Check for updates

Received: 7 August 2025 | Revised: 5 February 2026 | Accepted: 27 February 2026

DOI: 10.1002/dug2.70101

RESEARCH ARTICLE

DUSE WILEY

Explainable hybrid stacking ensemble method for hard rock pillar stability prediction and engineering applications

Ning Wang[1,2] | **Guoyan Zhao**[1] | **Meng Wang**[3] | **Taraka Pattela**[4]

[1]School of Resources and Safety Engineering, Central South University, Changsha, China

[2]Department of Applied Mathematics, ETS Ingenieros de Caminos, Universidad Politécnica de Madrid, Madrid, Spain

[3]Department of Mining and Materials Engineering, McGill University, Montreal, Canada

[4]Department of Earth, Environment and Physical Sciences, University of Siena, Siena, Italy

Correspondence

Meng Wang, Department of Mining and Materials Engineering, McGill University, Montreal, QC H3A 0E8, Canada.
Email: mwanglh@csu.edu.cn

Funding information

Scientific Research Foundation of Hunan Provincial Education Department, Grant/Award Number: CX20230167; Fundamental Research Funds for Central Universities of the Central South University, Grant/Award Number: 2024zzts0084; China Scholarship Council, Grant/Award Numbers: 202506370210, 202406370221

Abstract

Accurate prediction of hard rock pillar stability is vital for ensuring safety in underground mining operations. This study presents a novel hybrid stacking ensemble framework that integrates six machine learning models optimized by the Sparrow Search Algorithm (SSA) at both base and meta levels. A comprehensive dataset of 331 pillar cases was compiled from published case histories reported in 10 underground mines, characterized by five key geotechnical features. Seventy percent of the data were used for model training and the remaining 30% for testing. Performance was evaluated using five classification metrics, including accuracy, precision, recall, $F_{1\text{-score}}$, and kappa. Statistical comparisons using the Friedman test and Nemenyi post-hoc test confirmed the LightGBM-meta stacking model as the top-performing approach, achieving 93.0% accuracy, 0.937 $F_{1\text{-score}}$, and a kappa of 0.858. SHapley Additive exPlanations (SHAP) identified pillar stress and uniaxial compressive strength (*UCS*) as the most influential features, enhancing model interpretability and engineering insight. A user-friendly graphical user interface (GUI) was developed to enable practical deployment by mining engineers. The model was further validated on 11 field pillar cases from the Sanshandao Gold Mine, with parameters obtained via 3D laser scanning, in-situ stress measurement, and laboratory *UCS* testing. All 11 field cases were correctly classified. Overall, the proposed SSA-optimized stacking framework bridges data-driven prediction and field deployment, offering practical support for pillar design and stability risk management.

KEYWORDS

engineering application, hard rock pillar, model interpretability, pillar stability prediction, stacking ensemble learning

Highlights

- A hybrid stacking model optimized by Sparrow Search Algorithm shows the best performance with Light Gradient Boosting Machine as a meta-learner.
- SHapley Additive exPlanations analysis reveals pillar stress and uniaxial compressive strength as dominant factors for model decisions.
- A graphical user interface tool was built and applied to 11 Sanshandao cases; all predictions proved accurate.

1 | INTRODUCTION

In underground hard rock mining, pillars are vertical rock masses intentionally preserved during excavation to provide structural support for the surrounding strata (Ren et al., 2025; Yin et al., 2017; Zhang, et al., 2015; Zhao et al., 2024). These pillars play a vital role in maintaining the integrity of underground openings, ensuring roof stability, and enabling the safe continuation of mining operations (Potvin, 2009; Wagner, 1980; Wang et al., 2006). Pillar instability can result in progressive rock mass failure, localized collapse, or even large-scale subsidence, each of which poses significant risks to personnel, equipment, and production continuity (Maybee, 2000; Mortazavi et al., 2009; Zhang et al., 2024). Historical failures have demonstrated that misjudging pillar behavior can lead to devastating consequences, including mine abandonment and loss of life (Idris et al., 2015;

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2026 The Author(s). *Deep Underground Science and Engineering* published by John Wiley & Sons Australia, Ltd on behalf of China University of Mining and Technology.

Li et al., 2018; Li et al., 2014). Therefore, improving the accuracy of pillar stability assessment remains a fundamental goal in underground geomechanics and mine design (Esterhuizen et al., 2011; Li et al., 2021, 2024).

Empirical and semi-empirical approaches have long been used as primary tools for pillar stability evaluation (Zhang, Stead, et al., 2015; Zipf Jr, 1910; Zipf Jr & Mark, 1997). These methods often rely on the calculation of a safety factor, defined as the ratio between estimated pillar strength and applied stress, based on simplified assumptions (Mark, 2016). Notable contributions include the design formula developed by Hedley and Grant (1972), which incorporated the width-to-height ratio and uniaxial compressive strength (*UCS*) of rock to predict pillar strength. Later, Lunder (1994) proposed a stability classification system using a stability graph, which enabled visualization of pillar performance trends. Martin and Maybee (2000) combined the empirical method with the ductile failure behavior of rock mass and proposed the ultimate strength range of hard rock pillars. Despite their practicality and ease of use, empirical models are inherently constrained by the conditions of the original datasets used in their formulation. These methods typically fail to capture the complexity of rock mass behavior under varying geological, geometric, and stress conditions, limiting their applicability to broader scenarios (Aksoy & Onargan, 2006; Deng et al., 2003; Ranjith et al., 2017; Wagner, 2019).

Numerical modeling techniques have long been employed in pillar stability analysis (Itasca Consulting Group, Inc., 2023; Valencia, 2022). Early applications of the finite element method (FEM) in the 1970s, pioneered in part by Brady and Brown (2006), focused primarily on linear elastic analysis. Subsequent advancements saw the incorporation of elasto-plastic and damage models to simulate progressive failure mechanisms (Martin & Maybee, 2000). In the 1990s, the finite difference method (FDM) gained popularity for its ability to simulate large deformations and time-dependent creep behavior, although it continued to treat the rock mass as a continuous medium (Munjiza, 2004). To better capture the mechanical behavior of jointed and fractured rock, Cundall and Strack (1979) introduced the discrete element method (DEM), enabling the modeling of crack initiation, block interaction, and separation. Building upon these methods, Munjiza (2004) developed the combined finite-discrete element method (FDEM), which offers a unified framework for simultaneous stress analysis and explicit fracture simulation. Despite their improved realism and predictive capability, these numerical approaches are often computationally demanding and heavily reliant on detailed geological and mechanical input parameters, which constrain their practical application in routine engineering design (Guo & Zhao, 2014; Wang, 2022).

With the increasing availability of mining datasets, machine learning (ML) has become a promising alternative for pillar stability prediction (Gao et al., 2023; Wang et al., 2025). Algorithms such as support vector machines (SVM), random forests (RF), gradient boosting machines (GBM), and artificial neural networks (ANN) have been applied in numerous studies (Lawal & Kwon, 2021; Peng et al., 2025; Saadati et al., 2025). Zhou et al. (2011) utilized SVM to classify pillar outcomes with improved accuracy, while Ghasemi et al. (2017) implemented decision trees and SVM for probabilistic risk estimation. Ding et al. (2018) applied stochastic gradient boosting to evaluate multi-class pillar states, and Liang et al. (2020) demonstrated GBM-based methods outperforming classical regression in real mines. Recently, SVM-based models optimized by metaheuristics like Sparrow Search Algorithm (SSA), Whale Optimization Algorithm (WOA), and Gray Wolf Optimizer (GWO) have shown improved predictive performance in hard rock pillar stability evaluation (Li et al., 2023). Despite these advancements, most ML-based studies rely on single learners, which are prone to overfitting and poor generalizability, especially across heterogeneous geological conditions. Furthermore, most models lack transparency, limiting engineers' ability to interpret predictions and derive physical insight.

Stacking ensemble learning has emerged as a highly promising meta-learning strategy in recent years, demonstrating strong potential in solving complex, nonlinear, and high-dimensional prediction tasks (Wolpert, 1992). It has been successfully applied across diverse domains, including structural health monitoring, disease diagnosis, remote sensing, and hydrological forecasting (Al Shamsi & Abdallah, 2023; Lazzarini et al., 2023; Liu et al., 2024; Mu et al., 2024). However, stacking ensembles have not yet been applied to the field of hard rock pillar stability prediction. Stacking ensemble learning offers a compelling solution by integrating the outputs of multiple base learners via a meta-learner (Granata et al., 2022; Xu et al., 2022). Unlike bagging or boosting, stacking allows heterogeneous models to contribute complementary strengths, making it particularly well-suited for complex, nonlinear problems. Due to its layered architecture, stacking involves a large number of hyperparameters across base and meta-learners, which significantly influence model performance. Effective hyperparameter tuning is therefore critical to avoid overfitting and ensure generalization (Chatzimparmpas et al., 2020; Kardani et al., 2021). In this study, we employ the SSA, a recent swarm intelligence algorithm inspired by sparrows' foraging behavior, to optimize these parameters. SSA offers strong global search capability, rapid convergence, and low computational cost (Xue & Shen, 2020).

To address the challenges of accurate pillar stability prediction, this study proposes an interpretable hybrid stacking ensemble method enhanced by the SSA. A dataset of 331 pillar cases from 10 hard rock mines was compiled using five indicators. Six ML algorithms, namely RF, SVM, k-nearest neighbors (KNN), gradient boosting decision tree (GBDT), light gradient boosting machine (LightGBM), and eXtreme gradient boosting (XGBoost), were employed as base learners and were also used as meta-learners to construct six stacking models. SSA, combined with five-fold cross-validation, was applied to optimize hyperparameters at both levels. All models were comprehensively evaluated using five metrics and statistically compared using the Friedman test and Nemenyi post-hoc analysis. Model interpretability was achieved via SHapley Additive exPlanations (SHAP) analysis, and a graphical user interface (GUI) was developed to support practical deployment. Field

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

validation was conducted on 11 cases from the Sanshandao Gold Mine through 3D laser scanning, in situ stress measurement, laboratory testing, and detailed site investigation, confirming the robustness and engineering applicability of the proposed method.

2 | DATASET AND ANALYSIS

A total of 331 hard rock pillar cases were compiled from foundational studies conducted across 10 underground metal mines. The dataset was built from peer-reviewed empirical research published between 1961 and 2013. Among these, González-Nicieza et al. (2006) reported 61 cases from the Marble mine in Southern Spain. Schubert and Villaescusa (1998) contributed 49 cases from the McArthur River Mine in Canada. Potvin et al. (1988) contributed 47 cases from open stope mines in Canada. Von Kimmelmann et al. (1984) provided two large datasets of 47 cases each, drawn from Selebi Phikwe mines in South Africa. From Canada, Lunder (1994) documented 31 cases in the H-W mine, and Hedley and Grant (1972) contributed 28 cases from the Elliot Lake uranium mines. Other case records include 29 examples from Indian mines described by Murali Mohan et al. (2001), 18 cases from US stone mines by Esterhuizen et al. (2011), 12 cases from the Lo Tacón Mine in Spain presented by Alvarez-Garcia et al. (2013), and 9 cases from the Zinkgruvana Mine in Sweden detailed by Sjoberg (1992). Given that the pillar case histories were collected from multiple sources spanning several decades, potential heterogeneity in reporting standards and failure definitions exists. To ensure consistency, all cases were harmonized prior to analysis by unifying original binary or ternary labels into a conservative binary classification, where cases reported as "unstable" were reclassified as "failed." This treatment aligns with conservative design practices, in which ambiguous structural states are typically treated as unstable for safety assurance. The detailed information is shown in Figure 1.

The dataset includes five input variables selected based on mechanical relevance and availability. These are pillar width (w), pillar height (h), width-to-height ratio (w/h), UCS, and pillar stress (P_S). Width and height were measured in situ by mine engineers using field equipment. UCS was obtained from laboratory testing of core samples extracted near or from the pillar location. P_S was estimated using empirical methods based on mining depth, the local in situ stress regime, and the spatial configuration of surrounding pillars. It is included to represent the external loading condition acting on the pillar, and it complements geometry and UCS by capturing the load–capacity balance. Although w, h, and w/h are mathematically related, w/h is retained as a dimensionless indicator of pillar slenderness, which captures geometric proportion rather than absolute size and provides information not fully represented by w or h alone. The input features were selected based on cross-mine data availability and consistency in published case histories. In contrast, other potentially influential factors, including joint orientation, groundwater conditions, mining sequence, and backfill characteristics, are often site-specific and not consistently documented and were therefore not included. Descriptive statistics for each variable are presented in Table 1.

Out of the 331 cases, 163 were classified as Stable and 168 as Failed. Correlation analysis was performed using Pearson's method to assess linear relationships among variables. As shown in Figure 2, P_S displays a strong negative correlation with the w/h ratio and UCS, indicating that slender or weaker pillars are exposed to greater stress levels. A moderate positive correlation between w and UCS suggests that stronger pillars are often associated with larger geometries; h exhibited weak correlations, implying potential nonlinear interactions. Distribution histograms alongside the correlation matrix reveal noticeable skew in several variables, which justifies the need for preprocessing. Prior to model development, all variables were standardized using Z-score normalization to eliminate scale effects. The dataset was then randomly split into a training set (232 samples) and a testing set (99 samples) using a 70%–30% ratio to ensure reliable model training and validation.

3 | METHODOLOGY

3.1 | Stacking ensemble algorithm

Stacking is a supervised ensemble learning framework originally introduced by (Wolpert, 1992). It combines multiple predictive models to improve generalization

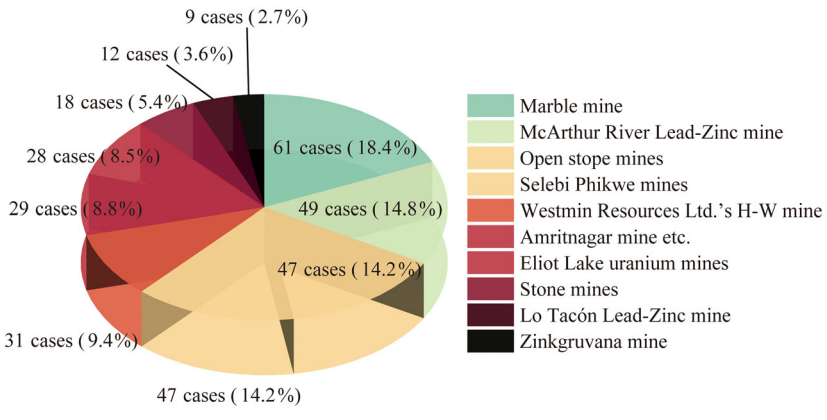


FIGURE 1 The source of the database and related information.

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Statistics of each feature.

Feature	Account	Mean	Median	Minimum	Maximum	Standard error	Standard deviation
w (m)	331	10.30	7.20	1.90	61.00	0.45	8.10
h (m)	331	10.20	5.00	1.80	53.00	0.55	9.90
w/h	331	1.28	1.10	0.30	8.30	0.04	0.80
UCS (MPa)	331	118.50	94.00	19.00	316.00	3.65	66.40
P_S (MPa)	331	39.79	33.00	0.56	127.60	1.49	27.10

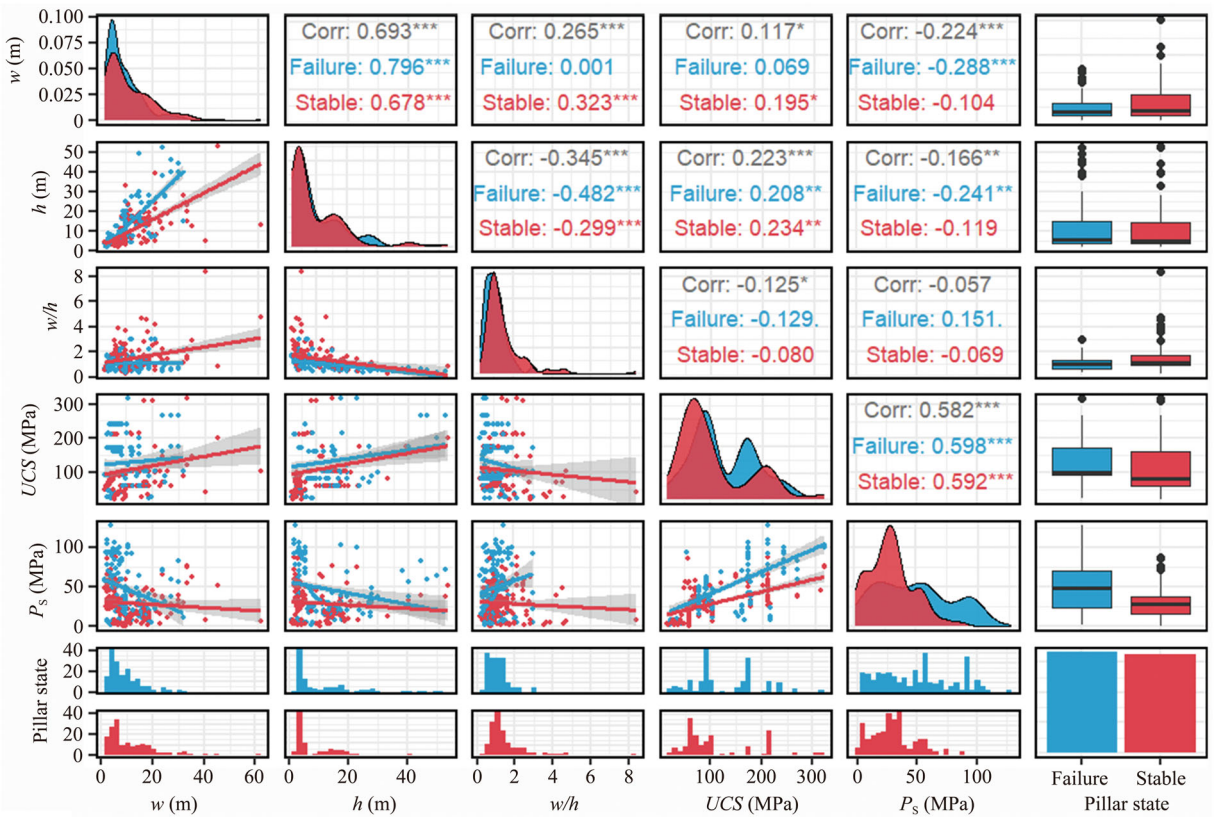


FIGURE 2 Distribution of dataset features and correlation matrix.

performance on complex learning tasks. Unlike traditional ensembles that use models of the same type, stacking allows for heterogeneous base learners. Each learner captures different structural patterns in the data, and their outputs are combined by a higher-level model, referred to as the meta-learner. In the context of hard rock pillar stability, where stress conditions and geometric features are coupled in complex ways, stacking offers a flexible and accurate modeling solution.

3.2 | Base and meta leaners

3.2.1 | KNN

KNN was originally proposed by Cover and Hart (1967) as a nonparametric and instance-based classification method. It does not require model training; instead, it classifies new samples based on their similarity to previously observed cases. The principle of the KNN classifier is displayed in Figure 3a. The KNN procedure consists of three main steps: (1) compute the distances between the input sample and all training samples using a defined metric; (2) identify the k closest samples by ranking the distances in ascending order; (3) assign the class label to the input sample based on the majority class among these k neighbors.

3.2.2 | Support vector machine

SVM was developed by Cortes and Vapnik (1995) as a supervised learning model for classification tasks. It constructs a decision boundary that maximizes the margin between two classes in the feature space. The principle of the SVM classifier is illustrated in Figure 3b. The SVM workflow involves four steps: (1) standardize the input features to ensure consistent scaling; (2) project the data into a higher-dimensional space using a kernel function if linear separation is not feasible; (3) solve an

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

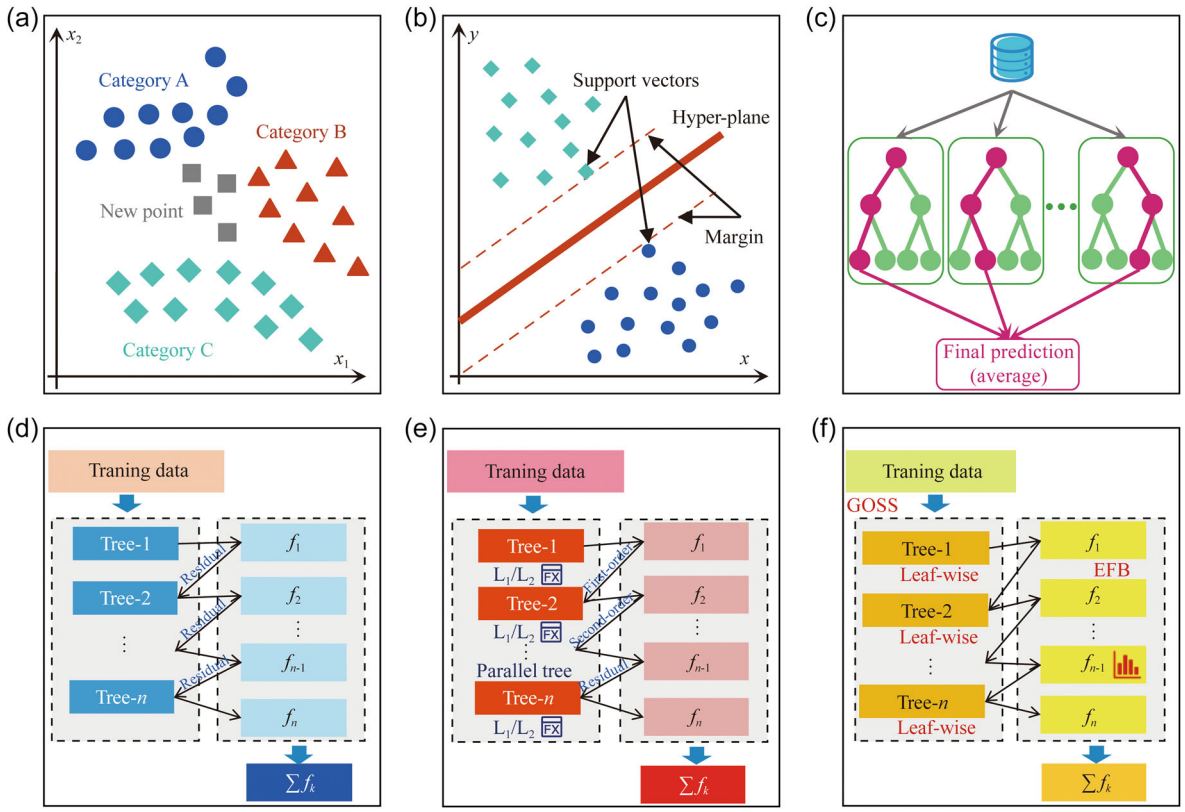


FIGURE 3 Diagram of the six base models. (a) K-nearest neighbors, (b) support vector machines, (c) random forest, (d) gradient boosting decision tree, (e) XGBoost, and (f) light gradient boosting machine.

optimization problem to identify the hyperplane that maximizes class separation; (4) classify new samples based on their position relative to this hyperplane.

3.2.3 | Random forest

RF was proposed by Breiman (2001) as an ensemble learning method that combines multiple decision trees to improve classification performance. The principle of the RF classifier is illustrated in Figure 3c. The model is constructed in three stages: (1) generate multiple bootstrap samples from the original dataset, each used to train an independent decision tree; (2) at each split of a tree, select a random subset of features and determine the optimal division based on impurity criteria; (3) predict new instances by aggregating the outputs of all trees through majority voting.

3.2.4 | Gradient boosting decision tree

GBDT was introduced by Friedman (2001) as a boosting-based ensemble method that builds a sequence of decision trees to progressively improve model performance. The principle of the GBDT classifier is illustrated in Figure 3d. The learning process follows an additive strategy: (1) initialize the model with a constant prediction based on the training target; (2) at each iteration, compute the residuals between the predicted and actual values; (3) fit a regression tree to these residuals, producing a new tree that approximates the gradient of the loss function; (4) update the model by adding the new tree, scaled by a learning rate, to the existing ensemble; (5) repeat the process until a predefined number of trees is reached or the improvement plateaus.

3.2.5 | Extreme gradient boosting

XGBoost was introduced by Chen and Guestrin (2016) as a scalable and regularized version of gradient boosting. It extends the GBDT framework by integrating advanced optimization techniques that improve both model performance and training efficiency. The principle of the XGBoost classifier is illustrated in Figure 3e. XGBoost enhances the boosting process through several mechanisms: (1) it incorporates L_1 and L_2 regularization terms into the objective function to control model complexity; (2) it employs second-order derivatives to improve gradient approximation; (3) it supports parallelized tree construction and optimized memory access for faster computation.

3.2.6 | Light gradient boosting machine

LightGBM was proposed by Microsoft in 2017 as a highly efficient gradient boosting framework optimized for speed and memory usage (Ke et al., 2017). The principle of the LightGBM is illustrated in Figure 3f. The training process of LightGBM introduces: (1) Gradient-based One-Side Sampling (GOSS), which retains instances with large gradients and randomly samples from the rest, improving efficiency without sacrificing accuracy; (2) Exclusive Feature Bundling (EFB), which

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

reduces feature dimensionality by combining mutually exclusive features to lower computational costs. In contrast to level-wise tree growth, LightGBM adopts a leaf-wise strategy that selects the leaf with the maximum loss reduction at each iteration.

3.3 | Sparrow search algorithm

In ensemble learning, model performance depends strongly on hyperparameter settings. Metaheuristic algorithms offer a flexible and effective alternative with strong global search ability (Gharehchopogh et al., 2023). This study adopts SSA to optimize hyperparameters across all classifiers.

The SSA introduced by Xue and Shen in 2020 is a population-based metaheuristic inspired by the foraging and anti-predation behavior of sparrows (Xue & Shen, 2020). The population is divided into three behavioral roles—discoverers, followers, and scouts—which update positions through distinct strategies to ensure a balance between global exploration and local exploitation. The workflow of SSA is illustrated in Figure 4, and the detailed steps are as follows:

Step 1: Initialize a population of N individuals in D-dimensional space. Evaluate their fitness values and sort them in ascending order.

Step 2: Assign the top-performing individuals as explorers X_E. Explorers perform a global search. When the environmental safety value is low, they update positions using an exponential decay rule. If a threat is detected, they escape through random perturbations. Their update mechanism is given as:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \cdot \exp\left(-\dfrac{i}{\alpha \cdot iter_{\max}}\right) & \text{if} \quad R_2 < ST, \\ X_{i,j}^{t} + Q \cdot L & \text{if} \quad R_2 \geq ST, \end{cases} \tag{1}$$

where X_{ij}^{t} represents the position of the i-th individual in the j-th dimension at iteration t, $iter_{\max}$ denotes the maximum number of iterations, α is a constant controlling the exponential decay rate, R_2 is a random number in the interval (0, 1), ST denotes the safety threshold, Q is a random number following a normal distribution, and L is a vector of ones with dimension D.

Step 3: Individuals ranked in the middle are classified as followers X_F. Those with lower ranks receive large-scale perturbations influenced by the worst individual. Others follow the leader with adaptive adjustments. The update rule for followers is:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{X_{\text{worst}} - X_{i,j}^{t}}{i^2}\right) & \text{if} \quad i > N/2, \\ X_{\text{p}}^{t+1} + \left|X_{i,j}^{t} - X_{\text{p}}^{t+1}\right| \cdot A^{+} \cdot L & \text{otherwise}, \end{cases} \tag{2}$$

where X_{worst} represents the worst position in the population, X_{p}^{t} denotes the current best position, A is a matrix

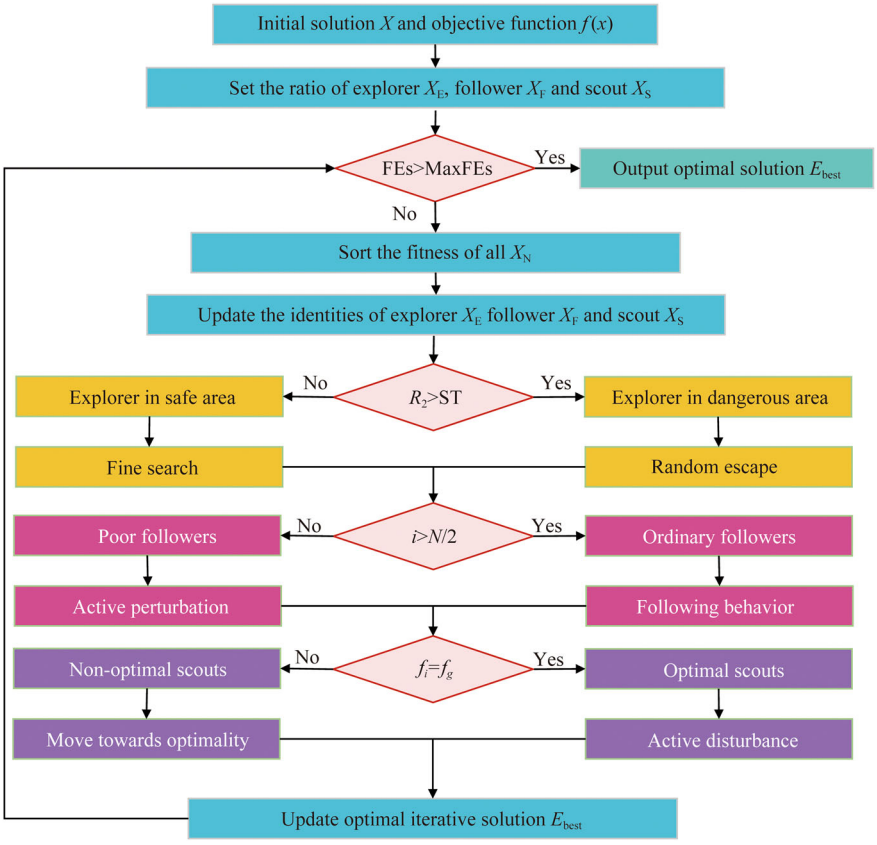


FIGURE 4 Calculation process of the Sparrow Search Algorithm.

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

whose elements are randomly generated as either 1 or −1, N is the population size.

Step 4: The lowest-ranked individuals act as scouts X_S. They enhance the population's diversity and sensitivity to stagnation. If a scout's fitness is worse than the global best, it moves toward the optimal position. If equal, it explores around the worst solution using a normalized perturbation:

$$X_{i,j}^{t+1} = \begin{cases} X_{\text{best}}^{t} + \beta \cdot \left| X_{i,j}^{t} - X_{\text{best}}^{t} \right| & \text{if} \quad f_i > f_g, \\ X_{i,j}^{t} + K \cdot \left(\frac{\left| x_{i,j}^{t} - x_{\text{worst}}^{t} \right|}{(f_i - f_w) + \varepsilon} \right) & \text{if} \quad f_i = f_g, \end{cases} \tag{3}$$

where X_{best} represents the global optimal position, X_{worst} represents the worst position, f_i denotes the fitness value of the i-th individual, f_g represents the global best fitness, f_w represents the worst fitness value, β and K are random numbers in the interval (0, 1), ε is a small constant used to avoid division by zero.

Step 5: Evaluate the new fitness of the population and update the global best solution. Repeat Steps 2–5 until the maximum number of iterations is reached.

3.4 | Model evaluation metrics

To comprehensively evaluate model performance within the stacking ensemble framework, a set of standard classification metrics is adopted. These include accuracy, precision, recall, $F_{1\text{-score}}$, and kappa, which are computed based on the confusion matrix elements: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

The primary metrics include accuracy, precision, recall, $F_{1\text{-score}}$, and kappa. They measure, respectively, the overall proportion of correct predictions, the reliability of positive predictions, the ability to detect true positives, the harmonic means of precision and recall, and the agreement between predicted and actual classes beyond chance. Their mathematical definitions are as follows:

$$A = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4}$$

$$P = \frac{TP}{TP + FP}, \tag{5}$$

$$R = \frac{TP}{TP + FN}, \tag{6}$$

$$F_1 = \frac{2PR}{P + R}, \tag{7}$$

$$k = \frac{p_o - p_e}{1 - p_e} = \frac{N\sum_{i=1}^{n} X_{ii} - \sum_{i=1}^{n}(X_{i+} \cdot X_{+i})}{N^2 - \sum_{i=1}^{n}(X_{i+} \cdot X_{+i})}, \tag{8}$$

where A, P, R, F_1, and κ denote accuracy, precision, recall, $F_{1\text{-score}}$, and kappa coefficient, respectively; p_o is the observed agreement between the predicted and actual classes; p_e is the expected agreement due to chance; N is the total number of samples; n is the number of classes; X_{ii} represents the number of samples in the i-th class that are correctly classified, i.e., the diagonal element of the confusion matrix; and X_{i+} and X_{+i} denote the marginal totals of the i-th row and the i-th column, respectively.

To further assess the discriminative ability of classifiers, the receiver operating characteristic (ROC) curve and its area under the curve (A_{UC}) are applied. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at various thresholds. Their formulas are given as:

$$TPR = \frac{TP}{TP + FN}, \tag{9}$$

$$FPR = \frac{FP}{TP + TN}. \tag{10}$$

In order to statistically validate differences in model performance across multiple evaluation metrics, the Friedman test is adopted, followed by the Nemenyi post-hoc test for pairwise comparison (Fatty et al., 2023; Lundberg & Lee, 2017). The Friedman test statistic is computed as:

$$x_{\mathbf{f}}^2 = \frac{12}{\lambda(\lambda + 1)}\left[\sum_{j=1}^{\lambda} R_{0j}^2 - \frac{\lambda(\lambda + 1)^2}{4}\right], \tag{11}$$

where λ denotes the number of models being compared, and R_{0j} denotes the average rank of the j-th model. If the null hypothesis of no significant difference is rejected, the critical difference (CD) between average ranks is calculated as:

$$CD = q_\alpha \cdot \sqrt{\frac{\lambda(\lambda + 1)}{6N}}, \tag{12}$$

where q_α denotes the critical value that corresponds to a given significance level α and the number of models under consideration. A pair of models is considered significantly different if the difference in their average ranks exceeds the CD value. This procedure ensures that performance comparisons are both statistically valid and practically meaningful.

3.5 | Framework hybrid stacking ensemble method

This study proposes an integrated framework for predicting the stability of hard rock pillars by combining stacking ensemble learning with intelligent hyperparameter optimization and interpretability techniques. The overall procedure is illustrated in Figure 5.

1. A dataset comprising 331 historical pillar cases from ten underground mines was compiled. All input features were standardized using Z-score normalization, and the data were randomly split into 70% for training and 30% for testing.
2. Six base classifiers, including KNN, SVM, RF, GBDT, XGBoost, and LightGBM, were optimized individually using the SSA. Each model's hyperparameter space was explored through iterative evaluation guided by five-fold cross-validation.

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

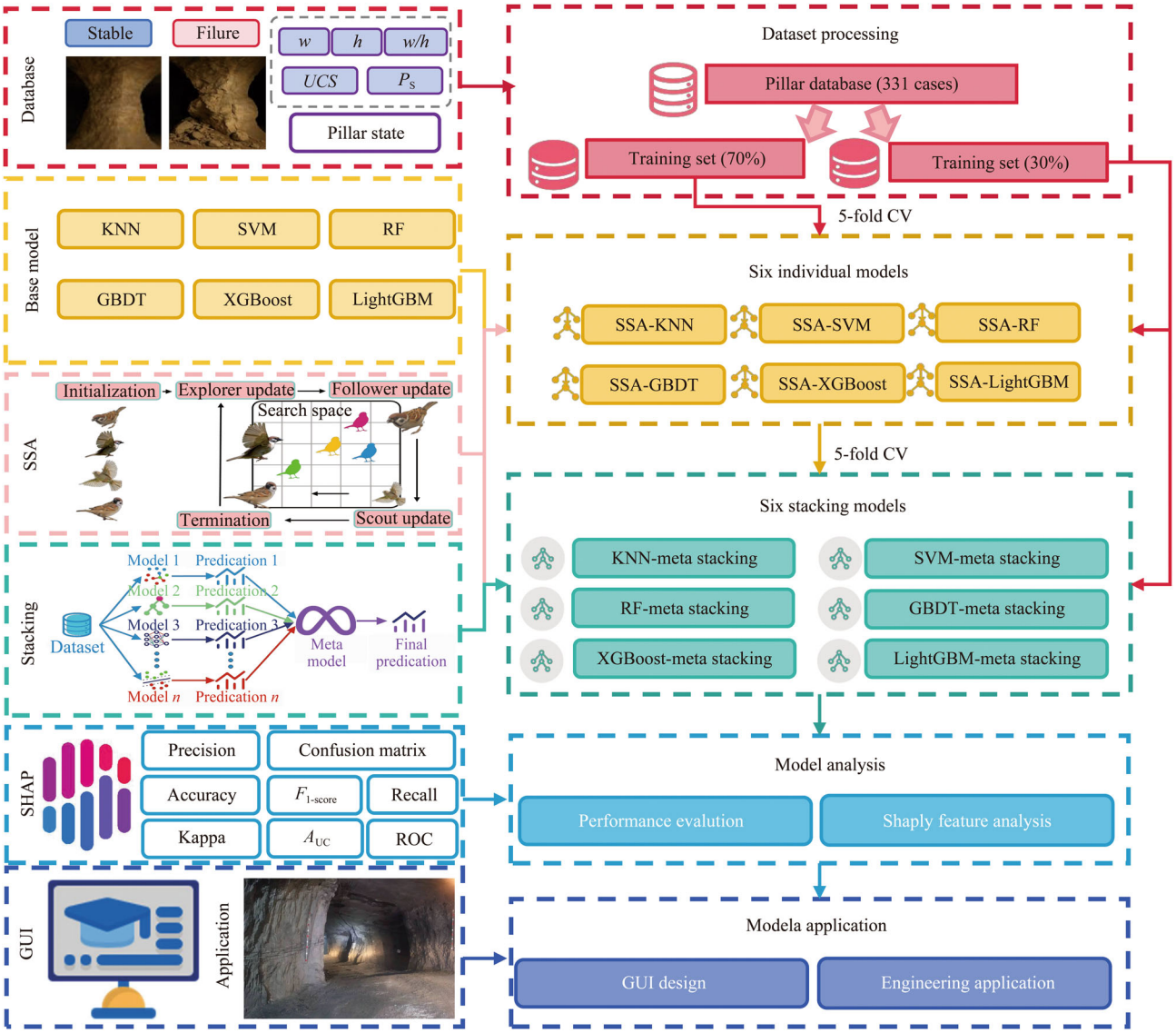


FIGURE 5 The process of constructing and applying the proposed method.

3. Stacking ensemble models were constructed using six optimized individual learners. Each model was assigned as the meta-learner in turn. SSA was further employed to fine-tune the hyperparameters at the meta-level.
4. All models were comprehensively evaluated using accuracy, precision, recall, $F_{1\text{-score}}$, kappa, and the area under the ROC curve. To assess statistical differences in model performance, the Friedman test and the Nemenyi post-hoc test were applied across all evaluation metrics.
5. The best-performing model was interpreted using SHapley Additive exPlanations. Global feature importance rankings and pairwise interaction effects were visualized.
6. A GUI was developed to support practical deployment. The interface allows users to input pillar parameters, obtain stability predictions, and interpret results through SHAP-based visual explanations.
7. The proposed method was validated through field application at the Sanshandao Gold Mine. The work included case investigation, 3D laser scanning for point cloud generation, laboratory testing to determine *UCS*, and ground stress measurement.

In summary, the SSA-optimized stacking ensemble framework demonstrates both methodological innovation and engineering relevance, offering a robust and interpretable solution for underground pillar stability assessment.

4 | RESULTS AND DISCUSSION

4.1 | Hyperparameters setting

Prior to constructing the SSA-enhanced stacking ensemble models, a unified hyperparameter optimization strategy was established for both base learners and meta-learners within the stacking ensemble framework. The hyperparameters to be optimized and their corresponding search ranges for all base and meta learners are summarized in Table 2, with ranges determined based on commonly adopted values in previous studies and preliminary testing. Hyperparameter optimization was conducted using the SSA. Within the stacking ensemble framework, SSA was first applied to optimize the hyperparameters of all base learners, after which the optimized base learners were used to construct stacking models, and SSA was further

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Results of hyperparameter tuning.

Model type	Model	Hyperparameter	Search range	Optimized value
Optimized base models	SSA-KNN	n_neighbors	(1, 20)	6
		p	(1.000, 3.000)	1.003
	SSA-SVM	C	(0.100, 100.000)	3.156
		gama	(0.010, 1.000)	0.011
	SSA-RF	n_estimators	(1, 500)	434
		max_depth	(3, 20)	9
		min_samples_split	(2, 20)	2
		min_samples_leaf	(1, 10)	1
	SSA-GBDT	N_estimators	(1, 500)	496
		learning_rate	(0.01, 0.30)	0.23
		max_depth	(1, 10)	10
		min_samples_split	(2, 20)	10
	SSA-XGBoost	n_estimators	(10, 500)	417
		learning_rate	(0.01, 0.30)	0.30
		max_depth	(3, 20)	10
	SSA-LightGBM	N_estimators	(10, 500)	441
		learning_rate	(0.01, 0.30)	0.24
		max_depth	(1, 20)	8
		num_leaves	(5, 20)	9
		min_data_in_leaf	(5, 20)	7
Stacking ensemble models	KNN-meta stacking	n_neighbors	(1, 20)	18
		p	(1.00, 3.00)	2.74
	SVM-meta stacking	C	(0.100, 100.000)	3.173
		gama	(0.010, 1.000)	0.011
	RF-meta stacking	n_estimators	(1, 500)	465
		max_depth	(3, 20)	9
		min_samples_split	(2, 20)	3
		min_samples_leaf	(1, 10)	2
	GBDT-meta stacking	N_estimators	(1, 500)	403
		learning_rate	(0.01, 0.30)	0.30
		max_depth	(1, 10)	10
		min_samples_split	(2, 20)	9
	XGBoost-meta stacking	n_estimators	(10, 500)	237
		learning_rate	(0.01, 0.30)	0.30
		max_depth	(3, 20)	8
	LightGBM-meta stacking	N_estimators	(10, 500)	500
		learning_rate	(0.01, 0.30)	0.06
		max_depth	(1, 20)	10
		num_leaves	(5, 20)	10
		min_data_in_leaf	(5, 20)	10

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

employed to tune the hyperparameters of the meta-learners. Identical SSA settings were adopted for all models, with a population size of 100 individuals and a maximum of 200 iterations as the stopping criterion.

To reduce stochastic bias inherent in metaheuristic optimization, SSA was integrated with five-fold cross-validation. The fitness function was defined as the average value of $1-A_{UC}$ across the five folds, thereby alleviating the influence of single data splits or single-run evaluations. Model performance was further evaluated on an independent hold-out test set that was not involved in training or optimization.

The fitness evolution of both base learners and stacking models during SSA optimization is visualized in Figure 6a,b, respectively. Figure 6a demonstrates that most individual models exhibit rapid fitness improvement in the early stages. Notably, LightGBM and GBDT achieve faster convergence and reach lower final fitness values, indicating more stable optimization behavior and superior learning capacity. In contrast, models such as KNN and SVM require more iterations to stabilize. These results indicate a more rugged search space that demands prolonged exploration to avoid premature convergence.

As illustrated in Figure 6b, similar trends are observed for stacking models. Configurations such as RF-meta and LightGBM-meta converge relatively quickly, typically within 120 iterations, while others exhibit slower, oscillatory convergence. Importantly, the LightGBM-meta stacking model achieved the lowest final fitness value among all ensembles.

The optimized hyperparameters for all models are presented in Table 2. Considerable variation is observed across learners. For instance, SSA-GBDT adopted an optimal n_estimators value of 496, while SSA-XGBoost settled at 417. KNN and SVM models were optimized toward more regularized structures, with parameters such as k = 6 and C = 3.156, respectively. Additionally, comparisons between base and meta configurations reveal interesting patterns. The LightGBM-meta stacking model, for example, uses a considerably smaller learning rate (0.06) than its base counterpart (0.24), suggesting that finer learning granularity is preferred when LightGBM operates at the meta-level. Differences in max_depth, subsample, and min_data_in_leaf across stacking models reflect the variation in optimal model complexity and generalization capability.

4.2 | Comparative analysis of model performance

To systematically evaluate the predictive capability of all models, 12 SSA-optimized classifiers were tested on the test set. These include six base models and six hybrid stacking ensemble variants. Five standard evaluation metrics were used: accuracy, recall, precision, $F_{1\text{-score}}$, and kappa. The results are summarized in Table 3. Overall, stacking models demonstrated higher performance across all metrics. Among them, the LightGBM-meta stacking model achieved the best performance, with an accuracy of 0.930, recall of 0.929, precision of 0.929, $F_{1\text{-score}}$ of 0.937, and kappa of 0.858.

To statistically validate differences among models, a Friedman test was performed, which is suitable for comparing multiple classifiers across multiple datasets or folds without assuming normality (Gao et al., 2024). The resulting model rankings are shown in Figure 7. We further applied the Nemenyi post-hoc test for pairwise comparisons at a significance level of $\alpha = 0.05$. In the *CD* diagram, the *CD* value ($CD = 0.849$) represents the minimum difference in average ranks required for two models to be considered significantly different under the Nemenyi test. Accordingly, models that are connected by a common horizontal line are not significantly different, whereas models that are not connected differ significantly at the 95% confidence level ($\alpha = 0.05$). Based on the rankings, the LightGBM-meta stacking model achieved the best average rank.

To further analyze classification behavior, confusion matrices for all 12 models are presented in Figure 8. These matrices show the number of *TP, TN, FP*, and *FN*. Among the base models, SSA-RF correctly predicted the most samples, with 91 correct classifications out of 99. For the stacking models, LightGBM-meta achieved the highest classification accuracy, correctly predicting 95 out of 99 samples, including 47 out of 49 failed cases (*TP*) and 48 out of 50 stable cases (*TN*).

The discriminative power of each model was further evaluated using ROC curves and A_{UC} values, as

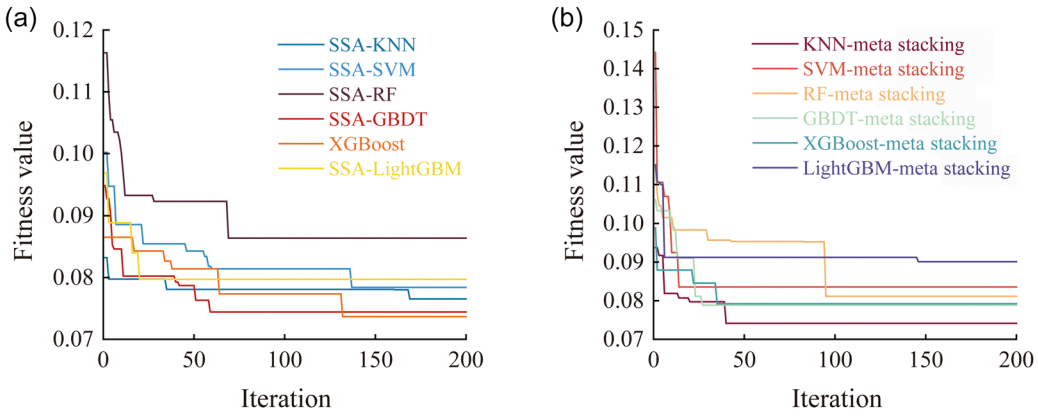


FIGURE 6 Iterative process of individual models and stacked models being optimized by SSA. (a) Individual model optimized by SSA and (b) stacking model optimized by SSA.

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Evaluation results of each model.

Model type	Model	Accuracy	Recall	Precision	$F_{1\text{-score}}$	Kappa
Optimized base models	SSA-KNN	0.860	0.891	0.860	0.875	0.716
	SSA-SVM	0.870	0.873	0.889	0.881	0.738
	SSA-RF	0.880	0.891	0.891	0.891	0.758
	SSA-GBDT	0.890	0.891	0.907	0.899	0.778
	SSA-XGBoost	0.870	0.873	0.889	0.881	0.738
	SSA-LightGBM	0.900	0.909	0.909	0.909	0.798
Stacking ensemble models	KNN-meta stacking	0.880	0.927	0.864	0.895	0.756
	SVM-meta stacking	0.870	0.873	0.889	0.881	0.738
	RF-meta stacking	0.900	0.927	0.895	0.911	0.797
	GBDT-meta stacking	0.890	0.928	0.879	0.903	0.776
	XGBoost-meta stacking	0.860	0.855	0.887	0.87	0.718
	LightGBM-meta stacking	0.930	0.929	0.929	0.937	0.858

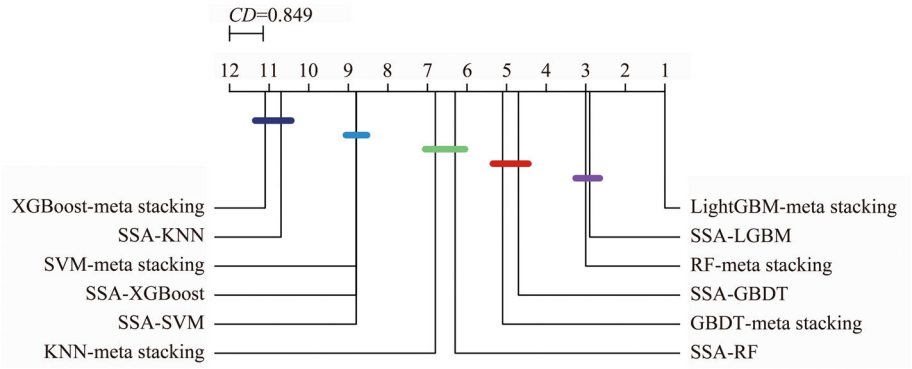


FIGURE 7 Comprehensive comparison of models.

illustrated in Figure 9. All models produced A_{UC} values greater than 0.85, reflecting high classification reliability under various thresholds. The LightGBM-meta stacking model achieved the highest A_{UC} score, further confirming its strong performance in balancing sensitivity and specificity across the decision space.

4.3 | SHAP

In high-risk geotechnical engineering applications, model interpretability plays a vital role in improving reliability and ensuring confidence in real-world deployment (Lundberg & Lee, 2017; Razavi-Termeh et al., 2025). To investigate how the LightGBM-meta stacking model makes predictions, SHAP was employed to interpret feature contributions. It should be emphasized that SHAP does not define physical or engineering failure thresholds. Instead, SHAP provides a quantitative decomposition of the model output by attributing the prediction to individual input features in terms of magnitude and direction.

Figure 10 presents two SHAP visualizations: a summary bar plot on the left and a beeswarm plot on the right. The bar plot ranks features according to their mean absolute SHAP values, reflecting their average contribution to the model's output. The beeswarm plot shows the distribution of SHAP values across all samples for each feature, highlighting both the direction and magnitude of influence depending on the feature's value. From the summary bar plot, P_s emerged as the most influential feature with a mean SHAP value of 3.21. P_S accounts for approximately 28% of the total SHAP contribution. It was followed by *UCS* at 2.75 and the *w*/*h* at 2.14. These features indicate that mechanical strength and geometric proportions are the dominant factors affecting pillar stability. In contrast, the *h* and *w* of the pillar had relatively lower importance, with SHAP values of 2.02 and 1.39. Notably, all selected features contribute to the model predictions with non-negligible effects across samples. This balanced contribution indicates that the model does not rely excessively on any single variable and supports the rationality of the selected feature set. In the beeswarm plot, red points represent high feature values, while blue points represent low values. The SHAP value along the horizontal axis quantifies the effect on the model output. For example, high values of P_S tend to have negative SHAP values, which push the model prediction toward the "failure" class. Conversely, low stress values typically yield positive SHAP values,

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

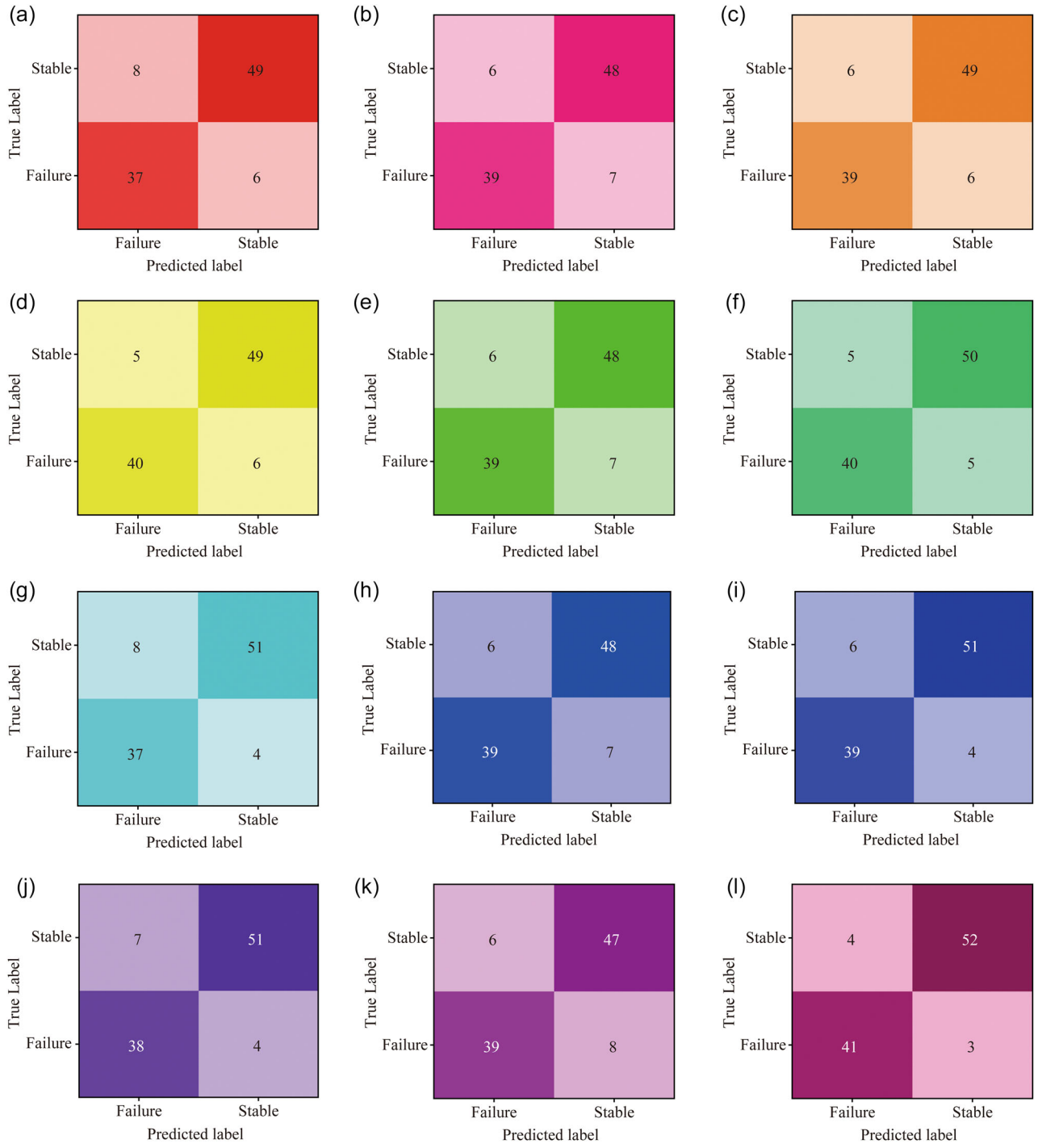


FIGURE 8 Confusion matrix of the models in the test set. (a) SSA-KNN, (b) SSA-SVM, (c) SSA-RF, (d) SSA-GBDT, (e) SSA-XGBoost, (f) SSA-LightGBM, (g) KNN-meta stacking, (h) SVM-meta stacking, (i) RF-meta stacking, (j) GBDT-meta stacking, (k) XGBoost-meta stacking, and (l) LightGBM-meta stacking.

favoring a "stable" classification. This pattern aligns with engineering intuition: excessive stress increases the likelihood of structural failure.

Figure 11 illustrates SHAP interaction plots between features to reveal non-linear interdependencies. Notably, the interaction between P_S and *w*/*h* shows that when the *w*/*h* is low, even moderate stress can drive predictions toward instability. However, when the ratio is high (e.g., greater than 1.5), the same stress level results in predictions skewed toward stability. This reflects the geometric reinforcement effect that benefits taller and wider pillars. Similarly, a strong interaction was observed between *UCS* and *w*, where higher *UCS* combined with greater width further supports stable classification. These interaction effects not only enhance model transparency but also offer actionable insights for engineering design. For example, improving rock strength, reducing P_S, or optimizing pillar geometry can substantially improve structural stability.

From an engineering perspective, SHAP is intended to explain why a given prediction is obtained and which parameters most strongly drive instability. In practical application, engineers can first examine the predicted failure probability provided by the model, and then use SHAP values to identify the dominant negative contributors for a

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

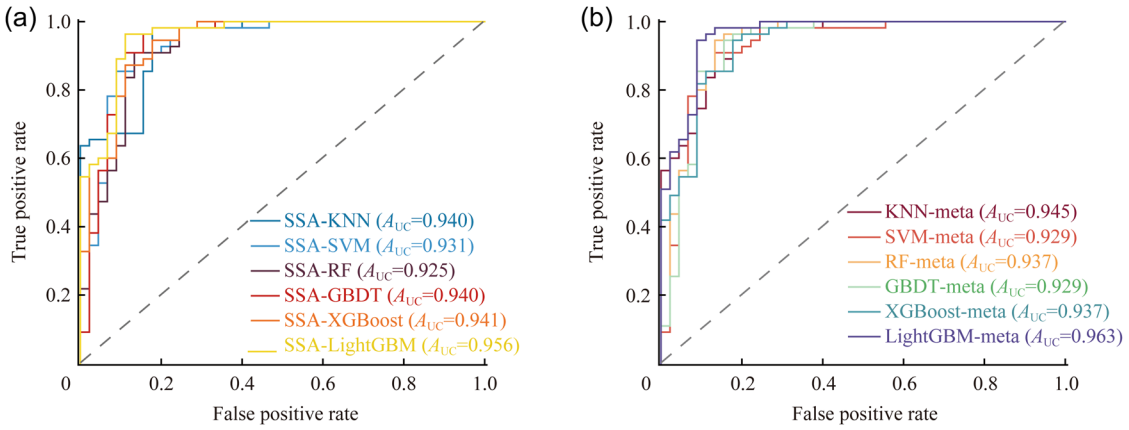


FIGURE 9 ROC curves and A_{UC} value of each model. (a) Optimized base model and (b) stacking ensemble models.

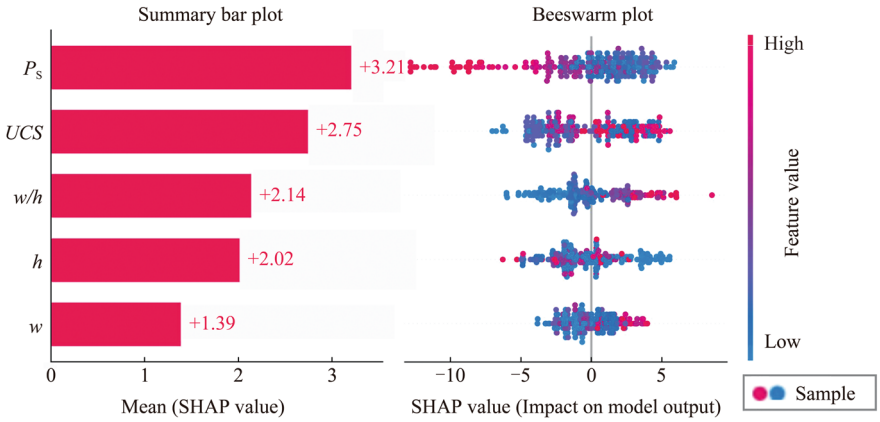


FIGURE 10 Contribution and influence of features on the output of the LightGBM-meta stacking model.

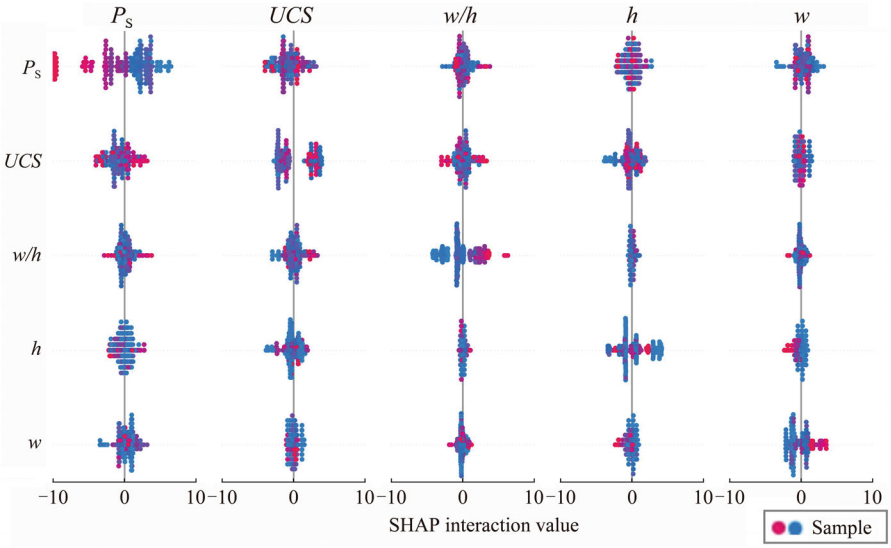


FIGURE 11 SHAP interaction diagram of LightGBM-meta stacking model.

specific pillar case. Targeted "what-if" adjustments can subsequently be performed on controllable variables, such as increasing the *w/h* ratio through geometric modification, improving effective *UCS* via reinforcement or support, or reducing P_S through layout or sequence optimization. By adjusting these parameters and observing the resulting change in prediction, SHAP supports informed and transparent engineering decision-making.

4.4 | Comparison with previous studies

A comparative overview is provided in Table 4, where prior studies are grouped by their original label setting (ternary or binary), and accuracies are interpreted within each setting. Early studies, such as Tawadrous et al. and Zhou et al., reported high accuracies using MLPNN and SVM, but their limited sample sizes restrict

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

generalizability. Later works, including Ding et al. and Liang et al., expanded datasets and applied boosting methods, achieving moderate accuracies between 83.0% and 90.0%. More recent research by Li et al. introduced swarm intelligence-based SVMs, improving accuracy to 91.3%, though still relying on single learners with limited interpretability.

In contrast, the proposed SSA-based stacking framework utilizes a larger and more diverse dataset, incorporates multiple learners, and jointly optimizes both base and meta models. The LightGBM-meta stacking model achieved 93.0% accuracy, demonstrating competitive performance compared with prior approaches under the corresponding label setting. Through SHAP-based interpretability and successful engineering deployment, this study transcends predictive performance by also ensuring practical applicability, interpretability, and field validation—elements largely absent from earlier works.

4.5 | Graphical user interface (GUI) design

To bridge the gap between ML techniques and practical engineering usage, a GUI was developed to facilitate real-time application of the proposed LightGBM-meta stacking model. As shown in Figure 12, the GUI is tailored for mining engineers, particularly those without a background in ML. It enables users to assess pillar stability based on five key input parameters: P_S, *UCS*, *w*, *h*, and *w*/*h*. The interface supports data import, prediction execution, case result visualization, and database management. Upon prediction, the GUI also generates SHAP-based interpretability plots to help users understand which parameters most influenced the model's output. The design incorporates modular extensibility, allowing future updates such as algorithm replacement or adaptation to imbalanced datasets without overhauling the core system. Notably, this GUI has already been deployed in an actual engineering scenario described in the following section. Field engineers successfully used it to evaluate underground pillar stability and reported that the system was accurate, intuitive, and highly practical.

5 | ENGINEERING APPLICATION

5.1 | Project background

The Sanshandao Gold Mine, located in Shandong Province, China, is one of the largest underground gold producers in Asia. The deposit lies within the Jiaodong metallogenic belt and is hosted in fractured granite rock masses, with steeply dipping ore bodies intersected by multiple faults and joints. These geological conditions, combined with the mine's large-scale operations, present significant challenges for underground stability control.

In mining zones located above −600 m, a downward horizontal drift-and-fill mining method is employed. This technique involves first completing the stope excavation, followed by delayed backfilling. As a result, inter-panel pillars are left exposed for prolonged periods, typically ranging from 6 months to 1 year, before being reinforced by fill. These unsupported pillars must maintain

TABLE 4 Comparison with previous studies.

Classification task	Reference	Model	Input	Dataset size	Pillar label	Accuracy (%)
Ternary classification (S–U–F)	Zhou et al. (2015)	SVM	*h*, *w*, *w*/*h*, *UCS*, P_S	251	S, U, F	82.4
		RF				82.4
	Ghasemi et al. (2017)	J48	*w*/*h*, P_S/*UCS*	178	S, U, F	84.8
		SVM				82.0
	Ding et al. (2018)	SGB	*h*, *w*, *w*/*h*, *UCS*, P_S	205	S, U, F	90.0
		RF				85.0
	Li et al. (2023)	GWO-SVM	*h*, *w*, *w*/*h*, *UCS*, P_S	306	S, U, F	90.2
		WOA-SVM				86.9
		SSA-SVM				91.3
	Kadkhodaei et al. (2025)	DT-SVM	*h*, *w*, *UCS*, P_S	236	S, U, F	91.4
Binary classification (S–F)	Tawadrous and Katsabanis (2007)	MLPNN	MT, DO, OBH, PT, L, SH, BH, RMR	60	S, F	93.0
	Zhou et al. (2011)	SVM	*h*, *w*, *w*/*h*, *UCS*, P_S	46	S, F	97.5
	Liang et al. (2020)	GBDT	*h*, *w*, *w*/*h*, *UCS*, P_S	236	S, F	83.1
		XGBoost				83.1
		LightGBM				81.9
	Li et al. (2022)	LMT	*H*, *h*, *w*, B, *w*/*h*	351	S, F	80.5
	This study	Stacking model	*h*, *w*, *w*/*h*, *UCS*, P_S	331	S, F	93.0

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

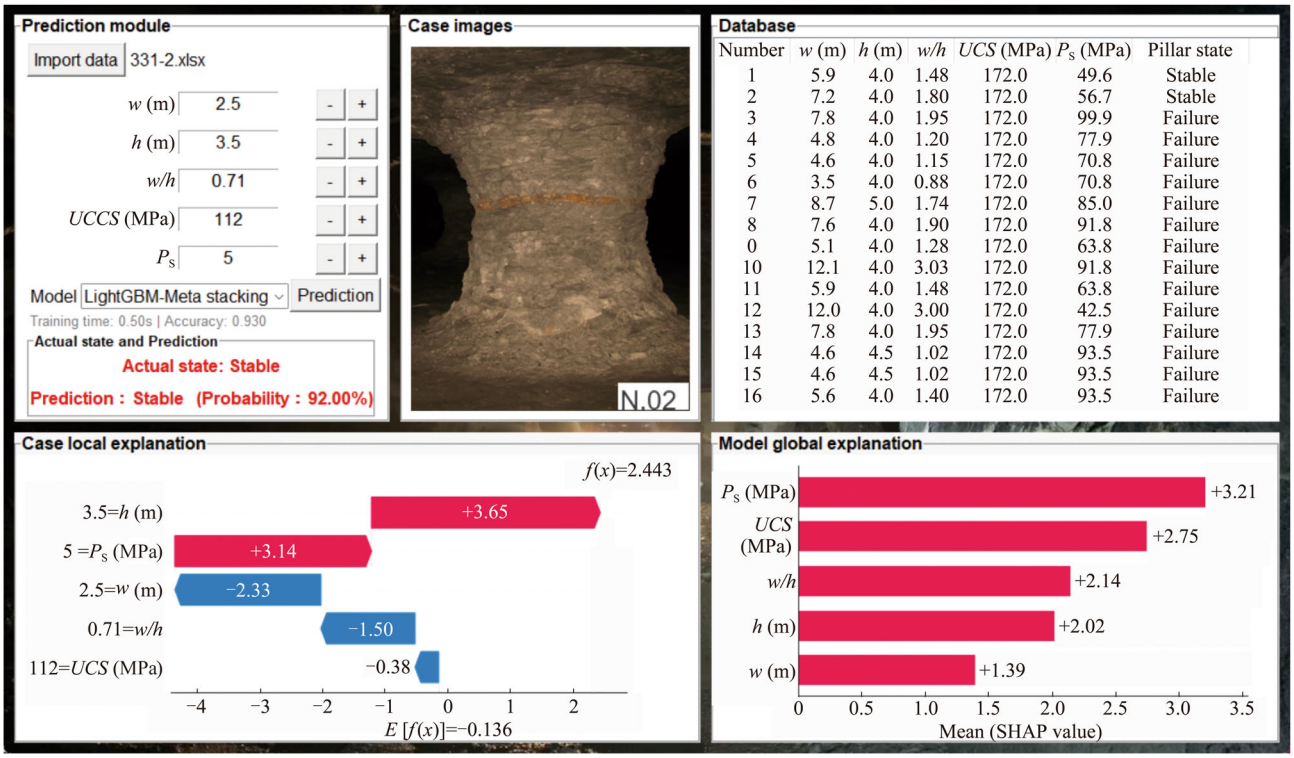


FIGURE 12 GUI interface integration of the proposed method.

structural integrity throughout the exposure duration, making reliable stability prediction critically important for both safety and production continuity. To assist field engineers in rapid decision-making, the developed GUI-based assessment tool was applied at this site.

5.2 | Verification results

To evaluate the model in a practical setting, an on-site investigation was conducted at the Sanshandao Gold Mine. As illustrated in Figure 13, 11 pillar cases were examined across different mining levels. The field survey included detailed recording of stability conditions and data acquisition using advanced measurement techniques.

The truth pillar state labels were determined primarily based on onsite stability investigation, including visual inspection of spalling, cracking, and deformation. Three-dimensional laser scanning was used to generate point cloud models of each pillar. This non-contact method enabled precise calculation of pillar width, height, and cross-sectional area while ensuring personnel safety. In addition, core samples were collected near each pillar to conduct laboratory tests for *UCS*. The in-situ stress of the mining area was measured using the stress relief method. These data, together with the spacing and dimensions of surrounding pillars, were used to estimate the P_S based on mechanical principles.

All five input parameters were entered into the GUI for prediction using the SSA-optimized stacking ensemble model. The GUI outputs both the predicted stability label and the corresponding stable probability, which represents the model's confidence in the prediction. In this study, a default decision threshold of 0.5 was adopted, such that cases with a stable probability greater than or equal to 0.5 were classified as Stable, and those below this threshold were classified as Failure. As shown in Table 5, the LightGBM-meta stacking model correctly predicted the stability outcomes for all 11 cases. This investigation demonstrates the model's reliability under complex field conditions. The GUI also proved practical in real-world usage, enabling engineers to efficiently assess pillar stability using measurable input parameters, without requiring ML expertise.

6 | RESEARCH SIGNIFICANCE AND LIMITATIONS

Given the limitations of previous methods and the pressing need for reliable, interpretable, and deployable pillar stability models, this study delivers several notable advancements, which are outlined below:

1. A novel hybrid stacking ensemble method is proposed by integrating SSA with a stacking ensemble learning, enabling efficient and global hyperparameter optimization for both base and meta classifiers.
2. Statistical comparison through Friedman and Nemenyitests confirmed the superiority of the proposed method over previous ML-based models in terms of accuracy, robustness, and generalization.
3. A comprehensive model interpretability strategy is employed using SHAP analysis, revealing consistent physical insights and enhancing engineering trust.
4. An engineering-friendly GUI is developed and validated through field investigation in Sanshandao

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

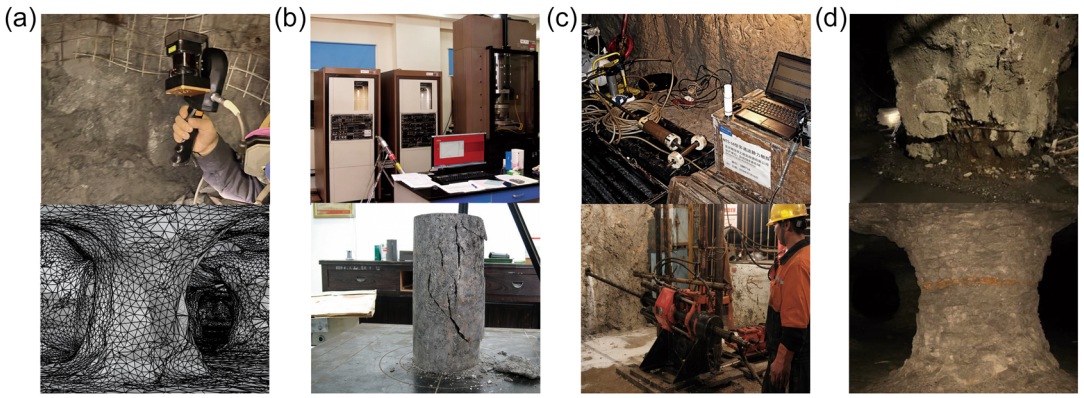


FIGURE 13 Case investigation process. (a) 3D scanning to obtain geological point cloud data, (b) measurement of uniaxial compressive strength, (c) stress relief method to measure ground stress, and (d) pillar stability investigation.

TABLE 5 Prediction results of 11 cases.

Case	Parameter					Actual state	LightGBM-meta Stacking	
	w (m)	*h* (m)	*w*/*h*	*UCS* (MPa)	P_S (MPa)		Predicted state	Stable probability
1	2.4	3.5	0.69	112	5.3	S	S	0.920
2	2.5	3.5	0.71	112	5.0	S	S	0.918
3	2.6	3.5	0.74	112	4.7	S	S	0.930
4	3.7	5.1	0.73	87	5.0	F	F	0.130
5	3.8	5.3	0.72	87	4.3	F	F	0.130
6	6.0	5.2	1.15	132	7.5	S	S	0.530
7	6.1	7.0	0.87	132	11.0	F	F	0.900
8	6.2	7.0	0.89	132	9.5	F	F	0.008
9	7.4	5.1	1.45	87	25.0	S	S	0.012
10	7.4	5.0	1.48	87	26.0	S	S	0.002
11	7.6	5.2	1.46	87	21.0	S	S	0.015

Gold Mine, providing practical and intuitive decision support for underground mining operations.

Despite the strong performance of the proposed SSA-based stacking ensemble model, several technical limitations remain that point to opportunities for further development:

1. The model relies on five input parameters that are readily obtainable in practical mining environments. While this facilitates broad applicability, it may limit the model's sensitivity to complex geological conditions. Important rock mass characteristics such as joint orientation, joint persistence, and groundwater conditions were not considered, yet these factors can significantly influence pillar behavior in fractured rock systems. Accordingly, the proposed model is positioned as a practical screening tool rather than a fully physics-resolved representation.
2. The current dataset is composed entirely of hard rock cases. This narrows the model's representativeness and leaves its applicability to soft or weakly cemented formations unverified. Extending the database to include more varied lithologies is necessary to enhance generalizability across geomechanical settings.
3. The model does not explicitly account for mining-induced disturbances such as stress redistribution caused by extraction sequences, blasting vibrations, or time-dependent degradation around exposed pillars. These dynamic effects are known to influence pillar stability and may require integration of process-aware variables or time-sequenced data in future model development.

7 | CONCLUSIONS

This study proposed a hybrid stacking ensemble method for predicting hard rock pillar stability, integrating six machine learning models with the SSA and cross-validation to enhance both accuracy and interpretability.

1. The models used included KNN, SVM, RF, GBDT, XGBoost, and LightGBM, all serving as both base learners and meta learners within the stacking architecture. Hyperparameters were optimized using SSA with five-fold cross-validation. A total of 331 pillar cases were collected to support model development and evaluation. The LightGBM-meta stacking model

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

achieved the highest overall performance, with an accuracy of 0.930, precision of 0.917, recall of 0.931, $F_{1\text{-score}}$ of 0.924, and a Cohen's kappa value of 0.858. These results were statistically validated using the Friedman test and the Nemenyi post-hoc test.
2. Feature importance analysis based on SHAP values revealed that pillar stress and *UCS* had the greatest impact on model predictions, followed by *w*/*h*, *w*, and *h*. A GUI was developed to facilitate non-expert use. It enables parameter input, real-time prediction, SHAP-based interpretability analysis, and visual feature mapping for individual pillar cases. These functions support intuitive understanding and rapid engineering decisions.
3. The model was applied to a real-world case study at the Sanshandao Gold Mine. Eleven pillars were investigated using 3D laser scanning, *UCS* testing, and in situ stress measurement. The LightGBM-meta stacking model correctly predicted the stability outcome for all cases. The GUI was deployed on-site and used by engineers, demonstrating strong applicability in actual mining environments.

In summary, the proposed framework provides a reliable, interpretable, and field-adaptable solution for pillar stability prediction. It bridges data-driven modeling and engineering applications, and offers potential for further expansion with dynamic monitoring data and geological complexity in future studies.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Laboratory of Mine Design and Numerical Modeling at McGill University for providing methodological guidance and technical resources. Computational support was generously provided by the Department of Applied Mathematics at the Technical University of Madrid. The authors also express sincere appreciation to the engineering team at the Sanshandao Gold Mine for their assistance and collaboration throughout the field investigation phase of this study. This research was supported by the Hunan Provincial Department of Education Foundation (Grant no. CX20230167), the Fundamental Research Funds for the Central Universities of Central South University (Grant no. 2024zzts0084), and the China Scholarship Council (Grant nos. 202506370210 and 202406370221).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Guoyan Zhao https://orcid.org/0000-0001-7416-3126
Meng Wang https://orcid.org/0000-0003-4596-6338

REFERENCES

Aksoy CO, Onargan T. Sizing room and pillar by numerical modeling for underground marble quarries in Turkey. *J Min Sci*. 2006;42(5):483-489. doi:10.1007/s10913-006-0077-9

Al Shamsi AA, Abdallah S. Ensemble stacking model for sentiment analysis of Emirati and Arabic dialects. *J King Saud Univ Comp Inform Sci*. 2023;35(8):101691. doi.org/10.1016/j.jksuci.2023.101691

Alvarez-Garcia IN, Ramos-Lopez FL, Gonzalez-Nicieza C, Alvarez-Fernandez MI, Alvarez-Vigil AE. The mine collapse at Lo Tacón (Murcia, Spain), possible cause of the Torre Pacheco earthquake (2nd May 1998, SE Spain). *Eng Fail Anal*. 2013;28:115-133. doi:10.1016/j.engfailanal.2012.09.009

Brady BH, Brown ET. *Rock Mechanics: For Underground Mining*. Springer Science & Business Media; 2006. doi:10.1007/978-1-4020-2116-9

Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

Chatzimparmpas A, Martins RM, Kucher K, Kerren A. StackGenVis: alignment of data, algorithms, and models for stacking ensemble learning using performance metrics. *IEEE Trans Vis Comput Graph*. 2021;27(2):1547-1557. doi:10.1109/TVCG.2020.3030352

Chen T, Guestrin C. Xgboost: a scalable tree boosting system. *Proc 22nd ACM Sigkdd Int Conf Knowl Disc Data Min*. 2016:785-794. doi:10.1145/2939672.2939785

Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297. doi:10.1007/BF00994018

Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory*. 1967;13(1):21-27. doi:10.1109/TIT.1967.1053964

Cundall PA, Strack ODL. A discrete numerical model for granular assemblies. *Géotechnique*. 1979;29(1):47-65. doi:10.1680/geot.1979.29.1.47

Deng J, Yue ZQ, Tham LG, Zhu HH. Pillar design by combining finite element methods, neural networks and reliability: a case study of the Feng Huangshan copper mine, China. *Int J Rock Mech Min Sci*. 2003;40(4):585-599. doi:10.1016/s1365-1609(03)00042-x

Ding H, Li G, Dong X, Lin Y. Prediction of pillar stability for underground mines using the stochastic gradient boosting technique. *IEEE Access*. 2018;6:69253-69264. doi:10.1109/access.2018.2880466

Esterhuizen GS, Dolinar DR, Ellenberger JL. Pillar strength in underground stone mines in the United States. *Int J Rock Mech Min Sci*. 2011;48(1):42-50. doi:10.1016/j.ijrmms.2010.06.003

Fatty A, Li AJ, Qian ZG. An interpretable evolutionary extreme gradient boosting algorithm for rock slope stability assessment. *Multimed Tools Appl*. 2023;83(16):46851-46874. doi:10.1007/s11042-023-17445-9

Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189-1232. doi:10.1214/aos/1013203451

Gao L, Pastor M, Li T, et al. A framework coupled neural networks and SPH depth integrated model for landslide propagation warning. *Acta Geotech*. 2023;18(7):3863-3888. doi:10.1007/s11440-022-01774-4

Gao W, Zang M, Mei G. 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*. 2024;83(1):2. doi:10.1007/s10064-023-03466-z

Gharehchopogh FS, Namazi M, Ebrahimi L, Abdollahzadeh B. Advances in sparrow search algorithm: a comprehensive survey. *Arch Comput Methods Eng*. 2023;30(1):427-455. doi:10.1007/s11831-022-09804-w

Ghasemi E, Kalhori H, Bagherpour R. Stability assessment of hard rock pillars using two intelligent classification techniques: a comparative study. *Tunnel Undergr Space Technol*. 2017;68:32-37. doi:10.1016/j.tust.2017.05.012

González-Nicieza C, Álvarez-Fernández MI, Menéndez-Díaz A, Álvarez-Vigil AE. A comparative analysis of pillar design methods and its application to marble mines. *Rock Mech Rock Eng*. 2006;39(5):421-444. doi:10.1007/s00603-005-0078-z

Granata F, Di Nunno F, de Marinis G. Stacked machine learning algorithms and bidirectional long short-term memory networks for multi-step ahead streamflow forecasting: a comparative study. *J Hydrol*. 2022;613:128431. doi:10.1016/j.jhydrol.2022.128431

Guo N, Zhao J. A coupled FEM/DEM approach for hierarchical multiscale modelling of granular media. *Int J Numer Meth Eng*. 2014;99(11):789-818. doi:10.1002/nme.4702

Hedley D, Grant F. *Stope-and-Pillar Design for the Elliot Lake Uranium Mines*. Vol. 2, No. 273. Canadian Institute of Mining, Metallurgy and Petroleum.

Idris MA, Saiang D, Nordlund E. Stochastic assessment of pillar stability at Laisvall mine using artificial neural network. *Tunnel Undergr Space Technol*. 2015;49:307-319. doi:10.1016/j.tust.2015.05.003

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Itasca Consulting Group, Inc. *FLAC3D—Fast Lagrangian Analysis of Continua in Three-Dimensions, Ver. 9.0*. Itasca Consulting Group, Inc.; 2023.

Kadkhodaei MH, Ghasemi E, Zhou J, Zahraei M. Evaluation of underground hard rock mine pillar stability using gene expression programming and decision tree-support vector machine models. *Deep Undergr Sci Eng*. 2025;4(1):18-34. doi:10.1002/dug2.12115

Kardani N, Zhou A, Nazem M, Shen S-L. Improved prediction of slope stability using a hybrid stacking ensemble method based on finite element analysis and field data. *J Rock Mech Geotech Eng*. 2021;13(1):188-201. doi:10.1016/j.jrmge.2020.05.011

Ke G, Meng Q, Finley T, et al. Lightgbm: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;30:3146-3154.

Von Kimmelmann M, Hyde B, Madgwick R. The use of computer applications at BCL Limited in planning pillar extraction and the design of mining layouts. *Design Perform Undergr Excavat*, 1984:53-63.

Lawal AI, Kwon S. Application of artificial intelligence to rock mechanics: an overview. *J Rock Mech Geotech Eng*. 2021;13(1):248-266. doi:10.1016/j.jrmge.2020.05.010

Lazzarini R, Tianfield H, Charissis V. A stacking ensemble of deep learning models for IoT intrusion detection. *Knowledge-Based Syst*. 2023;279:110941. doi:10.1016/j.knosys.2023.110941

Li C, Zhou J, Armaghani DJ, Cao W, Yagiz S. Stochastic assessment of hard rock pillar stability based on the geological strength index system. *Geomech Geophys Geo-Energy Geo-Res*. 2021;7(2):47. doi:10.1007/s40948-021-00243-8

Li C, Zhou J, Du K, Dias D. Stability prediction of hard rock pillar using support vector machine optimized by three metaheuristic algorithms. *Int J Min Sci Technol*. 2023;33(8):1019-1036. doi:10.1016/j.ijmst.2023.06.001

Li N, Zare M, Yi C, Jimenez R. Stability risk assessment of underground rock pillars using logistic model trees. *Int J Environ Res Public Health*. 2022;19(4):2136. doi:10.3390/ijerph19042136

Li Y, Li K, Feng X, Cai M. Development and evaluation of artificial expandable pillars for hard rock mining. *Int J Rock Mech Min Sci*. 2018;110:68-75. doi:10.1016/j.ijrmms.2018.07.014

Li Z, Dou L, Cai W, et al. Investigation and analysis of the rock burst mechanism induced within fault–pillars. *Int J Rock Mech Min Sci*. 2014;70:192-200. doi:10.1016/j.ijrmms.2014.03.014

Liang W, Luo S, Zhao G, Wu H. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics*. 2020;8(5):765. doi:10.3390/math8050765

Liu L, Zhao G, Liang W, Jian Z. Hybrid stacking ensemble algorithm and simulated annealing optimization for stability evaluation of underground entry-type excavations. *Undergr Space*. 2024;17:25-44. doi:10.1016/j.undsp.2023.11.002

Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Adv Neural Inf Process Syst*. 2017;30:4765-4774.

Lunder PJ. *Hard Rock Pillar Strength Estimation an Applied Empirical Approach*. PhD thesis. University of British Columbia; 1994.

Mark C. Science of empirical design in mining ground control. *Int J Min Sci Technol*. 2016;26(3):461-470. doi:10.1016/j.ijmst.2016.02.008

Martin CD, Maybee WG. The strength of hard-rock pillars. *Int J Rock Mech Min Sci*. 2000;37(8):1239-1246. doi:10.1016/S1365-1609(00)00032-0

Maybee WG. *Pillar Design in Hard Brittle Rocks*. MSc thesis. Laurentian University; 2000.

Mortazavi A, Hassani FP, Shabani M. A numerical investigation of rock pillar failure mechanism in underground openings. *Comp Geotech*. 2009;36(5):691-697. doi:10.1016/j.compgeo.2008.11.004

Mu G, Wei Q, Xu Y, Zhang H, Zhang J, Li Q. Capacity estimation for lithium-ion batteries based on heterogeneous stacking model with feature fusion. *Energy*. 2024;313:133881. doi:10.1016/j.energy.2024.133881

Munjiza A. *The Combined Finite-Discrete Element Method*. 1st edn. Wiley; 2004. doi:10.1002/0470020180

Murali Mohan G, Sheorey PR, Kushwaha A. Numerical estimation of pillar strength in coal mines. *Int J Rock Mech Min Sci*. 2001;38(8):1185-1192. doi:10.1016/s1365-1609(01)00071-5

Peng K, Liu X, Huang W, He H, Ren J, Luo S. The role of minimum principal stress in tunnel strainburst considering spatial structure effect. *Theor Appl Fract Mech*. 2025;139:105085. doi:10.1016/j.tafmec.2025.105085

Potvin Y. Strategies and tactics to control seismic risks in mines. *J South Afr Inst Min Metall*. 2009;109(3):177-186.

Potvin Y, Hudyma M, Miller H. Rib pillar design in open stope mining. *CIM Bullet*. 1988;81(912):31-36.

Ranjith PG, Zhao J, Ju M, De Silva RVS, Rathnaweera TD, Bandara AKMS. Opportunities and challenges in deep mining: a brief review. *Engineering*. 2017;3(4):546-551. doi:10.1016/J.ENG.2017.04.024

Razavi-Termeh SV, Sadeghi-Niaraki A, Naqvi RA, Choi S-M. Investigating the efficacy of physics-based metaheuristic algorithms in combination with explainable ensemble machine-learning models for landslide susceptibility mapping. *Stochast Environ Res Risk Assess*. 2025;39(3):1109-1141. doi:10.1007/s00477-025-02911-7

Ren J, Zhao Y, Sun Z, Liu S. Risk mapping for oil–gas pipeline under mining-induced subsidence through analytical methods. *Int J Coal Sci Technol*. 2025;12(1):53. doi:10.1007/s40789-025-00788-z

Saadati G, Javankhoshdel S, Mohebbi Najm Abad J, Mett M, Kontrus H, Schneider-Muntau B. AI-powered geotechnics: enhancing rock mass classification for safer engineering practices. *Rock Mech Rock Eng*. 2025;58:11319-11349. doi:10.1007/s00603-024-04189-7

Schubert C, Villaescusa E. An approach to hard rock pillar design at the McArthur River Mine. *AusIMM Annu Conf*. 1998:255-259.

Sjoberg JS. *Failure Modes and Pillar Behaviour in the Zinkgruvan Mine: Vol. The 33rd U.S. Symposium on Rock Mechanics*. USRMS; 1992:491-500.

Tawadrous AS, Katsabanis PD. Prediction of surface crown pillar stability using artificial neural networks. *Int J Numer Anal Methods Geomech*. 2007;31(7):917-931. doi:10.1002/nag.566

Valencia JJM. *A Risk-Based Pillar Design Approach for Improving Safety in Underground Stone Mines*. PhD thesis. Virginia Polytechnic Institute and State University; 2022.

Wagner H. Pillar design in coal mines. *J South Afr Inst Min Metall*. 1980;80(1):37-45.

Wagner H. Deep mining: a rock engineering challenge. *Rock Mech Rock Eng*. 2019;52(5):1417-1446. doi:10.1007/s00603-019-01799-4

Wang M. Modeling Time-Dependent Deformation Behavior of Jointed Rock Mass. PhD thesis, Laurentian University; 2022.

Wang M, Mitri HS, Zhao G, Wang S. Data-based assessment of rock strengths and cuttability using the monitored parameters while drilling, tunneling, and mining. *Tunnel Undergr Space Technol*. 2025;165:106901. doi:10.1016/j.tust.2025.106901

Wang SY, Lam KC, Au SK, Tang CA, Zhu WC, Yang TH. Analytical and numerical study on the pillar rockbursts mechanism. *Rock Mech Rock Eng*. 2006;39(5):445-467. doi:10.1007/s00603-005-0075-2

Wolpert DH. Stacked generalization. *Neural Net*. 1992;5(2):241-259. doi:10.1016/S0893-6080(05)80023-1

Xu C, Ding J, Qiao Y, Zhang L. Tomato disease and pest diagnosis method based on the stacking of prescription data. *Comput Electron Agric*. 2022;197:106997. doi:10.1016/j.compag.2022.106997

Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm. *Syst Sci Contr Eng*. 2020;8(1):22-34. doi:10.1080/21642583.2019.1708830

Yin W, Miao X, Zhang J, Zhong S. Mechanical analysis of effective pressure relief protection range of upper protective seam mining. *Int J Min Sci Technol*. 2017;27(3):537-543. doi:10.1016/j.ijmst.2017.03.021

Zhang Q, Zhang J, Guo S, Gao R, Li W. Design and application of solid, dense backfill advanced mining technology with two pre-driving entries. *Int J Min Sci Technol*. 2015;25(1):127-132. doi:10.1016/j.ijmst.2014.12.008

Zhang Y, Qi H, Li C, Zhou J. Enhancing safety, sustainability, and economics in mining through innovative pillar design: a state-of-the-art review. *J Saf Sustain*. 2024;1(1):53-73. doi:10.1016/j.jsasus.2023.11.001

Zhang Y, Stead D, Elmo D. Characterization of strength and damage of hard rock pillars using a synthetic rock mass method. *Comput Geotech*. 2015;65:56-72. doi:10.1016/j.compgeo.2014.12.002

Zhao Y, Xie R, Wang S, et al. Coal bursting liability determination by needle penetration test: empirical criterion and machine learning. *Int J Coal Sci Technol*. 2024;11(1):87.

Zhou J, Li X, Mitri HS. Comparative performance of six supervised learning methods for the development of models of hard rock pillar stability prediction. *Nat Hazards*. 2015;79(1):291-316. doi:10.1007/s11069-015-1842-3

Zhou J, Li X, Shi X, Wei W, Wu B. Predicting pillar stability for underground mine using Fisher discriminant analysis and

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SVM methods. *Trans Nonferrous Met Soc China*. 2011;21(12): 2734-2743. doi:10.1016/s1003-6326(11)61117-5

Zipf Jr RK. *Toward Pillar Design to Prevent Collapse of Room-and-Pillar Mines*. National Institute for Occupational Safety and Health; 1910.

Zipf Jr RK, Mark C. Design methods to control violent pillar failures in room-and-pillar mines. *Trans Insti Min Metallur Sect A Min Indust*. 1997;106:A124-A132.

AUTHOR BIOGRAPHY



Meng Wang received the PhD degree from Central South University, China. He was also a joint PhD visiting scholar at McGill University, Canada. His research interests include deep learning, rock mechanics, deep mining, and geotechnical engineering. He has authored multiple peer-reviewed articles in leading journals such as *Rock Mechanics and Rock Engineering, Energy, Tunnelling and Underground Space Technology, Transportation Geotechnics*, and the *Journal of Building Engineering*.

How to cite this article: Wang N, Zhao G, Wang M, Pattela T. Explainable hybrid stacking ensemble method for hard rock pillar stability prediction and engineering applications. *Deep Undergr Sci Eng*. 2026;1-19. doi:10.1002/dug2.70101

27701328, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/dug2.70101 by Universit£Â Degli Studi Di, Wiley Online Library on [12/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License