Intelligent Surrounding Rock Grade Identification Combining XGBoost Algorithm and Drilling Parameters of Drill Jumbo

Guoqiang Huang[1], Chengjin Qin[1](✉), Feixiang Liu[2], Ruihong Wu[1], Jianfeng Tao[1], Chengliang Liu[1], and Jinjun Liao[2]

[1] State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
qinchengjin@sjtu.edu.cn

[2] China Railway Construction Heavy Industry Group Co., Ltd., Changsha 410100, China

Abstract. Surrounding rock classification represents distinguishing the different grades of surrounding rock according to the hardness and integrity of surrounding rock. Accurately obtaining the surrounding rock grade of drill jumbo working face is not only the basis for selecting the tunnel position and support type, but also the key to ensure the safety of the drill jumbo's construction site. As the traditional classification methods, engineering drilling and geological mapping are time-consuming and labor-intensive. Aiming at this situation, this paper proposes an intelligent identification method of surrounding rock grade combine drilling parameters with machine learning algorithm XGBoost. Firstly, adequately analyse the correlation between drilling parameters and rock label, and select six drilling parameters as feature vectors for surrounding rock grade recognition. Then outlier processing and data screening are carried out on the data recorded by the drill jumbo. Next, we construct a model based on XGBoost to realize the rapid and accurate identification of surrounding rock grade. Finally, the effectiveness and superiority of the proposed method are demonstrated by the actual data collected by the drill jumbo in Gao Jiaping tunnel, and mix the partial data of Alianqiu tunnel together to construct 5 datasets to compare the identification performance of other classical algorithms. The results show that the recognition capability of the proposed method is superior to those of other algorithms, and the recognition accuracy of surrounding rock along the tunnel can reach 99.68%.

Keywords: drilling parameters · Surrounding rock grade · XGBoost · Intelligent identification

1 Introduction

In China, the scale of railway construction is increasing year by year, and the demand for ultra-long railway tunnels with high altitude and high burial depth increases as well. In the early days of the founding of country, there were only 429 railway tunnels in China, with a total length of 112 km. By 2020, China's railway operating mileage reached

L. Yan and J. Na (Eds.): ICFPMCE 2022, AHE 10, pp. 96–114, 2023.
https://doi.org/10.2991/978-94-6463-022-0_10

145,000 km, among which 16,798 railway tunnels were in operation and the total length reaches to 19,630 km. Drill jumbo is a kind of tunnelling engineering machinery constructed by drilling and blasting method, which is widely used in high-speed railway and highway tunnel projects. With the continuous improvement of labour costs, intelligent and unmanned tunnel construction is the inevitable trend of tunnel construction in the future. At present, the intelligent level of drill jumbo is gradually improving, and more data could be collected in the process of tunnel construction. However, the data collected by projects has not been fully used to guide the construction, which has hindered the further improvement of the tunnel construction level. In the construction process, the surrounding rock grade is not only the basis for selecting the tunnel position and support type, but also the key to ensure the safety of the drill jumbo's construction site. In the past, it was necessary to know the grade of surrounding rock by engineering drilling and geological mapping, which will significantly decrease construction efficiency. Using artificial intelligence to identify the surrounding rock grade of the working face can reduce the time for geologists to judge during construction. In addition, it can find bad geology in advance and warn to guarantee the safety of the construction site.

In the early stage, scholars mainly recognized the surrounding rock grade based on the physical characteristics of surrounding rock. Ma [1] proposed the controllable source audio-frequency magneto telluric method, which adopts the artificial source electromagnetic sounding method in the frequency domain. This method of classifying tunnel surrounding rock has strong applicability for deep and large tunnel investigation. Wu et al. [2] found that though the strength of carbonate rock is greatly reduced due to the recementation of calcite, the integrity of rock mass is still good. Therefore, the classification of surrounding rock of carbonate rock can be improved compared with that of non-soluble rock. Qiu et al. [3] selected the index of rock's uniaxial compressive strength, rock quality index, discontinuous structural plane state and filling condition and so on as inputs to propose a classification method based on rough set and ideal point, which was applied to actual engineering. Li [4] used the finite element numerical simulation method and ANSYS to simulate the process of tunnel construction, excavation and support. He established a geological generalized model to accurately evaluate of the deformation stability of surrounding rock and the effect of support measures. Liu et al. [5] selected compressive strength of rock mass, rock quality index RQD, joint spacing, joint condition, groundwater condition, influence of joint and fissure strike as inputs to establish a fuzzy comprehensive evaluation model to distinguish the grade of surrounding rock. Chen [6] designed and compiled an expert system applied to highway tunnel according to the 04 edition of the standard method, RMR classification method, Q classification method and fuzzy comprehensive evaluation method. The expert system can identify the surrounding rock grade with multiple methods and angles and provide effective reference for practical engineering. Xue et al. [7] applied the extension theory to the surrounding rock classification, indicating that the extension surrounding rock classification method has good applicability and accuracy in the loess tunnel sur-rounding rock classification. Rong et al. [8] used advanced geological prediction methods such as TGP detection method, SIR-20 geological radar detection method and palm-surface catalog method to dynamically predict the grade of tunnel surrounding rock.

Compared with geological mapping or classification regression algorithm, machine learning algorithm has high recognition accuracy and efficiency. Thus, Researchers are gradually trying to apply machine learning to this task. Yao et al. [9, 10] studied the correlation between various drilling parameters of drill jumbo and surrounding rock grade, and then selected feed speed, strike pressure, propulsion pressure, rotary pressure as inputs of SVM, which achieved good identification accuracy. Wen et al. [11] used genetic algorithm (GA) to optimize the key parameter of support vector machine (SVM), and took the extracted common factor as the input variable to establish a GA-SVM model rely on factor analysis. B. Rajesh Kumar et al. [12] used artificial neural network to predict the strength, intensity, dry density, p-wave velocity, tensile strength, Young's modulus and porosity of rock. Tian et al. [13] proposed that when drilling conditions are determined, specific energy of drilling can be used to classify surrounding rock. Yang et al. [14] selected RQD, Rc, rock integrity coefficient, structural plane strength coefficient and groundwater seepage as inputs to establish a generalized neural network (GRNN) model for the classification of tunnel surrounding rock.

Studies have shown that the drilling parameters are closely related to the grade of surrounding rock on the working face. However, most of the existing researches on the classification of surrounding rock are focused on the physical properties of rock mass such as uniaxial compressive strength of rock mass, rock quality index RQD, rock chemical composition and so on. Few people study the data recorded during the construction of construction machinery. Yao et al. [9, 10] constructed an intelligent identification model of surrounding rock grade based on drilling parameters of drill jumbo, but the selected drilling parameters included only feed speed, strike pressure, propulsive pressure and rotary pressure. In fact, more than four drilling parameters are recorded, and there are still geological features hidden in the drilling parameters that have not been excavated, leading to problems such as low model recognition accuracy and poor generalization ability.

Therefore, in order to compensate for the above-mentioned shortcomings, this paper proposes an intelligent identification method of surrounding rock grade based on drilling parameters and machine learning algorithm XGBoost. Firstly, we use Pearson correlation coefficient method to analyse the correlation between drilling parameters and rock label, and calculate the correlation coefficient between each parameter and rock label. The drilling parameters with strong correlation were selected as the input of intelligent classification model. Then, we clear the abnormal data recorded during tunnel construction, screen the number of samples and split training set and test set. Next, we put the selected characteristic parameters into proposed model to realize the classification of tunnel surrounding rock. Finally, we use actual construction data with the Stochastic Gradient Descent Classifier (SGDC), K Near Neighbor (KNN), Linear Support Vector Classifier (LSVC), Multilayer Perceptron (MLP) and integrated classification model of Random Forest and other existing models to verify the performance of proposed intelligent recognition method of surrounding rock grade. The results verify that the proposed method has high prediction accuracy, good robustness and fast training speed in the intelligent identification task of tunnel surrounding rock grade.

2 Materials

The drilling data of the drill jumbo used in this paper are from the Gao Jiaping tunnel project of Zheng-Wan High-speed Railway. Gao Jiaping tunnel is located in Limiao Town, Nanzhang County, Xiangyang City. The central location of this tunnel is DK451+786, and the starting to stopping location is DK449+037–DK454+535, with a total length of 5498 m and a maximum burial depth of about 320 m. The tunnel adopts import-export two-way tunneling, and the inlet location is DK449+037–DK452+300. The inlet section is mostly v-class surrounding rock, mainly medium-strong weathered shale, with soft rock broken and poor self-stability, which is a typical weak surrounding rock tunnel with complicated engineering geology, difficult construction and high safety risk. In the drilling process, Gao Jiaping tunnel project adopts a fully computer three-arm wheeled drill jumbo, which is mainly driven by hydraulic motor or diesel engine. As shown in Fig. 1, it has sensitive and rapid movement, sensitive steering and braking adjustment. But its climbing ability is not strong, and it is difficult to pass complex road surface.

Fig. 1. Wheeled Dirll Jumbo

Fig. 2. Construction site of rock drill jumbo in Gao Jiaping tunnel [16]

At present, the application of automatic induction graphics technology and laser scanning technology realizes the automatic orientation and automatic drilling of rock drill jumbo according to the drilling map [15]. Figure 2 shows the construction site of Gao Jiaping tunnel. The highly automated hydraulic drill jumbo realizes the automatic acquisition and processing of intelligent drilling and drilling parameters, which provides important conditions for the intelligent identification of surrounding rock grade based on drilling parameters.

3 Methodology

3.1 Parameter Selection

3.1.1 Parameter Introduction

The highly intelligent drill jumbo can record various physical parameters in real time through sensors in the process of drilling, and then automatically generate original data sets. The drill jumbo used in Gao Jiaping tunnel records nine drilling parameters, i.e., drilling depth (m), feed speed (m/min), water flow rate (L /min), strike pressure (bar), thrust pressure (bar), rotary pressure (bar), rotary speed (m/min), water pressure (bar) and interval Time (ms).

Feeding speed refers to the movement speed of the drill bit along the direction of the hole. After drilling begins, the photoelectric sensor is interrupted once every certain distance. At the same time, the timer records the drilling time, and the feed speed can be calculated. The hardness and strength of rock with different rock grades vary greatly, so the resistance of drill bit will inevitably increase when working with high strength rock. So, the feeding speed will decrease when the input power is the same. Therefore, feeding speed is a drilling parameter that can reflect geological conditions.

Strike pressure refers to the oil pressure inside the hydraulic cylinder of drilling arm when the drill jumbo breaks the surrounding rock, whose value is measured by pressure sensors distributed in the hydraulic cylinder. Obviously, breaking rocks with high strength and hardness requires higher strike pressure [17]. In addition, the better the integrity of the rock mass is, the higher strike pressure will be needed. Strength, hardness and rock integrity are the key factors to determine the rock grade. Therefore, the strike pressure is closely related to the rock grade and it is a very common parameter used in the identification task of surrounding rock.

Propulsive pressure refers to the oil pressure inside the hydraulic cylinder during the propulsive movement of the drill jumbo, which is used to keep the drill bit in close contact with the surrounding rock, whose value is measured by pressure sensors distributed in the hydraulic cylinder. As the rock strength is positively correlated with the surface hardness, the propulsive pressure required to keep in close contact when drilling the rock with high strength and hardness will be larger. In addition, the higher the integrity of the rock mass is, the greater the reverse impact on the bit and the greater the average propulsive pressure required. Therefore, the propulsive pressure is closely related to the rock grade, and it is a very common parameter used in the identification of surrounding rock.

Rotary pressure refers to the oil pressure of hydraulic oil when the drill jumbo performs rotary movement. After measuring the value of the strike pressure and the flow rate, rotary pressure could be calculated by the specified formula. The purpose of the rotary movement is to cut the broken rock down. Because the percussion action of the drill jumbo usually cannot completely break the rock, there are still broken rock attached to the rock wall to be cut and broken. Due to the rock uniaxial compressive strength and shear strength were positively related, the rotary pressure required must be higher when drilling the surrounding rock with high strength and hardness. In addition, even if the surrounding rock hardness is the same, more complete rock mass will produce more rock blocks to be cut. Therefore, the rotary pressure value is closely related to the grade of surrounding rock, and it is also a very common parameter used in the identification of surrounding rock.

Rotary speed refers to the rotary speed at which the drill bit rotates to cut rock. The drill pipe sends out a pulse for every turn, and the counter records the interrupt times of the timer in one turn. Then the time taken for one turn of the drill bit can be calculated, and take the inverse to get the rotary speed.

Water pressure refers to the pressure of water needed by drill jumbo to wash cuttings, and its value can be measured directly by pressure sensor. Water flow rate refers to the water flow rate of rock drill jumbo washing debris, and its value can be measured directly through the flowmeter. In some cases, water pressure and flow rate are closely related to geological conditions, when the drill jumbo replaces the rotary cutting movement of drill pipe with high-pressure water flushing of crushed rock. Therefore, whether to consider water pressure and water flow as effective characteristic parameters depends on the actual situation.

Drilling depth is used to record the drilling depth of the drill bit, and its value is measured by the displacement sensor. The depth of all holes in the same working face varies very little, so the drilling depth is not correlated with geological conditions.

3.1.2 Parameter Selection

Pearson's linear correlation coefficient can reflect the degree of linear correlation between two variables. In supervised learning tasks, the correlation coefficients be-tween features and labels can be calculated to determine whether the extracted features and categories are positively correlated or negatively correlated. The Pearson correlation coefficient between each characteristic parameter and label can be used as the basis for selecting characteristic parameters in the intelligent identification task of surrounding rock grade. Pearson linear correlation coefficient can be calculated by formula (1). The absolute value of the correlation coefficient is close to 1 means that the correlation between this parameter and surrounding rock type is strong; the closer it is to 0, the weaker the correlation between this parameter and surrounding rock type is.

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{1}$$

Put the measured data into formula (1) to calculate the results, which are showed in Table 1. As can be seen from Table 1: Drilling depth, rotation speed and interval time have little or even no correlation with the type of surrounding rock. The feed speed is positively correlated with surrounding rock class, while the percussive pressure, propulsion

Table 1. Correlation coefficient between drilling parameters and surrounding rock types

	drilling depth (m)	feed speed (m/min)	strike pressure (bar)	thrust pressure (bar)	rotary pressure (bar)	rotary speed (m/min)	water pressure (bar)	water flow rate (L /min)	interval time (ms)
correlation coefficient	−0.0427	0.2142	−0.4341	−0.4780	−0.1022	0	−0.7695	−0.3294	−0.0097

pressure, rotation pressure, water pressure and water flow rate are negatively correlated with surrounding rock class, which is consistent with the previous analysis results. The correlation between water pressure and surrounding rock type is the strongest. The correlation is ranked as: water pressure > propulsion pressure > strike pressure > water flow rate > feed speed > depth of drilling > interval time > rotary speed.

According to TB-10003-2005 railway tunnel design code, the grade of surrounding rock is a comprehensive index, which mainly reflects the uniaxial saturation compressive strength of surrounding rock and the integrity of surrounding rock. The basic quality index of surrounding rock mass [BQ] from large to small is: I-grade > II-grade > III-grade > IV-grade > V-grade [18]; Therefore, the comprehensive strength of the surrounding rock will decrease when it transitions from III-grade to V-grade, and the corresponding strike pressure, propulsion pressure, rotary pressure, pressure, water pressure, water flow will decrease, so they are a negative correlation. On the contrary, the feed speed will increase, so it is a positive correlation. The conclusions obtained are consistent with the theoretical analysis results. Further analysis of the data shows that the values of rotation speed are all 0, so it has nothing to do with the type of sur-rounding rock. The correlation coefficient is 0. There is little difference in the drilling depth of each type of surrounding rock, and also little difference in the average drilling depth of three types of surrounding rock, so the correlation is very low. As for the interval time t, it can be found that there is a strong mapping relationship between the two by studying the distribution of feed speed and interval time, as shown in Fig. 3. According to literature research, it is found that the feed speed is calculated by re-cording the time taken by drill jumbo for each 0.02 m, which is interval time t. The relation between feed speed Vj and interval time t satisfies Formula (2). Therefore, in parameter selection, feed speed with higher correlation coefficient can be selected and redundant interval time of drilling parameters can be abandoned.

$$Vj * t = 0.02\,m \tag{2}$$

From the perspective of parameter acquisition, water pressure and water flow rate reflect the machine state of the drill jumbo itself, which has nothing to do with geological conditions. However, the quantity, shape and even physical properties of cut-tings produced after drilling in different grades of surrounding rock are completely different, and the pressure and flow of flushing water required for cleaning cuttings will obviously be different. In Gao Jiaping tunnel engineering, the correlation between water pressure and surrounding rock grade is the highest among all drilling parameters, and the correlation between water flow and surrounding rock grade is also higher than the feed speed and rotary pressure. Therefore, water pressure and water flow must be regarded as important

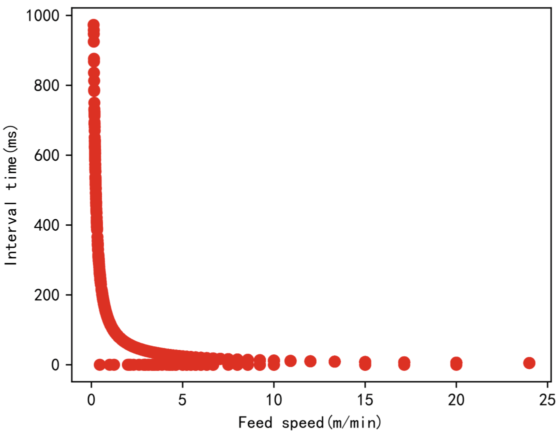

Fig. 3. The distribution of interval time and feed speed

parameters for surrounding rock grade identification when considering the input of the model.

3.2 Drilling Data Pre-processing

The drilling process of a single borehole includes two stages: pre-drilling and normal drilling. The pre-drilling stage has three characteristics: The drilling depth does not exceed 1 m; The propulsion pressure and strike pressure are less than the normal drilling stage; The propulsion pressure gradually increases to normal working pressure from artificial control. The propulsion pressure and strike pressure are relatively constant in the normal drilling stage. Some abnormal data will be generated during the pre-drilling stage, such as a feed speed of 40 m/min, 60 m/min or even 120 m/min. Therefore, before dividing the data set, it is necessary to clear the abnormal data in the sample data set. First of all, calculate the data average value of each parameter separately. Then, visualize the data distribution of each drilling parameter and observe the numerical distribution of each drilling parameter, clear sample data that deviates too much from the average value so that the data distribution of each feature parameter is relatively centralized. The degree of concentration can be compared by the sample standard deviation. The smaller the standard deviation, the more centralized the data distribution, which also indicates that the rock drill jumbo is in a normal state of drilling. Formula (3) and (4) can be used to calculate the data average value and standard deviation of each drilling parameter respectively.

$$\overline{x_j} = \frac{1}{n}\sum_{i=1}^{n} x_{ij} \tag{3}$$

$$\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x_j}\right)^2} \tag{4}$$

Table 2. Part of drilling data of drill jumbo after screening

number	drilling depth (m)	feed speed (m/min)	strike pressure (bar)	thrust pressure (bar)	rotary pressure (bar)	rotary speed (m/min)	water pressure (bar)	water flow rate (L /min)	interval time (ms)	grade
1	0.472	2.105	134	37	39	0	27	89	57	III
2	2.472	2.927	142	46	64	0	29	90	41	III
3	3.532	2.927	144	49	72	0	29	90	41	III
4	3.552	2.857	147	52	70	0	29	90	42	III
5	3.572	2.857	149	54	78	0	29	90	42	III
6	3.592	2.857	150	58	82	0	29	90	42	III
7	3.612	2.857	150	57	82	0	29	90	42	III
…	…	…	…	…	…	…	…	…	…	…
73436	3.973	2.611	117	65	107	0	27	69	46	V
73437	3.993	2.401	119	65	109	0	27	69	50	V
73438	4.013	2.611	117	64	105	0	27	70	46	V
73439	4.033	2.731	117	65	107	0	27	70	44	V
73440	4.053	2.501	118	64	108	0	27	70	48	V
73441	4.073	2.501	117	64	104	0	27	69	48	V
73442	4.093	2.931	117	64	104	0	27	69	41	V

where n represents the amount of data, x_{ij} represents the ith data in the jth drilling parameter, and $\overline{x_j}$ represents the average value of the jth drilling parameter.

According to the location mileage of the obtained data files, corresponding to the original geological prospecting information map of Gao Jiaping tunnel construction drawing design (vertical section information), the corresponding surrounding rock grade of each borehole data can be obtained. The drilling data of the working face with the location of 450824 m–451060 m and 451700 m–451813 m correspond to the III-grade, with the location of 449250 m–449666 m correspond to the V-grade, with the location of 449889 m–449902 m correspond to the IV-grade. These three types of surrounding rock are also the most common types of rock in practical engineering, and accurate identification of these three types of surrounding rock contributes greatly to the actual construction of the project. In addition to data outlier elimination, it is also necessary to reduce the variance in the amount of data available for the various labels, which can reduce the impact of unbalanced sample data and improve the identification effect of the model. In the data obtained, the number of grade 4 surrounding rock samples is the least, so the number of IV-grade surrounding rock samples after screening is taken as the benchmark, and then selecting the suitable number of III-grade and V-grade surrounding rock to form a high-quality data set. Table 2 shows part of the drilling data in the data.

Table 3 records the mean and standard deviation (std) of drilling parameters for each rock class before screening, those of drilling parameters after the removal of abnormal

Table 3. Comparison of surrounding rock samples before and after screening

grade		number		feed speed (m/min)	strike pressure (bar)	thrust pressure (bar)	rotary pressure (bar)	water pressure (bar)	water flow rate (L /min)
III	original	348332	mean	2.84	145.29	75.32	80.68	27.34	88.23
			std	1.78	13.14	22.62	25.38	1.815	5.28
	normal	347402	mean	2.82	145.40	75.38	80.73	27.38	88.34
			std	0.99	12.67	22.59	25.34	1.55	4.72
	screened	23975	mean	2.90	142.92	78.34	84.01	27.68	89.75
			std	0.93	14.17	20.97	23.13	1.39	4.37
IV	original	55689	mean	3.94	135.40	45.27	64.13	28.42	76.51
			std	2.98	20.58	14.84	17.79	15.66	10.19
	screened	26324	mean	3.76	140.92	48.45	55.41	25.88	76.87
			std	0.86	14.45	13.35	12.85	1.66	10.89
V	original	109092	mean	3.41	117.46	50.78	79.44	6.57	80.39
			std	3.01	20.05	14.19	16.36	9.85	18.60
	normal	107609	mean	3.32	118.43	50.82	79.63	6.47	80.86
			std	1.54	17.77	14.15	15.69	9.80	17.80
	screened	23143	mean	3.51	122.67	53.46	78.80	6.54	77.96
			std	1.45	19.89	14.20	15.64145	9.68	19.98

data, and those of drilling parameters after screening. By observing the data in the table, it can be found that the standard deviation of the data after screening is lower than that before screening, which indicates that the distribution of the screened borehole data is more concentrated, so as to avoid abnormal data affecting the training effect of the model and reducing the recognition performance.

4 The Proposed Method

4.1 Principle of XGBoost

XGBoost algorithm, also known as limit gradient lifting algorithm [19], is a typical ensemble learning algorithm. A strong classifier is formed by integrating several weak classifiers. XGBoost based on tree model is suitable for scenarios with low dimension of input data, such as monitoring data of various sensors of small devices. It can automatically make use of CPU multithreading for parallel computation, which is fast. In addition, XGBoost can effectively prevent over-fitting by adding regularization items to the algorithm, making the training results more stable. Therefore, this paper adopts XGBoost algorithm as the classification algorithm for surrounding rock grade recognition. The principle of XGBoost algorithm is as follows:

Set the maximum number of iterations as K, and XGBoost can be stated as:

$$f(x) = \sum\nolimits_{i=1}^{K} g_i(x) \tag{5}$$

where $g_i(x)$ is the basic classifier model generated during each iteration, and the CART tree selected as the basic classifier training set samples are $I = \{(x_i, y_i)\}\left(|I| = n, x_i \in R^l, y \in R\right)$. In each iteration, defining the optimization objective function for training the basic classifier:

$$L^{(t)} = \sum\nolimits_{i=1}^{n} L\left(y_i, \hat{y}_i^{(t-1)} + g_t(x_i)\right) + \Omega(f_t) + c \tag{6}$$

where c is a constant, t is the number of current iterations, $\hat{y}_i^{(t-1)}$ is the prediction result of strong classifier during t-1 iterations, and *Ω(ft)* is the regularization term, which can be expressed as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \sum\nolimits_{j=1}^{m} w_j^2 \tag{7}$$

where T means the number of leaf nodes, and w_j means the weight of the jth leaf node. For the decision tree Q (x) with T leaf nodes, let the weight of each leaf node be w_j (j = 1, 2…, T). Using Taylor's expansion formula (8) to expand optimization objective function formula (6), formula (9) is obtained.

$$f(x + \Delta x) \approx f(x) + f^{'(x)} \Delta x + \frac{1}{2} f^{''}(x) \Delta x^2 \tag{8}$$

$$L^{(t)} \approx \sum_{i=1}^{n} \left[L\left(y_i, \hat{y}^{t-1}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + \mathrm{c} \tag{9}$$

where $g_i = \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)$, $h_i = \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)$ are the first-step and second-order gradients of loss functions $L\left(y_i, \hat{y}^{t-1}\right)$ to $y_i^{(t-1)}$; The optimal weight solution w_j^* is shown in Formula (10):

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{10}$$

where $G_j = \sum_{i \in I_i} g_i$, $H_j = \sum_{i \in I_j} h_i$. Substitute w_j^* into the simplified formula of the objective function to obtain the optimal loss function $L^{(t)*}$, as shown in Formula (11).

$$L^{(t)*} = -\frac{1}{2} \sum\nolimits_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T + c \tag{11}$$

In order to generate the optimal decision tree in each round of training, the loss function value should be minimized when the left and right subtrees I_L and I_R are generated. Define the evaluation function as:

$$L_{\text{split}} = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \tag{12}$$

If all $L_{\text{split}} \le 0$, stop generating subtrees, then this round of training ends, otherwise continue to split subtrees. When the number of repeated training iterations reaches K, the training ends.

4.2 Identification Process

Aiming at the identification task of surrounding rock grade in tunnel construction by drill jumbo, this paper proposes an intelligent identification method of surrounding rock grade combined drilling parameters of drill jumbo with XGBoost algorithm. Firstly, collect the drilling parameters of multiple channels in real time by the drill jumbo data acquisition system. Then, select strong correlation parameters screened out by Pearson correlation coefficient method: feed speed, strike pressure, propulsion pressure, rotary pressure, water pressure, water flow rate to form the feature vector of surrounding rock grade recognition. Next, preprocess the data recorded during tunnel construction, remove the data of pre-drilling stage and the abnormal data under the normal drilling condition. After that, screen the number of samples and divide the training set and test set. Finally, normalize the data as input to the XGBoost model, the output result of the model is the grade of surrounding rock on the working face of the drill jumbo. The whole intelligent identification process is shown in Fig. 4.

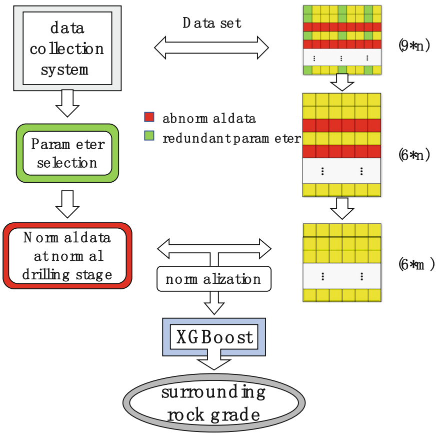

Fig. 4. Intelligent identification flow chart of surrounding rock grade

5 Validation and Analysis

5.1 Dataset Construction

The identification performance of the proposed method was verified by drilling data of Gao Jiaping tunnel with location of 449208 m, 449250 m, 449666 m, 449889 m–449902 m, 450824 m–451060 m and 451700 m–451813 m. After all the sample data were preprocessed, combine the surrounding rock data of grade III, IV and V after screening to form a total data set. The total number of samples was 73450. According to engineering experience, training set and test set were divided by 9:1, that means, 66105 samples were used to construct training set, and 7345 samples were used to construct test set. To test the generalization ability of the proposed model, the training set and the test set should not be in the same hole, so they should be randomly divided when dividing the data set. The specific partition results are shown in Table 4.

Feed speed, strike pressure, propulsive pressure, rotary pressure, water pressure and water flow rate were selected as the input of XGBoost, and surrounding rock grade was used as the classification label of XGBoost classifier. Import sample data of training set into the model for training. After the training, import the data in the test set into the trained model for prediction, and compare the prediction results with the real surrounding rock grade to obtain the performance indexes of XGBoost classifier in the intelligent recognition of surrounding rock grade.

5.2 Model Evaluation Index

Accuracy, Recall, Precision and F-1 score are four commonly used evaluation indexes in supervised learning tasks, and their calculation formulas are as follows:

$$Accuracy = \frac{1}{n}\sum\nolimits_{i=1}^{n} I\left(y_i^{\text{predict}} = y_i^{\text{actual}}\right) \tag{13}$$

$$Recall = \frac{1}{m}\sum\nolimits_{i=1}^{m} \frac{TP_i}{TP_i + FN_i} \tag{14}$$

$$Precision = \frac{1}{m}\sum\nolimits_{i=1}^{m} \frac{TP_i}{TP_i + FP_i} \tag{15}$$

$$F = \frac{1}{m}\sum\nolimits_{i=1}^{m} \frac{2P_iR_i}{P_i + R_i} \tag{16}$$

Table 4. The composition of the sample library (9:1)

grade	training set	test set	total
III	21879	2431	24310
IV	23697	2633	26330
V	20529	2281	22810
total	66105	7345	73450

where n represents the total number of samples, m represents the number of labeled species, y_i^{predict} is the grade of surrounding rock predicted by the model, y_i^{actual} is the actual grade of surrounding rock, TP_i represents the number of correctly predicted i-grade surrounding rock samples, FN_i is the number of i-grade surrounding rock samples predicted to be other types, FP_i is the number of other types of rock samples predicted to be i-grade surrounding rock. Accuracy is a common index in classification problems, that is, the proportion of correctly classified samples to total samples. However, accuracy often cannot effectively assess model performance over extremely unbalanced data sets. In a balanced dataset, the accuracy and recall are equal. This study considers four parameters as model evaluation indexes to validate the identification effectiveness of the proposed method.

5.2.1 Results and Discussion

The performance of the trained model to identify the test set is shown in Table 5. From the data in the table we can see that the average prediction accuracy of this method is very high, and the other three performance indexes are close to 1, which indicates that the proposed method can be used for the intelligent identification task of surrounding rock grade of drill jumbo tunnel construction.

In order to confirm the superiority of proposed method in the surrounding rock classification identification task, The identification performance of six intelligent identification algorithms including SGDC, LSVC [20], KNN, MLP, integrated classification model RFC and XGBoost were compared on the test set. Select the same drilling parameters as the input of the model, and train all methods for 10 times. Figure 5 shows the mean values of 4 indexes of all algorithms. From the figure we can see that XGBoost classifier get the highest value in all four indicators, which proves the effectiveness and superiority of XGBoost algorithm for classification of tunnel surrounding rock.

The confusion matrix of the six algorithms on the test set is shown in Fig. 6. The x-axis represents the rock grade predicted by the trained model and the y-axis represents the true surrounding rock grade. The correct prediction data for each type of surrounding rock is the most to XGBoost, which indicates that XGBoost not only has higher overall prediction accuracy than other algorithms, but also has the highest identification accuracy for each type of surrounding rock, indicating that XGBoost has the best comprehensive performance in surrounding rock identification task of drill jumbo.

To further evaluate generalization effectiveness and robustness of the proposed method in tunnel surrounding rock grade recognition task, this paper combines the

Table 5. Prediction results of XGBoost classifier

	Precision (%)	Recall (%)	F-1 score (%)	Accuracy (%)	total
III	99.79	99.71	99.75	99.69	2431
IV	99.66	99.51	99.58		2633
V	99.61	99.87	99.74		2281
Macro average	99.69	99.70	99.69		7345

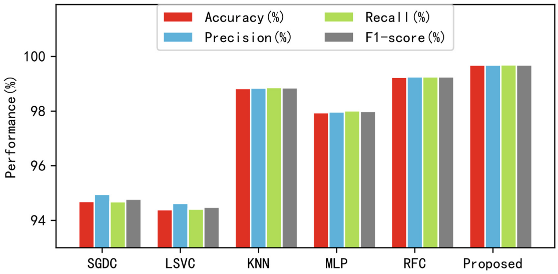

Fig. 5. Performance indicators of SGD, LSVC, KNN, MLP, RF and XGBoost classifiers

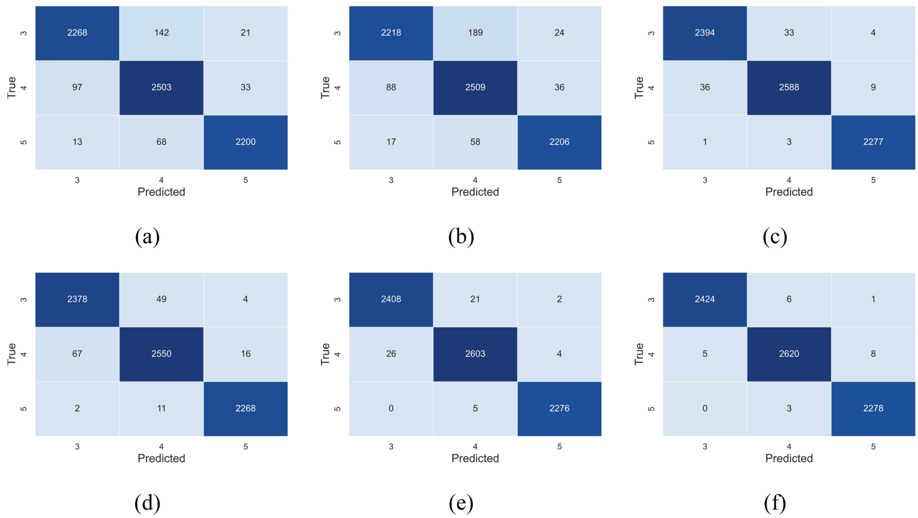

Fig. 6. The confusion matrix of each algorithm on the test set: (**a**) SGD; (**b**) LSVC; (**c**) KNN; (**d**) MLP; (**e**) Random Forest; (**f**) XGBoost classifier.

II-grade surrounding rock sample data from the Alianqiu tunnel, and constructs 5 data sets of tunnel surrounding rock grade identification with different orders of magnitude of sample data. Eventually test the comprehensive performance of six algorithms in all the data sets. Table 6 shows the sample quantity of 5 data sets and the division of training set and test set.

5 data sets represent:

DS1 is the original data of the drill jumbo during the construction of Gao Jiaping tunnel. The total amount of data is 302436, and the partition ratio is 9:1. The characteristics of DS1 are that the sample data are not cleaned and the amount of sample data of different types of surrounding rock varies greatly. DS2 is the data set used to test the identification performance of the proposed method before. The total data was 73450, and the partition ratio is 9:1. DS2 is characterized by the fact that abnormal data have been deleted and there is little difference in the number of samples of grade III, IV and V surrounding rocks, so it is a relatively high-quality data set. On the basis of DS2, DS3

Table 6. The division of Training set and Test set

Data set	surrounding rock grade	training set samples	test set samples	total
DS1 (9:1)	3	124200	13800	301070
	4	49662	5518	
	5	97101	10789	
DS2 (9:1)	3	21879	2431	73450
	4	23697	2633	
	5	20529	2281	
DS3 (9:1)	2	19701	2189	95540
	3	21672	2408	
	4	23166	2574	
	5	21447	2383	
DS4 (8:2)	2	168	42	800
	3	152	38	
	4	176	44	
	5	144	36	
DS5 (8:2)	2	1124	281	5000
	3	1044	261	
	4	896	224	
	5	936	234	

integrates borehole data of II-grade rock of Alianqiu Tunnel, and the volume of II-grade rock sample data is similar to that of IV-grade surrounding rock of Gao Jiaping tunnel. The total data was 95,540, and the partition ratio is 9:1. DS3 is characterized by the integration of the II-grade surrounding rock data of another project and the addition of label types. Because the surrounding rock category may not only be grade III, IV or V in the construction process, the trained classification model can adapt to the engineering practice of more scenarios. Moreover, there is little difference in the sample data of the four types of surrounding rocks. DS4 is composed of 800 sample data randomly selected from the DS3, and the partition ratio 8:2. There are210 II-grade rock samples, 190 III-grade rock samples, 220 IV-grade rock samples and 180 V-grade rock samples in DS4. DS5 is composed of 5000 sample data randomly selected from the DS3, and the partition ratio is 8:2. There are 1405 II-grade rock samples, 1305 III-grade rock samples, 1120 IV-grade rock samples and 1170 V-grade rock samples in DS5.

Similarly, train all models and test for 10 times on these 5 datasets, and take the mean values as the final performance results. The recognition performance of the six classifiers on the test set is shown in Fig. 7.

Firstly, by comparing the prediction performance of each algorithm on DS1 and DS2, it can be seen that sample data screening helps to improve significantly the recognition

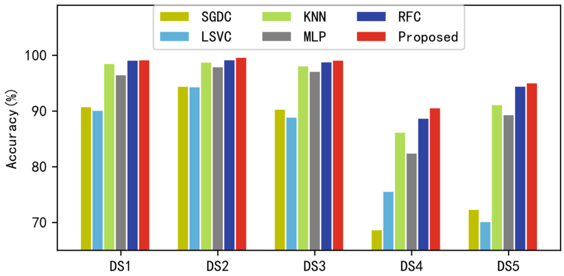

Fig. 7. Prediction accuracy of SGD, LSVC, KNN, Random Forest and Proposed method on different data sets

accuracy of intelligent classifier. Secondly, from the prediction performance of each classifier in DS2, DS4 and DS5, the number of samples in the dataset has impact on the prediction accuracy, but the XGBoost classifier can still maintain the recognition accuracy of more than 90% in the case of small sample data. Thirdly, the recognition rate of the proposed method reaches 99.17% on the DS3 which combines surrounding rock samples from different tunnel projects. It further verifies the comprehensive ability of the proposed method. Finally, the prediction accuracy of the proposed method in 5 data sets is more than 90%, which indicates that compared with the existing algorithm, it can be better qualified for the intelligent classification task of tunnel surrounding rock during the construction of drill jumbo.

6 Conclusions

Based on the research on the identification of surrounding rock grade in tunnel construction, this paper proposes an intelligent identification method of surrounding rock grade combined drilling parameters with machine learning algorithm XGBoost. Using Gao Jiaping tunnel engineering data to verify the superiority of this method. The main conclusions and contributions made are as follows:

Comprehensively analyze and excavate the correlation between rock grade and drilling parameters, and fully consider the strong correlation between water pressure, water flow rate and surrounding rock grade. Finally, select six drilling parameters as the input of the intelligent classification model, compared with the conventional method of using four drilling parameters, the identification accuracy of surrounding rock grade is greatly improved.

Establish an intelligent recognition model of surrounding rock grade based on XGBoost. Using 5 data sets to evaluate the comprehensive effectiveness of the proposed method on surrounding rock recognition tasks under different conditions, and the comprehensive comparison was made with existing models. The results show that the recognition effectiveness of the proposed method is superior to those of existing algorithm models, and the recognition accuracy of surrounding rock along the tunnel can reach 99.68%. In the Alianqiu tunnel data set, the proposed method still achieves high recognition accuracy, which further verifies the robustness and generalization ability of the proposed method.

Acknowledgments. This paper was supported by the National Key R & D Plan of China (Grant No. 2019YFB1705203) and Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0102).

Authors' Contributions. Conceptualization, G.H. and C.Q.; methodology, C.Q. and J.T.; software, R.W. and G.H.; validation, G.H., C.Q. and J.T.; formal analysis, C.Q.; investigation, C.Q.; resources, G.H., R.W., C.Q. F.L and J.L; data curation, F.L and J.L; writing—original draft preparation, G.H., R.W. and C.Q.; writing—review and editing, G.H. and C.Q.; visualization, R.W.; supervision, X.X.; project administration, C.Q., J.T. and C.L.; funding acquisition, C.Q.

References

1. Ma, X.: Application of controlled-source audio-frequency magnetotelluric method for classification of surrounding rock in deep and large tunnels in tunnel investigation. In: Petroleum Engineering Committee of Chinese Petroleum Society (2020). https://doi.org/10.26914/c.cnkihy.2020.050150
2. Wu, Z., Zhang, J.: Geological structure of karst tunnel and classification of surrounding rock. J. Railw. Eng., 6–12 (2012)
3. Zhang, L., Qiu, D., Li, S., Zhang, D.: Study on tunnel surrounding rock classification based on rough set and ideal point method. Rock Soil Mech., 171–175 (2011). https://doi.org/10.16285/j.rsm.2011.s1.083
4. Li, C.: Study on surrounding rock classification and monitoring measurement of New Austrian Tunnelling in Dabie Mountain. Wuhan University of Technology (2009)
5. Liu, Z., Feng, F., Wang, J., Hua, H., Wang, R.: Application of fuzzy comprehensive evaluation method in quality classification of mine rock mass. J. Wuhan Univ. Technol., 129–134 (2014)
6. Chen, A.: Study on expert system of roadway tunnel surrounding rock classification. Chongqing University (2013)
7. Xue, X., Zhang, J., Yao, G.: Study on classification method of loess tunnel surrounding rock based on extension theory. J. Undergr. Space Eng., 651–657 (2017)
8. Rong, Y., Shun, B., Shun, Y.: Application of advanced geological forecast in dynamic grading prediction of tunnel surrounding rock. J. Chongqing Jiaotong Univ. (Nat. Sci. Ed.), 18–23 (2017)
9. Yao, M.: Research on intelligent classification method of tunnel surrounding rock based on drilling parameters of rock drilling rig. Southwest Jiaotong University (2020). https://doi.org/10.27414/d.cnki.gxnju.2020.000812
10. Yao, M.: Study on intelligent classification of surrounding rock based on drilling parameters and SVM. Sichuan Architecture, 149–151 (2020)
11. Wen, T., Yu, F., Shao, L., Tian, Y.: Classification of tunnel surrounding rock based on GA-SVM. Highway Transp. Technol., 63–70 (2018)
12. Kumar, B.R., Vardhan, H., Govindaraj, M., et al.: Artificial neural network model for prediction of rock properties from sound level produced during drilling. Geomech. Geoeng., 53–61 (2013). https://doi.org/10.1080/17486025.2012.661469
13. Tian, H., Li, S., Xue, Y., Qiu, D., Su, M., Wang, K.: Identification of tunnel tuff stratum interface and surrounding rock classification method based on drilling energy theory. Rock Soil Mech., 2457–2464 (2012). https://doi.org/10.16285/j.rsm.2012.08.026
14. Yang, Z., Xue, X.: Study on GRNN model of tunnel surrounding rock classification. Western Explor. Eng., 161–162 (2018)
15. Li, L.: Design and simulation of automatic levelling system for drilling arm of drilling rig. Hebei Normal University of Science and Technology (2021). https://doi.org/10.27741/d.cnki.ghbkj.2021.000071

16. Cao, J.: Research on application of novel method in Gaojiaping tunnel construction. Shijiazhuang Tiedao University (2019). https://doi.org/10.27334/d.cnki.gstdy.2019.000363
17. Xu, L., Li, Y., Cheng, C.: Determination of deformation failure type and grade of surrounding rock in highway tunnel. J. Chongqing Jiaotong Univ., 16–20 (2002)
18. He, G., Wang, S.: Research and development of surrounding rock classification method for railway tunnel. In: Proceedings of Chinese Railway Engineering Geology (2005)
19. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACMSIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016)
20. Ding, S., Qi, B., Tan, H.: A review of support vector machine theory and algorithms. J. Univ. Electron. Sci. Technol. China, 2–10 (2011)