based Syst* 96:120–133. <https://doi.org/10.1016/j.knsys.2015.12.022>
- Moses D, Shimada H, Sasaoka T et al (2020) Rock slope stability analysis by using integrated approach. *World J Eng Technol* 08:405–428. <https://doi.org/10.4236/wjet.2020.83031>
- Niu P, Zhou A, Huang H (2020) Assessing model of highway slope stability based on optimized SVM. *China Geol* 3:339–344
- Parsa AB, Movahedi A, Taghipour H et al (2020) Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. *Accid Anal Prev* 136:105405. <https://doi.org/10.1016/j.aap.2019.105405>
- Pastor M, Tayyebi SM, Hernández A et al (2024) Modeling fast flows with variable water content: a depth-integrated SPH approach. *Comput Geotech* 174:106581. <https://doi.org/10.1016/j.compgeo.2024.106581>
- Pham K, Kim D, Park S, Choi H (2021) Ensemble learning-based classification models for slope stability analysis. *CATENA* 196:104886. <https://doi.org/10.1016/j.catena.2020.104886>
- Qi C, Tang X (2018a) Slope stability prediction using integrated metaheuristic and machine learning approaches: a comparative study. *Comput Ind Eng* 118:112–122. <https://doi.org/10.1016/j.cie.2018.02.028>
- Qi C, Tang X (2018b) A hybrid ensemble method for improved prediction of slope stability. *Int J Numer Anal Methods Geomech* 42:1823–1839. <https://doi.org/10.1002/nag.2834>
- Qiu Y, Zhou J (2023) Short-term rockburst damage assessment in burst-prone mines: an explainable XGBOOST hybrid model with SCSO algorithm. *Rock Mech Rock Eng* 56:8745–8770. <https://doi.org/10.1007/s00603-023-03522-w>
- Qiu Y, Li C, Huang S et al (2024) An ensemble model of explainable soft computing for failure mode identification

- in reinforced concrete shear walls. *J Build Eng* 82:108386. <https://doi.org/10.1016/j.jobe.2023.108386>
- Ray A, Kumar RESC, Bharati AK et al (2020) Hazard chart for identification of potential landslide due to the presence of residual soil in the Himalayas. *Indian Geotech J* 50:604–619. <https://doi.org/10.1007/s40098-019-00401-6>
- Salmasi F, Jafari F (2019) A simple direct method for prediction of safety factor of homogeneous finite slopes. *Geotech Geol Eng* 37:3949–3959. <https://doi.org/10.1007/s10706-019-00884-3>
- Segoni S, Nocentini N, Barbadori F et al (2025) A novel prototype national-scale landslide nowcasting system for Italy combining rainfall thresholds and risk indicators. *Landslides*. <https://doi.org/10.1007/s10346-024-02452-0>
- Shen Y, Wang X, Dai K et al (2024) Inference of creep landslide slip surface by InSAR technology and improved particle swarm optimization. *Landslides*. <https://doi.org/10.1007/s10346-024-02426-2>
- Tang J, Liang J, Han C et al (2019) Crash injury severity analysis using a two-layer stacking framework. *Accid Anal Prev* 122:226–238. <https://doi.org/10.1016/j.aap.2018.10.016>
- Tayyebi SM, Pastor M, Stickler MM et al (2022) Two-phase SPH modelling of a real debris avalanche and analysis of its impact on bottom drainage screens. *Landslides* 19:421–435. <https://doi.org/10.1007/s10346-021-01772-9>
- Terzaghi K, Peck RB, Mesri G (1996) *Soil mechanics in engineering practice*. John Wiley & Sons, Hoboken
- Tiwari S, Das SK (2025) Explainable artificial intelligence-driven ensemble learning for liquefaction susceptibility analysis. *Transp Infrastruct Geotechnol* 12:171
- Tiwari S, Das SK, Mohanty M (2025) A critical performance assessment of the machine learning algorithms for slope stability analysis. *Neural Comput Appl* 37:9735–9757
- Ullah S, Khan MU, Rehman G (2020) A brief review of the slope stability analysis methods. *Geol Behav* 4:73–77. <https://doi.org/10.26480/gbr.02.2020.73.77>
- Wang HB, Xu WY, Xu RC (2005) Slope stability evaluation using back propagation neural networks. *Eng Geol* 80:302–315. <https://doi.org/10.1016/j.enggeo.2005.06.005>
- Wang L, Wu C, Tang L et al (2020) Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. *Acta Geotech* 15:3135–3150. <https://doi.org/10.1007/s11440-020-00962-4>
- Wang G, Zhao B, Wu B et al (2023) Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases. *Int J Min Sci Technol* 33:47–59. <https://doi.org/10.1016/j.ijmst.2022.07.002>
- Yang C, Liu L-L, Huang F et al (2023a) Machine learning-based landslide susceptibility assessment with optimized ratio of landslide to non-landslide samples. *Gondwana Res* 123:198–216. <https://doi.org/10.1016/j.gr.2022.05.012>
- Yang Y, Zhou W, Jiskani IM et al (2023b) Slope stability prediction method based on intelligent optimization and machine learning algorithms. *Sustainability* 15:1169. <https://doi.org/10.3390/su15021169>
- Zhang H, Wu S, Zhang X et al (2022a) Slope stability prediction method based on the margin distance minimization selective ensemble. *Catena* 212:106055. <https://doi.org/10.1016/j.catena.2022.106055>
- Zhang W, Li H, Han L et al (2022b) Slope stability prediction using ensemble learning techniques: a case study in Yunyang County, Chongqing, China. *J Rock Mech Geotech Eng* 14:1089–1099. <https://doi.org/10.1016/j.jrmge.2021.12.011>
- Zhou J, Li E, Yang S et al (2019) Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Saf Sci* 118:505–518. <https://doi.org/10.1016/j.ssci.2019.05.046>
- Zhou J, Huang S, Qiu Y (2022) Optimization of random forest through the use of MVO, GWO and MFO in evaluating the stability of underground entry-type excavations. *Tunn Undergr Space Technol* 124:104494. <https://doi.org/10.1016/j.tust.2022.104494>
- Zhu C (2005) Slope stability analysis and evaluation - taking the slope of Yuanmo Expressway as an example. Master's thesis, Kunming University of Science and Technology, China (in Chinese)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.