

Real-time prediction of tunnel face conditions using XGBoost Random Forest algorithm

Lei-jie WU^a, Xu LI^{a*}, Ji-dong YUAN^b, Shuang-jing WANG^a

^a Key Laboratory of Urban Underground Engineering of Ministry of Education, Beijing Jiaotong University, Beijing 100044, China

^b School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

*Corresponding author. E-mail: xuli@bjtu.edu.cn

© The Author(s) 2023. This article is published with open access at link.springer.com and journal.hep.com.cn

ABSTRACT Real-time perception of rock conditions based on continuously collected data to meet the requirements of continuous Tunnel Boring Machine (TBM) construction presents a critical challenge that warrants increased attention. To achieve this goal, this paper establishes real-time prediction models for fractured and weak rock mass by comparing 6 different algorithms using real-time data collected by the TBM. The models are optimized in terms of selecting metric, selecting input features, and processing imbalanced data. The results demonstrate the following points. (1) The Youden's index and area under the ROC curve (*AUC*) are the most appropriate performance metrics, and the XGBoost Random Forest (XGBRF) algorithm exhibits superior prediction and generalization performance. (2) The duration of the TBM loading phase is short, usually within a few minutes after the disc cutter contacts the tunnel face. A model based on the features during the loading phase has a miss rate of 21.8%, indicating that it can meet the early warning needs of TBM construction well. As the TBM continues to operate, the inclusion of features calculated from subsequent data collection can continuously correct the results of the real-time prediction model, ultimately reducing the miss rate to 16.1%. (3) Resampling the imbalanced data set can effectively improve the prediction by the model, while the XGBRF algorithm has certain advantages in dealing with the imbalanced data issue. When the model gives an alarm, the TBM operator and on-site engineer can be reminded and take some necessary measures for avoiding potential tunnel collapse. The real-time prediction model can be a useful tool to increase the safety of TBM excavation.

KEYWORDS Tunnel Boring Machine, fractured and weak rock mass, machine learning model, real-time early warning, tunnel face rock condition

1 Introduction

The construction of long tunnels is often a key control project in large-scale infrastructure construction such as water conservancy, railways, and highways. For example, the Ya'an to Linzhi section of the Sichuan-Tibet Railway has a total length of 1011 km, of which the tunnel length is 851 km, including 72 tunnels. Compared with traditional drill and blast methods, the use of Tunnel Boring Machine (TBM) construction has many advantages such as safety, efficiency, environmental protection, automation, and high degree of informatization. Although the TBM construction method has many advantages, it can

also encounter problems when encountering adverse geological conditions, such as machine jamming and rock collapse. Therefore, perception technology of the tunnel face rock condition in the TBM construction process presents a bottleneck and an urgent problem that needs to be solved in order to ensure safe and efficient excavation of TBM in complex geological conditions [1–3].

Almost all tunnel engineering projects encounter weak and fragmented rock masses and fault zones. During route planning, designers will try to avoid obvious and large-scale adverse geology based on preliminary geological survey data. However, due to the limitations of project budgets and timelines, preliminary geological survey data are often at the macroscopic level, and therefore the situation on the tunnel face is unknown until

TBM construction. In addition, the TBM cutterhead and shield are huge, making it difficult for construction personnel to directly observe the specific situation on the tunnel face. In this situation, traditional detection, and *in situ* testing techniques [4–8] do not meet the requirements for TBM's continuous and efficient excavation. Therefore, real-time surrounding-rock perception technology has become a key technology for TBM construction.

TBM excavation is a process of interaction between the TBM cutterhead and the rock. Therefore, TBM excavation can be seen as a large-scale torsion-shear test, which means that a series of parameters are recorded by sensors as the TBM excavates under different rock conditions [9]. These parameters include TBM operating parameters, response parameters, and on-site conditions recorded by geologists, which largely provide rock mass information. Use of real-time data during the TBM excavation process to perceive the current conditions of the tunnel face offers significant advantages.

Many models for predicting TBM performance have been proposed [10–12]. Many early models were theoretical models based on laboratory indentation tests and linear cutting tests [13–17]. For example, these studies have revealed the mechanism of rock fracture in disc cutters. In addition, many scholars have applied these theoretical models to specific projects and have developed TBM excavation models that are suitable for current projects based on rock characteristics, including rock strength, rock integrity, and rock brittleness [18–23]. These models can be called semi-empirical and semi-theoretical models. However, the application of these models is limited due to the difficulty in obtaining a large amount of experimental data on rock characteristics and the complex and varied geological conditions of different tunnels or even of the same tunnel.

In addition to the models mentioned earlier, there are also artificial intelligence algorithms that can analyze the large amount of data collected by TBM sensors, in order to construct models for predicting the surrounding rock conditions. Compared to traditional statistical regression methods, artificial intelligence methods have obvious advantages in solving complex nonlinear problems [24–28]. The widespread availability of commercial codes for various machine learning algorithms have provided opportunity for researchers to approach this new area. Zhang et al. [29] proposed a rock mass type classifier with a prediction accuracy of 98% using the support vector classifier (SVC) algorithm and preprocessed TBM operation data. Zhu et al. [30] used a new performance evaluation index to evaluate the training effect of three types of classification algorithms on the imbalance data set of rock mass classification, including binary classifier, multi-class classifier and error-classification cost-sensitive classifier. Meanwhile, Hou et al. [31,32] proposed two models, one using random

forest algorithms and the other using long short-term memory network algorithms, with good performance for rock classification problem and collapse problem based on the TBM data set collected during the YinSong project.

Despite the good performance of these studies, there is an important issue that needs to be addressed. How can real-time and rapid predictions be made based on the data continuously collected by TBMs to meet the requirements of continuous TBM construction?

To achieve real-time perception of the rock condition in the tunnel face, this paper establishes real-time prediction models for fractured and weak rock mass (FWM) by comparing 6 different algorithms using real-time data collected by the TBM. The models are optimized from the perspectives of metric selection, input feature selection, and imbalanced data processing.

The flowchart and structure of this study is shown in Fig. 1, and is as follows. Section 2 presents a brief project overview and describes the data collection, including the adverse geological information, raw TBM data processing, label definition for classification, and feature extraction. Section 3 details the model building process, including selection of appropriate input features, optimal hyperparameter combinations, and evaluation metrics selection. Finally, Section 4 discusses the comparison results and prediction effects of the models.

2 Project overview and data collection

With the development of sensor technology, various types of sensors have been installed on TBM, and the data generated during the rock breaking process can be stored in real time. Supported by the Chinese 973 program, the high-quality TBM construction data from the YinSong water diversion project has provided the possibility for development of TBM rock mass real-time perception technology. This section mainly introduces the basic geological and data information of the project.

2.1 Geological conditions and Tunnel Boring Machine data profile

This study is based on the TBM construction data from the TBM3 Lot of the YinSong project, a large-scale water diversion project aimed at solving the urban water supply problem in the central region of Jilin Province, China. The TBM3 Lot tunnel is 23 km long, with a diameter of 7.9 m and an overburden depth of 40–250 m. The open TBM used in the TBM3 Lot was manufactured by the China Railway Engineering Equipment Group Co., Ltd. (CREG), with a maximum cutterhead diameter of 7.93 m, 56 disc-cutters, and a maximum thrust cylinder stroke of 1.8 m.

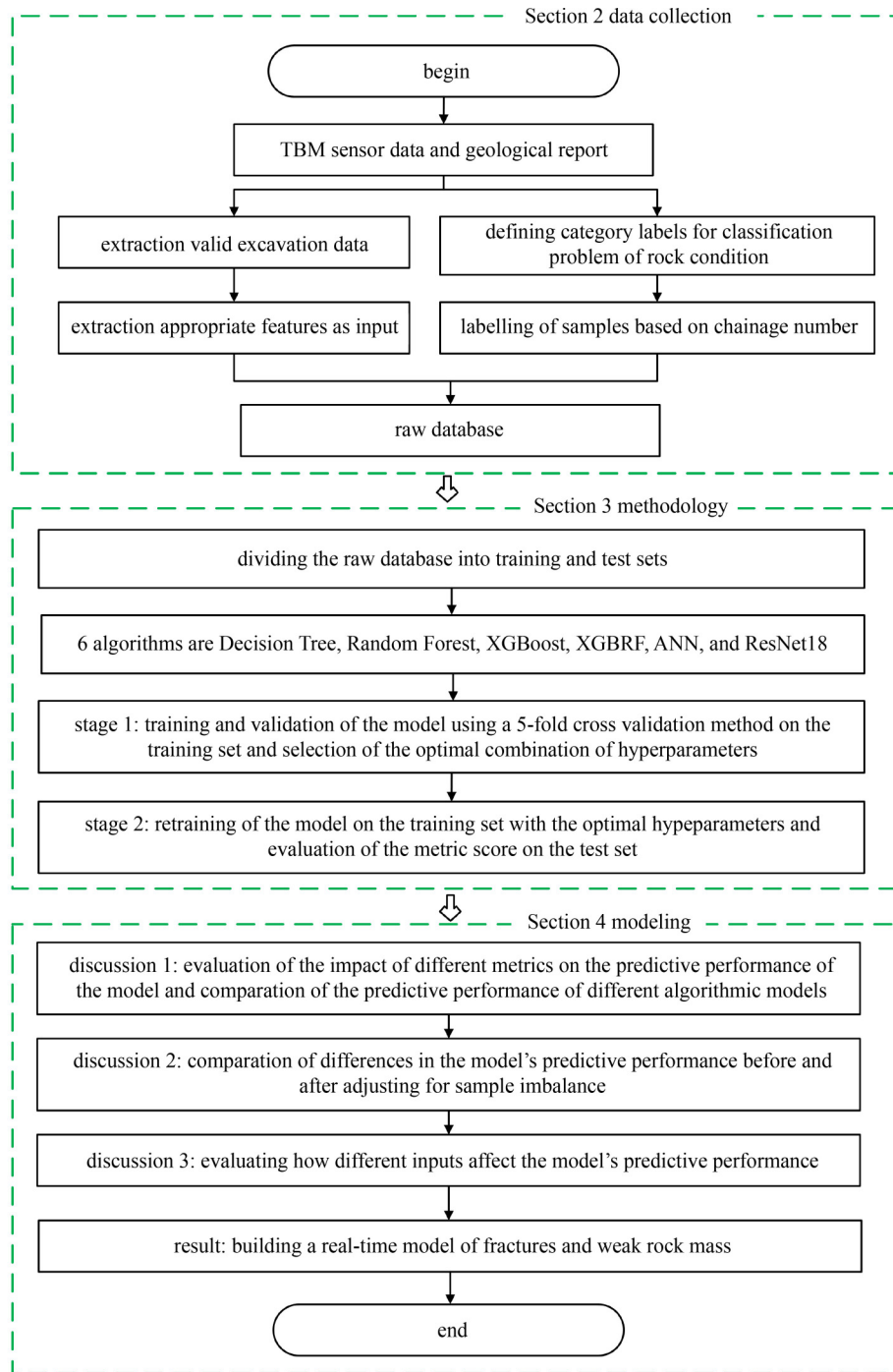


Fig. 1 Flowchart and structure for this study.

The data used in this study was collected from the area of chainage numbers 71476-51475 along the TBM tunneling direction. Figure 2 illustrates the geological situation of some areas. The adverse geology encountered during TBM tunneling mainly includes fault zones, caverns, weak surrounding rock, and water inrush. These geological factors can cause various difficulties during TBM construction. Neglect of timely support measures can lead to collapse and machine jamming accidents.

Detailed records of these unfavorable geological

information, reinforcement support locations, and multiple collapse locations are documented in the construction logs. In addition to this geological information, data collected by the TBM sensors are saved in csv format files daily. The csv files contain information and characteristics of the TBM recorded at 1 s interval. In total, 802 d of valid excavation data were collected during the tunneling process of the TBM3 Lot.

The selection of TBM control parameters is typically based on the rock mass conditions at the tunnel face and

is made by the TBM operator. This selection process includes determination of the cutterhead advance speed v (mm/min) and cutterhead rotation speed n (r/min). Once the control parameters are set, the tunneling response parameters, such as cutterhead torque T (kN·m) and total thrust force F (kN), typically vary with the rock conditions. These four parameters are basic in the rock breaking process of TBM construction and depend on the rock conditions.

Figure 3(a) shows the basic data recorded by TBM in one day, consisting of 10 working steps. Since the thrust cylinder of the TBM has a limited stroke and timely support requirement, the actual TBM tunneling process is performed step-by-step. In this study, a tunneling cycle is defined as a working step in the TBM tunneling process, and the data recorded in a tunneling cycle covers the entire rock breaking process of the TBM from start-up to

shut-down. A total of 12570 tunneling cycles were extracted from the raw data for the TBM3 Lot for this research.

As shown in Fig. 3(b), based on the loading process of the electric motor, the rock fragmentation data in a tunneling cycle can be divided into four phases: free running, loading, stable boring, and ending the boring. During the loading phase, the TBM cutterhead advances and contacts the tunnel face. Penetration is defined as the cutting depth of the disc cutters for each turn of the cutterhead, as shown in Eq. (1):

$$p = \frac{v}{n}, \quad (1)$$

where n is usually set at a fixed value, while v is gradually increased to the operator's intended value, and p gradually increases and is then maintained at a stable

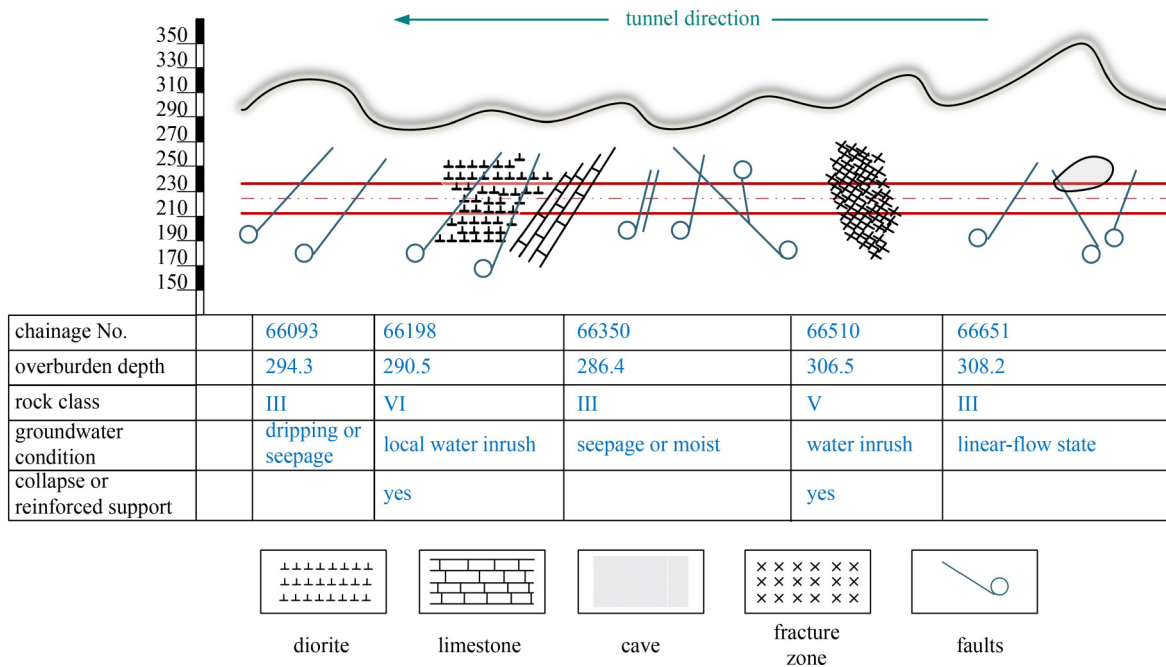


Fig. 2 Part of the geological profile records in the YinSong project.

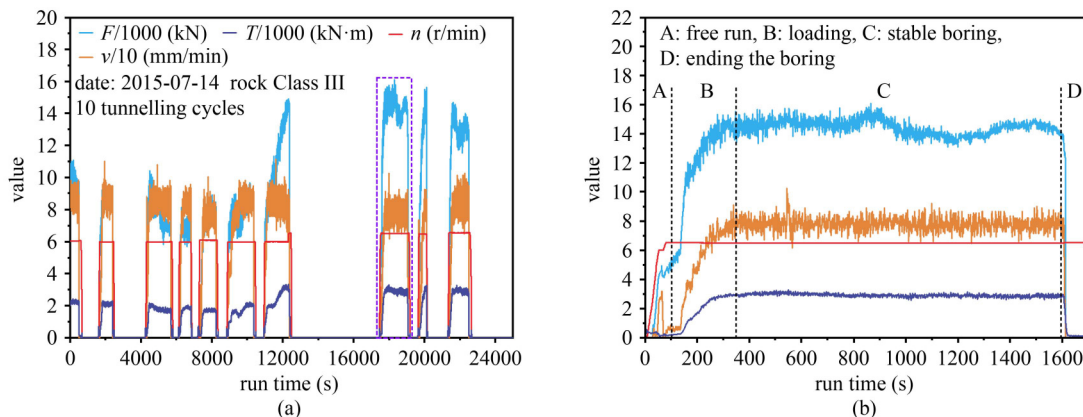


Fig. 3 TBM basic data: (a) working steps of one day; (b) a tunneling cycle.

value during the loading phase.

Therefore, the loading phase can be viewed as a continuous torsional shear test of TBM equipment at various penetration levels, providing rich geological information. In addition, the loading phase typically lasts 200 s, making it ideal for real-time rock condition perception.

2.2 Definition labels

As shown in Fig. 2, the geological profile record is a very detailed construction log that records information such as rock mass classification, whether collapse occurred, whether support was strengthened, and groundwater conditions during the construction process, based on the chainage number.

In this study, each sample was labeled based on the two conditions of “whether collapse occurred” and “whether support was strengthened” as provided by the full geological profile records. For our convenience, the rock mass that may cause collapse if not supported in time is called FWM. To solve the problems of collapse caused by FWM, this study labeled the original data samples as FWM or non-FWM based on the geological profile records.

In the following sections, serious supervised learning algorithms will be employed to predict FWM. This is a binary classification problem, so we denote the labels of the two categories, non-FWM and FWM, as 0 and 1, respectively, as shown in Eq. (2):

$$Label = \begin{cases} 1, & \text{FWM,} \\ 0, & \text{non-FWM.} \end{cases} \quad (2)$$

2.3 Feature extraction

Appropriate feature selection often directly determines the performance of a model. The YinSong project includes 199 sensor parameters, many of which are constant, low variance, and highly correlated data [33]. Obviously, data with such characteristics should not be selected as input features. In many studies, only four basic parameters are used as input features for the model [29,34,35], while some studies have used some physically meaningful rock breaking indicators such as torque penetration index (*TPI*) and field penetration index (*FPI*) [19,36,37].

To further validate the impact of feature selection on the model performance, Table 1 summarizes the commonly used parameters as candidates, based on previous studies. The specific details of the feature selection scheme are elaborated in Subsection 3.1.

The basic fracture parameters (*T*, *F*, *n*, *v*) are accessible from raw data. Since the data in the stable boring phase fluctuate less, the mean value ($Mean(X)$) and the

coefficient of variation ($C.V(X)$) in the stable boring phase are selected as input features. The calculation methods are shown in Eqs. (3)–(5). In addition, the key rock fracture indices can be calculated by Eqs. (6)–(10), incorporating *TPI*, *FPI*, work ratio (*WR*), the parameters in an *F*–*p* relationship proposed by Jing et al. [38], and the parameters in a *T*–*p*–*F* relationship. Here, Eqs. (9) and (10) are used to fit the loading phase data to calculate *AF*, *BF*, $R^2(A)$, I_c , I_f , and $R^2(B)$. The $Mean(X)$ and $C.V(X)$ of other indices are calculated separately in the loading or the stable boring phases (as shown in Table 1).

Statistical analyses were performed on these 26 features of two class samples and the results are shown in Table 1. Among them, X_1 – X_8 belong to the basic rock fracture parameters and are input features for data-driven models. X_9 – X_{26} belong to the key rock fracture indicators and are input features for knowledge-driven models. It is important to note the timeliness of parameter acquisition. In TBM construction, the duration of the loading phase is short and X_{15} – X_{26} can usually be obtained within a few minutes after the disc cutter contacts the rock mass, whereas the parameters X_1 – X_{15} calculated over the stable boring phase can usually not be obtained until the excavation is complete.

$$Mean(X) = \frac{1}{m} \sum X_i, \quad (3)$$

$$Std(X) = \sqrt{\frac{\sum_{i=1}^m (X_i - Mean(X))^2}{m}}, \quad (4)$$

$$C.V(X) = \frac{Std(X_i)}{Mean(X_i)}, \quad (5)$$

$$TPI = \frac{T}{p}, \quad (6)$$

$$FPI = \frac{F}{p}, \quad (7)$$

$$WR = 2\pi \times 10^3 \frac{T \cdot n}{F \cdot v}, \quad (8)$$

$$F = AF \cdot p + BF, \quad (9)$$

$$T = I_c \cdot p + I_f \cdot F, \quad (10)$$

where *X* represents a parameter, such as *T*, *F*, *n*, *v*, *TPI*, *FPI*, and *WR*. *m* is the total amount of data in the loading or stable boring phases. The fitting coefficients for the loading phase data using Eqs. (9) and (10) are denoted as *AF*, *BF*, I_c , and I_f , which represent the weights of each

Table 1 Statistical description of features

input variables	calculation method	feature No.	feature	FWM		non-FWM		
				mean	std.	mean	std.	
basic rock fracture parameters	calculate from stable boring phase data	X_1	Mean (T)	1482.2	757.8	2338.1	689.1	
		X_2	C.V (T)	0.052	0.032	0.036	0.021	
		X_3	Mean (F)	8601.0	2986.3	11318.8	3343.4	
		X_4	C.V (F)	0.04	0.02	0.044	0.021	
		X_5	Mean (v)	63.7	12.1	65.2	11.5	
		X_6	C.V (v)	0.075	0.019	0.086	0.019	
		X_7	Mean (n)	5.3	0.9	6.3	0.9	
		X_8	C.V (n)	0.010	0.021	0.008	0.017	
key rock fracture indices	calculate from stable boring phase data	X_9	Mean ($TPIs$)	130.3	78.7	236.1	89.4	
		X_{10}	C.V ($TPIs$)	0.092	0.026	0.093	0.021	
		X_{11}	Mean ($FPIs$)	760.5	374.1	1267.1	525.2	
		X_{12}	C.V ($FPIs$)	0.092	0.024	0.104	0.024	
		X_{13}	Mean (WRs)	90.0	32.9	117.8	23.6	
		X_{14}	C.V (WRs)	0.094	0.025	0.096	0.020	
		calculate from loading phase data	X_{15}	Mean ($TPIu$)	116.6	65.8	207.6	78.8
			X_{16}	C.V ($TPIu$)	0.234	0.153	0.229	0.097
	X_{17}		Mean ($FPIu$)	1033.3	533.5	1697.6	714.2	
	X_{18}		C.V ($FPIu$)	0.324	0.139	0.325	0.120	
	X_{19}		Mean (WRu)	90.5	34.1	121.5	27.2	
	X_{20}		C.V (WRu)	0.250	0.174	0.224	0.108	
	X_{21}		AF	223.2	218.7	510.3	317.0	
	X_{22}		BF	5999.1	1260.5	7021.0	1838.8	
	X_{23}		$R^2(A)$	0.52	0.29	0.61	0.24	
	X_{24}		I_c	88.9	88.1	171.0	102.4	
	X_{25}	I_f	0.038	0.076	0.025	0.055		
	X_{26}	$R^2(B)$	0.69	0.25	0.78	0.16		

Notes: Mean (X) and C.V (X) represent the mean and coefficient of variable X , respectively, in each boring phase; $TPIu$, $FPIu$, and WRu are the TPI , FPI , and WR in the loading phase, respectively; $TPIs$, $FPIs$, and WRs are the TPI , FPI , and WR in the stable boring phase, respectively.

variable in the regression equations. $R^2(A)$ and $R^2(B)$ measure the degree to which Eqs. (9) and (10) fit the loading phase data.

data set using the optimal hyperparameters combination and evaluating the final classifier's performance on the testing data set.

3 Methodology

In Section 2, we prepared the raw data set, this section presents the main training process of classifier for fractured and weak surrounding rock with the flowchart and structure being depicted in Fig. 4. The primary process comprises three stages: (a) selecting appropriate input features from the raw data set and dividing them into training and test data sets; (b) determining the optimal combination of hyperparameters by training and validating the model on the training data set using 5-fold cross-validation; (c) retraining the model on the training

3.1 Test scheme for input feature selection

From the perspective of data-driven and knowledge-driven approaches, the 26 features can be divided as follows. (1) The basic rock fracture which is original data collected by the TBM, X_1 to X_8 ; (2) the key rock fracture indices which is feature cross of rock fracture parameters based on human understanding of the rock fracture mechanism, X_9 to X_{26} . To compare the performance of the data-driven and knowledge-driven approaches, the corresponding features were selected as inputs, denoted as scheme 1, as shown in Table 2.

In TBM construction, the duration of the loading phase

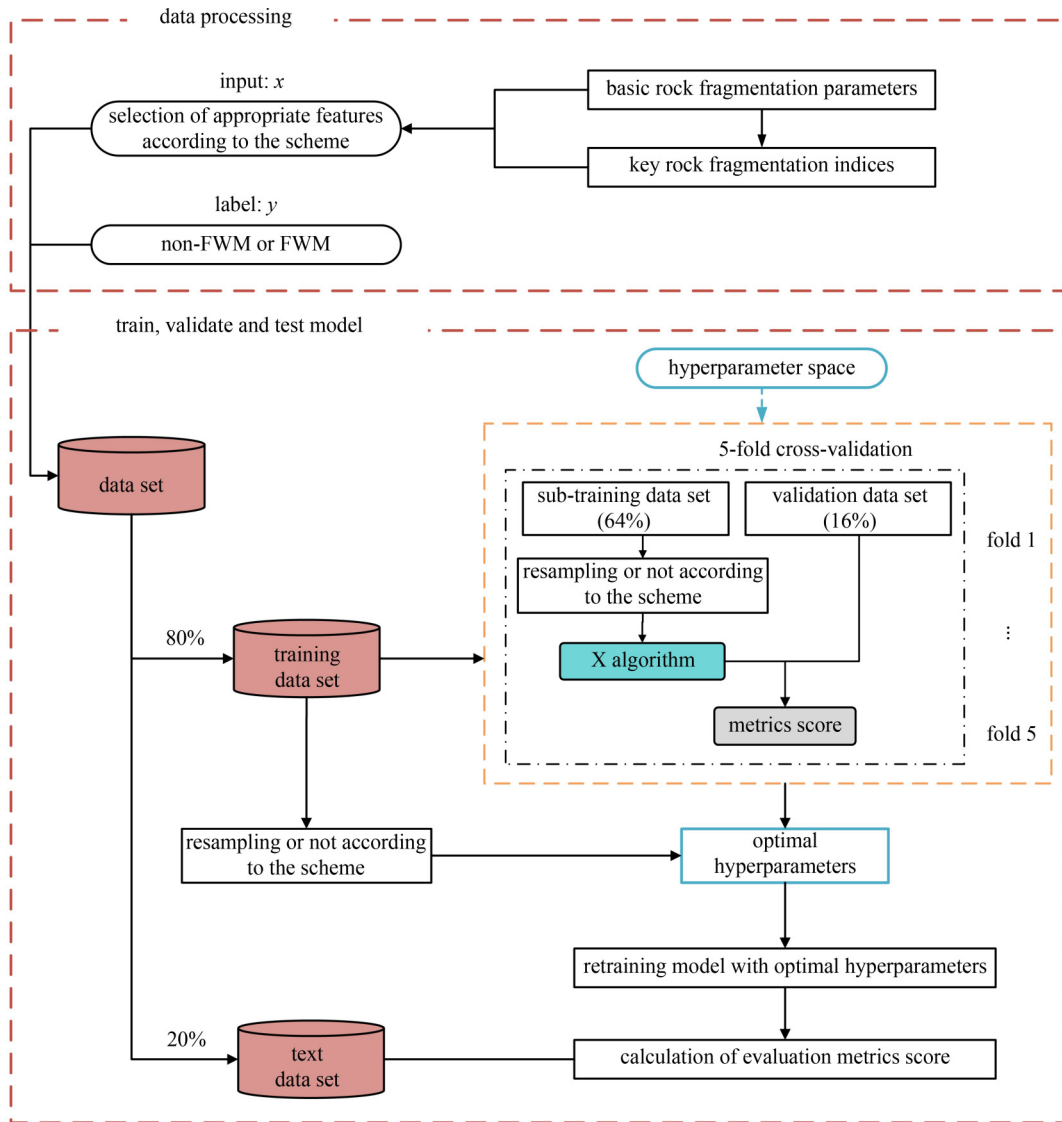


Fig. 4 Flowchart of building the FWM classifier.

is short and X_{15} to X_{26} can usually be obtained within a few minutes after the disc cutter contacts the rock mass. To establish a real-time perception model, feature selection is performed according to scheme 2 in Table 2.

The new data set is formed according to the schemes in Table 2, and split it into 80% training data set and 20% test data set. The results of this part will be shown in Subsection 4.3.

3.2 Imbalanced data processing method

For classification tasks, the balance of different categories of sample sizes is an issue that needs to be considered. According to the construction records of the YinSong project, there were 848 tunneling cycles labeled as FWM and 11720 tunneling cycles labeled as non-FWM, with a ratio of 1:13.8 between the two sample categories. When training a model on an unprocessed data set, there is a risk of high overall accuracy but poor prediction

Table 2 Test scheme for feature selection

No.	model	input features
scheme 1	data-driven model	X_1-X_8
	knowledge-driven model	X_9-X_{26}
scheme 2	real-time perception model	$X_{15}-X_{26}$
	dual-driven model	X_1-X_{26}

performance on FWM samples, as discussed in Subsection 4.2. This is because the model has learned prior information that “the number of non-FWM samples is much larger than the number of FWM samples”. Relying on this information, the model can achieve a seemingly good performance by classifying as many samples as possible as non-FWM, which deviates from the safety requirements of construction. In TBM construction, collapse can cause significant economic loss and personnel injury, so we expect the prediction model to pay more attention to FWM samples.

To address this issue, this paper adopts a comprehensive resampling technique to process the training data set samples. On the one hand, this technique uses the Synthetic Minority Over-sampling Technique (SMOTE) method to expand the FWM samples. The basic idea of the SMOTE algorithm is to generate a new sample by considering its k nearest neighbors. On the other hand, an under-sampling method called Edited Nearest Neighbors (ENN) will remove some non-FWM samples that do not meet the requirements of the method.

One detail to note is that this paper only processes the training set data, which can avoid another issue that needs to be considered in machine learning: “data leakage”. This is because the final trained model needs to be validated using a data set that better reflects the real situation.

3.3 Model training and validation

In this stage, different algorithm models are trained and validated. When evaluating different settings (“hyperparameters”) for algorithm, there is a risk of overfitting on the test data set because the parameters can be tweaked until the estimator performs optimally. This way, information about the test data set can “leak” into the model and the evaluation metrics become no longer credible. To find the optimal hyperparameters, 5-fold cross-validation is applied to the training data set. This method splits the training data set into sub-training data set (64% of the raw data set) and validation data set (16% of the raw data set), then trains the model on the sub-training data set, and finally evaluates it on the validation data set.

3.3.1 Algorithm introduction

In this paper, 6 algorithms are used to predict the rock condition in tunnel face, including Decision Tree (DT), Random Forest (RF), Extreme Gradient Boosting Decision Trees (XGBoost), and XGBoost Random Forest (XGBRF) algorithm which combines RF and XGBoost. Due to the popularity of the neural network algorithms, we also compared the results of ANN and ResNet18. The main purpose of this paper is to apply these algorithms to predict the rock mass. DT, RF, XGBoost, and ANN algorithms are not introduced further in this paper, and the principles can be referred from Refs. [39,40].

DT and RF are implemented by the scikit learn package for Python. XGBoost and XGBRF are implemented by the XGBoost package for Python. ANN and ResNet18 are implemented by the Pytorch package with Python. Moreover, the training and testing of all classifiers were processed by a computer with a CPU of Intel I CoITM i7-10875H @ 2.30 GHz in a Windows environment.

(1) XGBoost Random Forest

The XGBRF algorithm combines the benefits of both RF and gradient boosting. As is the case for RFs, XGBoost RF builds multiple DTs on random subsets of the training data and features, which helps to reduce overfitting and increase accuracy.

But unlike the traditional RFs, XGBRF builds DTs sequentially, as in a gradient boosting process. At each iteration, the algorithm fits a DT to the residual errors of the previous iteration and adds it to the ensemble of trees. This process continues until a specified number of trees or a stopping criterion is reached. Figure 5 shows the structure of the XGBRF algorithm.

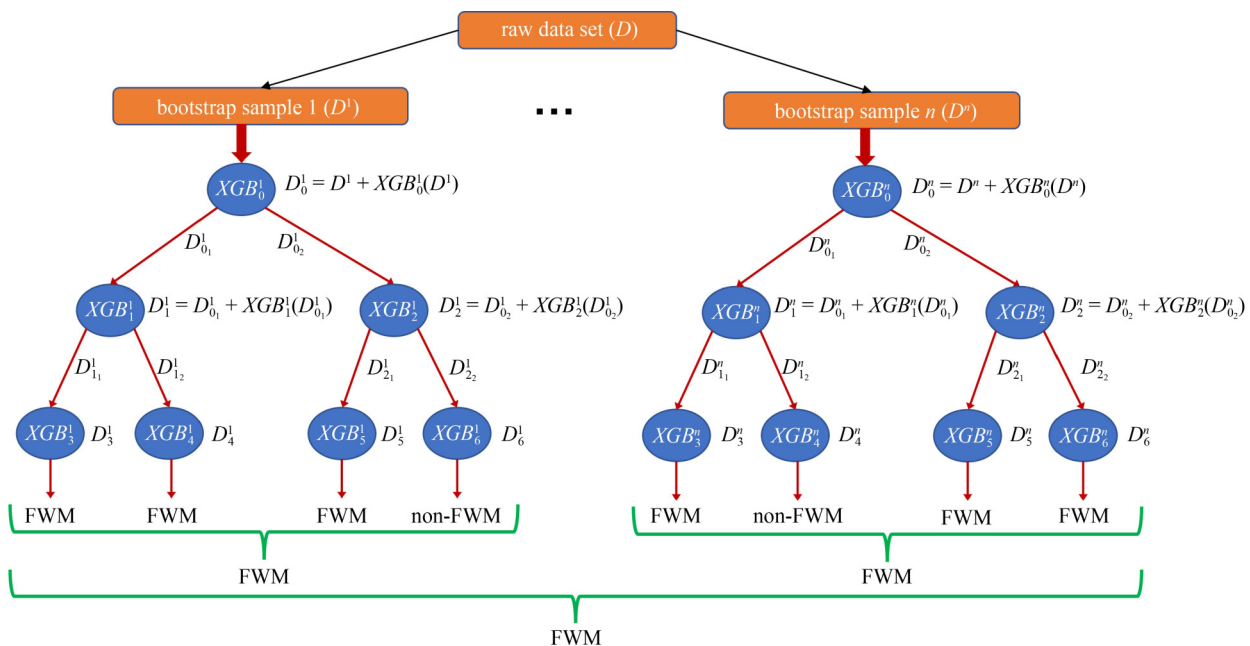


Fig. 5 Diagram of XGBRF model structure.

XGBRF also uses a gradient boosting-like loss function to optimize the objective during the training process. This loss function takes into account the prediction errors of the previous iterations to improve the accuracy of the model.

(2) Modified ResNet18

ResNet18 is an optimized convolutional neural network (CNN) architecture that was introduced by He et al. [41]. The name ResNet stands for “Residual Network” which refers to the network’s use of residual connections to address the problem of vanishing gradients in deep neural networks.

The ResNet18 architecture consists of 18 layers, including a convolutional layer, a max-pooling layer, and several residual blocks. Each residual block contains two convolutional layers, and the input to each block is passed through a skip connection that allows the network to learn residual representations. The skip connection also allows gradients to flow more easily through the network, which can help prevent vanishing gradients.

In our study, we employed ResNet18 as the backbone architecture for our specific task. However, we made notable modifications to the original architecture to accommodate the unique characteristics of our dataset. We replaced the standard two-dimensional convolutional layers with one-dimensional convolutional layers to handle our input, which is represented as a two-dimensional array. Additionally, we incorporated batch normalization and ReLU activation functions within each residual block to enhance the model’s learning capabilities and enable better convergence.

The utilization of ResNet18 in our research highlights its effectiveness in extracting meaningful features and

capturing intricate patterns within our specific domain. By leveraging the power of residual connections and our tailored modifications, we aimed to achieve superior performance and accurate predictions for our target task.

For a visual representation of the modified ResNet18 architecture, please refer to Fig. 6, which provides a comprehensive overview of the network’s structure and the flow of information within its layers.

3.3.2 Hyperparameters space for various algorithm models

In the field of machine learning, hyperparameters are parameters that are preset before model training and cannot be learned from the data. Their purpose is to govern various aspects of the training process and can significantly impact the performance of the models.

This study aimed to compare the predictive performance of various algorithms in FWM prediction. Although hyperparameter tuning was not a primary focus of the study, Table 3 presents some of the hyperparameters for the six models considered, while other hyperparameters were kept at their default values.

To search for the best hyperparameter, the Optuna library was utilized. This library offers efficient hyperparameter optimization by performing multi-dimensional searches in the hyperparameter space, incorporating advanced techniques such as pruning and early stopping to ensure computational efficiency [42].

3.3.3 Evaluation metrics

After training a model on a sub-training data set, it is

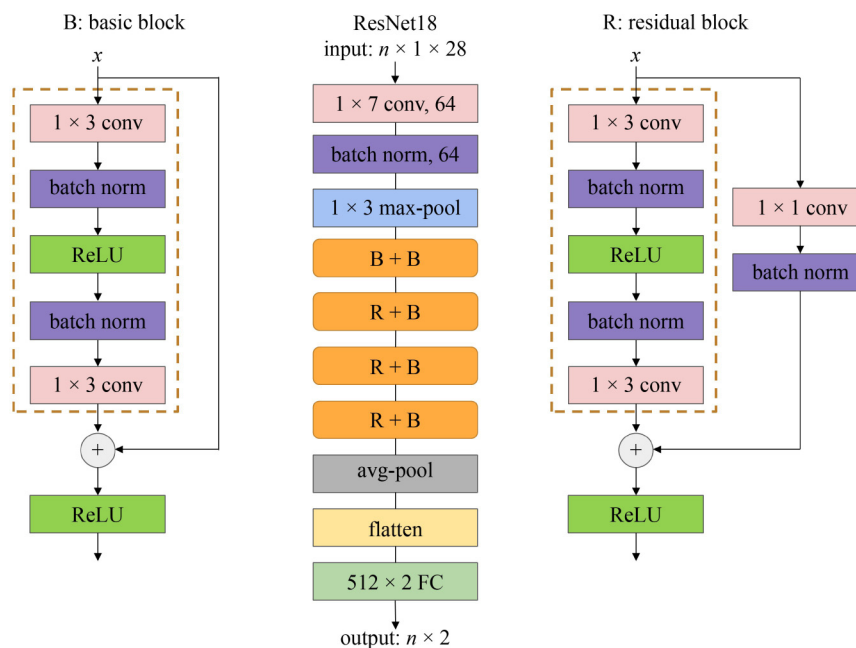


Fig. 6 Diagram of ResNet18 model structure.

Table 3 Hyperparameter space for various algorithm models

algorithm	parameters	value range	algorithm	parameters	value range
XGBoost	n_estimators	[10, 400], step = 2	XGBRF	n_estimators	[10, 400], step = 2
	learning_rate	$[10^{-3}, 10^{-1}]$, step = 10^{-1}		learning_rate	$[10^{-3}, 10^{-1}]$, step = 10^{-1}
	max_depth	[5, 100], step = 1		max_depth	[5, 100], step = 1
	gamma	[0, 1], step = 0.1		gamma	[0, 1], step = 0.1
	reg_alpha	[0, 1], step = 0.01		reg_alpha	[0, 1], step = 0.01
	reg_lambda	[0, 1], step = 0.01		reg_lambda	[0, 1], step = 0.01
DT	max_depth	[5, 100], step = 1	RF	n_estimators	[10, 400], step = 2
	max_features	['sqrt', 'log2', 'auto']		max_depth	[5, 100], step = 1
	class_weight	[None, 'balanced']		max_features	['sqrt', 'log2', 'auto']
ANN	epochs	50	ResNet18	class_weight	[None, 'balanced']
	lr	$[10^{-5}, 10^{-1}]$, log-uniform sampling		epochs	50
	batch_size	[16, 32, 64]		lr	$[10^{-5}, 10^{-1}]$, log-uniform sampling
	optimizer	['Adam', 'SGD']		batch_size	[16, 32, 64]
	momentum	[0, 1], step = 0.1		optimizer	['Adam', 'SGD']
	activation function	ReLU		momentum	[0, 1], step = 0.1
	loss function	cross entropy loss		activation function	ReLU
	number of hidden layers	[1, 5], step = 1		loss function	cross entropy loss
	number of nodes per hidden layer	[32, 64, 128]			

Notes: The parameter names correspond to the libraries of machine learning algorithms as introduced at the start of Subsubsection 3.3.1; the libraries can be found on the official websites of each algorithm.

common to validate the trained model using a separate validation data set. An accurate evaluation metric is crucial for determining optimal hyperparameters. Thus, this paper compares the impact of various metrics on the final model's performance, and the specific results can be found in Subsection 4.1.

For binary classification, the classifier's prediction results consist of four possible scenarios, true positive (TP), false negative (FN), false positive (FP), and true negative (TN). Table 4 presents some commonly used evaluation metrics, among which true positive rate (TPR), also known as recall, signifies the proportion of actual positive samples correctly predicted as positive. False positive rate (FPR), also referred to as the false alarm rate, indicates the proportion of actual negative samples incorrectly predicted as positive. False negative rate (FNR), alternatively termed miss rate, denotes the proportion of actual positive samples erroneously predicted as negative. The prediction result of the classifier for a certain sample is in the form of the probabilities of two categories, 'non-FWM' and 'FWM'. For instance, if a classifier predicts [0.6, 0.4] for a sample, the probability of it being FWM is 0.4. Typically, if the classification threshold is 0.5, the sample is classified as 'non-FWM'. However, if the threshold is set to 0.3, the sample is classified as 'FWM'.

Different classification thresholds correspond to different predictions made by the classifier. A Receiver

Operating Characteristic (ROC) curve is generated by plotting the TP rate and the FP rate at various threshold settings. The Youden's index is a measure for determining the optimal classification threshold value, which is obtained when the index reaches its maximum value, i.e., the optimal threshold value. Moreover, the two more effective metrics for imbalanced data are Matthew's Correlation Coefficient (*MCC*) and Area Under the ROC Curve (*AUC*). In this paper, *MCC* is used with optimal threshold. Table 4 presents these metrics along with their calculation equations.

3.4 Retraining and evaluation of model

At this stage, to assess the classifier's ability to generalize, the model is retrained using the training data set and the optimal hyperparameters, and subsequently evaluated using the test data set. The results presented in the next section are based on the classifier's performance on the test data set.

In addition to the three metrics mentioned in Subsubsection 3.3.3, two metrics that more directly reflect construction safety are of particular interest to construction workers: the false alarm rate and the miss rate. According to Eq. (10), the Youden's index is negatively correlated with the sum of the FP and FN rates; i.e., the larger the Youden's index, the better the predictive performance of a model.

4 Real-time perception modeling of rock condition

In the preceding sections, we have described the necessary preparations for developing a rock mass condition perception model, including data set preparation, algorithm selection, and evaluation metric establishment. In this section, we first demonstrate the impact of different metrics on hyperparameter selection results and compare the performance of various algorithmic models. Subsequently, we compare the differences in the model’s predictive performance before and after addressing the problem of sample imbalance. Finally, we investigate the effects of different inputs on the model’s predictive performance and develop a real-time perception model for FWM.

4.1 Prediction performance of various algorithms

During the 5-fold cross-validation process, the selection of hyperparameters was performed simultaneously with the evaluation of model’s performance. Initially, we selected a specific evaluation metric, denoted as Y to assess the model’s predictive ability on the validation data set. The evaluation metric Y included Youden’s index (J), AUC , and MCC . Through the comparison of the average Y scores obtained from the validation results of the 5 folds, we were able to identify the optimal hyperparameter combination, denoted as HPC (Y), for a particular algorithm within its corresponding hyperparameter space (refer to Table 3 for details). Subsequently, we retrained the algorithm model using the HPC (Y) and obtained the evaluation scores on the test data set, which are presented in Table 5.

4.1.1 Influence of different metrics on hyperparameter selection results

To explore the influence of three metrics on the selection of hyperparameters and corresponding model performance, Figs. 7(a)–7(f) displays the final evaluation scores of

classifiers trained with three different hyperparameter combinations across six algorithms. As explained in Subsubsection 3.3.3, the formula derivation indicates that the Youden’s index holds the highest priority as a metric. Consequently, we highlight the hyperparameter combination with the highest J score in the subgraphs. From Table 5 and Fig. 7, we deduce the following observations.

(1) The trend lines for AUC scores and J scores exhibit similar patterns, implying that, in this study, AUC offers equally good evaluation effect.

(2) The trend line for the MCC score does not align with that of the J score, as indicated in Figs. 7(b), 7(c), and 7(e). For instance, in Fig. 7(c), the XGBoost classifier with HPC (J) exhibits the highest J score, the lowest MCC score, and 17.2% miss rate. In contrast, the XGBoost classifier with HPC (MCC) has the lowest J score, highest MCC score, and 25.3% miss rate. However, if the total error rate is similar, the classifier with HPC (J) is more appropriate due to potential safety hazards associated with omitted fractured rock mass. Therefore, the Youden’s index offers a more comprehensive evaluation of the classifier’s performance.

(3) If the classifier with HPC (X) also performs well on the test data set, it indicates good generalization performance. The hyperparameter combinations corresponding to classifiers with the largest J score have been highlighted in the figures. For example, in the case of the XGBRF algorithm, the classifier with HPC (AUC) has the highest J and AUC scores, indicating that the XGBRF algorithm has good generalization ability.

In summary, the Youden’s index and AUC are the most appropriate performance metrics for this data set, and the XGBRF algorithm exhibits superior generalization performance.

4.1.2 Comparison of prediction performance of different algorithm models

The aim of this study is to compare and evaluate the performance of various algorithmic models regarding the perception of fractured and weak rock. In Subsubsection

Table 4 Equations for evaluation metrics

metrics	definition	Eq. No.
TPR	$TPR = \frac{TP}{TP+FN}$	(6)
FPR	$FPR = \frac{FP}{FP+TN}$	(7)
FNR	$FNR = \frac{FN}{TP+FN} = 1 - Recall$	(8)
MCC	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$	(9)
AUC	area under the ROC curve	
Youden’s index	$J = TPR - FPR = 1 - FNR - FPR$	(10)

Notes: TPR = recall; FPR = false alarm rate; FNR = miss rate

4.1.1, we drew conclusions and selected the optimal HPC for each algorithm to enhance their performance. Here, we present the prediction results of six classifiers on the test data set samples in the form of a confusion matrix, in Fig. 8. The test data set consisted of 2340 non-fractured

and weak rock (non-FWM) samples and 174 fractured and weak rock (FWM) samples, with a ratio of 13.4:1. By examining the results of the six classifiers, we infer the following. The ResNet18, DT, and ANN algorithm classifiers exhibit good recognition results for non-FWM

Table 5 Performance of the classifiers trained under different HPCs

algorithm	hyperparameter combination	evaluation metric score					
		<i>AUC</i>	optimal threshold	<i>J</i>	<i>MCC</i>	<i>FPR</i>	<i>FNR</i>
DT	HPC (<i>J</i>)	0.791	0.454	0.476	0.278	21.9	30.5
	HPC (<i>AUC</i>)	0.814	0.547	0.505	0.327	15.6	33.9
	HPC (<i>MCC</i>)	0.830	0.730	0.506	0.331	15.0	34.5
RF	HPC (<i>J</i>)	0.874	0.316	0.599	0.342	22.9	17.2
	HPC (<i>AUC</i>)	0.873	0.282	0.596	0.328	26.6	13.8
	HPC (<i>MCC</i>)	0.875	0.290	0.607	0.337	25.6	13.8
XGBoost	HPC (<i>J</i>)	0.873	0.080	0.590	0.334	23.8	17.2
	HPC (<i>AUC</i>)	0.868	0.156	0.582	0.352	18.8	23.0
	HPC (<i>MCC</i>)	0.866	0.329	0.572	0.353	17.5	25.3
XGBRF	HPC (<i>J</i>)	0.865	0.484	0.584	0.311	30.7	10.9
	HPC (<i>AUC</i>)	0.867	0.493	0.601	0.340	23.8	16.1
	HPC (<i>MCC</i>)	0.866	0.489	0.590	0.330	24.9	16.1
ANN	HPC (<i>J</i>)	0.812	0.608	0.504	0.333	14.5	35.1
	HPC (<i>AUC</i>)	0.816	0.705	0.521	0.313	19.7	28.2
	HPC (<i>MCC</i>)	0.809	0.700	0.504	0.287	23.8	25.9
ResNet18	HPC (<i>J</i>)	0.826	0.715	0.542	0.359	14.2	31.6
	HPC (<i>AUC</i>)	0.826	0.636	0.541	0.331	18.3	27.6
	HPC (<i>MCC</i>)	0.832	0.471	0.532	0.297	25.5	21.3

Note: The term ‘‘HPC (*Y*)’’ refers to the hyperparameter combination selected based on the highest *Y* metric score on the validation data set, where *Y* can represent *J*, *AUC*, or *MCC*.

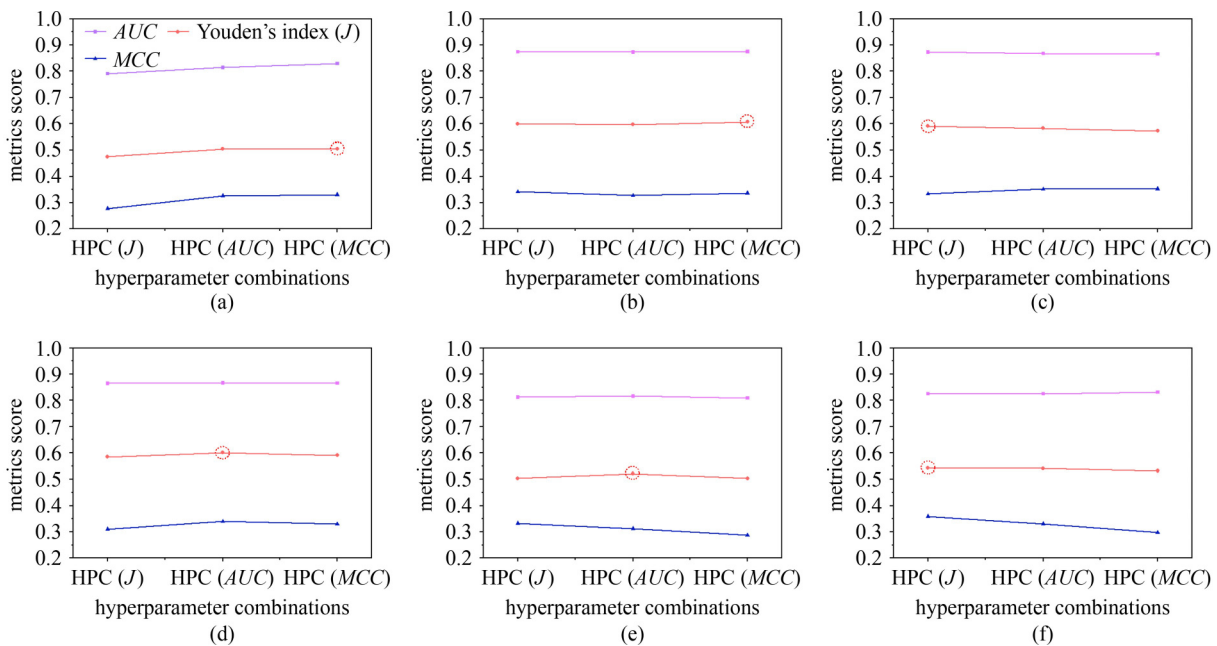


Fig. 7 Metric scores of classifiers trained with different HPCs: (a) DT; (b) RF; (c) XGBoost; (d) XGBRF; (e) ANN; (f) ResNet18.

samples, with ResNet18 achieving the highest recognition rate of 85.8%. On the other hand, the RF, XGBRF, and XGBoost algorithm classifiers have good recognition results for FWM samples, with the RF model having the highest recognition rate of 86.2%. However, based on the confusion matrix results alone, it is challenging to determine which algorithm is better.

To visually compare of the models' predictions, we plot some of the indicator scores in Fig. 9. According to Fig. 9(a), the RF, XGBRF, and XGBoost classifiers exhibit higher J scores, but XGBoost also has a higher miss rate than the other two classifiers. Thus, we analyze the ROC curves and optimal thresholds of the other two classifiers, as shown in Fig. 9(b). Based on the ROC curves, the overall performance of the two classifiers is

similar. Still, the optimal threshold of the XGBRF classifier is 0.493, while the optimal threshold of the RF classifier is only 0.28. This finding suggests that the XGBRF classifier performs better for imbalanced data.

Through a comparison analysis, we find that the XGBRF algorithm exhibits excellent predicting and generalizing performance, and the final hyperparameter is HPC (AUC) through hyperparameter screening. The 'n_estimators' is 286, the 'max_depth' is 67, the 'learning_rate' is 0.039, the 'gamma' is 1.0, and the 'reg_lambda' is 0.02.

4.2 Influence on whether to process imbalanced data

The issue of sample imbalance can often lead to

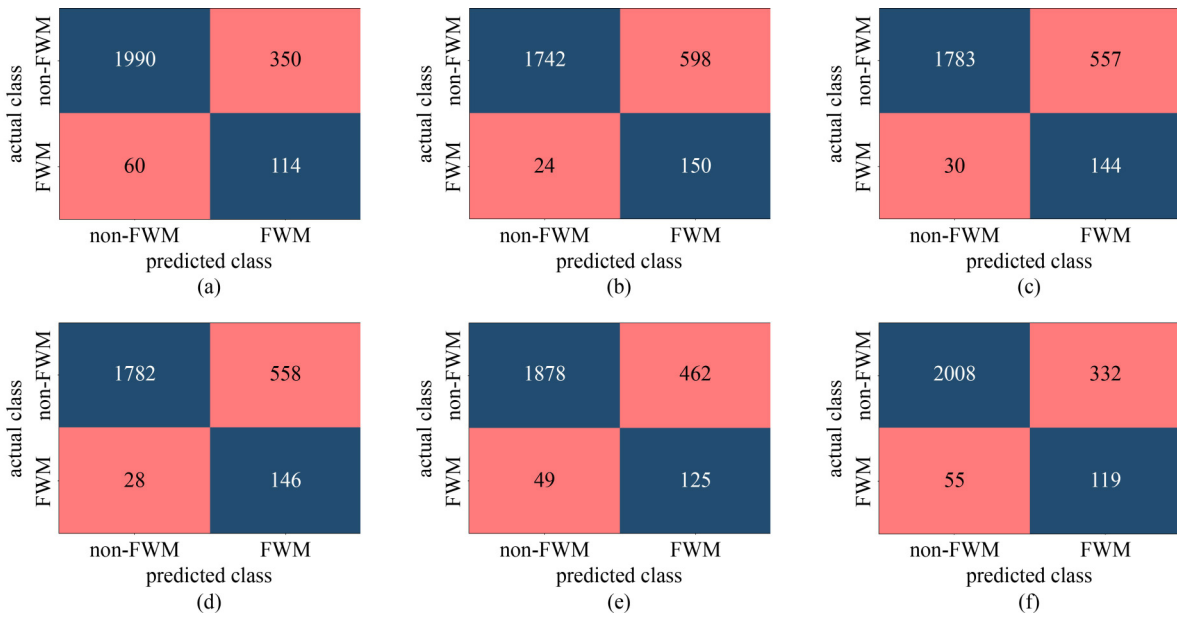


Fig. 8 Confusion matrices of six algorithm classifiers on test dataset: (a) DT; (b) RF; (c) XGBoost; (d) XGBRF; (e) ANN; (f) ResNet18.

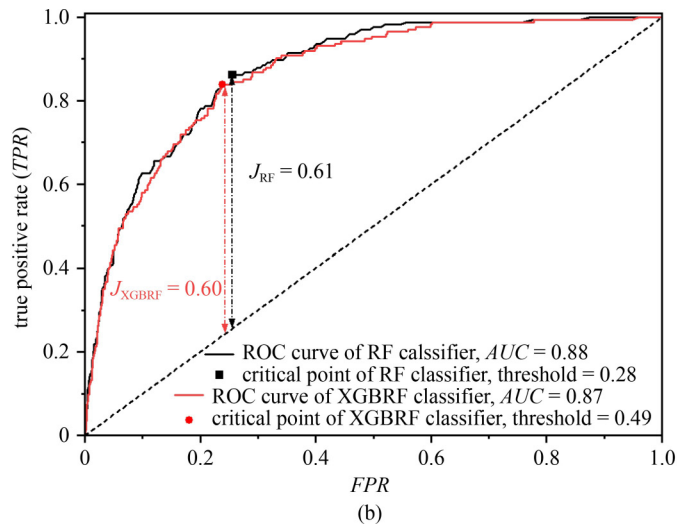
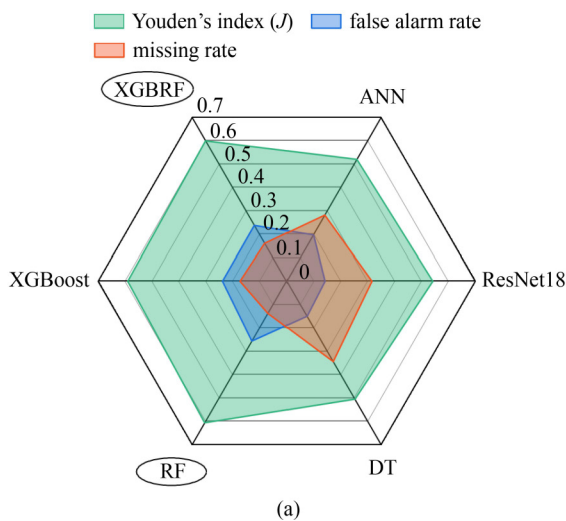


Fig. 9 Metric scores and ROC curves: (a) metric score for various classifiers; (b) ROC curves for RF and XGBRF classifier.

disastrous prediction results. In this study, we seek to investigate the impact of resampling on the prediction performance of the model, and we compare the predictive performance of the two cases according to the process in the flowchart (Fig. 4). Our findings, presented in Fig. 10, show that the classifier trained without resampling tends to predict non-FWM samples as high probability events, indicating a preference for non-FWM.

To further examine the impact of resampling, we analyzed the statistics of false alarm rates and miss rates of RF and XGBRF algorithm models under three conditions: (a) when the data used in the training model was not resampled and the threshold was set at 0.5; (b) when the training model used resampled data and the threshold was set at 0.5; (c) when the training model used resampled data and an optimal threshold was applied. As shown in Fig. 11, the false alarm rate of both the RF and XGBRF classifiers was very low when the data used in the training model was not resampled and the threshold was set at 0.5. However, the miss rate was higher for RF (75.3%) compared to XGBRF (54.6%), indicating that XGBRF may perform better in addressing data imbalance issues.

Moreover, our study observed an interesting phenomenon that the optimal threshold of the XGBRF classifier was closer to 0.5 than that of the RF classifier, demonstrating its advantage in dealing with data sets with

extremely imbalanced sample classes. After resampling the training data and setting the threshold value at 0.5, the miss rates of the XGBRF and RF classifiers decreased to 33.9% and 34.5%, respectively. This suggests that resampling can effectively improve the model's performance in response to sample imbalance.

In conclusion, our study suggests that resampling can significantly enhance the prediction performance of machine learning models. Additionally, the XGBRF algorithm may have certain advantages in dealing with the data imbalance issue, as demonstrated by its performance in this study.

4.3 Impact of different input features on model performance

The preceding section focused on the evaluation of the performance of each algorithmic model, utilizing a set of 26 input features. In this section, we aim to demonstrate the importance of individual input features on the predictive capability of the model. Additionally, we analyze the effect of selecting different input variables to train the model.

4.3.1 Input features importance

Different input features exert varying degrees of

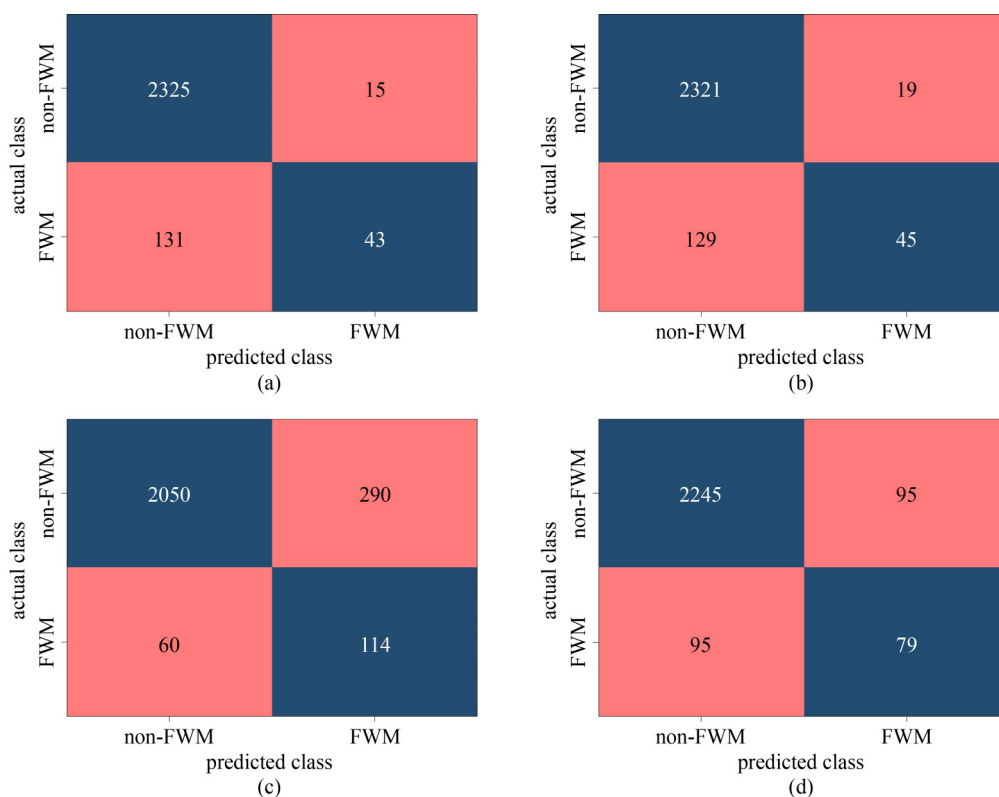


Fig. 10 Confusion matrices of different classifier with a threshold of 0.5: (a) RF without resampling; (b) XGBRF without resampling; (c) RF with resampling; (d) XGBRF with resampling.

influence on predictive performance. In the context of data analysis, it is beneficial to eliminate features that exhibit relatively insignificant contributions to the model's generalization performance. As depicted in Fig. 12, there are four features with a contribution exceeding 3.0%, and with a total contribution of 57.1%. These features, ranked in order of descending contribution, are Mean (*TPIs*), Mean (*n*), Mean (*TPIu*), and C.V (*TPIs*); Mean (*TPIs*) alone accounts for 41.4% of the contribution. These features are highlighted in the figure for emphasis.

Furthermore, Fig. 13 portrays the distribution curves of features across different class samples. The red dashed lines correspond to the parameter distribution of the FWM samples, while the black solid lines represent the parameters distribution of the non-FWM samples. A comparison of the two subplots reveals that the Mean (*FPIs*) value distribution ranges of the FWM and non-FWM samples are distinctly different, indicating that the FWM characteristics can be identified by utilizing this feature.

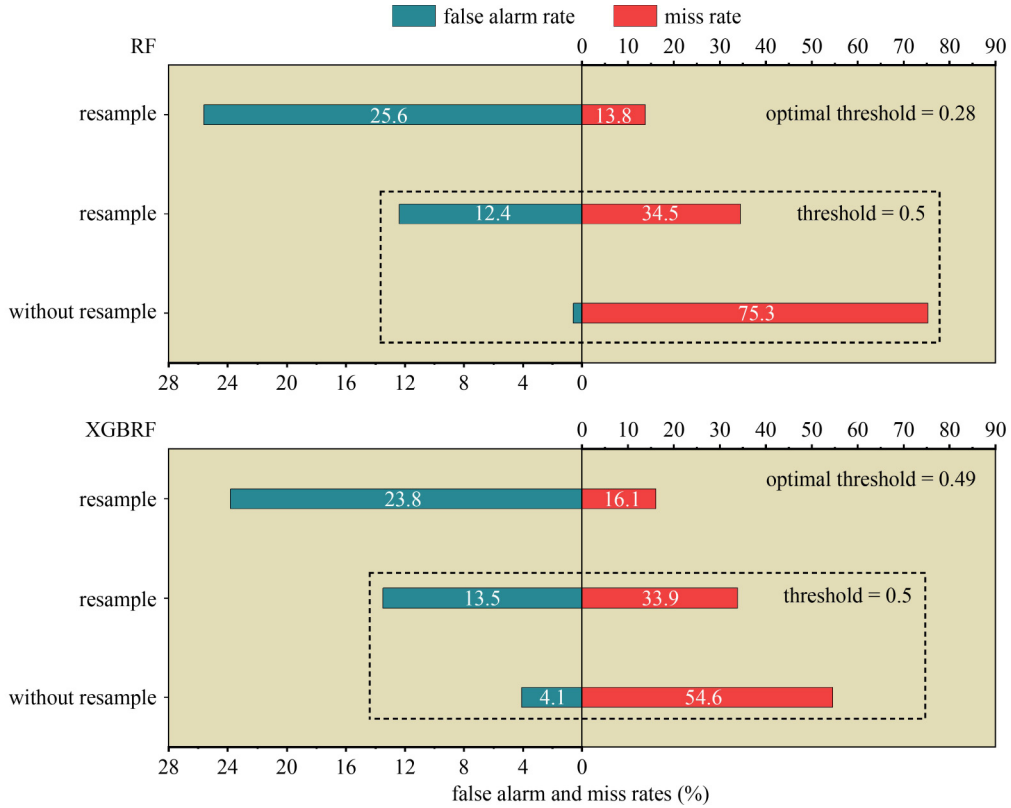


Fig. 11 False alarm and miss rates of RF and XGBRF classifiers.

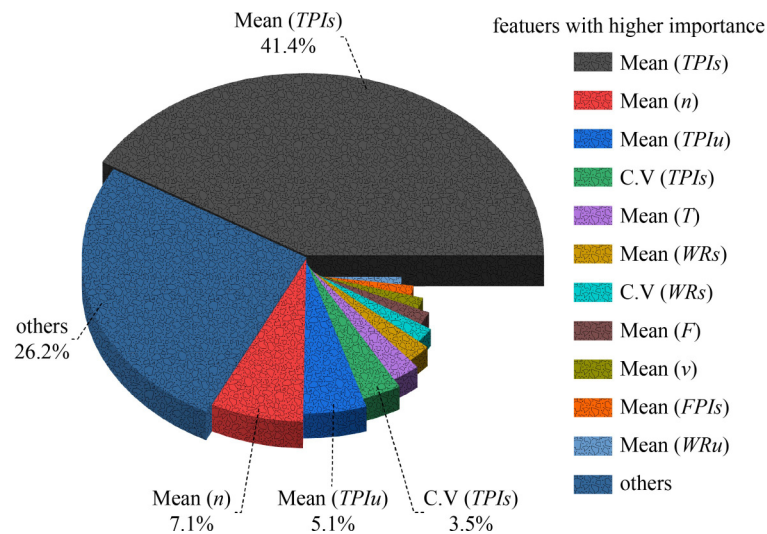


Fig. 12 Importance of input features based on XGBRF model.

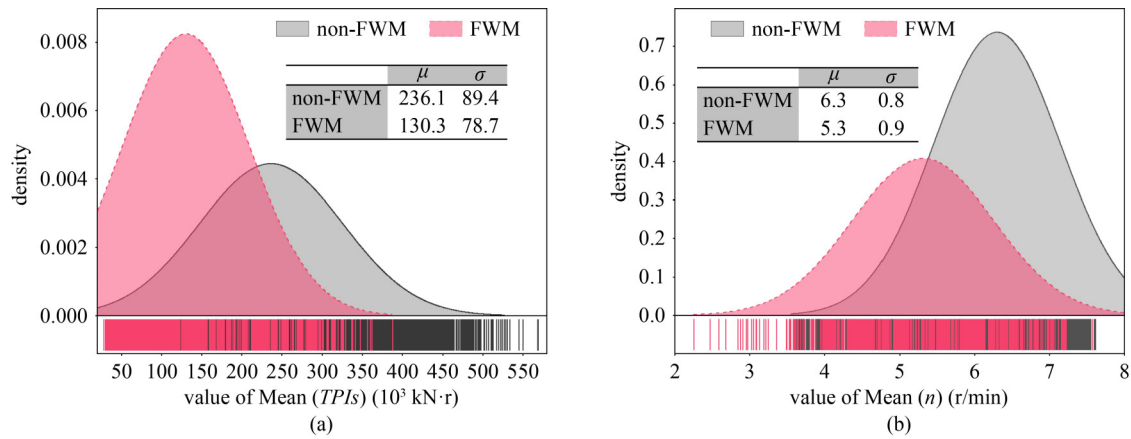


Fig. 13 Distribution of features across different class samples: (a) Mean (*TPIs*); (b) Mean (*n*).

4.3.2 Data-driven model vs. knowledge-driven model

Data-driven models are constructed using only the fundamental sensor parameters obtained from TBM without incorporating any expert knowledge of rock-breaking behaviors. In contrast, knowledge-driven models are developed by including input features that consist of variables infused with expert knowledge of rock-breaking behaviors.

To compare the discrepancies between data-driven, knowledge-driven, and dual-driven models, we trained the XGBRF algorithm using the three different input approaches mentioned in Table 6. To ensure the validity of our results, each model trained autonomously and selected its optimal parameters according to the procedure illustrated in Fig. 4.

Table 6 Performance of data-driven versus knowledge-driven models

model	input features	sum of feature importance (%)	<i>J</i>	<i>FPR</i> (%)	<i>FNR</i> (%)
data-driven	X_1-X_8	23.5	0.538	28.4	17.8
knowledge-driven	X_9-X_{26}	76.5	0.580	22.5	19.5
dual-driven	X_1-X_{26}	100.0	0.601	23.8	16.1

Notes: Results in this table are based on optimal threshold conditions for *FPR* and *FNR*; *FPR* = false alarm rate; *FNR* = miss rate.

Specifically, the data-driven model utilized input features X_1-X_8 with an aggregate importance of 23.5%. Knowledge-driven model used input features X_9-X_{26} with an aggregate importance of 76.5%. Finally, the dual-driven model utilized all the parameters as input features with a total feature importance score of 100%. By comparing the results of the three models, the following phenomena were observed.

Firstly, the false alarm rates of the data-driven classifier and the knowledge-driven classifier were 28.4% and 22.5%, respectively, while the miss rates of the two classifiers were relatively similar. This indicates that the knowledge-driven model has better perceptual results than the data-driven model.

Secondly, the dual-driven classifier has the highest *J* score and the lowest miss rate. It can be concluded that adding expert knowledge variables to the input features can effectively improve the model’s detection of FWMs.

4.3.3 Real-time prediction model for fractured and weak rock mass

Real-time prediction is a critical requirement in TBM construction to prevent accidents such as cutterhead jamming or collapses by accurately assessing and responding to the surrounding rock conditions. Despite the establishment of a well-performing model, the timeliness of parameter acquisition was not considered in selecting input parameters. For instance, parameters such as *TP_{lu}*, *FP_{lu}*, and *WR_u*, are calculated based on loading phase data, while *TP_{ls}*, *FP_{ls}*, and *WR_s*, etc. are calculated based on stable boring phase data. In TBM construction, the duration of the loading phase is relatively short and data can be obtained within a few minutes after the cutterhead contacts the rock mass, whereas stable boring data can typically be obtained after excavation is complete.

To achieve real-time prediction, a real-time approach was chosen to train the XGBRF model. This means that only the loading phase features were used to train the model, and the predicted samples only needed to provide the loading phase features. Table 7 illustrates the predictive performances of the real-time model and the dual-driven model. The false alarm and miss rates of the real-time classifier are 24.6% and 21.8%, respectively, indicating that the real-time prediction model can meet

Table 7 Performance of real-time predictive models

model	input features	sum of feature importance (%)	<i>J</i>	<i>FPR</i> (%)	<i>FNR</i> (%)
dual-driven	X_1-X_{26}	100.0	0.601	23.8	16.1
real-time	$X_{15}-X_{26}$	22.8	0.536	24.6	21.8

Notes: Results in this table are based on optimal threshold conditions for *FPR* and *FNR*; *FPR* = false alarm rate; *FNR* = miss rate.

the TBM construction safety early warning requirements effectively. However, the dual-driven model has a significantly lower miss rate than the real-time model.

In conclusion, the real-time model can offer practical real-time prediction during TBM construction, while the dual-driven model can continuously correct the real-time model's prediction results during continuous excavation. This combination ensures both the real-time prediction capability and the prediction's accuracy, thereby enhancing TBM construction safety.

5 Conclusions

The paper presents a novel model for predicting the rock conditions of tunnel face using six distinct algorithms and the data collected by TBM sensors. Specifically, the paper proposes a real-time perception method for FWM based on the XGBRF algorithm, which is compared to other algorithms.

Initially, this study compares the effects of different metrics on hyperparameter selection and the predictive performance of various algorithms. Subsequently, the study discusses the model's performance before and after processing imbalanced data. Finally, the authors explore the variations in model performance that arise from using different input data and establish a real-time prediction model for weak and broken surrounding rock. The specific conclusions are drawn as follows.

1) The most appropriate performance metrics for this data set are the Youden's index and *AUC*, and the XGBRF algorithm exhibits superior generalization performance.

2) After a comparison analysis, we found that the XGBRF algorithm demonstrates excellent prediction and generalization performance. Taking the *AUC* as the metric, the final selection from the hyperparameter screening included 'n_estimators' set to 286, 'max_depth' set to 67, 'learning_rate' set to 0.039, 'gamma' set to 1.0, and 'reg_lambda' set to 0.02.

3) Resampling imbalanced data sets can significantly improve the prediction performance of machine learning models. Additionally, the XGBRF algorithm may have certain advantages in dealing with the imbalanced data problem.

4) Different input features exert varying degrees of influence on the predictive performance of a model. For the XGBRF classifier, four features with a contribution exceeding 3.0% are identified, with a total contribution of 57.1%. These features, ranked in descending order of contribution, are Mean (*TPIs*), Mean (*n*), Mean (*TPIu*), and C.V (*TPIs*), with Mean (*TPIs*) alone accounting for 41.4% of the contribution.

5) In the case of similar miss rates, the false alarm rate of knowledge-driven classifier is significantly lower than

that of data-driven classifier. This indicates that knowledge driven models have better perceptual results than data driven models.

6) The miss rates of the real-time classifier and dual-driven classifier are 21.8% and 16.1%, respectively, indicating that the real-time prediction model can meet the TBM construction safety early warning requirements effectively.

The proposed real-time prediction model can offer practical real-time prediction during TBM construction, while the dual-driven model can continuously correct the real-time model's prediction results during continuous excavation. This combination ensures both the real-time prediction capability and the prediction's accuracy, enhancing TBM construction safety. When the model issues an alarm, the TBM operator and on-site engineer can take necessary measures to avoid potential collapse.

Acknowledgements We sincerely give our thanks to the data support from the National Program on Key Basic Research Project of China (No. 2015CB058100), China Railway Engineering Equipment Group Corporation and the Survey and Design Institute of Water Conservancy of Jilin Province. This work was supported by the Natural Key R&D Program of China (No. 2022YFE0200400).

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

Conflict of Interests The authors declare that they have no conflict of interest.

References

- Gong Q M, Yin L J, Ma H S, Zhao J. TBM tunnelling under adverse geological conditions: An overview. *Tunnelling and Underground Space Technology*, 2016, 57: 4–17
- Rostami J. Performance prediction of hard rock Tunnel Boring Machines (TBMs) in difficult ground. *Tunnelling and Underground Space Technology*, 2016, 57: 173–182
- Zheng Y L, Zhang Q B, Zhao J. Challenges and opportunities of using tunnel boring machines in mining. *Tunnelling and Underground Space Technology*, 2016, 57: 287–299
- Yokota Y, Yamamoto T, Shirasagi S, Koizumi Y, Descour J, Kohlhaas M. Evaluation of geological conditions ahead of TBM

- tunnel using wireless seismic reflector tracing system. *Tunnelling and Underground Space Technology*, 2016, 57: 85–90
5. Li S C, Liu B, Xu X J, Nie L C, Liu Z Y, Song J, Sun H F, Chen L, Fan K R. An overview of ahead geological prospecting in tunneling. *Tunnelling and Underground Space Technology*, 2017, 63: 69–94
 6. Li S C, Nie L C, Liu B. The practice of forward prospecting of adverse geology applied to hard rock tbn tunnel construction: The case of the Songhua river water conveyance project in the middle of Jilin province. *Engineering*, 2018, 4(1): 131–137
 7. Yang S L, Wang Z F, Wang J, Cohn A G, Zhang J Q, Jiang P, Nie L C, Sui Q M. Defect segmentation: Mapping tunnel lining internal defects with ground penetrating radar data using a convolutional neural network. *Construction & Building Materials*, 2022, 319: 125658
 8. Wang J S, Yang S L, Xu X J, Jiang P X, Ren Y X, Du C X, Du S L. 3C–3D tunnel seismic reverse time migration imaging: A case study of Pearl River Delta Water Resources Allocation Project. *Journal of Applied Geophysics*, 2023, 210: 104954
 9. Li J B, Jing L J, Zheng X F, Li P Y, Yang C. Application and outlook of information and intelligence technology for safe and efficient TBM construction. *Tunnelling and Underground Space Technology*, 2019, 93: 103097
 10. Hassanpour J, Rostami J, Zhao J. A new hard rock TBM performance prediction model for project planning. *Tunnelling and Underground Space Technology*, 2011, 26(5): 595–603
 11. Farrokh E, Rostami J, Laughton C. Study of various models for estimation of penetration rate of hard rock TBMs. *Tunnelling and Underground Space Technology*, 2012, 30: 110–123
 12. Rostami J. Study of pressure distribution within the crushed zone in the contact area between rock and disc cutters. *International Journal of Rock Mechanics and Mining Sciences*, 2013, 57: 172–186
 13. Liu Q S, Pan Y C, Liu J P, Kong X X, Shi K. Comparison and discussion on fragmentation behavior of soft rock in multi-indentation tests by a single TBM disc cutter. *Tunnelling and Underground Space Technology*, 2016, 57: 151–161
 14. Ma H S, Gong Q M, Wang J, Yin L J, Zhao X B. Study on the influence of confining stress on TBM performance in granite rock by linear cutting test. *Tunnelling and Underground Space Technology*, 2016, 57: 145–150
 15. Smith J V. Assessing the ability of rock masses to support block breakage at the TBM cutter face. *Tunnelling and Underground Space Technology*, 2016, 57: 91–98
 16. Yin L J, Miao C T, He G W, Dai F C, Gong Q M. Study on the influence of joint spacing on rock fragmentation under TBM cutter by linear cutting test. *Tunnelling and Underground Space Technology*, 2016, 57: 137–144
 17. Pan Y C, Liu Q S, Liu J P, Huang X, Liu Q, Peng X X. Comparison between experimental and semi-theoretical disc cutter cutting forces: Implications for frame stiffness of the linear cutting machine. *Arabian Journal of Geosciences*, 2018, 11(11): 1–20
 18. Hamidi J K, Shahriar K, Rezai B, Rostami J. Performance prediction of hard rock TBM using Rock Mass Rating (RMR) system. *Tunnelling and Underground Space Technology*, 2010, 25(4): 333–345
 19. Hassanpour J, Rostami J, Khamehchiyan M, Bruland A, Tavakoli H R. TBM performance analysis in pyroclastic rocks: A case history of Karaj Water conveyance tunnel. *Rock Mechanics and Rock Engineering*, 2010, 43(4): 427–445
 20. Hassanpour J, Vanani A G, Rostami J, Cheshomi A. Evaluation of common TBM performance prediction models based on field data from the second lot of Zagros water conveyance tunnel (ZWCT2). *Tunnelling and Underground Space Technology*, 2016, 52: 147–156
 21. Delisio A, Zhao J, Einstein H H. Analysis and prediction of TBM performance in blocky rock conditions at the Löttschberg Base Tunnel. *Tunnelling and Underground Space Technology*, 2013, 33: 131–142
 22. Dudt J P, Delisio A. The “penalty factors” method for the prediction of TBM performances in changing grounds. *Tunnelling and Underground Space Technology*, 2016, 57: 195–200
 23. Pan Y C, Liu Q S, Liu Q, Bo Y, Liu J P, Peng X X, Cai T. Comparison and correlation between the laboratory, semi-theoretical and empirical methods in predicting the field excavation performance of tunnel boring machine (TBM). *Acta Geotechnica*, 2022, 17(2): 653–676
 24. 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. *Tunnelling and Underground Space Technology*, 2022, 120: 104285
 25. Zhang Q L, Zhu Y W, Ma R, Du C X, Du S L, Shao K, Li Q B. Prediction method of TBM tunneling parameters based on PSO-Bi-LSTM model. *Frontiers in Earth Science*, 2022, 10: 854807
 26. Qiu D H, Fu K, Xue Y G, Tao Y F, Kong F M, Bai C H. TBM tunnel surrounding rock classification method and real-time identification model based on tunneling performance. *International Journal of Geomechanics*, 2022, 22(6): 04022070
 27. Li J B, Chen Z Y, Li X, Jing L J, Zhang Y P, Xiao H H, Wang S J, Yang W K, Wu L J, Li P Y, Li H B, Yao M, Fan L T. Feedback on a shared big dataset for intelligent TBM Part I: Feature extraction and machine learning methods. *Underground Space*, 2023, 11: 1–25
 28. Li J B, Chen Z Y, Li X, Jing L J, Zhang Y P, Xiao H H, Wang S J, Yang W K, Wu L J, Li P Y, Li H B, Yao M, Fan L T. Feedback on a shared big dataset for intelligent TBM Part II: Application and forward look. *Underground Space*, 2023, 11: 26–45
 29. Zhang Q L, Liu Z Y, Tan J R. Prediction of geological conditions for a tunnel boring machine using big operational data. *Automation in Construction*, 2019, 100: 73–83
 30. Zhu M Q, Gutierrez M, Zhu H H, Ju J W, Sarna S. Performance Evaluation Indicator (PEI): A new paradigm to evaluate the competence of machine learning classifiers in predicting rockmass conditions. *Advanced Engineering Informatics*, 2021, 47: 101232
 31. Hou S K, Liu Y R, Li C Y, Qin P X. Dynamic prediction of rock mass classification in the tunnel construction process based on random forest algorithm and TBM *in situ* operation parameters. In: *IOP Conference Series: Earth and Environmental Science*. Beijing: IOP Publishing Ltd., 2020, 052056
 32. Hou S K, Liu Y R. Early warning of tunnel collapse based on Adam-optimised long short-term memory network and TBM

- operation parameters. *Engineering Applications of Artificial Intelligence*, 2022, 112: 104842
33. Li J H, Li P X, Guo D, Li X, Chen Z Y. Advanced prediction of tunnel boring machine performance based on Big Data. *Geoscience Frontiers*, 2021, 12(1): 331–338
 34. Liu B, Wang R, Zhao G, Guo X, Wang Y, Li J, Wang S. Prediction of rock mass parameters in the TBM tunnel based on BP neural network integrated simulated annealing algorithm. *Tunnelling and Underground Space Technology*, 2020, 95: 103103
 35. Feng S X, Chen Z Y, Luo H, Wang S Y, Zhao Y F, Liu L P, Ling D S, Jing L J. Tunnel boring machines (TBM) performance prediction: A case study using big data and deep learning. *Tunnelling and Underground Space Technology*, 2021, 110: 103636
 36. Gong Q M, Zhao J, Jiang Y S. *In situ* TBM penetration tests and rock mass boreability analysis in hard rock tunnels. *Tunnelling and Underground Space Technology*, 2007, 22(3): 303–316
 37. Chen Z Y, Zhang Y P, Li J B, Li X, Jing L J. Diagnosing tunnel collapse sections based on TBM tunneling Big Data and deep learning: A case study on the YinSong Project, China. *Tunnelling and Underground Space Technology*, 2021, 108: 103700
 38. Jing L J, Li J B, Yang C, Chen S, Zhang N, Peng X X. A case study of TBM performance prediction using field tunnelling tests in limestone strata. *Tunnelling and Underground Space Technology*, 2019, 83: 364–372
 39. Breiman L. Random Forests. *Machine Learning*, 2001, 45(1): 5–32
 40. Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM, 2016, 785–794
 41. He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2016*. Las Vegas, NV: IEEE, 2016, 770–778
 42. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, NY: ACM, 2019, 2623–2631