



OPEN ACCESS

EDITED BY
Jun Yang,
Northeastern University, China

REVIEWED BY
Jixiang Liu,
Xiamen University, China
Wei Lv,
Wuhan University of Technology, China

*CORRESPONDENCE
Wanqi Wang,
176540570@qq.com

SPECIALTY SECTION
This article was submitted to
Geoscience and Society,
a section of the journal
Frontiers in Earth Science

RECEIVED 23 September 2022
ACCEPTED 31 October 2022
PUBLISHED 17 January 2023

CITATION
Song S, Xu G, Bao L, Xie Y, Lu W, Liu H
and Wang W (2023), Classifying the
surrounding rock of tunnel face using
machine learning.
Front. Earth Sci. 10:1052117.
doi: 10.3389/feart.2022.1052117

COPYRIGHT
© 2023 Song, Xu, Bao, Xie, Lu, Liu and
Wang. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Classifying the surrounding rock of tunnel face using machine learning

Shubao Song¹, Guangchun Xu², Liu Bao^{1,2}, Yalong Xie¹,
Wenlong Lu¹, Hongfeng Liu¹ and Wanqi Wang^{1*}

¹China Academy of Railway Sciences Corporation Limited, Beijing, China, ²Faculty of Geosciences and Environmental Engineering, Southwest Jiaotong University, Chengdu, China

Accurately classifying the surrounding rock of tunnel face is essential. In this paper, we propose a machine learning-based automatic classification and dynamic prediction method of the surrounding rocks of tunnel face using the data monitored by a computerized rock drilling trolley based on the intelligent mechanized construction process for drilling and blasting tunnels. This method provides auxiliary support for the intelligent decision of dynamic support at the construction site. First, this method solves the imbalance in the classification of the surrounding rock samples by constructing the Synthetic Minority Oversampling Technique (SMOTE) algorithm using 500 samples of drilling parameters covering different levels and lithologies of a tunnel. Second, it filters the importance of the characteristic samples based on the random forest method. Third, it uses the XGBoost algorithm to model the processed data and compare it with AdaBoost and BP neural network models. The results show that the XGBoost model achieves a higher accuracy of 87.5% when the sample size is small. Finally, we validate the application scenarios of the above algorithm/model regarding the key aspects of the tunnel construction process, such as surrounding rock identification, design interaction, construction supervision, and quality evaluation, which facilitates the upgrading of intelligent tunnel construction.

KEYWORDS

tunnel construction, digital twin, SMOTE, drill, machine learning

1 Introduction

The common methods of tunnel construction include drilling and blasting, shield construction, and immersed tube construction, among which over 80% of tunnel construction use the drilling and blasting method (Wang, 2010, 2020). Rock drilling rigs with hydraulic mechanical arms have been used in tunnels since the 1980s, which marks the beginning of mechanized tunnel construction. In the 21st century, as we entered the age of intelligence (Zhao et al., 2017), new opportunities and challenges for the development of technological innovation in railway tunnel construction has emerged (Yang et al., 2022), which has attracted the attention of the world's leading tunnel construction countries. In the future, the worldwide competition in railway tunnel

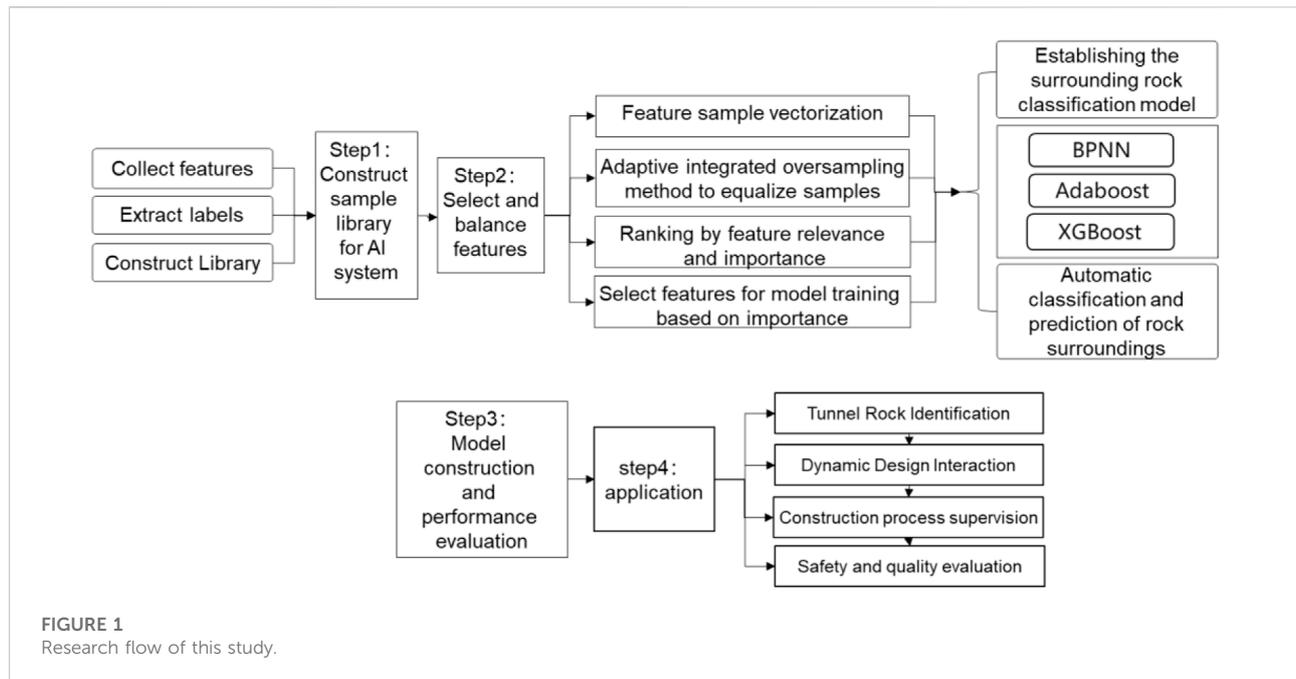
construction technology level directly depends on the breadth and depth of the application of intelligent technology in tunnel construction. Intelligent construction is an essential reflection of the level of intelligent construction technology. The number of mechanical tools used in the construction process and the depth of participation represent the technical level of railway tunnel construction. Intelligent equipment is the premise and core node of intelligent tunnel construction, and the full analysis and use of data generated by intelligent equipment is an important part of achieving the digital twin of the tunnel. More importantly, using the intelligent method to accurately identify the surrounding rock parameters of the tunnel working face and the application of the intelligent classification method of surrounding rock can provide timely geological condition feedback, which is conducive to the early detection and early warning of the adverse geology ahead.

At present, the cross-section of the railway tunnel excavation can reach 160 m², which is subject to faults, dense joints, local weathering, and stratigraphic divisions. Local optimization and adjustment of the design parameters must be made in time (Zhao, 2019). So far, we have been mainly relying on geologists to identify changes in the rock level *in situ*, which is subjective, time-consuming, and heavily affected by the technical level of the geologist (Liu et al., 2018). At this stage, intelligent rock drill trucks can carry out over-support drilling and anchor drilling and produce corresponding construction log information (Yu et al., 2018). However, although the truck has intelligent functions such as automatic positioning, automatic marking of drilling positions, and automatic data transmission, most of the information is discrete sensing data recording the operating status of the machine itself. Meanwhile, the data grows fast, with the structures varying greatly, and the representation forms are diverse. Besides, the data interaction format and data storage schemes differ greatly. Therefore, the collected data cannot be directly used for intelligent classification of the surrounding rock of tunnel face and guide the adaptive adjustment of support structure types and parameters. The dynamic control of the current state of construction at a later stage is mainly based on engineering experience and manual input of basic parameters and then matching. This will have a direct impact on the safety, speed, efficiency, and quality of the tunnel support measures and the cavern support measures in place, which in turn affects the stability of tunnel face and the quality of the tunnel construction.

Machine learning is a branch of artificial intelligence that is undergoing the most rapid development. Machine learning provides methods that can “learn” and uses sample data to make predictions or decisions and resolve complicated questions of reality without being explicitly “programmed.” Machine learning technologies are used in various applications (Yang et al., 2021a; Liu et al., 2021; Xiao et al., 2021). Recently, machine learning technologies have undergone rapid advances.

Machine learning technologies have also been applied to a few rock quality analyses. Wedge et al. (2019) used convolutional neural networks and drilling parameters collected in mineral exploration to determine the lithology of strata and stratigraphic partitioning information and compare them with manual judgment results. Yi et al. (2021) used a support vector machine (SVM) and two neural network models to describe the identification of significant heterogeneity of surrounding rock on the face of large cross-section rock tunnels. Nishitsujiy and Exley (2019) compared the performance of SVM, deep learning, linear classifier, and Bayesian classifier in the classification of lithology. They concluded that deep learning might become the main method of lithology classification in the future. Valentin et al. (2019) used ultrasonic and micro-resistivity imaging logs as input to construct a classification model for borehole image data and identified four lithologies: calcareous, gabbro, shale, and siltstone, using a deep residual network. Cai (2002) selected seven types of parameters, such as rock strength, self-weight stress, rock integrity, and mining influence, as input to the neural network to identify the stability state of the surrounding rock of the roadway project. The research of the above scholars shows that it is feasible to use machine learning theory combined with drilling parameters to identify the geological structure information such as formation lithology, rock thickness, and joint development. However, it still requires intensive investigation on how to further quantitatively identify the surrounding rocks in different areas of tunnel face in practical engineering to guide the adjustment of design parameters.

In this paper, we selected tunneling sites with complex lithology for sample collection and focused on monitoring and analyzing the drilling parameters of computerized rock drill trucks. By using the random forest algorithm without significantly reducing the accuracy of the surrounding rock classification results or affecting the classification distribution, the correlation degrees between the surrounding rock and the drilling parameters—including propulsion speed, impact pressure, propulsion pressure, and rotary pressure—were obtained to be 82%, 63%, 50%, and 40%, respectively. This results in the ranking in order of the importance of each characteristic parameter with a strong correlation, affecting the classification results of the surrounding rock. For the selected tunnels, which mainly exhibit Class III, Class IV, and some Class V rock, the Synthetic Minority Oversampling Technique (SMOTE) algorithm was applied to analyze and simulate the few samples characterizing the mechanical parameters and add new samples to the data set. This solves the problem to some extent that the classifier emphasizes the majority classes and ignores the minority classes due to the differences in tunnel construction progress, the inconsistent number of surrounding rock grades, and the imbalance of samples. Third, we established a machine learning-based classification model for the surrounding rock of tunnel face. By drawing on the idea of integrated learning, we constructed the Back-Propagation Neural



Network, AdaBoost, and XGBoost algorithms to predict the processed data, respectively. It was found that with continuous adjustment of the learning rate and other hyperparameters, the prediction accuracy of the XGBoost algorithm (ensemble-tree-based) is the highest, reaching the optimal performance of 87.5%.

Figure 1 shows the research flow of this paper, which focuses on the key issues of perception, analysis, and decision-making of the drilling data acquired by intelligent equipment in real time. The investigation is to support the key aspects of the intelligent tunnel construction process, such as surrounding rock identification, design interaction, construction supervision, and quality evaluation, thus transforming and upgrading from the traditional working mode to the intelligent mode.

The contributions of this study are as follows: (1) establishing a prediction algorithm for automated surrounding rock classification based on the operation data of mechanical construction; (2) proposing a machine learning-based method for automatic classification and dynamic prediction of tunnel working face perimeter rock and (3) supporting the transformation and upgrading of the tunnel construction process from traditional working mode to intelligent working mode through machine learning technology.

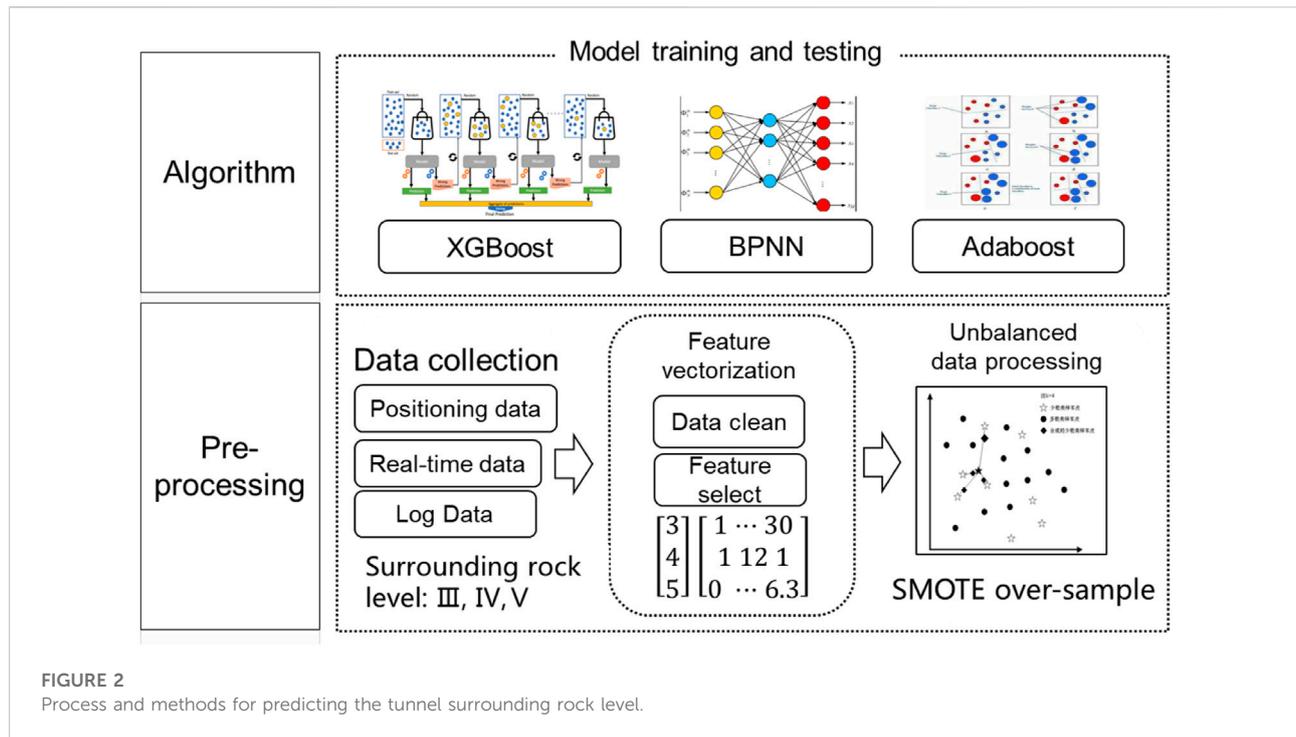
2 Materials and methods

The classification of the tunnel surrounding rock is important for identifying the nature of the surrounding rock, determining the stability of the tunnel envelope, selecting the type of tunnel support, ensuring the safety and health of

construction workers, and guiding safe construction. In general, the classification of surrounding rock is determined by a combination of two methods, namely qualitative classification and quantitative index. Qualitative classification refers to the use of an on-site geological sketch of the tunnel face to obtain a qualitative description of the rock hardness and rock integrity; while quantitative refers to the use of a rock rebound test, rock compressive strength test, rock wave velocity test, and rock body wave velocity test, and the introduction of groundwater, ground stress and the main structure of the surface production indicators, to obtain its surrounding rock classification index (Ranjbarnia et al., 2018).

In engineering practice, the determination of the surrounding rock level takes qualitative study as a primary tool and quantitative study as the secondary one. In this paper, we focus on the surrounding rock level data qualitatively obtained from the on-site tunnel face sketch results and use them as sample data for machine learning to solve classification problems. In addition, as the correctness of the surrounding rock level label will affect the accuracy of prediction, we verified the labeling of the surrounding rock level given by the field geological engineers through a small amount of rock rebound tests, rock compressive strength tests, rock wave velocity tests and rock body wave velocity tests.

Many factors influence the accurate identification of the surrounding rock of tunnel face, such as the geological analysis during the preliminary survey and design, advanced geological forecasting, the construction of the borehole camera, spectral imaging, 3D digital photography, laser scanning, and drilling measurement. They are of significance for reference to



the surrounding rock parameters evaluation. With the promotion of large supporting mechanized equipment, the real-time rock drilling parameters collected by the machine's self-awareness system can play a critical role and provide rapid response feedback (Zhao and Lu, 2018) for the determination of the surrounding rock level. Consequently, in this paper, we focus on the drilling parameters generated by a specific model of a computerized three-arm rock drilling rig during the rock drilling process in a tunnel with complex rock quality. The drilling parameters include propulsion speed, rotary pressure, propulsion pressure, rotary velocity, impact pressure, etc. A machine-learning sample library was then constructed using the above drilling parameters as well as the surrounding rock levels identified from the geological sketch of the tunnel face. The process and methodology for the prediction of tunnel surrounding rock levels are shown in Figure 2.

- 1) Data perception: The intelligent rock drilling trolley is used to collect drilling parameters such as the surrounding geological environment, operating conditions, and equipment information of the tunnel being constructed by the drill and blast method.
- 2) Data cleaning and collation: the collected raw data are cleaned and collated, and the data features are then vectorized. Finally, the data set is balanced using the SMOTE algorithm.
- 3) Model construction: The balanced dataset is used as the input of the XGBoost model for model training. XGBoost is an optimized distributed gradient boosting method that

implements machine learning under the gradient boosting framework and solves numerous data-related problems in a rapid and accurate manner. It is an improvement over the gradient boosting decision tree with higher prediction accuracy and training efficiency.

The prediction accuracy and other aspects of the model are compared with that of the BPNN (Back propagation neural network) and AdaBoost models to find the model with the best prediction effect and stability.

2.1 Data pre-processing analysis

Due to the differences in geological conditions and construction progress of the tunnels, the number of collected tunnel surrounding rock level samples was inconsistent and disproportionate. According to the data collected from the test tunnel, the samples of the surrounding rock levels were mainly divided into three types: III, IV, and V. There were 162 samples of III, 278 samples of IV, and 60 samples of V. The imbalance of the categories was high. However, it is generally considered that the ratio of data samples should be kept around 1:1 to make the classification model better reflect the classification effect (Liu et al., 2020). For this reason, in this paper, we employ the SMOTE method for processing to solve the data imbalance problem.

SMOTE algorithm was proposed by Chawla in 2002 and has been adopted by both academia and industry (Kam and Dick,

TABLE 1 Main data set for computerized rock drill rigs.

Serial number	Data category	Data Content	
1	Trolley positioning data	Current stake of the dolly	
2	Real-time status data	Name of data	Normal operating range (HC110)
		Promoting pressure	-10-230 (bar)
		Speed of propulsion	-50-50 (m/min)
		Impact pressure	-10-210 (bar)
		Slewing pressure	-10-210 (bar)
		water pressure	0-50 (bar)
		water flow	0-180 (L/min)
		Current operating status, total current, and voltage of the whole machine	
3	Recurring log data	Positioning mileage, number of holes, total drilling distance, drilling start and stop time, jamming time, flushing/other time, feed rate, impact pressure, thrust pressure, rotary pressure, water pressure, water flow	

2006). The general idea of SMOTE is to interpolate between minority class samples to generate additional samples. The method generates new synthetic samples based on the k nearest neighbor samples of the minority samples, which are random points on a line segment with endpoints corresponding to the two nearest neighbor minority class samples.

$$x_{new} = x + rand(0, 1) * ||\hat{x} - x|| \quad (1)$$

where x is a minority class sample, \hat{x} is the nearest neighbor sample, and $|| \cdot ||$ denotes the distance formula.

Because the SMOTE algorithm has the problem of a lack of diversity, many improved algorithms have been proposed, such as BorderlineSMOTE proposed by Han et al. (2005) and adaptive synthetic (ADASYN) sampling proposed by He et al. (2008). BorderlineSMOTE will only generate synthetic data for minority class samples adjacent to the boundary, which leads to weaker model generalization. On the contrary, ADASYN generates minority-class data samples adaptively based on the distribution of minority-class data samples, and minority samples that are harder to learn will generate more synthetic data than minority samples that are easier to learn. In this paper, the ADASYN algorithm is adopted for data augmentation of the imbalanced sample data.

2.2 Feature selection

The drilling parameters of the intelligent rock drill rig reflect the response of the rock drill to different surrounding rocks under the action of constant impact energy. The intelligent rock drill trolley automatically collects a series of operating process data of the rock drill in real-time and records in detail the measured values and parameters, operating status, and other information during the operation of the intelligent construction

equipment. The trolley has the characteristics of high collection frequency and a large amount of data. It possesses the most detailed process record data during the whole operation process. According to the operating characteristics of intelligent construction equipment, the data is sampled, quantified, and coded with a certain collection frequency by means of each data interface, and the data is cleaned according to some specific rules to form a data format that meets the needs of business functions such as real-time dynamic, intelligent grading of the surrounding rock.

The main data set information collected by the computerized rock drill rig is shown in Table 1.

The total energy output of the mechanical power system of the intelligent rock drill rig during the construction operation condition with normal main motor current and voltage can well reflect the quality and level of the surrounding rock (Jiang and Shen, 2018). It is generally believed that the lower the energy required to break the rock, the worse the quality of the surrounding rock and the higher the level of the surrounding rock; the higher the total energy output of the required drilling rig power system, the better the quality of the surrounding rock and the lower the level of the surrounding rock. Different mechanical arms of the multi-arm rock drill rig have different drilling parameters, and it is extremely difficult for us to consider all the influencing factors one by one in the actual engineering prognosis. On the one hand, too many parameters will bring trouble to the actual engineering site data collection. On the other hand, the excess parameters will make the construction of the neural network model complicated and prolong the training time. Meanwhile, the parameters of these influencing factors are not independent of each other, but there is a certain coupling relationship. Considering all of them may produce the problem of overfitting and be unfavorable to the prediction results.

TABLE 2 Feature name, description, and derived importance.

Feature name	Characterization	Importance
Propulsion pressure (bar)	Numerical type: feed pressure to the rock drill load-bearing mechanism to push the cylinder	0.50
Propulsion speed (m/min)	Numerical type: rock drill travel speed	0.82
Impact pressure (bar)	Numerical type: feed pressure to the rock drill impact hydraulic mechanism	0.63
Slewing pressure (bar)	Numerical type: oil feed pressure to the drilling tool rotation mechanism of the rock drill	0.40
Water pressure (bar)	Numerical type: rock driller inlet line pressure	0.20
Water flow rate (L/min)	Numerical type: rock driller inlet line flow	0.15

TABLE 3 List of hyperparameters of each classification model.

Model	List of Parameters
BPNN	batch=32, stddev=0.1(variance), lr=0.1 (learning rate), epochs=500 (number of rounds), hide_layer_num=3
XGBoost	base_score=0.5 (global bias), booster='gbtree' (base classifier model), learning_rate=0.05 (learning rate), max_depth=8 (maximum depth of tree), n_estimators=50 (number of trees)
AdaBoost	scoring='accuracy', cv=5-fold, error_score='raise'

In order to meet the research requirements and improve data quality, the raw data were cleaned with the goal of accuracy, completeness, and consistency.

Regarding duplicate values, since the ID is a unique identifier for each sample, no duplicate rejection operation is required after a duplicate lookup of the data. Regarding the missing values, for continuous variables, the missing values are filled with the mean value of the overall data on the variable; for discrete variables, the value with the highest frequency of the overall data on the variable is filled; when the number of missing variables is too large, the data is directly rejected.

The characteristic variables were initially filtered to see the distribution of values on the overall data for each variable, in turn, by the statistical function. In particular, the positioning mileage (current stake of the dolly) was removed because it does not provide useful information for the identification of the surrounding rock; the working status was removed because it only has a unique value on the whole data (normal working without warning status); and the variables such as the number of holes and the total length of the holes are not relevant to the output variable "surrounding rock level" and were removed; for The discrete variables with too many values for the four time values of drilling start/stop time, jamming time, and flushing/other time will produce a sparse matrix with too many dimensions, which will affect the effect of learning for classification, so they were also removed.

Second, the data was standardized to eliminate the effect of magnitude. Different variables often have different magnitudes and may differ in order of magnitude, and features with larger values tend to receive higher weights in the classification. In order

to avoid the bias of the classifier among different features, the data was normalized to scale the value interval of the features to a specific range so that different feature variables would have the same weight in the classification and improve the efficiency of the model.

z-score normalization, also known as standard deviation normalization, is the most common method used in data normalization. The mean of each dimensional feature after processing is 0, and the standard deviation is 1. For each value v_i of the specific dimensional feature, the standardization formula is as follows.

$$v'_i = \frac{v_i - \text{mean}}{\text{std}}$$

where *mean* is the mean of the original data in this dimension, and *std* is the standard deviation of the original data in this dimension.

After the dimensionality reduction of the data by eliminating the data of low relevance and redundant features, the size of the data set was reduced, which can effectively improve machine learning efficiency.

In addition to the above analysis, after data filtering of the original data, we then use the random forest algorithm (Yang et al., 2021b; Wei et al., 2022) to obtain the importance score of each attribute in the process of classifying and predicting the surrounding rock level, which can measure the value of the features in the model. The top six important features for the prediction model of the surrounding rock classification include propulsion speed, propulsion pressure, impact pressure, slewing pressure, water pressure, and water flow.

The list of features after completing feature selection is described in Table 2.

2.3 Training and testing of the classification model

To verify the applicability of the algorithms to different classification models and their influences on the effect of the surrounding rock level classification, we select several classification learning models commonly used in existing research for comparison experiments, including BPNN (Back

TABLE 4 Model training and testing results.

Surrounding rock level	Actual Sample Size/Sample	Number of Accurate Samples Identified by the Model/Sample			Accuracy/%		
		BPNN	XGBoost	AdaBoost	BPNN	XGBoost	AdaBoost
III	265	227	242	168	85.7%	91.3%	63.1%
IV	278	218	254	184	78.4%	91.4%	66.2%
V	257	188	213	153	73.2%	82.9%	59.5%

TABLE 5 The confusion matrix.

	Predicted condition	
	Positive (PP)	Negative (PN)
Actual condition	Positive (P)	True Positive (TP)
	Negative (N)	False Positive (FP)
		True Negative (TN)
		False Negative (FN)

propagation neural network), XGBoost, and AdaBoost algorithmic models.

To ensure the stability of the experimental results, the model training process is based on k-fold cross-validation (Zhao et al., 2020, 2021). The data set is divided into five mutually exclusive equal subsets, and five rounds of training tests are conducted, with one subset taken as the test set and the other four subsets as the training set in each round without repetition. The evaluation index results after five rounds of training tests are averaged, and the final evaluation results are the output.

3 Results

The hyperparameters of each classification model are sought by GridSearchCV. The list of hyperparameters for each model is described in Table 3.

Adopting the training set and prediction set samples after pre-processing and feature selection mentioned above, the calculated prediction set discrimination results are shown in Table 4. In the case that the training set and prediction set were the same, XGBoost had the best prediction accuracy of 87.5%. The grading accuracy of the BP neural network for surrounding rock classification of the prediction set was 79.2%. AdaBoost had a much longer training time and was sensitive to the sample. Abnormal samples in the iteration may get higher weights, thus affecting the final prediction accuracy of strong learners. AdaBoost performed the worst this time, with an accuracy of 62.9%.

No matter how efficient a model we select, its prediction results will always be subject to some errors. Therefore, we analyze and evaluate the results and performance of these

classification models on the test set based on a confusion matrix (Table 5). In this confusion matrix, four categories include TP (true positive), False Positive (FP), True Negative (TN), and False Negative (FN). TP case is a positive case that was correctly classified. FP case is a negative case that was incorrectly classified as positive. FN case is a positive case that was incorrectly classified as negative. TN case is a negative case that is correctly classified.

Precision is the proportion of true positive examples among all examples classified as positive. The closer its value is to 1, the better the classification performance for positive examples.

$$Precision = \frac{TP}{TP + FP}$$

Recall is the proportion of all true positive examples that are correctly classified as positive examples. The closer its value is to 1, the better the classification performance for positive examples.

$$Recall = \frac{TP}{TP + FN}$$

The F-measure (F) is the harmonic mean of *Precision* and *Recall*. The closer its value is to 1, the better the combined classification performance for positive samples. Its formula is shown below.

$$F = \frac{2}{Recall^{-1} + Precision^{-1}}$$

Or equivalently,

$$F = \frac{2Recall*Precision}{Recall + Precision}$$

TABLE 6 Classification model evaluation results.

Classifier	Precision	Recall	F value
BPNN	0.83	0.80	0.81
XGBoost	0.90act	0.88	0.89
AdaBoost	0.78	0.81	0.80

The evaluation results of the three classification models, namely BPNN, XGBoost, and AdaBoost, are depicted in Table 6. We find that our machine learning-based approach performs well in the studied problem.

With only the operating state of the machine itself known, the construction of the above-mentioned different classification models has well predicted the level of the surrounding rock with fast response. This also shows that the tunnel surrounding rock level is closely related to the drilling and blasting machinery construction parameters. By establishing a relationship between the drilling parameters of rock drilling machinery and the classification of the surrounding rock levels, the drilling parameters can be further optimized based on the predicted surrounding rock levels, which is of great significance for the study of intelligent visualization of tunnel construction.

4 Discussion

4.1 Discussion of the prediction results

The above predictions analyze the relationship between the change in the surrounding rock level and the mechanical drilling parameters during the construction of the rock drilling platform from different angles. Overall, the model meets the basic requirements for prediction accuracy. However, due to the lack of drilling parameter data and the inadequate classification of corresponding surrounding rock levels in practical analysis, the training data obviously cannot accurately fit the real surrounding rock level changes during the tunnel-boring process. The correlation between the construction drilling parameters and the parameters themselves has not been considered. In addition, we did not consider the differences in geological conditions and tunnel geometry. We think some improvements, such as optimizing the selection of input parameters, expanding the number of samples, diversifying the study area, and introducing more advanced and reasonable prediction methods, can be made to achieve better prediction results (Moore et al., 2022; Xu et al., 2022; Zhang et al., 2022). Furthermore, big data analysis based on heterogeneous monitoring data is suggested to help decision-making from the traditional construction method based on the

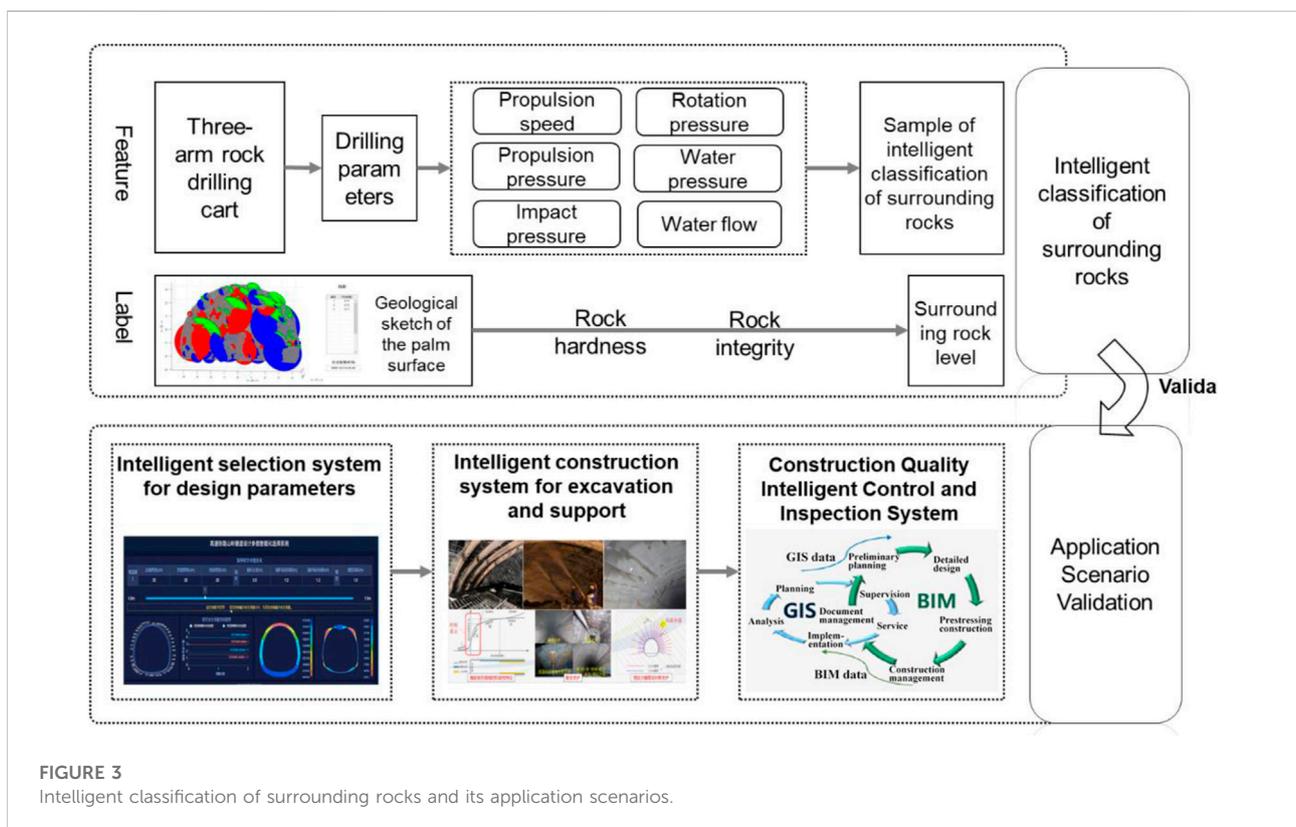


FIGURE 3 Intelligent classification of surrounding rocks and its application scenarios.

model of physical entities and the new decision model based on artificial intelligence.

4.2 Discussion of application scenarios

The research is based on machine learning and the design of algorithm models to achieve automatic collection, analysis, and classification of information from the complex geological environment of tunnels in difficult mountainous areas (Figure 3). The virtual simulation training model and the construction site can interact dynamically in the real-time field and share data with each other, leading to self-learning and self-optimization driven by the algorithm model. By using intelligent feedback analysis for forecasting and big data monitoring based on tunnel construction machinery, we can effectively identify the cases of poor stability of tunnel surrounding rock, including over-deformation, over-damage, and effectiveness for reinforcement. In addition, we will be able to assess the tunnel stability under the supporting system and evaluate the rationality of the supporting parameters so as to realize an intelligent, refined, and dynamic design of the tunnel supporting structure. The developed model will provide accurate, efficient, and comprehensive auxiliary decision-making for construction and management personnel, thus effectively strengthening the quality control and safety management of tunnel as well as improving the mechanization level of tunnel construction, accelerating the project progress and enhancing the construction efficiency under the premise of ensuring quality and safety.

5 Conclusion

In this paper, we established a prediction algorithm for automated surrounding rock classification based on the operation data of mechanical construction. Firstly, the data set was pre-processed, including missing value processing, outlier processing, data standardization processing, data normalization processing, and data sampling by the ADASYN algorithm to address the imbalance of data categories. Then, feature screening was carried out on the data set, and the importance of data features was sorted through the Random Forest algorithm so as to retain the features that have a larger impact on the prediction results and eliminate the features that have a smaller impact, hence enhance the model generalization ability and reduce the risk of overfitting. The model was divided into a training set and a test set, and the training data were fitted with different classifiers by the cross-validation method. The optimal classifier parameters were determined by the grid search method to adjust the parameters and evaluate the experimental results. Through the intelligent classification

of “fast collection-real-time transmission-remote evaluation” of surrounding rock information, the intelligent connection between intelligent perception and intelligent equipment and background server is realized. The above-mentioned new artificial intelligence decision-making model from the original signal end of the equipment to the automated classification of tunnel envelope can be applied to key links and scenarios such as surrounding rock recognition, design interaction, construction supervision, and quality evaluation in the tunnel construction process. It supports the transformation and upgrading of the tunnel construction process from the traditional working mode to the intelligent one.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

SS: formal analysis, methodology, writing—original draft. GX: formal analysis, writing—review and editing. LB: validation, writing—review and editing. YX: validation and writing—review and editing. WL: validation and writing—review and editing. HL: validation and writing—review and editing. WW: conceptualization, supervision, and writing—original draft. All authors contributed to the article and approved the submitted version.

Conflict of interest

Authors SS, LB, YX, WL, HL and WW were employed by the company China Academy of Railway Sciences Corporation Limited.

The remaining author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Cai, S. (2002). *The stability classification for the wall rock in mining roadway based on artificial neural network and the bolt supporting research*. Chongqing: Chongqing University.
- Han, H., Wang, W. Y., and Mao, B. H. (2005). "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in International Conference on Intelligent Computing (Berlin, Heidelberg: Springer), 878.
- He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," in 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence), 1322.
- Jiang, Y., and Shen, B. (2018). Strengthened mechanized construction management and construction efficiency analysis of large cross-sectional tunnel. *Tunn. Constr.* 38 (8), 84.
- Kam, J., and Dick, S. (2006). Comparing nearest-neighbour search strategies in the SMOTE algorithm. *Can. J. Electr. Comput. Eng.* 31 (4), 203–210. doi:10.1109/cejce.2006.259180
- Liu, D., Yao, M., and Zhang, X. (2018). Stability evaluation and control measures for tunnel face of large cross-sectional tunnel on Zhengzhou-Wanzhou high-speed railway. *Tunn. Constr.* 38 (08), 1311.
- Liu, J., Wang, B., and Xiao, L. (2021). Non-linear associations between built environment and active travel for working and shopping: An extreme gradient boosting approach. *J. Transp. Geogr.* 92, 103034. doi:10.1016/j.jtrangeo.2021.103034
- Liu, Q., Wang, X., Huang, X., and Yin, X. (2020). Prediction model of rock mass class using classification and regression tree integrated AdaBoost algorithm based on TBM driving data. *Tunn. Undergr. Space Technol.* 106, 103595. doi:10.1016/j.tust.2020.103595
- Moore, C., Scott, D., Burberry, L., and Close, M. (2022). Using sequential conditioning to explore uncertainties in geostatistical characterization and in groundwater transport predictions. *Front. Earth Sci. (Lausanne)*. 10, 979823. doi:10.3389/feart.2022.979823
- Nishitsuji, Y., and Exley, R. (2019). Elastic impedance based facies classification using support vector machine and deep learning. *Geophys. Prospect.* 67 (4), 1040–1054. doi:10.1111/1365-2478.12682
- Ranjbaria, M., Rahimpour, N., and Oreste, P. (2018). A simple analytical approach to simulate the arch umbrella supporting system in deep tunnels based on convergence confinement method. *Tunn. Undergr. Space Technol.* 82, 39–49. doi:10.1016/j.tust.2018.07.033
- Valentin, M. B., Bom, C. R., Coelho, J. M., Correia, M. D., Márcio, P., Marcelo, P., et al. (2019). A deep residual convolutional neural network for automatic lithological facies identification in Brazilian pre-salt oilfield wellbore image logs. *J. Petroleum Sci. Eng.* 179, 474–503. doi:10.1016/j.petrol.2019.04.030
- Wang, M. (2010). *Tunnelling and underground engineering technology in China*. Beijing: China Communications Press.
- Wang, Z. (2020). *The construction technology of large cross-sectional tunnels in Zhengzhou-Wuhan high-speed railway with safety, rapidity and standardization*. Beijing: China Communications Press.
- Wedge, D., Hartley, O., McMickan, A., Green, T., and Holden, E. J. (2019). Machine learning assisted geological interpretation of drillhole data: Examples from the Pilbara Region, Western Australia. *Ore Geol. Rev.* 114, 103118. doi:10.1016/j.oregeorev.2019.103118
- Wei, D., Yang, L., Bao, Z., Lu, Y., and Yang, H. (2022). Variations in outdoor thermal comfort in an urban park in the hot-summer and cold-winter region of China. *Sustain. Cities Soc.* 77, 103535. doi:10.1016/j.scs.2021.103535
- Xiao, L., Lo, S., Liu, J., Zhou, J., and Li, Q. (2021). Nonlinear and synergistic effects of TOD on urban vibrancy: Applying local explanations for gradient boosting decision tree. *Sustain. Cities Soc.* 72, 103063. doi:10.1016/j.scs.2021.103063
- Xu, Z., Ding, Z., Gu, G., Jiang, J., Wang, L., and Niu, X. (2022). Deep exploration of Jiaodong type gold deposit, taking Shanhou gold deposit, southern part of Zhaoping fault as an example. *Front. Earth Sci. (Lausanne)*. 10, 939375. doi:10.3389/feart.2022.939375
- Yang, L., Ao, Y., Ke, J., Lu, Y., and Liang, Y. (2021b). To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. *J. Transp. Geogr.* 94, 103099. doi:10.1016/j.jtrangeo.2021.103099
- Yang, L., Liang, Y., He, B., Lu, Y., and Gou, Z. (2022). COVID-19 effects on property markets: The pandemic decreases the implicit price of metro accessibility. *Tunn. Undergr. Space Technol.* 125, 104528. doi:10.1016/j.tust.2022.104528
- Yang, L., Liang, Y., Zhu, Q., and Chu, X. (2021a). Machine learning for inference: Using gradient boosting decision tree to assess non-linear effects of bus rapid transit on house prices. *Ann. GIS* 27 (3), 273–284. doi:10.1080/19475683.2021.1906746
- Yi, W., Wang, Mi., Tong, J., Zhao, S., Li, J., Gui, D., et al. (2021). Inhomogeneity identification method for surrounding rock of large-section rock tunnel face based on support vector machine. *China Railw. Sci.* 9 (5), 112.
- Yu, L., Wang, Z., and Yang, N. (2018). Study of measurement and distribution characteristics of surrounding rock stress of large cross-sectional high-speed railway tunnel with mechanized construction. *Tunn. Constr.* 38 (08), 1342.
- Zhang, L., Li, J., Wang, W., Li, C., Zhang, Y., Jiang, S., et al. (2022). Diagenetic facies characteristics and quantitative prediction via wireline logs based on machine learning: A case of lianggaoshan tight sandstone, fuling area, southeastern sichuan basin, southwest China. *Front. Earth Sci. (Lausanne)*. 10, 1018442. doi:10.3389/feart.2022.1018442
- Zhao, G., Yang, D., Liu, J., Jia, J., and Wu, H. (2019). High pressure behavior of crystal [2,2'-bi(1,3,4-oxadiazole)]-5,5'-dinitramide: A dft investigation. *J. Mol. Graph. Model.* 41 (1), 87–93. doi:10.1016/j.jmglm.2019.04.005
- Zhao, H., and Lu, W. (2018). Application of triple-boom rock drilling jumbo to weak surrounding rock construction of Zhengzhou-Wanzhou high-speed railway tunnel. *Tunn. Constr.* 38 (8), 1342.
- Zhao, R., Yao, M., Yang, L., Qi, H., Meng, X., and Zhou, F. (2021). Using geographically weighted regression to predict the spatial distribution of frozen ground temperature: A case in the qinghai-tibet plateau. *Environ. Res. Lett.* 16 (2), 024003. doi:10.1088/1748-9326/abd431
- Zhao, R., Zhan, L., Yao, M., and Yang, L. (2020). A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5. *Sustain. Cities Soc.* 56, 102106. doi:10.1016/j.scs.2020.102106
- Zhao, Y., Tian, S. M., and Sun, Y. (2017). Development and planning of high-speed railway tunnels in China. *Tunn. Constr.* 37 (1), 11.