fractal and fractional

MDPI

Article

Data-Driven Approach for Intelligent Classification of Tunnel Surrounding Rock Using Integrated Fractal and Machine Learning Methods

Junjie Ma [1,2], Tianbin Li [1,*], Roohollah Shirani Faradonbeh [2,*], Mostafa Sharifzadeh [2], Jianfeng Wang [1], Yuyang Huang [1], Chunchi Ma [1], Feng Peng [1] and Hang Zhang [3]

[1] College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu 610059, China; majunjie0909@gmail.com (J.M.); wangjianfeng@stu.cdut.edu.cn (J.W.); hyycdut@cdut.edu.cn (Y.H.); machunchi17@cdut.edu.cn (C.M.); pengfeng@stu.cdut.edu.cn (F.P.)
[2] WA School of Mines: Minerals, Energy and Chemical Engineering, Curtin University, Kalgoorlie, WA 6430, Australia; m.sharifzadeh@curtin.edu.au
[3] Chongqing City Construction Investment (Group) Co., Ltd., Chongqing 400023, China; zhanghang_nn720@163.com
* Correspondence: ltb@cdut.edu.cn (T.L.); roohollah.shiranifaradonbeh@curtin.edu.au (R.S.F.)

Abstract: The degree of rock mass discontinuity is crucial for evaluating surrounding rock quality, yet its accurate and rapid measurement at construction sites remains challenging. This study utilizes fractal dimension to characterize the geometric characteristics of rock mass discontinuity and develops a data-driven surrounding rock classification (SRC) model integrating machine learning algorithms. Initially, the box-counting method was introduced to calculate the fractal dimension of discontinuity from the excavation face image. Subsequently, crucial parameters affecting surrounding rock quality were analyzed and selected, including rock strength, the fractal dimension of discontinuity, the discontinuity condition, the in-situ stress condition, the groundwater condition, and excavation orientation. This study compiled a database containing 246 railway and highway tunnel cases based on these parameters. Then, four SRC models were constructed, integrating Bayesian optimization (BO) with support vector machine (SVM), random forest (RF), adaptive boosting (AdaBoost), and gradient boosting decision tree (GBDT) algorithms. Evaluation indicators, including 5-fold cross-validation, precision, recall, F1-score, micro-F1-score, macro-F1-score, accuracy, and the receiver operating characteristic curve, demonstrated the GBDT-BO model's superior robustness in learning and generalization compared to other models. Furthermore, four additional excavation face cases validated the intelligent SRC approach's practicality. Finally, the synthetic minority over-sampling technique was employed to balance the training set. Subsequent retraining and evaluation confirmed that the imbalanced dataset does not adversely affect SRC model performance. The proposed GBDT-BO model shows promise for predicting surrounding rock quality and guiding dynamic tunnel excavation and support.

Keywords: data driven; fractal dimension; discontinuity; surrounding rock classification; machine learning; synthetic minority oversampling technique (SMOTE)

Citation: Ma, J.; Li, T.; Shirani Faradonbeh, R.; Sharifzadeh, M.; Wang, J.; Huang, Y.; Ma, C.; Peng, F.; Zhang, H. Data-Driven Approach for Intelligent Classification of Tunnel Surrounding Rock Using Integrated Fractal and Machine Learning Methods. *Fractal Fract.* **2024**, *8*, 677. https://doi.org/10.3390/fractalfract8120677

Academic Editor: Carlo Cattani

Received: 23 October 2024
Revised: 16 November 2024
Accepted: 18 November 2024
Published: 21 November 2024

1. Introduction

Tunneling is the optimal method for traversing complex and hazardous mountainous terrain, mitigating ground disaster risks to transportation infrastructure construction and operation. While tunnel boring machines (TBMs) and drilling and blasting (DB) techniques are primary excavation methods, geological conditions, technological limitations, and cost considerations often favor DB approaches. Excavation progressively reveals new geological features. The efficient utilization of newly exposed geological data is crucial for assessing surrounding rock quality and guiding dynamic excavation and support protocols.

Scholars have proposed various methods for assessing surrounding rock quality, including rock mass rating (RMR), Q-system, geological strength index (GSI), rock mass index (RMi), and basic quality (BQ) methods [1–5]. Rock strength and the degree of rock mass discontinuity are critical parameters in surrounding rock quality assessment. Rock strength can be indirectly measured at the construction site using point load and rebound test methods. The degree of rock mass discontinuity is a comprehensive indicator that measures geometric and physico-mechanical characteristics of discontinuity. The BQ method adopts the rock integrity index (K_v) to characterize the degree of rock mass discontinuity. K_v is typically obtained through elastic wave testing, which is costly and susceptible to interference. Rock quality designation (RQD) [6], proposed by Deere (1964), is widely used to characterize geometric characteristics of discontinuity. RMR and Q-system methods employ RQD and discontinuity conditions to characterize the degree of rock mass discontinuity jointly. However, the RQD calculation is controversial and fails to reflect rock mass anisotropy. Palmstrøm [7,8] introduced the volumetric joint count (J_v) to measure geometric characteristics of discontinuity and analyzed its empirical relationship with RQD. In the modified GSI method proposed by Sonmez and Ulusay [9], J_v is used to characterize the rock mass structure quantitatively. The RMR14 method uses J_v to replace RQD to characterize the geometric characteristics of discontinuity. However, the rapid and accurate measurement of J_v remains challenging at the tunnel construction site. Recent advancements in artificial intelligence (AI) have enabled image recognition methods to identify the degree of rock mass discontinuity [10,11]. These methods require high-quality images and complex preprocessing, resulting in high costs and limited practicality. Consequently, it is necessary to explore new methods or parameters for measuring the degree of rock mass discontinuity and evaluating surrounding rock quality.

Recent studies have increasingly applied fractal theory to analyze and quantify rock mass conditions and behaviors. Xiao et al. [12] evaluated the blasting damage of rock by calculating the fractal dimension of rock cracks. Y.B. Liu et al. [13] used fractal dimensions to characterize the evolution of mining-induced fractures quantitatively. H.Q. Liu et al. [14] employed an improved box-counting method to investigate the correlations between the rock pore fractal dimensions, porosity, and absolute permeability. Q. Zhang et al. [15] explored the quantitative relationship between fractal dimensions and K_v using the box-counting and Monte Carlo simulation methods. Q. Zhang et al. [16] proposed a method to measure the fractal dimension of outcrops and used it to predict the roughness of discontinuity. B.X. Li et al. [17] and Wu et al. [18] utilized the fractal dimension to characterize the roughness of rock fracture surfaces. L.C. Li et al. [19] adopted fractal dimensions to describe orientation, spacing, and the trace length of joints quantitatively. Hong et al. [20] quantified joint characteristics using fractal dimensions to refine the GSI. Feng et al. [21] explored the correlation between jointed rock mass strength and the fracture quantity distribution's fractal dimension. The above studies show that fractal theory has a high capability to characterize the geometric characteristics of discontinuities through a joint and fracture fractal dimension analysis. However, limited research has explored fractal dimensions for evaluating surrounding rock quality based on discontinuity geometric characteristics.

Furthermore, as deep-buried and ultra-long tunnels proliferate, conventional quantitative classification methods prove inadequate for real-time assessments of surrounding rock conditions during tunnel excavation. Machine learning (ML) algorithms are increasingly employed to develop intelligent classification models for surrounding rocks. These include adaptive boosting machines, artificial neural networks, Bayesian networks, categorical boosting, classification and regression trees, convolutional neural networks, extreme gradient boosting, extremely randomized trees, fuzzy c-means, graph convolutional networks, light gradient boosting, logistic regressions, long short-term memory networks, Markov chains, multiple layer perceptrons, random forests, and support vector machines [22–35]. However, the intelligent prediction of surrounding rock in DB tunnels faces challenges in obtaining discontinuity degrees and related input parameters. Inspired by the aforementioned related fractal applications and considering the statistical self-similarity of

discontinuity, exploring the fractal dimension of discontinuity combined with ML offers a promising approach for the intelligent prediction of surrounding rock quality grades.

This study developed an intelligent classification model for surrounding rock by integrating fractal and ML methods. An approach to obtain the fractal dimension of discontinuity using the box-counting method was proposed. Crucial factors influencing surrounding rock quality were analyzed and selected, including rock strength, the fractal dimension of discontinuity, discontinuity conditions, groundwater conditions, in-situ stress conditions, and excavation orientations. A database of 246 surrounding rock cases was compiled. Bayesian optimization (BO) was employed to determine optimal hyperparameters for SVM, RF, AdaBoost, and GBDT algorithms, establishing four SRC models. Model performance was evaluated using 5-fold cross-validation, confusion matrices (including precision, recall, F1-score, micro-average F1-score, macro-average F1-scores, and accuracy), and the receiver operating characteristic (ROC) curve. The practicality of the proposed intelligent SRC approach was verified using four additional cases. Finally, the synthetic minority over-sampling technique (SMOTE) was utilized to investigate the impact of imbalanced datasets on model performance.

2. Fractal Representation of Discontinuity

Fractal theory can quantify the complexity of discontinuity characteristics, which assists engineers in better understanding the nature of discontinuity and facilitates the efficient and accurate determination of the quality grade of the surrounding rock.

2.1. Fractal Dimension

Mandelbrot [36] proposed the fractal theory to describe complex phenomena in nature. Fractal theory is widely employed to study things with statistical self-similarity, such as protein, chemical structures, and time series data [37–40]. Fractal dimension D_f is a crucial parameter for quantifying fractal characteristics. Typical D_f calculation methods include the Hausdorff method, Higuchi method, box-counting method, Brownian motion method, and detrended fluctuation analysis method. Notably, the box-counting method, due to its simplicity and ease of use, is widely utilized for fractal characterization in geotechnical engineering. Researchers [14,18,41–44] have successfully applied the box-counting method to characterize rock mass conditions and behavior, demonstrating its efficacy in the fractal analysis of geological structures.

The core idea of the box-counting method is to cover the object with squares (i.e., boxes) of varying sizes (ε_i), calculating the minimum number of boxes ($N(\varepsilon_i)$) required for complete coverage at each size. For fractal objects, $N(\varepsilon_i)$–ε_i follows the relationship shown in Equation (1), where i takes a positive integer.

$$N(\varepsilon_i) \propto \varepsilon_i^{-D_f} \tag{1}$$

Equation (2) can be obtained by taking the logarithm of both sides of Equation (1):

$$D_f = -\lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon_i)}{\ln \varepsilon_i} \tag{2}$$

Multiple measurements of the same object yield a series of measurement scales and corresponding box counts. The linear regression analysis of the logarithms of Equation (2) yields a slope whose absolute value represents the fractal dimension D_f, as expressed in Equation (3):

$$\ln N(\varepsilon_i) = -D_f \ln \varepsilon_i + \ln C \tag{3}$$

2.2. Fractal Characteristics of Discontinuity

Rock mass discontinuity encompasses geometric and physico-mechanical characteristics. Geometric characteristics involve number, spacing, persistence length, and the orientation of discontinuities. Physico-mechanical attributes comprise aperture, rough-

ness, filling, and the weathering of discontinuities. Numerous studies demonstrate that rock mass discontinuities exhibit statistical self-similarity, which fractal theory effectively characterizes [15–18,41]. The fractal dimension increases with the degree of discontinuity development.

Determining the fractal dimension of rock mass discontinuity at the tunnel construction sites poses significant challenges. Therefore, this study integrated fractal theory with image processing techniques to characterize the geometric characteristics of rock mass discontinuity quantitatively. However, recognizing that two-dimensional images cannot accurately reflect the physico-mechanical characteristics of discontinuities, this study also incorporates field surveys to obtain discontinuity conditions. Figure 1 illustrates the three-step process for determining the fractal dimension of geometric features in rock mass discontinuities.

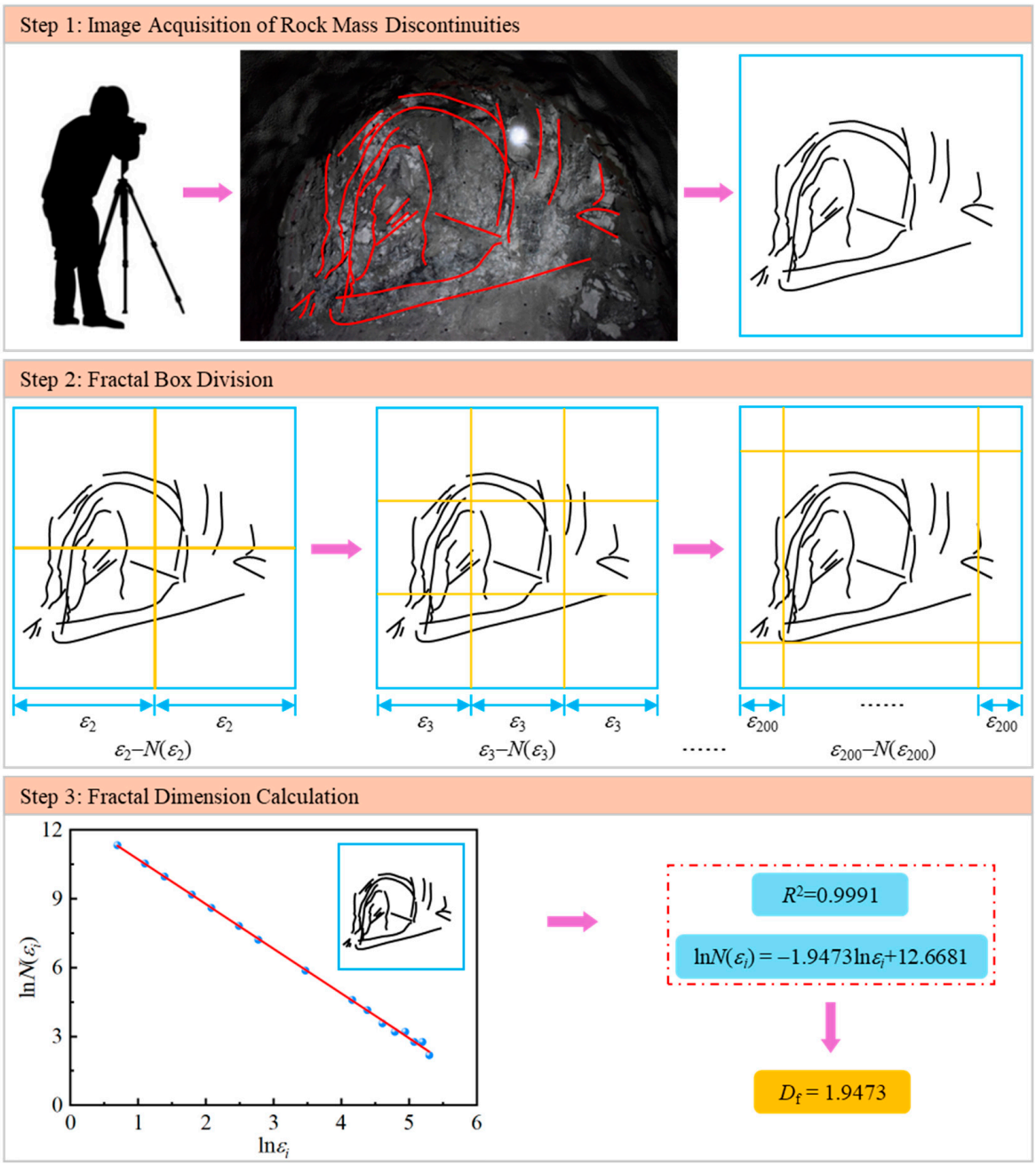

Figure 1. Flow chart of fractal dimension D_f calculation for geometric features of rock mass discontinuity.

Step 1 (Image acquisition): Capturing high-resolution images of the tunnel excavation face using an infrared camera or a high-performance mobile phone. Critical discontinuities affecting face quality and stability are manually delineated.

Step 2 (Fractal box division): Dividing square fractal boxes iteratively until the number of non-empty boxes stabilizes. Box sizes range from 2 to 200 in this study.

Step 3 (Fractal dimension calculation): The fractal dimension D_f is obtained through linear fitting of the logarithms of box size and box count, as shown in Equation (3).

Steps 2 and 3 can be executed directly in ImageJ 1.54g software. The discontinuity image from Step 1 is first binarized in ImageJ, followed by a fractal dimension calculation using the software's built-in function (i.e., Fractal box count).

3. Classification Indicator and Database

The surrounding rock quality grade is a comprehensive indicator reflecting the surrounding rock conditions. Therefore, determining the critical control factors of surrounding rock quality is essential for building a database and intelligent classification model.

3.1. Indicator Analysis and Selection

Factors influencing surrounding rock quality can be categorized into rock mass composition and correction factors. Rock mass composition encompasses rock strength and discontinuity characteristics. While high rock strength indicates the better bearing capacity of the surrounding rock, it also presents an elevated risk of rockburst under overstress conditions and excavation disturbance [45]. The uniaxial compressive strength of rock (R_c) is the most widely utilized indicator for characterizing rock strength. Discontinuity characteristics comprise geometric and physico-mechanical properties. As demonstrated in Section 2, the fractal dimension D_f effectively characterizes the geometric aspects of discontinuities. The opening, roughness, and filling conditions of discontinuities reflect their physico-mechanical properties. In the RMR, RMR14, HC, and BQ classification methods, the discontinuity condition serves as a proxy for these physico-mechanical characteristics. The surrounding rock structure is intimately linked to the distribution and development of discontinuities, which govern the bearing capacity, deformation, and failure behavior of the surrounding rock [4].

Correction factors for surrounding rock quality include groundwater conditions, in-situ stress conditions, and the relationship between the major weak structural plane and the tunnel axis direction. Groundwater can diminish rock strength and erode and expand discontinuities, thereby reducing surrounding rock quality. Underground projects in water-rich and weak surrounding rock areas are susceptible to water and mud inrush disasters [46]. Moreover, in high in-situ stress environments, hard rocks are prone to rockburst, while soft rocks tend to exhibit squeezing deformation [47–49]. Conversely, in low in-situ stress environments, weak and fractured rocks face an increased risk of falling blocks and collapse. The ratio of R_c to σ_1 (maximum principal stress) is frequently employed to quantify in-situ stress conditions. Additionally, rock mass behavior is related to the different combinations of excavation direction and the occurrence of major structural planes. Unfavorable discontinuity orientation can precipitate collapse or the sliding failure of surrounding rock [50]. The relationship between the major weak structural plane and the tunnel axis direction can be called excavation orientation. In particular, the RMR, RMR14, Q-system, HC, and BQ methods all consider the influence of groundwater conditions, in-situ stress conditions, and the excavation orientation on surrounding rock quality.

In summary, rock strength, the fractal dimension of discontinuity, discontinuity conditions, groundwater conditions, in-situ stress conditions, and the excavation orientation constitute a comprehensive indicator system for the intelligent classification of surrounding rock quality (see Figure 2). Table 1 presents the qualitative classification criteria for subcategories of discontinuity conditions, groundwater conditions, and the excavation orientation. More details about the determination methods of these parameters can be referred to Wu et al. [5].

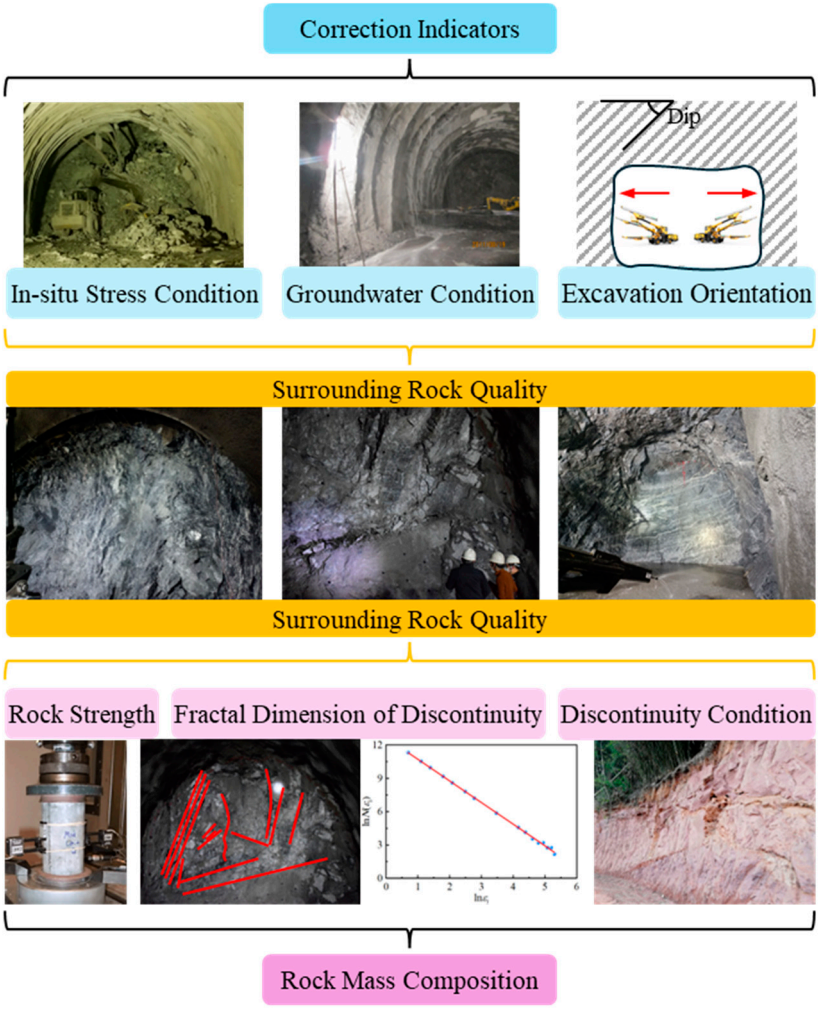

Figure 2. Indicator system of surrounding rock classification.

Table 1. Subcategories classes of discontinuity condition, groundwater condition, and excavation orientation.

Parameters	Intervals (Discrete Value)
Discontinuity condition	Good (1), moderate (2), poor (3), very poor (4)
Groundwater condition	Dry (1), damp (2), wet (3), dripping (4), rain-like dripping (5), linear-like flowing (6), tubular-like flowing (7), gushing-like flowing (8)
Excavation orientation	Very favorable (1), favorable (2), fair (3), unfavorable (4), very unfavorable (5)

3.2. Database Description

This study developed a database comprising 246 railway and highway tunnel cases based on the indicator system in Figure 2. Figure 3 illustrates the distribution of surrounding rock parameters within the database. The horizontal axis values of Figure 3c,d,f correspond to the subcategories and discrete values of discontinuity condition, groundwater condition, and excavation orientation, respectively, as outlined in Table 1. The following observations can be inferred from this figure.

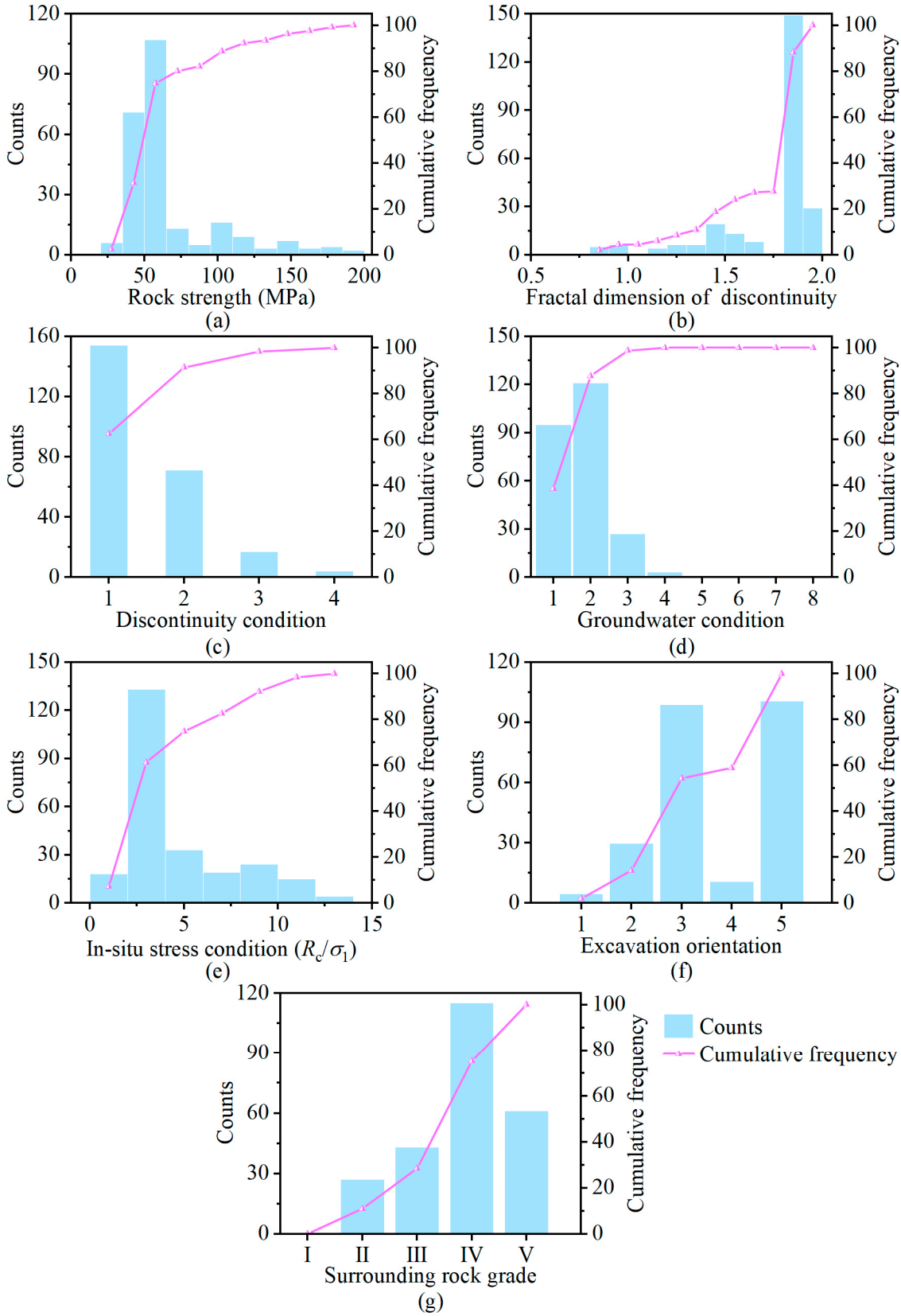

Figure 3. Statistical distribution of surrounding rock parameters in the database: (**a**) distribution of rock strength; (**b**) distribution of fractal dimension of discontinuity; (**c**) distribution of discontinuity condition; (**d**) distribution of groundwater condition; (**e**) distribution of in-situ stress condition; (**f**) distribution of excavation orientation; and (**g**) distribution of surrounding rock grade.

(1) Rock strength exhibits a wide range, predominantly between 40 and 60 MPa. (2) Fractal dimension clusters around 1.8–1.9, indicating moderate discontinuity in most samples. (3) Discontinuity conditions are primarily moderate to good. (4) Groundwater conditions are predominantly dry to damp. (5) The majority of cases display strength–stress ratios (R_c/σ_1) below 4, suggesting prevalent high in-situ stress conditions with cases of the database. (6) The relationship between the major weak structural plane and the tunnel axis direction shows significant uncertainty, influenced by discontinuity distribution. (7) Surrounding rock grades are unevenly distributed, with Classes IV and V predominating. It should be noted that surrounding rock grades were determined using the BQ method, detailed in Wu et al. [5].

4. Intelligent Model Training and Evaluation

Data-driven intelligent classification methods for surrounding rock can accurately and in real-time determine surrounding rock grades, which is significant for dynamic tunnel construction. Therefore, this study employed various machine learning algorithms to develop intelligent SRC models and determined the optimal classifier for surrounding rock grade through a standard model evaluation process.

4.1. Algorithms Description

4.1.1. Support Vector Machine (SVM)

The core principle of SVMs is to identify the optimal hyperplane that maximizes the margin between different classes in the feature space, thus enhancing generalization capability [51].

The Lagrange function is introduced to solve the optimal hyperplane problem, as expressed in Equation (4).

$$L(\boldsymbol{w}, b, \alpha) = \frac{\|w\|^2}{2} - \sum_{i=1}^{k} \alpha_i \left[y_i \left(\boldsymbol{w}^{\mathrm{T}} \cdot \boldsymbol{x}_i + b \right) - 1 \right] \tag{4}$$

where $\boldsymbol{w}$ is the weight vector, b is the bias, α_i is the Lagrange multiplier ($\alpha_i \geq 0$), and x_i is the sample features of the i-sample set ($i = 1, 2, \cdots, k$; $i, k \in \mathrm{N}^*$).

Subsequently, the primal problem can be transformed into a dual problem, with the decision function represented by Equation (5).

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[\sum_{i=1}^{k} \alpha_i^* y_i \left(\boldsymbol{x}_i^T \cdot \boldsymbol{x} \right) + b^* \right] \tag{5}$$

where sgn() is a symbolic function, α_i^* is the optimal Lagrange multiplier, and b^* is the optimal bias.

In particular, the kernel function $\mathrm{K}(\boldsymbol{x}_i{}^T, \boldsymbol{x})$ is introduced to avoid complex operations in high-dimensional space for multi-classification tasks. Equation (5) can be rewritten as Equation (6).

$$f(\boldsymbol{x}) = \operatorname{sgn}\left[\sum_{i=1}^{k} \alpha_i^* y_i \mathcal{K}\left(\boldsymbol{x}_i^T, \boldsymbol{x} \right) + b^* \right] \tag{6}$$

Linear, Polynomial, Rbf, and Sigmoid kernels are frequently used to train SVM models, with the Rbf kernel typically yielding superior performance in multi-classification tasks [52]. More details about the SVM algorithm and its application can be found in Cervantes et al. [51].

4.1.2. Random Forest (RF)

The RF, proposed by Breiman [53], is an ensemble learning algorithm that uses decision trees as base learners. It introduces randomness in the process of training data partitioning and node splitting, effectively reducing the variance of the model. The workflow of RF can be divided into four steps [54].

Step 1 (Bootstrap sampling): The RF performs multiple samplings with a replacement of the original dataset to generate multiple different training subsets.

Step 2 (Building decision trees): Each training subset generated in Step 1 is used to construct an individual decision tree.

Step 3 (Node splitting): Node splitting is performed simultaneously with decision tree construction. At each node, the RF selects a portion of features and calculates their information gain or Gini index to select the optimal splitting feature.

Step 4 (Tree voting): Each decision tree outputs a prediction independently. In classification tasks, the RF determines the final class label through a "majority vote" among the trees. In regression tasks, the final prediction is determined by averaging the predictions of all decision trees.

In the RF classification model, "n_estimators", "max_depth", and "min_samples_split" are the most critical hyperparameters, which play an essential role in improving the prediction accuracy of the model and preventing the model from overfitting.

4.1.3. Adaptive Boosting (AdaBoost)

AdaBoost algorithm combines multiple weak classifiers to build a strong classifier. It continuously adjusts the weights of weak classifiers through iteration to gradually improve the overall performance of the model. The workflow of AdaBoost can be divided into four steps.

Step 1 (Initializing weights): Assigning equal weights to all samples, $w_i = 1/N$, $i = 1, 2, \cdots, N$. N is the total number of samples.

Step 2 (Training weak classifier): In the mth iteration, fitting a weak classifier $h^{(m)}(x)$ based on the current sample weights ${w_i}^{(m)}$.

Then, calculating the classification error $E^{(m)}$ based on Equation (7) x_i is the input parameter, y_i is sample label. II() is an indicator function that equals 1 if the prediction is incorrect and 0 otherwise.

$$E^{(m)} = \sum_{i=1}^{N} w_i^{(m)} \cdot \mathbb{I}\left(h^{(m)}(x_i) \neq y_i\right) \tag{7}$$

Step 3 (Calculating and updating weights): Calculating the weight of weak classifier ($\alpha^{(m)}$) based on Equation (8).

$$\alpha^{(m)} = \frac{1}{2} \log\left(\frac{1 - E^{(m)}}{E^{(m)}}\right) \tag{8}$$

The smaller the classification error is, the larger the weight assigned to the weak classifier. Subsequently, the sample weights are updated based on the weight of the weak classifier, which is shown as Equation (9). Furthermore, the updated weights are normalized so that the sum of all weights is 1.

$$w_i^{(m+1)} = w_i^{(m)} \cdot \exp\left(-\alpha^{(m)} \cdot y_i \cdot h^{(m)}(x_i)\right) \tag{9}$$

Step 4 (Output): The final strong classifier ($H(x)$) is a weighted combination of all weak classifiers, which is shown as Equation (10). M is the total number of weak classifiers.

$$H(x) = \text{sign}\left(\sum_{m=1}^{M} \alpha^{(m)} \cdot h^{(m)}(x_i)\right) \tag{10}$$

AdaBoost was initially designed for binary classification tasks. Hastie et al. [55] extended it to multi-classification tasks, in which the hyperparameter "algorithm" needs to be set to "SAMME".

4.1.4. Gradient Boosting Decision Tree (GBDT)

The GBDT, proposed by Friedman [56], is the core idea of training a series of decision trees iteratively, where each tree attempts to correct the wrong predictions of the previous tree. The GBDT uses the gradient descent method to minimize the loss function, thereby gradually improving the model performance in continuous iterations. The workflow of GBDTs can be summarized in four steps.

Step 1 (Initializing model): The GBDT first initializes the model $F_0(x)$ by minimizing the loss function, which is shown as Equation (11). Typically, the initial model is a constant value, denoted by c, often the mean of the target variable y.

$$F_0(x) = \underset{c}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, c) \tag{11}$$

where argmin represents the parameter value for which a function achieves the minimum value in its domain. $L(y_i, c)$ is the loss function. The logarithmic loss function is usually employed in classification tasks. y_i is the target value for the ith sample.

Step 2 (Calculating residuals): In each iteration, GBDT determines the target of the next tree by calculating the negative gradient (residual) of the previous model (see Equation (12)). The residual reflects the difference between the current model prediction and the actual target value.

$$r_i^{(m)} = -\left[\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)}\right] \tag{12}$$

where $r_i^{(m)}$ denotes the residual of the ith sample in the mth iteration. $F_{m-1}(x_i)$ represents the predicted value of the model in the previous iteration on the ith sample. $L(y_i, F_{m-1}(x_i))$ represents the loss of the current model on the ith sample.

Step 3 (Fitting new decision tree): A new decision tree $T_m(x)$ is fitted based on the calculated residuals as the target, see Equation (13). The goal of the new tree is to minimize the current loss function.

$$T_m(x) = \underset{T}{\operatorname{argmin}} \sum_{i=1}^{N} \left(r_i^{(m)} - T(x_i)\right)^2 \tag{13}$$

where $T(x_i)$ represents the predicted value of tree T for sample x_i.

Step 4 (Updating model): The model is updated by adding the predictions of the new decision tree, scaled by a learning rate ρ, to the current model, see Equation (14). This step ensures that the loss function is minimized progressively with each iteration.

$$F_m(x) = F_{m-1}(x) + \rho \cdot T_m(x) \tag{14}$$

where $F_m(x)$ denotes the updated model. $F_{m-1}(x)$ represents the updated model in the previous iteration.

Repeat steps 2–4 until reaching the predetermined number of iterations (M) or the model converges. The final model is the weighted sum of all decision trees.

4.1.5. Bayesian Optimization (BO)

The core idea of BO is to iteratively update a prior probability model and narrow down the optimal value of a target function progressively. The basic procedure can be concluded as five steps.

Step 1 (Initialization): Start by selecting a prior distribution for the target function, typically using The Gaussian Process (GP) as the prior model. The GP is a non-parametric Bayesian model that defines the possible distributions of the target function across the input space. Then, randomly select a set of points to sample the target function and compute their values. These initial data serve as the foundation for constructing the preliminary surrogate model.

Step 2 (Surrogate model building): Leveraging the initial sampling data, the surrogate model undergoes Bayesian updating via the GP (see Equation (15)). The model updating involves adjusting the mean and covariance functions according to the new observations.

$$p(f(x)|D,x) = \frac{p(D|f(x),x)\cdot p(f(x))}{p(D|x)} \tag{15}$$

where x represents the sampling point. $f(x)$ represents the predictive distribution of the target function. $p(f(x)\,|\,D, x)$ represents the posterior distribution updated based on existing data D. $p(D\,|\,f(x),\ x)$ represents the likelihood of the data. $p(f(x))$ represents the prior distribution.

Step 3 (Acquisition function calculation): Select the acquisition function to decide where to sample next. Common acquisition functions include Expected Improvement (EI) and Upper Confidence Bound (UCB).

Step 4 (Sampling and update): Based on the value of the acquisition function, the next sampling point is selected to perform a new sampling of the objective function. After sampling, the surrogate model and the prior distribution are updated.

Step 5 (Iteration): Through continuous sampling and updates, the surrogate model gradually approaches the optimal value of the objective function. In each iteration, BO improves the model by gradually minimizing the loss function until the stopping condition, such as reaching the maximum number of iterations or the function converges.

Compared with random search and grid search optimization methods, BO will reduce the search space and improve search efficiency.

4.2. Model Construction

Figure 4 shows the overall framework for building the SRC model. The framework consists of three crucial modules. Module 1 involves random database division, with the ratio of training set to test set being 8:2 (196:50). Module 2 focuses on hyperparameter optimization in model training using the BO method. This study constructed four SRC models that combined the ML and BO algorithms, including SVM-BO, RF-BO, AdaBoost-BO, and GBDT-BO SRC models. Module 3 contains model evaluation, which is critical to determining the best SRC model.

In addition to excellent training samples, the best combination of hyperparameters is crucial to the performance of intelligent models. Table 2 shows the optimized hyperparameters of each SRC model using the BO method.

Table 2. List of SRC model hyperparameters.

Model	Hyperparameters
SVM-BO	kernel = 'rbf', C = 93.0236, gamma = 5.4510×10^{-3}
RF-BO	max_depth = 9, max_features = 1.0, n_estimators = 10, min_samples_split = 2, min_samples_leaf = 1
AdaBoost-BO	Estimator = DecisionTreeClassifier(max_depth = 5), learning_rate = 0.3985, n_estimators = 626, algorithm = SAMME
GBDT-BO	learning_rate = 0.1408, max_depth = 10, max_features = 0.8287, min_samples_leaf = 1, min_samples_split = 12, n_estimators = 921, subsample = 0.6274

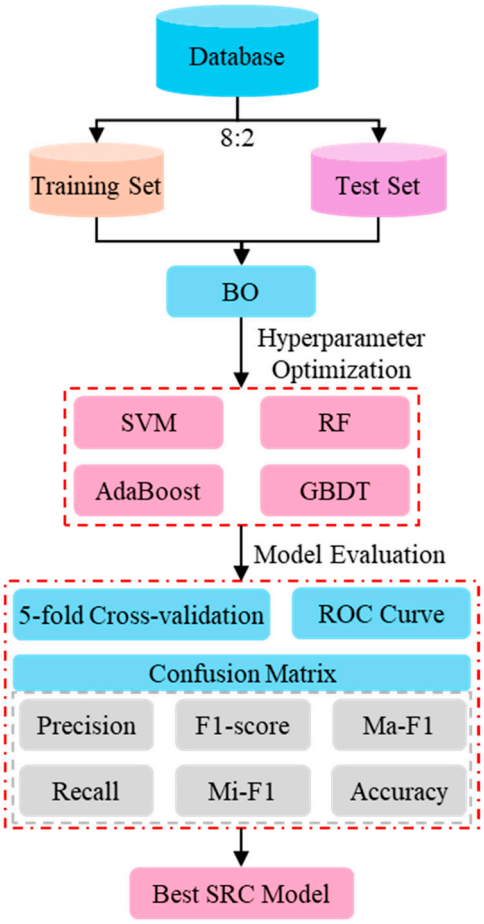

Figure 4. Flowchart of the SRC model building.

4.3. Model Evaluation

4.3.1. Evaluation Metrics

This study employed the 5-fold cross-validation, ROC curve, and confusion matrix to evaluate model performance. The 5-fold cross-validation technique can evaluate the stability and generalization ability of the model. The ROC curve reflects the classification ability of the model at different thresholds. A range of performance metrics can be derived from the confusion matrix, including precision, recall, F1-score, micro-average F1-score (Mi-F1), macro-average F1-scores (Ma-F1), and accuracy (see Equations (16)–(21)). These metrics with higher values indicate superior performance. Precision and recall measure the ability of the model to predict the majority and minority classes correctly. F1-score serves as a comprehensive indicator that considers both precision and recall to evaluate the predictive ability of the model for each surrounding rock grade. Mi-F1 reflects the overall performance of the model on the entire dataset, which is suitable for an unbalanced database. Ma-F1 reflects the balanced performance of the model in all classes, which is suitable for situations in which all classes are important. Accuracy reflects the overall prediction accuracy of the model.

In particular, SRC is a multi-classification task, and it should be first transformed into a binary classification problem in performance evaluation. Figure 5 illustrates this transformation process, using Class IV surrounding rock as an example.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

$$\text{F1 Score}_i = 2 \times \frac{\text{Precison}_i \times \text{Recall}_i}{\text{Precison}_i + \text{Recall}_i} \tag{18}$$

$$\text{Mi} - \text{F1} = 2 \times \frac{\sum_{i=1}^{n} \text{Precison}_i \times \sum_{i=1}^{n} \text{Recall}_i}{\sum_{i=1}^{n} \text{Precison}_i + \sum_{i=1}^{n} \text{Recall}_i} \tag{19}$$

$$\text{Ma} - \text{F1} = \frac{1}{k} \sum_{i=1}^{k} \text{F1 Score}_i \tag{20}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{21}$$

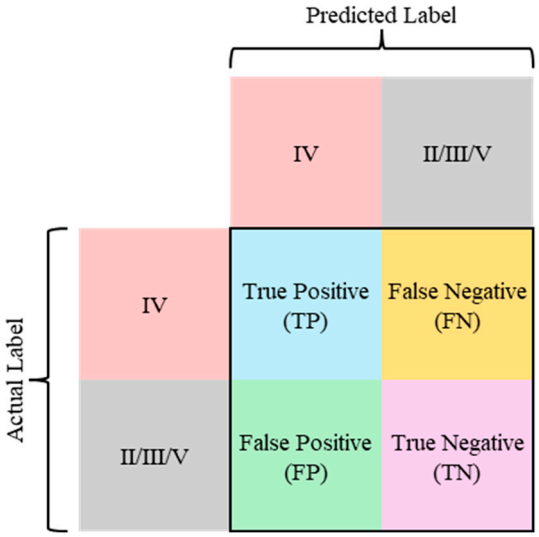

Figure 5. Binary confusion matrix (taking the prediction of class IV as an exemplar).

4.3.2. Results and Analysis

Table 3 shows the cross-validation results of the training set for SRC models. The results show that the average accuracy of the RF-BO, AdaBoost-BO, and GBDT-BO models all exceed 95%, among which the GBDT-BO model has the highest average accuracy of 97.45%. The difference between the highest and lowest accuracy in the validation set of SVM-BO, RF-BO, AdaBoost-BO, and GBDT-BO models is 10.25%, 7.69%, 7.69%, and 5.13%, respectively. The above results indicate that the GBDT-BO model performs better in the training set.

Table 3. The accuracy of the 5-fold cross-validation for each SRC model in the training stage.

Model	Accuracy					
	Validation Set 1	Validation Set 2	Validation Set 3	Validation Set 4	Validation Set 5	Average
SVM-BO	90.00%	89.74%	94.87%	87.18%	84.62%	89.28%
RF-BO	97.50%	94.87%	100.00%	100.00%	92.31%	96.94%
AdaBoost-BO	95.00%	97.44%	92.31%	94.87%	100.00%	95.92%
GBDT-BO	97.50%	97.44%	94.87%	97.44%	100.00%	97.45%

Figure 6 shows the confusion matrix of the test set predicted by each SRC model. The results show that SVM-BO, RF-BO, AdaBoost-BO, and GBDT-BO models mispredicted 4, 4, 3, and 2 samples, respectively. Based on the confusion matrix in Figure 6, other performance evaluation metrics are calculated, as shown in Table 4. The evaluation results of precision, recall, and F1-score reveal that the GBDT-BO models perform best on grades II and III surrounding rock. For grade IV, RF-BO, AdaBoost-BO, and GBDT-BO models demonstrate optimal performance. All models perform equally for grade V. The evaluation results of Mi-F1, Ma-F1, and accuracy show that the GBDT-BO model performs best on the test set, closely followed by the AdaBoost-BO model. The comprehensive analysis of the confusion

matrix and its derived evaluation metrics rank the model performance on the test set as follows: GBDT-BO model > AdaBoost-BO model > RF-BO model > SVM-BO model.

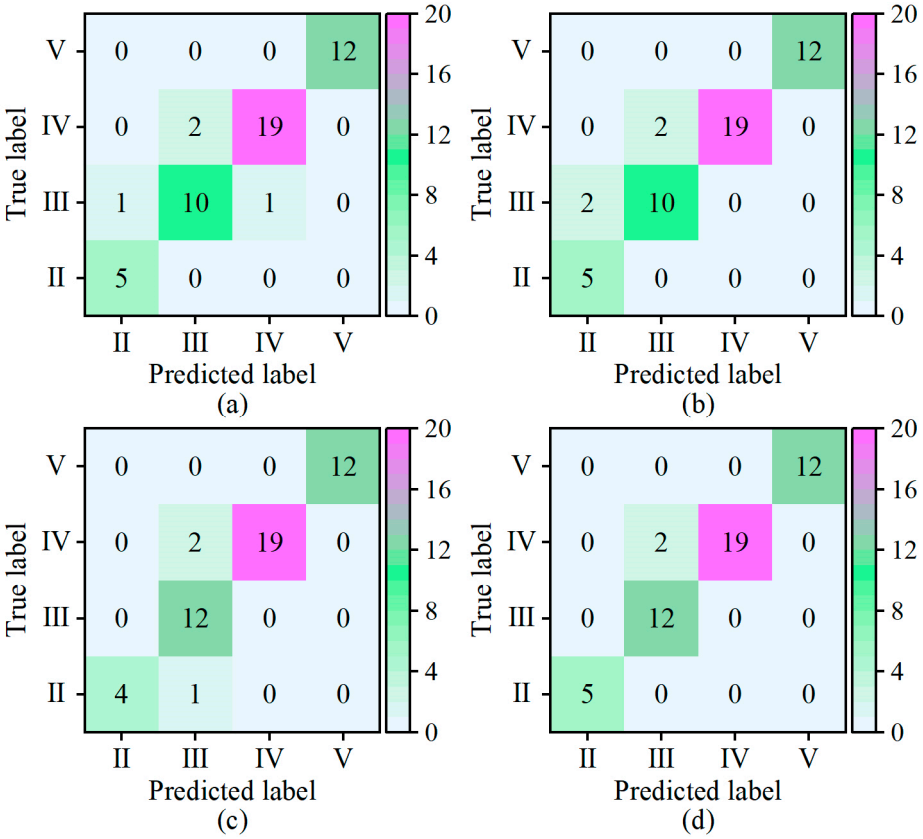

Figure 6. Confusion matrix of prediction results for each SRC model in the test stage: (**a**) SVM-BO model; (**b**) RF-BO model; (**c**) AdaBoost-BO model; and (**d**) GBDT-BO model.

Table 4. Precision, recall, F1-score, Mi-F1, Ma-F1, and accuracy values of each SRC model in the test stage.

Model	SRG	Evaluation Results					
		Precision	Recall	F1-Score	Mi-F1	Ma-F1	Accuracy
SVM-BO	II	0.8333	1.0000	0.9091	0.9200	0.9173	92.00%
	III	0.8333	0.8333	0.8333			
	IV	0.9500	0.9048	0.9268			
	V	1.0000	1.0000	1.0000			
RF-BO	II	0.7143	1.0000	0.8333	0.9200	0.9042	92.00%
	III	0.8333	0.8333	0.8333			
	IV	1.0000	0.9048	0.9500			
	V	1.0000	1.0000	1.0000			
AdaBoost-BO	II	1.0000	0.8000	0.8889	0.9400	0.9319	94.00%
	III	0.8000	1.0000	0.8889			
	IV	1.0000	0.9048	0.9500			
	V	1.0000	1.0000	1.0000			
GBDT-BO	II	1.0000	1.0000	1.0000	0.9600	0.9683	96.00%
	III	0.8571	1.0000	0.9231			
	IV	1.0000	0.9048	0.9500			
	V	1.0000	1.0000	1.0000			

Figure 7 shows the ROC curves of each model. The results show that, in the Micro-average and Macro-average ROC curves, the GBDT-BO model has the largest AUC (area

under the curve), followed by the RF-BO model. It underscores that the GBDT-BO model performs better in classification capabilities on the test set.

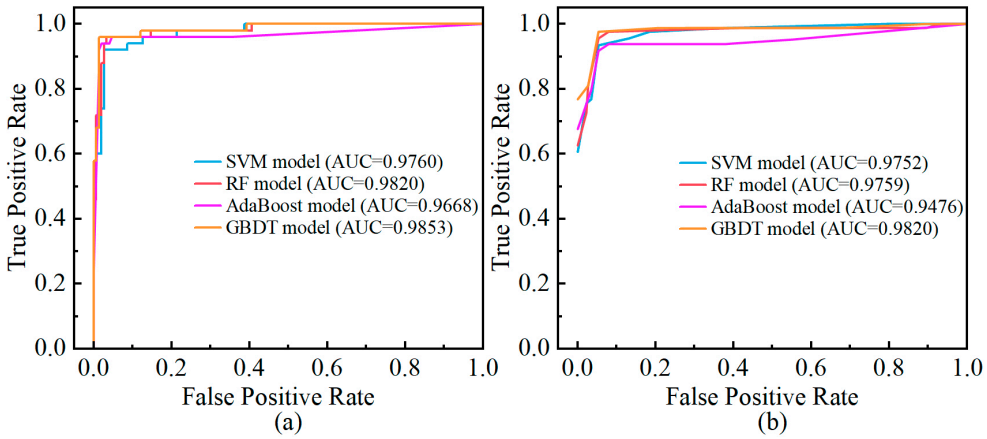

Figure 7. ROC curve and corresponding AUC value of each SRC model: (**a**) micro-average ROC curve; (**b**) macro-average ROC curve.

A comprehensive evaluation of the four SRC models on both training and test sets indicates that the GBDT-BO model outperforms others in classification accuracy, stability, and generalization. These results suggest GBDT-BO as the optimal choice for this SRC task.

5. Application and Discussion

Validation of the model's accuracy and applicability can be achieved through case studies of additional tunnels. Furthermore, the utilized database necessitates an investigation into the effects of data imbalance on the performance of intelligent SRC models.

5.1. Site Application

This study collected an additional four cases to validate the applicability of the GBDT-BO SRC model. Figure 8 shows the images of the excavation face and its corresponding discontinuity distribution condition, with manually drawn red lines indicating discontinuities. The lithologies of K36+072, K52+996, K55+315, and ZK92+850 are sandstone intercalated with muddy siltstone, muddy sandstone, basalt, and silty mudstone intercalated with sandstone, respectively.

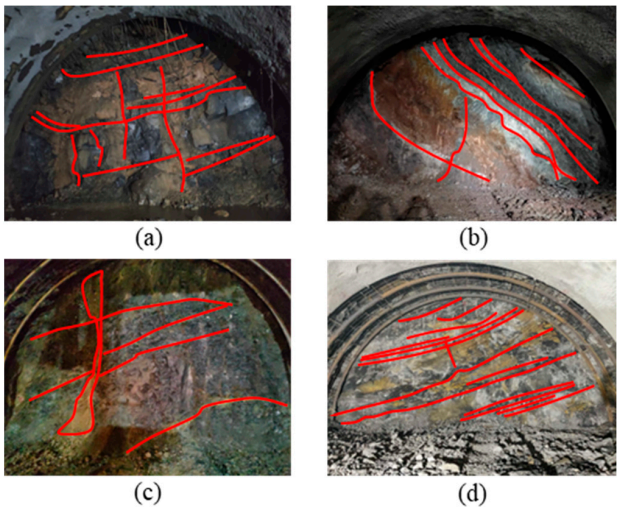

Figure 8. Images of application case: (**a**) Wuyiwan tunnel (K36+072); (**b**) Luoche tunnel (K52+996); (**c**) Yinchanggou tunnel (K55+315); and (**d**) Bimoyuan tunnel (ZK92+850).

The fractal dimensions of discontinuities in validation cases were calculated using the method described in Section 2.2. Figure 9 presents linear fitting results. The coefficient of determination (R^2) all exceed 0.99, indicating that the geometric characteristics of discontinuities are well represented by fractal dimensions. Furthermore, the results of Figure 9 demonstrate that fractal dimensions are closely related to the persistence length, quantity, spacing, and orientation of discontinuities.

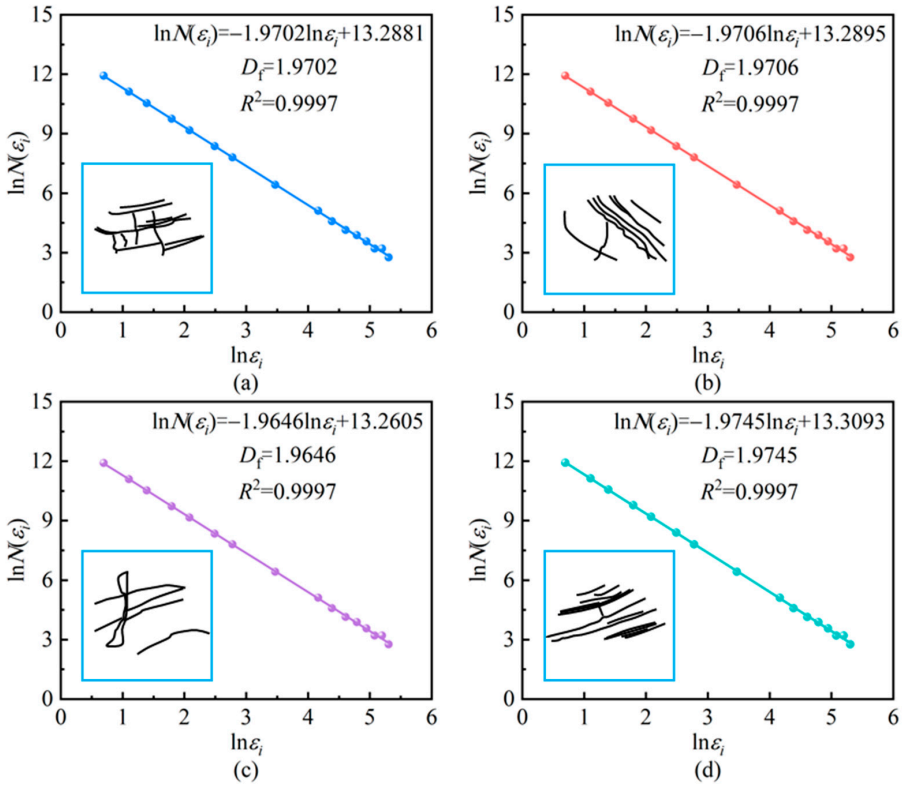

Figure 9. ln$N(\varepsilon_i)$–lnε_i curve of application cases: (**a**) K36+072; (**b**) K52+996; (**c**) K55+315; and (**d**) ZK92+850.

Table 5 presents additional surrounding rock information for each case. The actual surrounding rock grade corresponds to the support measures implemented on-site. The GBDT-BO model's predictions for the validation cases align with the actual classifications. The successful validation using four cases from different tunnels demonstrates the GBDT-BO model's adaptability and potential for application in various tunnel engineering projects.

Table 5. Input parameters and predicted results of application cases.

Case Study	Rock Strength (MPa)	Fractal Dimension	Discontinuity Condition	Groundwater Condition	In-Situ Stress Condition	Excavation Orientation	Predicted Grade	Actual Grade
K36+072	34.7	1.9702	poor	rain-like dripping	8.3	very unfavorable	IV	IV
K52+996	9.8	1.9706	poor	dripping	2.8		V	V
K55+315	73.9	1.9646	moderate	wet	10.5		III	III
ZK92+850	12.5	1.9745	moderate	dry	7.2		IV	IV

5.2. The Influence of Imbalance Data on Model Performance

Figure 3 illustrates the imbalanced nature of the sample distribution, with grade IV surrounding rock samples significantly exceeding other classes. Such imbalance typically leads to biased model predictions favoring the majority class, hindering the accurate identification of minority classes. Therefore, this study employed SMOTE to balance the training dataset and investigate the impact of sample imbalance on the SRC model performance.

SMOTE, based on K-Nearest Neighbors (KNNs) and interpolation principles, artificially generates minority class samples to achieve dataset balance. The algorithm operates in three steps. Firstly, for a randomly selected minority class sample x_i, its k nearest neighbors (x_1, x_2, …, x_k) are determined using Euclidean distance. Secondly, a neighbor x_j is randomly chosen from the k neighbors. Subsequently, a new sample x_{new} is generated using Equation (22). This procedure is carried out for each minority class until all classes achieve equal representation.

$$x_{\text{new}} = x_i + \text{rand}(0,1)\left|x_i - x_j\right| \tag{22}$$

Notably, SMOTE balancing was applied only to the training set to evaluate the generalization performance of the SRC model after SMOTE processing. The balanced training set contained 94 cases for each grade of surrounding rock. The same random_state parameter value was used to maintain identical initial training and test sets. SVM-SMOTE, RF-SMOTE, AdaBoost-SMOTE, and GBDT-SMOTE represent the SVM, RF, AdaBoost, and GBDT SRC models using the SMOTE and BO methods, respectively. Finally, the model evaluation methods described in Section 4.3.1 were employed to assess the impact of imbalanced samples on model performance.

Figure 10 illustrates the 5-fold cross-validation results of each model on the SMOTE-balanced training set. VS denotes the validation set. After SMOTE processing, the average accuracy increased by 0.87%, 2.00%, 1.16%, and 0.42% for SVM-SMOTE, RF-SMOTE, AdaBoost-SMOTE, and GBDT-SMOTE models, respectively.

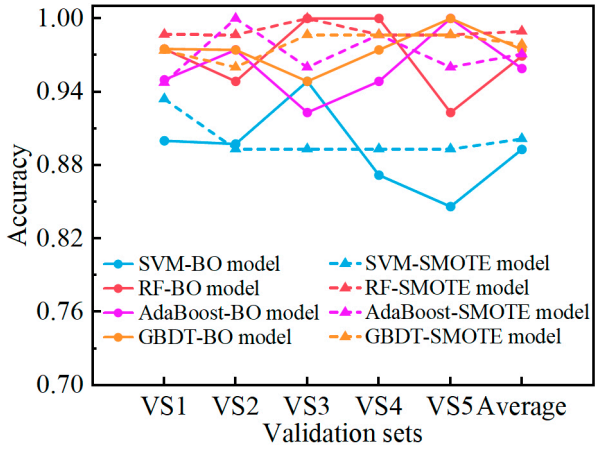

Figure 10. Comparison of 5-fold cross-validation results.

Figure 11 presents the confusion matrix of the predicted test set after SMOTE balancing. The SVM-SMOTE, RF-SMOTE, AdaBoost-SMOTE, and GBDT-SMOTE models mispredicted 4, 4, 3, and 3 samples, respectively.

Figure 12 compares the precision, recall, and F1-score of each SRC model. After SMOTE processing, the following occurred: (1) The SVM model's performance decreased for predicting grade II, improved for predicting grades III and IV, and remained unchanged for predicting V surrounding rocks. (2) The RF model's performance improved for predicting grades II and III, decreased for predicting grade IV, and remained unchanged for predicting grade V surrounding rocks. (3) The AdaBoost model's performance remained unchanged for predicting all grades. (4) The GBDT model's performance decreased for predicting grades II and III and remained unchanged for predicting grades IV and V surrounding rocks.

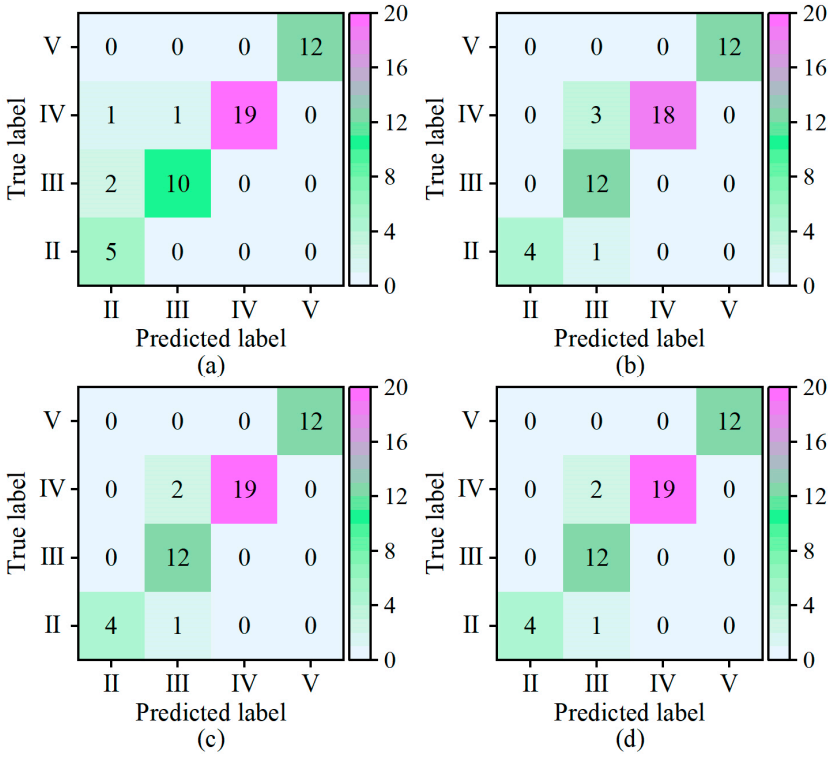

Figure 11. Confusion matrix of prediction results for each SRC model after SMOTE processing: (**a**) SVM-SMOTE model; (**b**) RF-SMOTE model; (**c**) AdaBoost-SMOTE model; and (**d**) GBDT-SMOTE model.

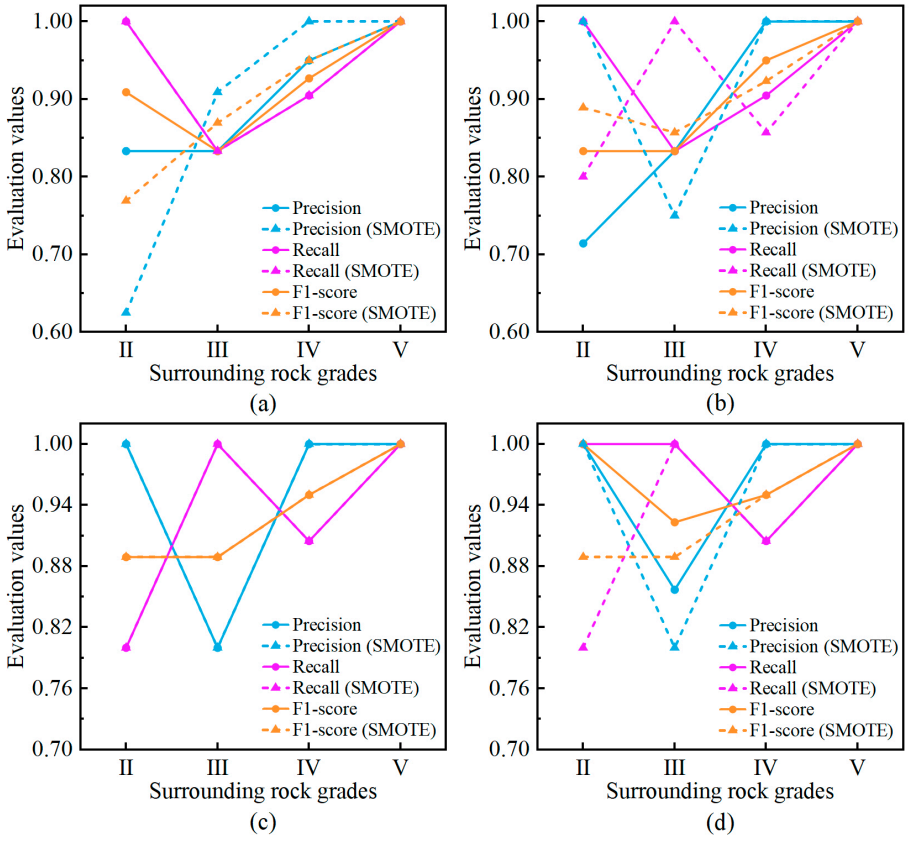

Figure 12. Comparison of precision, recall, and F1-score results: (**a**) SVM model; (**b**) RF model; (**c**) AdaBoost model; and (**d**) GBDT model.

Figure 13 compares the Mi-F1, Ma-F1, and accuracy of each SRC model. For the SVM and RF models, Mi-F1 and accuracy remained unchanged, while Ma-F1 of the SVM model decreased, and Ma-F1 of the RF model increased, indicating reduced overall performance of the SVM model and increased the overall performance of the RF model. The AdaBoost model's performance remained unchanged. For the GBDT model, Mi-F1 and accuracy decreased, while Ma-F1 increased, still resulting in a reduced overall performance.

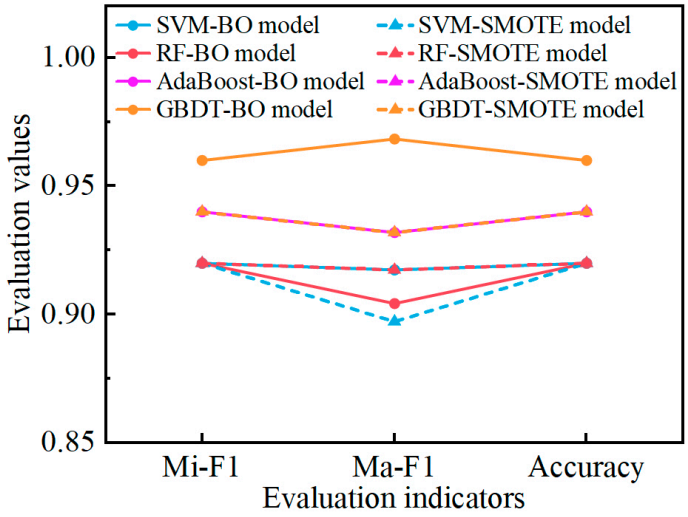

Figure 13. Comparison of Mi-F1, Ma-F1, and accuracy results.

Figure 14 compares the AUC values of the ROC curves for each SRC model. After SMOTE processing, the SVM and RF models' AUC slightly decreased, the AdaBoost model's AUC significantly increased, and the GBDT model's AUC significantly decreased.

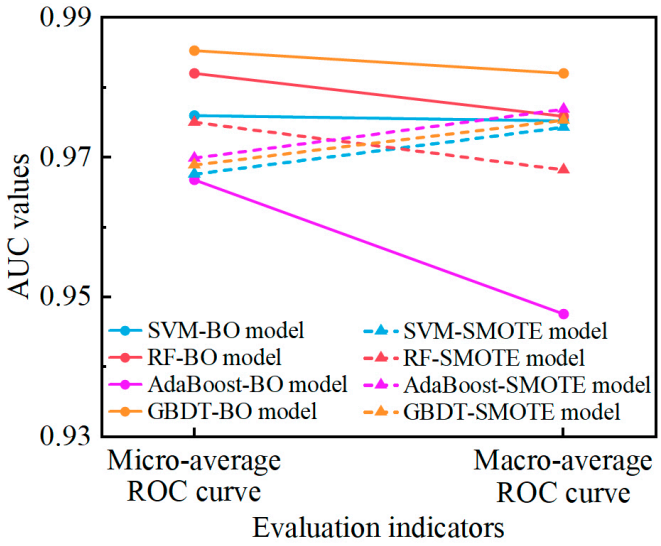

Figure 14. Comparison of AUC results.

In summary, SMOTE balancing decreased the SVM and RF models' performance slightly, left the AdaBoost model unchanged, and significantly decreased the GBDT model's performance. The results suggest that the imbalanced dataset did not adversely affect SRC models. The GBDT-BO model remains the best performer in this study.

6. Conclusions

This study utilized fractal dimensions to characterize the geometric characteristics of discontinuities, compiled a surrounding rock parameter database containing 246 cases, and developed intelligent SRC models combining ML algorithms with BO method. The main conclusions can be drawn as follows:

(1) The proposed fractal dimension calculation method can achieve an accurate characterization of the geometric characteristics of discontinuities, including the persistence length, spacing, quantity, and the orientation of discontinuities. The fractal dimension of discontinuities serves as an effective input parameter for intelligent SRC models.

(2) Four SRC models were constructed, combining SVM, RF, AdaBoost, and GBDT with BO methods. The GBDT-BO model demonstrated superior performance, achieving the average accuracy of 5-fold cross-validation, Mi-F1, Ma-F1, accuracy, and AUCs of micro-average and macro-average ROC curves were 97.45%, 0.9600, 0.9683, 96.00%, 0.9853, and 0.9820, respectively. Four additional tunnel cases validated the GBDT-BO model's practicality.

(3) SMOTE was employed to balance the training set, and the SRC models were constructed by recombining SVM, RF, AdaBoost, and GBDT with BO methods. Comprehensive evaluation using 5-fold cross-validation, the confusion matrix, and the ROC curve revealed decreased performance in SVM, RF, and GBDT models, while AdaBoost remained unchanged. It indicates that imbalanced data did not adversely affect SRC models in this study.

In conclusion, the proposed fractal dimension calculation method of discontinuities and the GBDT-BO model enable the accurate assessment of surrounding rock quality grade. Future research will focus on automating the acquisition of tunnel excavation face images and discontinuity fractal dimensions, aiming to achieve the real-time, intelligent surrounding rock classification for precise tunnel excavation and support guidance.

Author Contributions: Conceptualization, J.M. and T.L.; methodology, J.M.; software, J.W.; validation, J.W., Y.H. and C.M.; formal analysis, C.M.; investigation, F.P. and H.Z.; resources, J.M.; data curation, J.M. and T.L.; writing—original draft preparation, J.M.; writing—review and editing, T.L., R.S.F. and M.S.; visualization, J.W.; supervision, T.L., R.S.F. and M.S.; project administration, T.L.; funding acquisition, T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Natural Science Foundation of China, Grant Nos. U19A20111 and 42177173 and the Doctoral Direct Train Project of Chongqing Natural Science Foundation, Grant No. CSTB2023NSCQ-BSX0029.

Data Availability Statement: All tunneling data used in this paper are from confidential projects and can be provided on the condition of anonymity with restrictions.

Conflicts of Interest: The authors declare no conflicts of interest. The funding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Bieniawski, Z.T. Engineering classification of jointed rock masses. *Civ. Eng. S. Afr.* **1973**, *15*, 335–343. Available online: https://refhub.elsevier.com/S1674-7755(21)00081-0/sref5 (accessed on 9 June 2024).
2. Barton, N.; Lien, R.; Lunde, J. Engineering classification of rock masses for the design of tunnel support. *Rock Mech.* **1974**, *6*, 189–236. [CrossRef]
3. Hoek, E. Strength of rock and rock masses. *ISRM News J.* **1994**, *2*, 4–16.
4. Palmstrøm, A. RMi-A Rock Mass Characterization System for Rock Engineering Purposes. Ph.D. Thesis, University of Oslo, Oslo, Norway, 1995.
5. Wu, A.Q.; Zhao, W.; Zhang, Y.H.; Fu, X. A detailed study of the CHN-BQ rock mass classification method and its correlations with RMR and Q system and Hoek-Brown criterion. *Int. J. Rock Mech. Min. Sci.* **2023**, *162*, 105290. [CrossRef]
6. Deere, U.D. Technical description of rock cores for engineering purpose. *Rock Mech. Eng. Geol.* **1964**, *1*, 17–22.
7. Palmstrom, A. The volumetric joint count—A useful and simple measure of the degree of rock mass jointing. In Proceedings of the IV Congress International Association of Engineering Geology, New Delhi, India, 10–15 December 1982.

8. Palmstrom, A. Measurements of and correlations between block size and rock quality designation (RQD). *Tunn. Undergr. Space Technol.* **2005**, *20*, 362–377. [CrossRef]
9. Sonmez, H.; Ulusay, R. Modifications to the geological strength index (GSI) and their applicability to stability of slopes. *Int. J. Rock Mech. Min. Sci.* **1999**, *36*, 743–760. [CrossRef]
10. Zhou, M.L.; Chen, J.Y.; Huang, H.W.; Zhang, D.M.; Zhao, S.; Shadabfar, M. Multi-source data driven method for assessing the rock mass quality of a NATM tunnel face via hybrid ensemble learning models. *Int. J. Rock Mech. Min. Sci.* **2021**, *147*, 104914. [CrossRef]
11. Huang, H.W.; Wu, C.; Zhou, M.L.; Chen, J.Y.; Han, T.Z.; Zhang, L. Rock mass quality prediction on tunnel faces with incomplete multi-source dataset via tree-augmented naive Bayesian network. *Int. J. Min. Sci. Technol.* **2024**, *34*, 323–337. [CrossRef]
12. Xiao, C.L.; Yang, R.S.; Ding, C.X.; You, Y.Y.; Tang, W.D. Multi-scale research on blasting damage of rock based on fractal theory. *Rock Mech. Rock Eng.* **2024**, *57*, 5899–5911. [CrossRef]
13. Liu, Y.B.; Cheng, J.L.; Jiao, J.J.; Gao, Z.; Cheng, P. Influences of the hard rock proportion coefficient on the evolution pattern and fractal characteristics of mining fractures in a composite roof. *Int. J. Geomech.* **2024**, *24*, 04024038. [CrossRef]
14. Liu, H.Q.; Xie, H.P.; Wu, F.; Li, C.B.; Gao, R.B. A novel box-counting method for quantitative fractal analysis of three-dimensional pore characteristics in sandstone. *Int. J. Min. Sci. Technol.* **2024**, *34*, 479–489. [CrossRef]
15. Zhang, Q.; Shen, Y.X.; Pei, Y.C.; Wang, X.J.; Wang, M.H.; Lai, J.Q. Determination of integrity index K_v in CHN-BQ method by BP neural network based on fractal dimension *D*. *Fractal Fract.* **2023**, *7*, 546. [CrossRef]
16. Zhang, Q.; Pei, Y.C.; Shen, Y.X.; Wang, X.J.; Lai, J.Q.; Wang, M.H. A new perspective on predicting roughness of discontinuity from fractal dimension *D* of outcrops. *Fractal Fract.* **2023**, *7*, 496. [CrossRef]
17. Li, B.X.; Zhang, W.M.; Xue, Y.G.; Li, Z.Q.; Kong, F.M.; Kong, R.; Wang, G. Approach to characterize rock fracture surface: Insight from roughness and fractal dimension. *Eng. Geol.* **2023**, *325*, 107302. [CrossRef]
18. Wu, M.Y.; Wang, W.S.; Shi, D.; Song, Z.L.; Li, M.H.; Luo, Y.F. Improved box-counting methods to directly estimate the fractal dimension of a rough surface. *Measurement* **2021**, *177*, 109303. [CrossRef]
19. Li, L.C.; Wu, W.B.; El Naggar, M.H.; Mei, G.X.; Liang, R.Z. Characterization of a jointed rock mass based on fractal geometry theory. *Bull. Eng. Geol. Environ.* **2019**, *78*, 6101–6110. [CrossRef]
20. Hong, K.; Han, E.; Kang, K. Determination of geological strength index of jointed rock mass based on image processing. *J. Rock Mech. Geotech. Eng.* **2017**, *9*, 702–708. [CrossRef]
21. Feng, Z.C.; Zhao, Y.S.; Zhao, D. Investigating the scale effects in strength of fractured rock mass. *Chaos Solitons Fractals* **2009**, *41*, 2377–2386. [CrossRef]
22. Yao, W.M.; Li, C.D.; Zhan, H.B.; Zhou, J.Q.; Criss, R.E. Estimation of geological strength index through a Bayesian sequential updating approach integrating multi-source information. *Tunn. Undergr. Space Technol.* **2020**, *102*, 103426. [CrossRef]
23. Liu, Q.S.; Wang, X.Y.; Huang, X.; Yin, X. Prediction model of rock mass class using classification and regression tree integrated AdaBoost algorithm based on TBM driving data. *Tunn. Undergr. Space Technol.* **2020**, *106*, 103595. [CrossRef]
24. Hou, S.K.; Liu, Y.R.; Yang, Q. Real-time prediction of rock mass classification based on TBM operation big data and stacking technique of ensemble learning. *J. Rock Mech. Geotech. Eng.* **2022**, *14*, 123–143. [CrossRef]
25. Yin, X.; Liu, Q.S.; Huang, X.; Pan, Y.C. Perception model of surrounding rock geological conditions based on TBM operational big data and combined unsupervised-supervised learning. *Tunn. Undergr. Space Technol.* **2022**, *120*, 104285. [CrossRef]
26. Lu, H.; Kim, E.; Gutierrez, M. A probabilistic Q-system using the Markov Chain to predict rock mass quality in tunneling. *Comput. Geotech.* **2022**, *145*, 104689. [CrossRef]
27. Baghbani, A.; Choudhury, T.; Costa, S.; Reiner, J. Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. *Earth-Sci. Rev.* **2022**, *228*, 103991. [CrossRef]
28. Bo, Y.; Liu, Q.S.; Huang, X.; Pan, Y.C. Real-time hard-rock tunnel prediction model for rock mass classification using CatBoost integrated with sequential model-based optimization. *Tunn. Undergr. Space Technol.* **2022**, *124*, 104448. [CrossRef]
29. Ayawah, P.E.A.; Sebbeh-Newton, S.; Azure, J.W.A.; Kaba, A.G.A.; Anani, A.; Bansah, S.; Zabidi, H. A review and case study of artificial intelligence and machine learning methods used for ground condition prediction ahead of tunnel boring machines. *Tunn. Undergr. Space Technol.* **2022**, *125*, 104497. [CrossRef]
30. Fu, X.L.; Wu, M.Z.; Tiong, R.L.K.; Zhang, L.M. Data-driven real-time advanced geological prediction in tunnel construction using a hybrid deep learning approach. *Autom. Constr.* **2023**, *146*, 104672. [CrossRef]
31. Chen, C.; Seo, H. Prediction of rock mass class ahead of TBM excavation face by ML and DL algorithms with Bayesian TPE optimization and SHAP feature analysis. *Acta Geotech.* **2023**, *18*, 3825–3848. [CrossRef]
32. Xue, Y.D.; Luo, W.; Chen, L.; Dong, H.X.; Shu, L.S.; Zhao, L. An intelligent method for TBM surrounding rock classification based on time series segmentation of rock-machine interaction data. *Tunn. Undergr. Space Technol.* **2023**, *140*, 105317. [CrossRef]
33. Zhou, C.; Gao, Y.Y.; Chen, E.J.; Ding, L.Y.; Qin, W.B. Deep learning technologies for shield tunneling: Challenges and opportunities. *Autom. Constr.* **2023**, *154*, 104982. [CrossRef]
34. Ma, J.J.; Li, T.B.; Yang, G.; Dai, K.K.; Ma, C.C.; Tang, H.; Wang, G.W.; Wang, J.F.; Xiao, B.; Meng, L.B. A real-time intelligent classification model using machine learning for tunnel surrounding rock and its application. *Georisk* **2023**, *17*, 148–168. [CrossRef]
35. Ma, J.J.; Ma, C.C.; Li, T.B.; Yan, W.J.; Shirani Faradonbeh, R.; Long, H.T.; Dai, K.K. Real-time classification model for tunnel surrounding rocks based on high-resolution neural network and structure–optimizer hyperparameter optimization. *Comput. Geotech.* **2024**, *168*, 106155. [CrossRef]

36. Mandelbrot, B.B. *The Fractal Geometry of Nature*; Times Books: San Francisco, CA, USA, 1982.
37. Zhang, Z.; Liu, G.F.; Chang, P.; Wang, X.M.; Lin, J. Fractal characteristics for coal chemical structure: Principle, methodology and implication. *Chaos Solitons Fractals* **2023**, *173*, 113699. [CrossRef]
38. Brouty, X.; Garcin, M. Fractal properties, information theory, and market efficiency. *Chaos Solitons Fractals* **2024**, *180*, 114543. [CrossRef]
39. Sendker, F.L.; Lo, Y.K.; Heimerl, T.; Bohn, S.; Persson, L.J.; Mais, C.-N.; Sadowska, W.; Paczia, N.; Nußbaum, E.; del Carmen Sánchez Olmos, M.; et al. Emergence of fractal geometries in the evolution of a metabolic enzyme. *Nature* **2024**, *628*, 894–900. [CrossRef]
40. Zhang, Z.; Liu, G.F.; Wang, X.M.; Lv, R.S.; Liu, H.; Lin, J.; Barakos, G.; Chang, P. A fractal langmuir adsorption equation on coal: Principle, methodology and implication. *Chem. Eng. J.* **2024**, *488*, 150869. [CrossRef]
41. Ai, T.; Zhang, R.; Zhou, H.W.; Pei, J.L. Box-counting methods to directly estimate the fractal dimension of a rock surface. *Appl. Surf. Sci.* **2014**, *314*, 610–621. [CrossRef]
42. Ning, J.G.; Liu, X.S.; Tan, Y.L.; Wang, J.; Tian, C.L. Relationship of box counting of fractured rock mass with Hoek-Brown parameters using particle flow simulation. *Geomech. Eng.* **2015**, *9*, 619–629. [CrossRef]
43. Erharter, G.H. Rock mass structure characterization considering finite and folded discontinuities: A parametric study. *Rock Mech. Rock Eng.* **2024**, *57*, 5229–5249. [CrossRef]
44. Ma, J.J.; Li, T.B.; Zhang, Z.; Shirani Faradonbeh, R.; Sharifzadeh, M.; Ma, C.C. Novel multifractal-based classification model for the quality grades of surrounding rock within tunnels. *Undergr. Space* **2025**, *20*, 140–156. [CrossRef]
45. Davoodi, S.; Mehrad, M.; Wood, D.A.; Rukavishnikov, V.S.; Bajolvand, M. Predicting uniaxial compressive strength from drilling variables aided by hybrid machine learning. *Int. J. Rock Mech. Min. Sci.* **2023**, *170*, 105546. [CrossRef]
46. Kong, F.M.; Xue, Y.G.; Gong, H.M.; Jiang, X.D.; Song, Q.; Fu, Y.S.; Fu, K. The formation mechanism of dynamic water and mud inrush geohazard triggered by deep-buried tunnel crossing active fault: Insights from the geomechanical model test. *Tunn. Undergr. Space Technol.* **2023**, *142*, 105437. [CrossRef]
47. Shirani Faradonbeh, R.; Taheri, A. Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques. *Eng. Comput.* **2019**, *35*, 659–675. [CrossRef]
48. Liang, M.; Peng, H.; Xie, W.W.; Yu, B.; Han, Y.; Zhu, M.L.; Song, G.X.; Huang, N.H. Dynamic multiclass prediction of tunnel squeezing intensity with stacking model and Markov process. *Tunn. Undergr. Space Technol.* **2024**, *146*, 105632. [CrossRef]
49. Ma, K.; Peng, Y.L.; Liao, Z.Y.; Wang, Z.R. Dynamic responses and failure characteristics of the tunnel caused by rockburst: An entire process modelling from incubation to occurrence phases. *Comput. Geotech.* **2024**, *171*, 106340. [CrossRef]
50. Rahimi, B.; Sharifzadeh, M.; Feng, X.T. Ground behaviour analysis, support system design and construction strategies in deep hard rock mining—Justified in Western Australian's mines. *J. Rock Mech. Geotech. Eng.* **2020**, *12*, 1–20. [CrossRef]
51. Cervantes, J.; Garcia-Lamont, F.; Rodríguez-Mazahua, L.; Lopez, A. A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing* **2020**, *408*, 189–215. [CrossRef]
52. Patle, A.; Chouhan, D.S. SVM kernel functions for classification. In Proceedings of the 2013 International Conference on Advances in Technology and Engineering (ICATE), Mumbai, India, 23–25 January 2013.
53. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
54. Baghbani, A.; Shirani Faradonbeh, R.; Lu, Y.; Soltani, A.; Kiany, K.; Baghbani, H.; Abuel-Naga, H.; Samui, P. Enhancing earth dam slope stability prediction with integrated AI and statistical models. *Appl. Soft Comput.* **2024**, *164*, 111999. [CrossRef]
55. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-class AdaBoost. *Stat. Interface* **2009**, *2*, 349–360. [CrossRef]
56. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. Available online: https://www.jstor.org/stable/2699986 (accessed on 18 November 2024). [CrossRef]