Intelligent Identification of Coal-Rock Type Based on Boring Parameters of Dig Windlass and XGBoost

Guoqiang Huang, Chengjin Qin(✉), Ruihong Wu, Jianfeng Tao, and Chengliang Liu

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

Abstract. Coal is an important natural resource in China and plays an essential role in the development of industry and national economy. To realize unmanned mining, it is necessary to identify coal-rock type of working face accurately and efficiently. As the photographing is interfered by water mist, dust, air flow, lighting, vibration and other factors, the accuracy of image feature recognition methods are seriously affected. Therefore, this paper proposes an intelligent identification method based on boring parameters of dig windlass and XGBoost algorithm. Firstly, the coupling relationship between machine parameters recorded by dig windlass was analysed to remove a large number of redundant parameters, which reduces 22.7% training time of the model and 63.8% identification time. Secondly, remove the data recorded by the dig windlass under abnormal working conditions. Then, construct a model based on XGBoost algorithm and input the selected parameters and data into the model for training. Finally, the validity of the proposed method is verified by the data collected from the Project of Chai Jiagou Coal Mine in Tongchuan. The results show that the accuracy rate of coal rock type identification is more than 98% even when the training data is very little and the abnormal data under normal working condition is kept, which confirms the effectiveness and strong robustness of the proposed method.

Keywords: Boring parameters · Coal and rock identification · XGBoost · Dig windlass

1 Introduction

Coal is China's main energy source, which is basis to ensure the rapid development of the national economy [1]. At present, China's coal mining is mainly underground mining. By 2008, the fully mechanized mining degree of former state-owned key coal mines has reached 84.15%, and the per unit yield of working face has reached 53474 t per day. However, the development of comprehensive excavating machine is relatively slow, and the comprehensive excavating rate only increases from 1.71% in 1980 to 36.85% in 2008. The ratio of mining face (mining/excavation) has been maintained at about 1:3.1 in recent years [2]. At the same time, it appears gradually an annual output of

L. Yan and J. Na (Eds.): ICFPMCE 2022, AHE 10, pp. 162–178, 2023.
https://doi.org/10.2991/978-94-6463-022-0_15

several million tons as fully mechanized mining technology develops, or even ten million tons of super working face, so that the annual consumption of the number of mining roadway increased greatly. Therefore, roadway tunnelling has become the common and key technology of efficient and intensive production in coal mines.

Coal and rock identification technology is an essential part to realize mining automation. Firstly, reliable coal and rock identification technology can ensure that dig windlass always cut the soft coal seam, reduce the loss of equipment and the amount of maintenance; Secondly, it can ensure that the dig windlass will not damage the roof when driving along the roof and bottom plate, so as to reduce the risk of accidents; Thirdly, reduce the time for workers to identify coal seams and increase the speed of tunnelling. Field workers will be affected by a variety of factors at the mining site, and it's easy to cause misidentification, cutting head or motor damage and other faults. Finally, reduce the labor intensity of coal miners. As is known to all, the coal mine environment is not suitable for long stay and a little error from workers is likely to cause person casualties and other major accident. It is often reported in the news that some mines explode and many workers are buried and die. Considering from the perspective of personal safety, less or even no humanization is the inevitable development trend of coal mining industry in the future. Therefore, reliable coal and rock identification system contributes to improve economic benefits and ensuring mining safety.

Many scholars have made some research achievements on coal and rock identification of fully mechanized mining face. In the early stage, researchers mainly focused on feature analysis of coal and rock images collected at fully mechanized mining face, and then applied intelligent machine learning algorithm to train and realize coal and rock recognition at working face. Zhang et al. [3], based on the differences between the basic characteristics of coal and rock, proposed two types of image grey level "similarity" metric estimation model and hierarchical cluster recognition model for coal and rock boundary, which provides a new idea for coal and rock recognition. Si et al. [4] use improved U-NET network model to achieve coal and rock image recognition, with the average pixel accuracy of 95.81%, achieving high precision coal and rock recognition at working face. Xiong [5] took coal and rock images of mining face as the research object and applied deep learning to coal and rock recognition, achieving wonderful results. Zhang [6] implicates transform domain and Gaussian mixture model clustering with the help of image processing technology and feature extraction method. The content and texture information of coal and rock images are extracted by discrete cosine transform and discrete wavelet transform respectively, and the feature vectors are classified and identified by Gaussian mixture model clustering, which improves the recognition accuracy.

In addition to the image method, other recognition methods have also been proposed. Wei et al. [7] took hyper spectrum as the technical means of coal and rock identification. They collected the hyperspectral data of coal and rock, and designed an algorithm to recognize the rock type through the extraction of hyperspectral characteristic bands, with the recognition accuracy up to 91.3%. Zhang [8] proposed to apply the least square model and the robust extended local binary mode incorporating smooth filtering idea to the field of coal and rock identification, and the average recognition accuracy could reach 93.21% after selecting the best parameters. Lu et al. [9] used trial cutting saw blade to cut coal and rock, and analysed the change of hydraulic cylinder pressure during coal and rock cutting to identify coal and rock. This approach provides a new idea for the design of intelligent dig windlass. Meng [10] collected the reflectance spectral curve of samples through the ground object spectrometer. Took the reflectance spectral data of the samples and the composition content of the samples as independent variables, and the sample category as dependent variables to build coal and rock classification model of coal and rock classification, choose kernel principal component analysis combined with SVM model for the average recognition rate can reach 95%. Tian et al. [11] proposed a multi-sensor recognition method using data fusion theory. Four pin shaft sensors with equivalent strength treatment are arranged at the rocker arm and the pin shaft of the connecting frame, and the combined criterion of multi-sensor data fusion is used to identify the coal-rock interface. This method is complicated but still achieved some achievements.

However, Zhang [12] proposed that the production site of working face is an irregular superposition field in which various factors such as sound, light, electricity, magnetism, heat, wind, dust, water, fog and force intermix and influence each other. Any coal and rock identification scheme implemented in roadway should consider the actual situation of the site. In the complex superposition field of mining face, any identification method based on sensitive physical factors is not reliable. Although some progress has been made in image recognition, it cannot meet the actual needs.

Therefore, in order to remedy the above-mentioned shortcomings, this paper proposes an intelligent identification method of coal rock type based on the boring parameters of dig windlass and XGBoost algorithm. Firstly, extract feature parameters, 48 parameters recorded by dig windlass are fully analysed, non-state parameters such as pitch angle and heading angle of dig windlass are removed, and screen out effective parameters reflecting the driving state of dig windlass. Secondly, the correlation matrix between state parameters is analysed and a large number of redundant parameters are removed in the state parameters, eventually leave 10 independent effective characteristic parameters as the input of the subsequent model, which will greatly shorten the training time of the subsequent model. Then, based on the electrical working principle of dig windlass [13, 14], the sample data with zero current value of cutting motor and zero current value of scraper motor are cleared, which indicates that the dig windlass is in abnormal working state. Finally, the 10 filtered parameters are taken as the input of the model, and construct the data set. The results show that the accuracy of type identification on the test set can reach 100%. Considering the small number of training samples in practical engineering, this paper also extracted samples of different orders of magnitude from data sets recorded in different periods and constructed four data sets to test the validity, strong generalization and strong robustness of the proposed method.

2 Materials

2.1 Project Introduction

The data used in this paper come from the tunnel construction site of Chai Jiagou Coal Mine in Tongchuan. Located in Taian Town, Yijun County, Tongchuan City, Shaanxi Province. The Chai Jiagou Coal Mine is part of the Yulong Coal Field, with industrial

Fig. 1. Appearance of dig windlass

reserves of 21.57 million tons and an annual production capacity of 900,000 tons. At present, the coal mine has used a highly automated dig windlass machine, as shown in Fig. 1.

The highly automatic and intelligent dig windlass has realized the automatic acquisition, processing and output of boring data. At present, the data recorded by the dig windlass used in Chai Jiagou Coal mine include 48 parameters including date, cutting current, scraper current, etc. Under normal working conditions, more than 300,000 tunnelling data can be recorded in a day, which provides an important condition for coal rock type identification research of fully mechanized coal face based on boring parameters.

Table 1 shows some records of data information table of dig windlass in Tongchuan Chai Jiagou Coal Mine. It can be found that from February 22 to March 29, the fully-mechanized mining face of the dig windlass machine was all rock, while from April 1 to April 28, the fully-mechanized mining face was almost all coal seam. Therefore, the boring data of these periods provide the boring data of rock stratum and coal seam for the identification of coal and rock type of comprehensive driving face.

2.2 Introduction of Dig Windlass

Dig windlass is mainly composed of four core parts: cutting section, temporary support, hydraulic bolt drilling rig group, automatic feeding network and steel belt device. According to the function of the dig windlass is divided into six areas, namely cutting area, support area, timely anchorage area, stable supplementary support area, control area and supply area. To achieve one excavation, one protection, one network, one cooperative operation. The support area, timely anchorage area and stable supplementary support area consist of 7 jacking anchor arms and 6 side anchor arms [15].

The quick excavation process of dig windlass is as follows: safety inspection → unit cutting, loading, coal transport → temporary support → permanent support. The construction process includes cutting process, bolt support process and cable construction process. The traditional shearer bolt support construction and so on are completed by manual, but the boring and anchoring process of intelligent dig windlass are automatically completed.

Table 1. Data information recorded by dig windlass in Chai Jiagou Coal Mine

Date	Broken teeth	Amount of feed	Description
February 22	13	12	rock layer
February 23	10	8	rock layer
…	…	…	…
April 3	2	12	coal seam
April 4	0	10	coal seam
April 5	0	6	coal seam

3 Methodology

3.1 Parameter Selection

3.1.1 Parameter Introduction

The parameters collected by dig windlass are mainly from the real-time monitoring system of dig windlass attitude includes heading Angle, pitch Angle, pitch Angle deviation and heading Angle deviation; Driving system includes host voltage, three-phase current of cutting motor, three-phase current of oil pump motor, three-phase current of left and right load motor, three-phase current of left and right scraper motor, cutting power, pump power, scraper power. etc. Anchor drill arm control system includes fault code, cut PID protection, PID regulation, boom propulsion PWM, boom retracting PWM, boom telescopic feedback, lifting PID protection, PID regulation, boom lifting PWM, boom lowering PWM, boom lifting feedback. State detection system includes construction date, cutting temperature, oil pump temperature, left and right scraper temperature, scraper tensioning pressure, cutting gear oil temperature, cutting gear oil pressure, cutting height, cutting depth.

Three-phase current value of cutting motor, oil pump motor, left and right loading motor, left and right scraper motor are measured by the ammeter on the corresponding circuit [16]. The cutting motor is the power source of the cutting mechanism of dig windlass, which drives the cutting head to rotate and cut through the reducer. The cutting capacity of dig windlass is determined by cutting power, which is directly related to cutting motor current. The load on the cutting head changes with the change of the hardness of the cutting palm surface. The mechanical load on the cutting head is transferred to the output shaft of the motor in the form of resistance torque. The motor is operating under load, and the change of load will cause the change of the cutting motor current. Therefore, it is important to investigate the relationship between cutting motor current and coal rock type identification. Similarly, the oil pump motor drives the variable pump to operate and provide hydraulic oil for the whole hydraulic system. The greater the load, the greater the pressure of hydraulic oil. Therefore, it is very important to study the working current of the oil pump motor to identify coal and rock intelligently. The loading conveyer is used to collect and transport the cut coal to the scraper conveyer, which carries the coal out of the roadway. Scraper conveyor is a conveyer with scraper chain traction, conveying bulk material in the trough. Only scraper transport function to maintain continuous operation, can production be carried out normally. Otherwise, the whole coal mining face will appear to stop production, so that the entire production interruption. Voltage: coal mine power consumption is 1140 V, there will be a slight fluctuation in the work, the value is measured by the voltmeter in the circuit. Cutting temperature, oil pump temperature, left and right scraper temperature: these four parameters refer to the motor temperature during the operation of the corresponding motor. The motors used in coal mines are equipped with PT100 thermistor, and the motor temperature is calculated by measuring the resistance value of the thermistor. Before scraper conveyor works, a hydraulic cylinder is required to exert pretightening force, and the scraper tension pressure is obtained by measuring the oil pressure of the hydraulic cylinder through the pressure sensor. Excessive tension stress is easy to damage the scraper conveyor and it can't work normally with too small tension stress as well. Cutting gear oil temperature refers to the reducer internal lubricating oil temperature, measured by the temperature sensor installed in the reducer. Cutting gear oil pressure refers to the external to the reducer when the internal oil pressure, its value is measured by the pressure sensor. Cutting power, oil pump power, left and right scraper power are calculated by voltage, current and a variety of power factors.

The parameters recorded by the anchor and drill arm control system and the attitude real-time monitoring system of the dig windlass are to understand the driving state, driving direction and cutting head position of the dig windlass for the operator. The cutting height is the distance between the cutting head and the bottom plate, and the cutting depth is to record the feed of the cutting head, so as to accurately detect the position information of the heading head. GUC1000 mining linear displacement sensor is installed in the cutting head telescopic cylinder, cutting head lifting cylinder, cutting head horizontal rotary cylinder, etc. Considering the importance of position information of lifting cylinder and horizontal rotary cylinder to control, two lifting cylinders and two horizontal cylinders are equipped with linear displacement sensors. The heading Angle and pitch Angle are used to determine the driving direction of the windlass, and their values are measured by the gyroscope installed on the dig windlass. The rest of the parameters are used for detection and output feedback to protect the windlass system and prevent accidents.

3.1.2 Parameter Selection

First of all, during the construction of the dig windlass, due to the serious water seepage in the coal mine, the hydraulic pump of the loading conveyor was not driven by motor, which causes that the current value recorded by the loading motor was zero, so the three-phase current value of the left and right loading motor will not be considered. Secondly, the coal-rock type of the working face has nothing to do with the driving position and direction of the dig windlass, so the parameters recorded by the control system of the anchor and drill arm and the real-time monitoring system of the attitude of the windlass will not be considered.

After removing above parameters, the remaining 26 parameters still had redundant parameters. Pearson linear correlation coefficients between all parameters were calculated and listed as a matrix, as shown in Fig. 2.

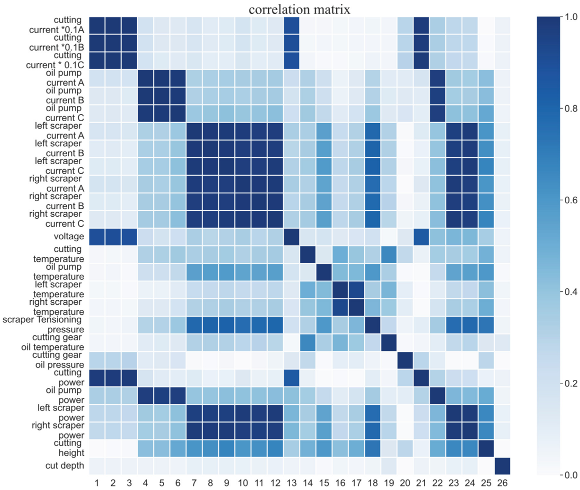


Fig. 2. Correlation coefficient matrix diagram

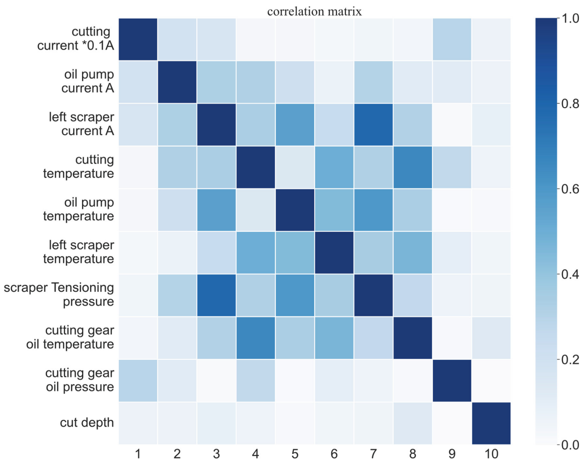


Fig. 3. Correlation coefficient matrix diagram after screening

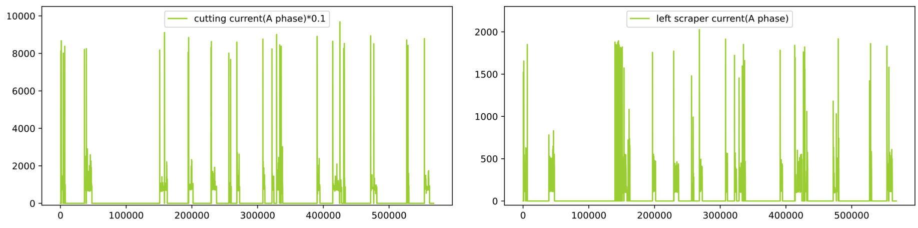
It can be seen from the figure that some of the squares are darker in color, indicating that there is a strong correlation between some boring parameters, which indicates that the parameters still contain redundant information. If the 26-dimensional parameters are directly introduced into the model, it is difficult for the model to extract accurate geological information from the operating parameters, which affects the recognition accuracy and prolongs the training time. First of all, there is a strong linear correlation between the three-phase current of each motor, and the load of the left scraper motor and the right scraper motor are the same, so the current value also changes synchronously. Therefore, only the A-phase current of the motor is selected as an independent characteristic parameter. At the same time, when the internal resistance of the motor is constant, there is a linear mapping between the motor current and the host voltage and motor power, so selecting the motor current as the input can remove the motor power and host voltage and other parameters. All parameters with a correlation coefficient exceeding 0.85 will be considered redundant parameters, and only one of these parameters will be retained after screening.

So ultimately select A-phase current of cutting motor, A-phase current of oil pump motor, A-phase current of left scraper motor, cutting temperature, oil pump temperature, left scraper temperature, scraper tension pressure, cutting gear oil temperature, cutting gear oil pressure, cutting depth as the characteristic parameters of coal and rock identification to input the subsequent intelligent identification model. The correlation coefficient matrix of these ten parameters is shown in Fig. 3. As can be seen from the figure, except for the squares on the diagonal, the colors in other places are very light, which indicates that the correlation degree between all parameters is low, thus achieving the purpose of eliminating characteristic redundant parameters. Therefore, these ten parameters can fully reflect the information of comprehensive excavation surface reflected by boring data.

3.2 Boring Data Preprocessing

According to the coal mine data statistics log, the initial data set of coal seam and rock layer is formed and the corresponding label is marked on the data, rock layer is 0, coal seam is 1.

Firstly, the oil pump motor will be turned on to provide oil for hydraulic pipeline. Then the cutting motor will be turned on for trial operation. After that, the scraper conveyor and the loading conveyor will be turned on. Only when all parts are ready, can coal mining begin. Cutting motor provides power for cutting drum, is the main power source of cutting coal. The loading conveyer is used to gather the cut coal and transport it to the scraper conveyer by which the coal is carried out of the roadway. Therefore, there is a time gap between the start of the oil pump motor, the cutting motor and the scraper conveyor, and the recorded data will have a lot of data with zero cutting current or zero scraper current. Take the data recorded in Chai Jiagou on March 1, 2021 as an example. The excavation sections on that day were all rock layer, and the data recorded the operation data of the dig windlass at every moment of the day. However, the construction status of the dig windlass was not recorded, such as stopped work, standby or normal work. In the process of coal cutting cycle in a day, the dig windlass will be in standby or stop state for many times in the middle of the process, as shown in Fig. 4(a), (b). Both cutting current and scraper current will have a trough, and zero means that no excavation will be carried out when the dig windlass is in shutdown state. In the pre-processing of boring parameters, it is necessary to remove the sample data recorded when the dig windlass does not work. Partial data of the filtered dataset are shown in Table 2.



(a) (b)

Fig. 4. **(a)** the distribution of cutting current. **(b)** the distribution of left scraper current.

Table 2. Filtered data set (part)

number	cutting current *0.1A	oil pump current A	left scraper current A	Cutting temperature	oil pump temperature	left scraper temperature	scraper Tensioning pressure	cutting gear oil temperature	cutting gear oil pressure	cut depth	label
1	600	56.6	10.7	123	57	36	16	56	1	0	0
2	600	58.5	10.7	123	57	37	16	56	1	0	0
3	615	54.6	9.7	123	57	36	16	56	1	0	0
4	615	54.6	10.7	123	57	36	16	56	1	0	0
5	615	54.6	10.7	123	57	36	16	56	1	0	0
6	615	54.6	11.2	123	57	36	16	56	1	0	0
…	…	…	…	…	…	…	…	…	…	…	…
218718	703	66.4	12.2	71	68	34	17	53	0	65535	1
218719	703	66.4	11.2	71	68	34	17	53	0	65535	1
218720	703	66.4	11.2	71	68	34	17	53	0	65535	1
218721	688	65.4	12.2	70	68	33	17	53	0	65535	1

4 The Proposed Method

4.1 Principle of XGBoost

XGBoost, which stands for eXtreme Gradient Boosting, is a powerful algorithm. It was proposed by Chen Tianqi [17] in 2016. XGBoost is an optimized distributed gradient enhancement library with many advantages. In addition, XGBoost algorithm based on tree model is more suitable for processing low-dimensional data. Therefore, this paper adopts XGBoost algorithm as coal and rock recognition algorithm. The XGBoost algorithm flow is as follows:

XGBoost can be stated as Eq. (1):

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

where $g_i(x)$ is the basic classifier model generated during each iteration, K is the maximum number of iterations. In each iteration, the optimization objective function for training the basic classifier can be defined as Eq. (2):

$$L^{(t)} = \sum\nolimits_{i=1}^{n} L\Big(y_i, \hat{y}_i^{(t-1)} + g_t(x_i)\Big) + \gamma T + \frac{1}{2}\lambda \sum\nolimits_{j=1}^{m} w_j^2 \tag{2}$$

where x_i represents the sample data in training set, y_i represents the label correspond to x_i in training set, t means the current iterations, $\hat{y}_i^{(t-1)}$ is the prediction result of strong classifier during t-1 iterations, and T stands the number of leaf nodes, and w_j stands the weight of the jth leaf node.

Use Taylor's expansion to expand optimization objective function Eq. (2) to obtain Eq. (3).

$$L^{(t)} \approx \sum\nolimits_{i=1}^{n} \left[L\left(y_i, \hat{y}^{t-1}\right) + \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) f_t(x_i) + \frac{1}{2} \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) f_t^2(x_i) \right] + \gamma T + \frac{1}{2}\lambda \sum\nolimits_{j=1}^{m} w_j^2 \tag{3}$$

The optimal weight solution w_j^* is shown in Eq. (4), and the optimal loss function $L^{(t)*}$ is shown in Eq. (5):

$$w_j^* = -\frac{\sum_{i \in I_j} \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)}{\sum_{i \in I_j} \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) + \lambda} \tag{4}$$

$$L^{(t)*} = -\frac{1}{2} \sum\nolimits_{j=1}^{T} \frac{\left(\sum_{i \in I_j} \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)\right)^2}{\sum_{i \in I_j} \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) + \lambda} + \gamma T \tag{5}$$

where $I_j = \{i|q(x_i) = j\}$. In order to generate the optimal decision tree in each round of training, the evaluation function should be defined to minimize the loss function value when generating left and right subtrees IL and IR. Define evaluation function as Eq. (6).

$$L_{\text{split}} = \frac{1}{2} \left[\frac{\left(\sum_{i \in I_L} \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)_i\right)^2}{\sum_{i \in I_L} \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) + \lambda} + \frac{\left(\sum_{i \in I_R} \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)\right)^2}{\sum_{i \in I_R} \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) + \lambda} - \frac{\left(\sum_{i \in I} \partial_{y_i^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right)\right)^2}{\sum_{i \in I} \partial^2_{y_t^{(t-1)}} l\left(y_i, y_i^{(t-1)}\right) + \lambda} \right] - \gamma \tag{6}$$

When each node generates the left and right subtrees, the feature and eigenvalue splitting subtrees that maximize L_{split} need to be selected. When all $L_{split} \leq 0$, subtree generation is stopped and this round of training ends. Repeat the above steps until the maximum number of iterations K is reached, and the whole training ends.

4.2 Identification Process

Take 48 channel data recorded by the data acquisition system of each part of dig windlass as the original data. Firstly, remove the parameters irrelevant to the dig windlass boring system according to the parameter types, and then further remove the redundant

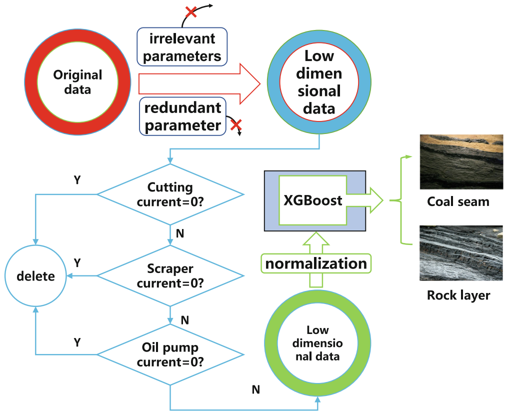


Fig. 5. Intelligent identification flow chart of coal-rock type

parameters according to the correlation coefficient matrix of the remaining parameters to complete the extraction of characteristic parameters, so the original high-dimensional data is transformed into low-dimensional data. Then, according to the cutting current, scraper current and oil pump current value to judge the working state of the dig windlass. Only when the values of the three are not 0, it indicates that the windlass is in working state. Therefore, the sample data with 0 value in these three parameters are removed, and the mean and standard deviation of the training set are used to standardize the recorded data to complete the data processing process. Finally, input the processed low-dimensional data into the XGBoost model, and the output of the model corresponds to the coal rock type of fully mechanized mining face. 0 represents coal seams and 1 represents rock strata. The intelligent identification flow chart of coal-rock type is presented in Fig. 5.

5 Validation and Analysis

5.1 Dataset Construction

This task uses the boring data recorded by the dig windlass in the construction of Chai Jiagou Coal Mine on March 5, April 13 and April 14. After data preprocessing, they are combined together to form a total data set DS1, with a total number of 218,721 samples. According to engineering experience, training sets and test sets are divided into 4:1, that is, 174,976 samples are used as training sets, and 43,745 samples are used as test sets. In order to enhance the generalization ability of the model, the data sets of training set and test set should not be in the same time period, so the training set and test set should be randomly divided when dividing the data set. The specific partition results are shown in Table 3.

Select the screened characteristic parameters of A-phase current of cutting motor, A-phase current of oil pump motor, A-phase current of left scraper motor, cutting temperature, oil pump temperature, left scraper temperature, scraper tension pressure, cutting gear oil temperature, cutting gear oil pressure, cutting depth as the input of XGBoost. 0

Table 3. The composition of the sample library

Data set	Rock layer	Coal seam	total
training set	94637	80339	174976
test set	23813	19932	43745
total	118450	100271	218721

and 1 serve as the output of the XGBoost classifier. Import the training set sample data to train the model. After the training is completed, import the data of the test set into the trained classification model for identification. Compared the identification results with the actual coal and rock categories to obtain various performance indicators of XGBoost classifier.

5.2 Model Evaluation Index

In order to evaluate the identification effectiveness of the model, introduce the accuracy rate to evaluate the model. The calculation formula of accuracy is as follow:

$$Accuracy = \frac{1}{n}\sum\nolimits_{i=1}^{n} I\left(y_i^{\text{predict}} = y_i^{\text{actual}}\right) \tag{7}$$

where n represents the total number of samples, m means the number of sample categories, y_i^{predict} is the sample point category predicted by the model, y_i^{actual} sample points to the actual category. In addition, the training time of the model and the recognition time on the test set are also taken as evaluation indexes. In the construction site, there is a high demand for model recognition speed. The faster the identification speed, the higher the efficiency of coal roadway excavation. The total time of ten times of model training and ten times of test were taken as evaluation indexes. The calculation formula are as follows

$$Training\ time = time1 - time0 \tag{8}$$

$$Test\ time = time2 - time1 \tag{9}$$

where time0 represents the beginning moment of model training, time1 represents the end moment of ten training sessions, and time2 represents the end moment of ten test sessions.

5.3 Results and Discussion

The results of XGBoost classifier tests on the test set are shown in Table 4. It can be seen from the data in the table that the recognition accuracy of XGBoost classifier is 100%. What should be mentioned is, the data processing process only clears the non-working data, but still retains the abnormal data in the working state, which indicates that the

Table 4. Prediction results of XGBoost classifier

	Precision (%)	Recall (%)	F-1score (%)	Accuracy	support
0	100	100	100	100	23813
1	100	100	100		19932
macro avg	100	100	100		43745

Table 5. Comparison of parameters before and after screening

	Training time(s/10)	Test time(ms/10)	Accuracy (%)
Before parameter screening	40.17	225	100
After parameter screening	51.94	622	100

proposed method is not sensitive to abnormal data, and the trained model has strong robustness, which is very helpful to the actual construction.

In order to verify the effectiveness of parameter screening, the recognition accuracy, training time and testing time of the model were compared in the case of redundancy parameter elimination and retention of all characteristic parameters. The specific results are shown in Table 5. The identification accuracy of the model does not decrease after the parameters screening, because only redundant parameters are removed, the rest parameters already include all effective information of the comprehensive excavation surface. More importantly, the training time of the model after parameter screening decreased from 51.94 s per 10 times to 40.17 s per 10 times, decreasing by 22.7%. The recognition time on the test set decreased from 622 ms per 10 times to 225 ms per 10 times, with a decrease of 63.8%. For the construction site, the reduction of identification time by 63.8% will greatly improve the construction efficiency and reduce the construction cost. Therefore, the validity and necessity of eliminating redundant parameters are verified.

Since the above results were obtained when the sample size reached more than 200,000, the sample size of the training set was large. In order to evaluate the generalization ability and comprehensive performance of XGBoost classifier in coal-rock type recognition task of dig windlass, this paper constructed four data sets with different orders of magnitude of sample data, and tested the comprehensive performance of XGBoost algorithm in the training sets with little sample data. Table 6 shows the sample quantity of four data sets and the division of training set and test set.

Among them, DS1 data set is the data set used for the above analysis, with a total number of 218721 samples. Training set and test set are divided into 4:1, characterized by large data volume. DS2 is a data set composed of 300 sample data randomly selected from screened coal seam data and screened rock strata data in DS1, with a total of 600 samples. The training set and test set are divided into 4:1, characterized by small sample data. DS3 is a data set composed of 5000 samples randomly selected from screened coal seam data and screened rock strata data in DS1 respectively, with a total of 10000 samples. The training set and test set are divided according to 9:1, characterized by

Table 6. The division of Training set and Test set

Data set	label	Number of training set samples	Number of test set samples
DS1 (4:1)	0	94637	23813
	1	80339	19932
DS2 (4:1)	0	236	64
	1	244	56
DS3 (9:1)	0	4501	499
	1	4499	501
DS4 (4:1)	0	7955	2045
	1	8045	1955

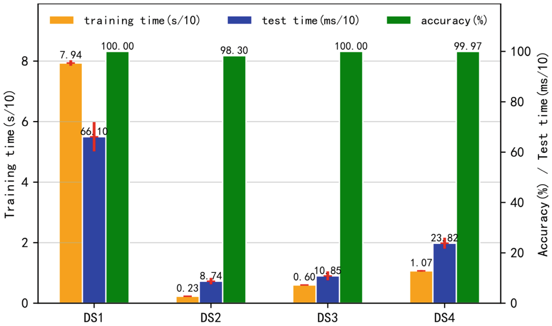


Fig. 6. Identification results of XGBoost classifiers on different data sets

medium sample data. DS4 is a data set composed of 5000 rock strata data on March 1 and 5000 coal seam data on April 23 on the basis of DS3. The source of DS4 is different from DS1, so they have different data characteristics.

Train the proposed method for 10 times in each dataset and take the means and standard deviations of test indexes as the final results. The test results after training with XGBoost classifier are shown in Fig. 6. As can be seen from the bar chart, the prediction accuracy of XGBoost classifier on the four data sets is over 98%, and it can still maintain a very high recognition rate with little training data. In addition, the recognition accuracy of the DS4 data set reached 99.98%. The strong generalization of XGBoost classifier in coal and rock identification of dig windlass is verified.

6 Conclusion

This paper fully studied the identification of coal and rock types in coal roadway construction site, and proposed an intelligent identification method of coal and rock types based on boring data of dig windlass and XGBoost algorithm. Based on the on-site data collected in Tongchuan Chai Jiagou Coal Mine for verification, the main conclusions and contributions are as follows:

(1) Comprehensively analyse and excavate the correlation between boring parameters, and extract independent effective features from the high-dimensional boring parameters to effectively eliminate redundant information in the features. Ensure that the identification accuracy does not decrease, while reducing 22.7% training time of the model and 63.8% identification time of the coal rock type; Establish a coal and rock type recognition model based on the boring parameters of dig windlass and XGBoost algorithm, and accurately identify the coal and rock types of the comprehensive face of dig windlass by using low-dimensional feature input. With the method proposed in this paper, the recognition accuracy can still reach 100% when there are abnormal data in the data set. The effectiveness and strong robustness of the proposed method are verified.
(2) Use four datasets to evaluate the comprehensive performance of the proposed method on coal and rock identification task under different conditions, including data sets from different sources. The results show that the identification accuracy rate of coal and rock along the roadway is more than 98%. The strong generalization ability of the proposed method is further verified.
(3) Most coal and rock type identification tasks are based on image feature recognition of the images collected at fully mechanized mining face. Currently, no researchers have conducted in-depth research on the boring parameters of dig windlass. Therefore, this is a new identification method of coal and rock, which provides a new solution for coal roadway construction site.

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. 2021 SHZDZX0102).

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. and C.Q.; data curation, G.H.; writing—original draft preparation, G.H., R.W. and C.Q.; writing—review and editing, G.H. and C.Q.; visualization, R.W.; supervision, C.Q., J.T. and C.L.; project administration, C.Q., J.T. and C.L.; funding acquisition, C.Q.

References

1. Xue, P., Xue, B., Zhang, C., Zhang, B., Bai, Z.: Composing the movement of Leaping development with responsibility–Huozhou Coal power Group Lvlin Energy Chemical Co., LTD. Pangpangta coal mine construction safety and efficient mine record. Brand (9), 54–63 (2010). https://doi.org/10.19373/j.cnki.14-1384/f.2010.09.011
2. Zhao, X.: Research on load and coal rock determination and identification technology based on multi-sensor information fusion. In: China University of Mining and Technology (Beijing) (2011)

3. Zhang, W., Fu, Y., Liu, X.: Research on coal and rock recognition algorithm and self-learning model. In: Coal Mine Safety, 2021, pp. 132–136 (2021). https://doi.org/10.13347/j.cnki.mkaq.2021.08.022
4. Si, L., Wang, Z., Xiong, X., Tan, C.: Coal and rock identification method of fully mechanized mining face based on improved U-NET network model. J. Coal 1–12 (2021). https://doi.org/10.13225/j.cnki.jccs.2020.1011
5. Xiong, X.: Research on coal rock identification method of fully mechanized mining face based on deep learning. In: China University of Mining and Technology (2020). https://doi.org/10.27623/d.cnki.gzkyu.2020.000891
6. Zhang, T.: Coal and rock identification method based on transform domain and Gaussian mixture model clustering. In: Coal Technology, 2018, pp. 320–323 (2018). https://doi.org/10.13301/j.cnki.ct.2018.11.120
7. Ren, W., Xu, L., Meng, X., Wu, J., Zhang, K.: Coal and rock identification method based on hyperspectral characteristic absorption peak. In: Spectroscopy and Spectral Analysis, 2021, pp. 1942–1948 (2021)
8. Zhang, R.: Discussion on coal rock identification method based on LSM and RELBP. In: Computing Technology and Automation, 2021, pp. 109–113 (2021). https://doi.org/10.16339/j.cnki.jsjsyzdh.202101021
9. Lu, Z., Gao, G., Wang, Z., Ma, Y.: Analysis and design of coal shearer for coal rock identification. In: Coal Mine Machinery, pp. 9–11 (2021). https://doi.org/10.13436/j.mkjx.202102003
10. Meng, X.: Experimental research on coal rock identification method based on visible-near-infrared spectroscopy. In: Anhui University of Science and Technology (2020). https://doi.org/10.26918/d.cnki.ghngc.2020.000080
11. Tian, L., Dai, B., Wang, Q.: Coal rock identification method based on multi-strain data fusion of shearer rocker pin. J. China Coal Soc. 1203–1210 (2020). https://doi.org/10.13225/j.cnki.jccs.2019.0364
12. Zhang, Q.: Discussion on solving the technical problems of coal and rock identification. In: Coal Technology, 2020, pp. 183–185. (2020). https://doi.org/10.13301/j.cnki.ct.2020.08.053
13. Mao, K.: Research and application of Electric System of roadheader based on standardization. In: Contemporary Chemical Industry Research, 2021, pp. 83–84 (2021)
14. Li, M.: Problems and improved design of scraper conveyer of roadheader. In: Coal Engineering, 2016, pp. 134–136 (2016)
15. Feng, W.: Research on the application of fast driving technology of anchor digger. In: Energy and Energy Conservation, 2021, pp. 194–195 (2021). https://doi.org/10.16643/j.cnki.14-1360/td.2021.05.081
16. He, Y., Tian, M., Song, J., Feng, J., Dong, Z.: Rock hardness identification based on cutting motor current of roadheader. In: Coal Mine Safety, 2020, pp. 159–162 (2020). https://doi.org/10.13347/j.cnki.mkaq.2020.12.032
17. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794 (2016). https://doi.org/10.1145/2939672.2939785