

Development and application of a floor failure depth prediction system based on the WEKA platform

Yao Lu^{1,3}, Liyang Bai², Juntao Chen^{*1,3}, Weixin Tong^{1,3} and Zhe Jiang^{1,2}

¹College of Energy and Mining Engineering, College of Safety and Environmental Engineering, Shandong University of Science and Technology, Qingdao 266590, China

²Lvliang University, Lvliang, 033000, Shanxi, China

³State Key Laboratory of Mining Disaster Prevention and Control Co-Founded by Shandong Province and the Ministry of Science and Technology, Shandong University of Science and Technology, Qingdao 266590, China

(Received July 5, 2020, Revised August 20, 2020, Accepted September 7, 2020)

Abstract. In this paper, the WEKA platform was used to mine and analyze measured data of floor failure depth and a prediction system of floor failure depth was developed with Java. Based on the standardization and discretization of 35-set measured data of floor failure depth in China, the grey correlation degree analysis on five factors affecting the floor failure depth was carried out. The correlation order from big to small is: mining depth, working face length, floor failure resistance, mining thickness, dip angle of coal seams. Naive Bayes model, neural network model and decision tree model were used for learning and training, and the accuracy of the confusion matrix, detailed accuracy and node error rate were analyzed. Finally, artificial neural network was concluded to be the optimal model. Based on Java language, a prediction system of floor failure depth was developed. With the easy operation in the system, the prediction from measured data and error analyses were performed for nine sets of data. The results show that the WEKA prediction formula has the smallest relative error and the best prediction effect. Besides, the applicability of WEKA prediction formula was analyzed. The results show that WEKA prediction has a better applicability under the coal seam mining depth of 110 m~550 m, dip angle of coal seams of 0°~15° and working face length of 30 m~135 m.

Keywords: floor failure depth; WEKA platform; the grey relational degree; optimal model; prediction system

1. Introduction

Mining activities inevitably result in the stress redistribution and the fracture failure of rock mass in underground engineering (Liu *et al.* 2020, Genis 2018, Sun *et al.* 2019). This failure greatly changes the permeability of surrounding rock, resulting in mine accidents in engineering. In order to prevent roof water gushing or roof water inrush during coal mining, filling mining is a common and effective method (Wang *et al.* 2020, Jiang *et al.* 2020, Zhang and Meng 2019). Mining-induced water inrush is one of the five major disasters in mining production (Bukowski 2011, Polak *et al.* 2016, Sato *et al.* 2000). The mining-induced water inrush is a coupled process of stress, failure, water pressure and other factors (Wu *et al.* 2018, Narain *et al.* 2010, Schäfer and Teschauer 2001).

In the field measurement of the failure depth of coal seam floor, the P-wave velocity and borehole leakage are used as inversion parameters to reflect the stress state, structural change and failure degree of floor rock mass through CT detection method and double-end water plugging device (Zhang *et al.* 2018). Through the actual

observation and analyses in the water inrush site, the characteristics of water inrush hazard have been revealed with relevant data, contributing to evaluate the risk of underground mining water inrush (Huang *et al.* 2016, Si *et al.* 2021).

In the process of tunneling, the problem of water inrush is also faced (Blümling *et al.* 2007, Butron *et al.* 2017). Fernandez and Moon (2010) proposed an analytical solution that uses mathematical derivation to evaluate the flow of groundwater flowing into the tunnel based on the hydraulic coupling effect. Eren *et al.* (2015) investigated the effect of horizontal in situ stress on failure mechanism around underground openings excavated in isotropic, elastic rock zones. Zhou *et al.* (2015) established an optimal classification method based on grey system theory (GST) and applied it to accurately predict the occurrence probability of water inrush in karst tunnels. In view of the complex disaster-causing mechanism and difficult quantitative predictions of water inrush, several theoretical methods have been adopted to realize dynamic assessment of water inrush in the progressive process of tunnel construction (Li *et al.* 2015). Yuan *et al.* (2016) presented a modified grey clustering method to systematically evaluate the risk of water inrush in karst tunnels.

Entering a new era of safe and green coal mining, the prevention and control of water inrush from coal floor requires more innovative theories and key technologies. Database mining is a comprehensive science containing

*Corresponding author, Ph.D.
E-mail: polariscjt@126.com

Table 1 Original data

N	Working face number	Mining depth/m	Dip/ degree	Thickness/m	Face length/m	Floor resistance failure potential	Floor failure depth/m
1	Huaibei mine 1031	474	8	4.5	183	0.7	16.2
2	Xingtai mine 7607	320	4	5.4	60	0.6	9.7
3	Xiezhuang mine 3414	750	20	2.1	90	0.8	13.1
4	Jingxing mine 4707	400	9	4	34	0.4	6
5	Xingdong mine 2121	1000	10	3.75	150	0.8	32.5
6	Jingxing mine 5701(2)	227	12	3.5	30	0.4	7
7	Wucun mine 32031(2)	640	13	1.5	196	0.6	26.4
8	Zhaoge mine 1237(2)	1000	30	2	200	0.6	38
9	Zhucun mine 54002	210	4	1.3	102	0.8	5
10	Hebi mine 128	230	26	3.5	180	0.4	20
11	Geting mine 11601	300	9	2	40	0.5	6
12	Xinzhuangzi mine 4303(2)	310	26	1.8	128	0.2	29.6
13	Suncunmine11121(east)	1018.6	18.9	2	125	0.2	33
14	Wangfeng mine 1951	123	15	1.1	100	0.2	13.4
15	Xingtai mine 7802	259	4	3	160	0.6	16.4
16	Fengfengtowmine2701(1)	145	16	1.5	120	0.4	14
17	A coal mine 7608	400	6	5.39	80	0.7	10.2
18	Fengfengfour mine 4804	110	12	1.4	100	0.4	10.7
19	Jingxingmine470 (little) 1	400	9	7.5	34	0.4	8
20	Baizhuang mine 7406	225	14	1.9	130	0.8	9.75
21	Chenghetwomine 22510	300	8	1.8	100	0.4	10
22	Shuanggou mine 1208	287	10	1	130	0.6	9.5
23	Fengfengthreemine3707	130	15	1.4	135	0.4	12
24	Wucun mine 32031(1)	375	14	2.4	70	0.6	9.7
25	Zhaoge mine 1237(1)	900	26	2	200	0.6	27
26	Jingxing mine 4707 (big)	400	9	4	45	0.4	6.5
27	Jingxingthree mine 5701(1)	227	12	3.5	30	0.4	3.5
28	Chensilou mine 21301	584	10	2.7	149	0.5	14
29	Shuanggou mine 1204	308	10	1	160	0.6	10.5
30	Liangzhuang 51101W	640	15	1.5	165	0.2	20.1
31	Wangfeng mine 1830	123	15	1.1	70	0.2	7
32	Wucun mine 3305	327	12	2.4	120	0.6	11.7
33	Caozhuang mine 9203	148	18	1.8	95	0.8	9
34	Caocun 11-014	200	10	1.6	100	0.2	8.5
35	Xinzhuangzi mine 4303(1)	310	26	1.8	128	0.2	16.8

various techniques and complex subjects. Data mining has become a hot spot of artificial intelligence research (Bai *et al.* 2017, Camoes *et al.* 2017, Jimenez *et al.* 2018, Kaveh *et al.* 2018).

This paper collects the relevant geological data about the failure depth of coal seam floor in China in recent years, comprehensively analyzes the influence of mining height, mining depth, mining thickness, coal seam dip angle, fault and floor failure resistance on the failure depth of coal seam floor. Through data mining and analysis on the WEKA platform, the optimal prediction model is obtained. Depending on the optimal model, the relationship between

the floor failure depth and the influencing factors is calculated. Finally, the prediction system of floor failure depth is developed using Java language, and the practical application is verified.

2. Processing of WEKA sample data

WEKA is a data mining and machine learning software that integrates data pre-processing, learning algorithms and evaluation. WEKA includes many classical learning and training methods, such as Bayes belief network, naive

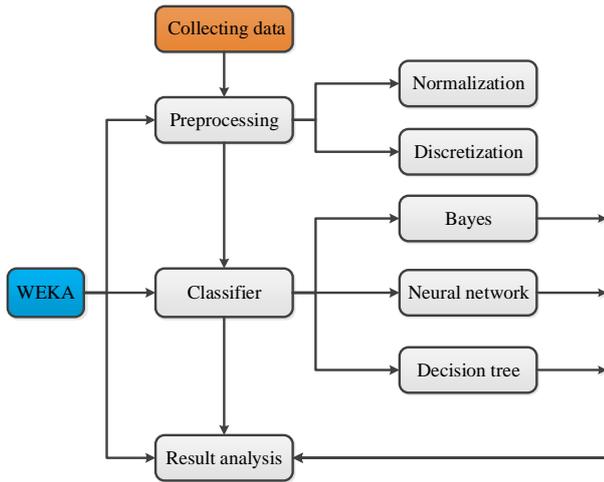


Fig. 1 Data processing flow chart

Table 2 Training samples

N	A1	A2	A3	A4	A5	B	N	A1	A2	A3	A4	A5	B	N	A1	A2	A3	A4	A5	B
1	2	1	2	3	3	2	10	1	1	3	1	3	2	19	1	1	1	3	3	2
2	1	1	3	1	3	1	11	1	1	3	1	1	1	20	2	2	1	3	1	3
3	3	2	1	2	3	2	12	1	1	1	2	1	2	21	1	2	1	1	1	1
4	3	1	2	3	3	3	13	1	1	1	2	3	1	22	1	1	1	2	3	2
5	2	2	1	3	3	3	14	1	2	1	2	1	2	23	1	2	1	2	3	1
6	1	3	2	3	1	3	15	1	2	1	1	3	1	24	1	1	1	2	1	1
7	1	1	1	1	2	1	16	3	3	1	3	3	3	25	1	3	1	2	1	2
8	1	3	1	2	1	3	17	1	1	2	1	1	1							
9	3	2	1	2	1	3	18	2	1	1	3	2	2							

Bayes network, SVM (support vector machine), C4.5 decision tree, multi-layer feedforward artificial neural network.

Because of the repeatability, noisy and high dimensionality of original data, it is necessary to preprocess the data to remove duplication and noisy data. The data processing is shown in Fig. 1. The data is first collected and preprocessed, and then trained with different algorithms. At the final, the results are analyzed.

Based on the measured data of floor failure in China, 44 sets of original data samples of floor failure are selected, including 6 main influencing factors such as mining depth, dip angle, mining thickness, working face length, floor failure resistance and faults. Due to the incomplete fault data, the fault influence on the failure depth is removed. The remaining five factors are taken as the research object. The modeling is established using 35 sets of data, as shown in Table 1 (Bai *et al.* 2017).

From Table 1, it can be seen that the mining depth of coal seam is mainly between 100 and 1100 m. The coal seam is mainly composed of the near horizontal seam and gently inclined seam. From the view of thickness, it is mainly composed of the thin seam and medium-thick seam. The original data are normalized, so that their values are concentrated between 0~1, and then the effective data are better retained. Labels of A1, A2, A3, A4, A5 and B are marked for the mining depth, dip angle, mining thickness,

working face length, floor failure resistance and failure depth, respectively. The normalization formula is shown as follows (Bai *et al.* 2017):

$$G_{ij} = \frac{X_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)} \quad (1)$$

where X_{ij} is the sample before normalization, G_{ij} is the sample after normalization, $\min(X_j)$ is the minimum of the original sample, $\max(X_j)$ is the maximum in the original sample.

The discretization in WEKA includes supervised and unsupervised discretization of numerical attributes, which are used to discretize some numerical attributes in the dataset into classification attributes. The normalized isometric 0~1 is divided into three parts. In this paper, A1 (mining depth), A2 (dip angle of coal seam), A3 (mining thickness), A4 (working face length) and A5 (failure resistance of floor) are discretized. The floor failure depth is also divided into three parts, namely 0~10 m is represented by 1 (low failure grade), 10~20 m by 2 (medium failure grade), >20 m by 3 (high failure grade). Then the repeated data can be found from the discretized data. The discretization equation is shown as follows:

$$L_{ij} = \begin{cases} Q = \frac{\max(G_j) - \min(G_j)}{3} \\ 0, & \min(G_j) < G_{ij} < \min(G_j) + Q \\ 1, & \min(G_j) + Q < G_{ij} < \min(G_j) + 2Q \\ 2, & \min(G_j) + 2Q < G_{ij} < \max(G_j) \end{cases} \quad (2)$$

where L_{ij} is the discretized sample, $\max(G_j)$ is the maximum normalized sample data, $\min(G_j)$ is the minimum normalized sample data, and Q is the step size.

After the discretization, the data are divided into three sections -inf-0.333333, 0.3333-0.66666667 and 0.666667-inf and then processed in Excel. Then repeated results are obtained, including set 18 and set 21; set 9 and set 22; set 15 and set 29; sets 4, 6, 26 and set 27; sets 14, 16 and set 23; set 20 and set 33; set 8 and set 28. Therefore, 10 sets of data such as sets 4, 6, 8, 9, 14, 15, 16, 18, 20 and 26 are removed, and the remaining 25 sets are taken as training samples, as shown in Table 2.

3. The optimal prediction model of WEKA

3.1 Grey relational analyses

Since there are complex factors affecting the floor failure depth, the grey correlation analysis, which can analyze the small sample data, is performed to obtain the main and minor factors. Furthermore, five factors are analyzed, including mining depth, dip angle, mining thickness and working face length and the failure resistance of the floor.

Firstly, the initial values of 25 sets of data are transformed. The mining depth, dip angle, mining thickness, working face length and floor failure resistance are expressed as $X_0^{(0)}(k)$, $X_1^{(0)}(k)$, $X_2^{(0)}(k)$, $X_3^{(0)}(k)$, $X_4^{(0)}(k)$

Table 3 The transformation results of initial value

Working face number	A1	A2	A3	A4	A5	Floor failure depth
Huaibei mine 1031	1.000	1.000	1.000	1.000	1.000	1.000
Xingta imine 7607	0.599	0.675	0.500	1.200	0.328	0.857
Xiezhuang mine 3414	0.809	1.582	2.500	0.467	0.492	1.143
Xingdong mine 2121	2.006	2.110	1.250	0.833	0.820	1.143
Wucun mine 32031(2)	1.630	1.350	1.625	0.333	1.071	0.857
Hebi mine 128	1.235	0.485	3.250	0.778	0.984	0.571
Geting mine 11601	0.370	0.633	1.125	0.444	0.219	0.714
Xinzhuangzimine4303(2)	1.827	0.654	3.250	0.400	0.699	0.286
Suncun mine 11121(east)	2.037	2.149	2.363	0.444	0.683	0.286
A coal mine 7608	0.630	0.844	0.750	1.198	0.437	1.000
Jingxing mine 4707 (little) 1	0.494	0.844	1.125	1.667	0.186	0.571
Chenghe two mine 22510	0.617	0.633	1.000	0.400	0.546	0.571
Zibo Shuanggou mine 1208	0.586	0.605	1.250	0.222	0.710	0.857
Fengfeng three mine 3707	0.741	0.274	1.875	0.311	0.738	0.571
Wucun mine 32031(1)	0.599	0.791	1.750	0.533	0.383	0.857
KailuanZhaoge mine1237(1)	1.667	1.899	3.250	0.444	1.093	0.857
Jingxing three mine 5701(1)	0.216	0.479	1.500	0.778	0.164	0.571
Chensilou mine 21301	0.864	1.232	1.250	0.600	0.814	0.714
Shuanggou mine in Zibo 1204	0.648	0.650	1.250	0.222	0.874	0.857
Liangzhuang 51101W	1.241	1.350	1.875	0.333	0.902	0.286
Handan Wangfeng mine 1830	0.432	0.259	1.875	0.244	0.383	0.286
Wucun mine 3305	0.722	0.690	1.500	0.533	0.656	0.857
Caozhuang mine 9203	0.556	0.312	2.250	0.400	0.519	1.143
Caocun 11-014	0.525	0.422	1.250	0.356	0.546	0.286
Xinzhuangzi mine 4303(1)	1.037	0.654	3.250	0.400	0.699	0.286

and $X_5^{(0)}(k)$ before the transformation. The transformation results are shown in Table 3.

The equation for calculating the change of initial values is shown as (Liu *et al.* 2010):

$$x^{(1)}(k) = \frac{x^{(0)}(k)}{x^{(0)}(1)} (x^{(0)}(1) \neq 0, k = 1, 2, 3, \dots, n) \quad (3)$$

The equation for grey correlation coefficient is shown as:

$$r(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{0i}(k) + \rho\Delta_{\max}} \quad (4)$$

where $\Delta_{0i}(k) = |x_0(k) - x_i(k)|$ is the absolute difference, $\Delta_{\min} = \min_i \min_k \Delta_{0i}(k)$ is the minimum difference of two grades, $\Delta_{\max} = \max_i \max_k \Delta_{0i}(k)$ is the maximum difference of two grades. $\rho \in (0, 1)$ is the variation coefficient, which is 0.5 in general. It is used to prevent data distortion caused by excessive absolute difference, enhancing the difference of correlation coefficients.

After calculating the correlation coefficient of each factor, the arithmetic mean value is calculated as follows:

$$r = \frac{1}{n} \sum_{k=1}^n r(x_0(k), x_i(k)) \quad (5)$$

Through calculation, it can be concluded that the minimum difference of two grades is 0, the maximum difference of two grades is 2.212963. The grey correlation degrees of mining depth, coal seam dip angle, mining thickness, working face length and floor failure resistance are 0.83, 0.61, 0.70, 0.82 and 0.74, respectively. Therefore, the influence of five factors on the floor failure depth is ordered as mining depth>working face length>floor failure resistance>mining thickness>coal seam dip.

3.2 Comparative analyses of training results

WEKA includes four methods of model evaluation, namely the Training set evaluation, Supplied test evaluation, Cross-validation and Percentage split. The application of training set or supplied test set aims to select a set of instances from data samples for testing. The cross-validation can decompose data into N copies, starting from the first datum to the end of N copies. In this paper, the training set evaluation is used. Accuracy analysis is performed by three models of Naive Bayes, neural network

Table 4 Confusion matrix

Naive Bayes			Neural network			Decision tree					
a	b	c	classified as	a	b	c	classified as	a	b	c	classified as
7	2	0	a=1	6	3	0	a=1	6	3	0	a=1
3	5	1	b=2	0	9	0	b=2	1	8	0	b=2
0	1	6	c=3	0	1	6	c=3	0	3	4	c=3

Comparing the confusion matrix with Classifier error, the highest classification accuracy is obtained by the artificial neural network model, 84%.

Table 5 Detailed accuracy of naive Bayes

Kappa	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	Class
	0.778	0.188	0.700	0.778	0.737	0.578	1
0.577	0.556	0.188	0.625	0.556	0.588	0.379	2
	0.857	0.056	0.857	0.857	0.857	0.802	3
Average weight	0.720	0.151	0.717	0.720	0.717	0.569	

Table 6 Detailed accuracy of artificial neural networks

Kappa	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	Class
	0.667	0.000	1.000	0.667	0.800	0.749	1
0.757	1.000	0.250	0.692	1.000	0.818	0.721	2
	0.857	0.000	1.000	0.857	0.923	0.901	3
Average weight	0.840	0.090	0.889	0.840	0.841	0.781	

Table 7 Detailed accuracy of decision tree

Kappa	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	Class
	0.667	0.063	0.857	0.667	0.750	0.646	1
0.571	0.889	0.375	0.571	0.889	0.696	0.497	2
	0.571	0.000	1.000	0.571	0.727	0.700	3
Average weight	0.720	0.158	0.794	0.720	0.724	0.607	

and decision tree, from the perspectives of the confusion matrix, detailed precision and node error rate analyses.

3.2.1 Confusion matrix

Confusion Matrix, known as error matrix, is a special matrix to reveal algorithm performance. When a class is mistaken for another, the confusion is presented. Confusion matrix of three algorithms is sorted out. Table 4 shows the results of the confusion matrix.

In this table, a, b and c represent three classes, each row of the confusion matrix represents the actual class, and each column represents the predicted results of the classifier.

From the confusion matrix of Naive Bayes classifier in Table 4, 9 sets of data are low failure grade, among which 2 sets are mistaken for medium failure grade; 9 sets are medium failure grade, among which 3 sets are mistaken for low and 1 set for high failure grade; 7 sets are failure grade, and 1 set is mistaken for medium failure grade. It can be seen that there are 18 correct classification, 7 misclassifications with the accuracy of 72% and error rate of 28%.

From the confusion matrix of the artificial neural

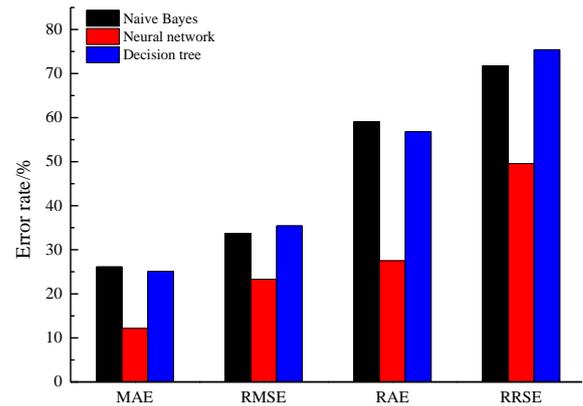


Fig. 2 Node error rate

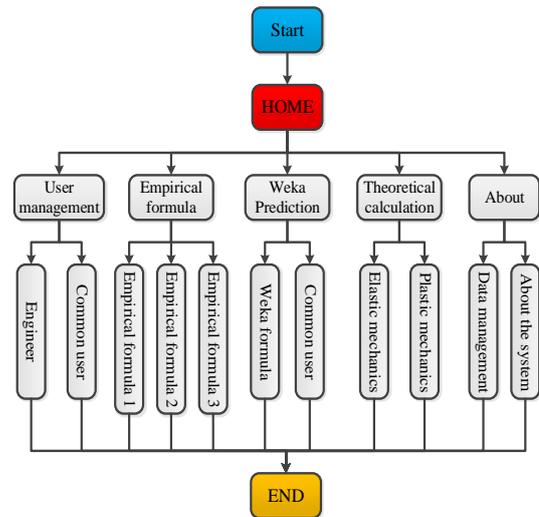


Fig. 3 Process design of the software

network classifier in Table 4, 9 sets of data are low failure grade, among which 3 sets are mistaken as medium; 9 sets are medium, whose predictions are all correct; 7 sets are high, and 1 set is mistaken for medium. There are 21 correct classifications, 4 misclassifications. The accuracy and error rate of classification are 84% and 16%, respectively.

From the confusion matrix of decision tree classifier in Table 4, 9 sets are low failure grade, among which 3 sets are mistaken for medium; 9 sets are medium, 1 set is mistaken for low; 7 sets are high failure grade, among which 3 sets are mistaken for medium. It can be seen that, there are 18 correct classifications, 7 misclassifications. The accuracy and error rate are 72% and 28%, respectively.

3.2.2 Detailed accuracy

Detailed accuracy is mainly embodied in TP Rate, FP Rate, Precision, Recall, F-Measure and MCC (Matthews correlation coefficient) and ROC Area (Receiver Operating Characteristic curve).

Weighted average weight refers to the weighted average of each parameter based on the type proportion in the actual classification.

Kappa statistics range from [-1,1]. When Kappa statistics are closer to 1, it suggests that the classifier has a better predictive effect. If K=1, it means that the predictive

value is identical to the real value.

Detailed accuracy of naive Bayes model, artificial neural network model and decision tree model is shown in Tables 5, 6 and 7.

The following conclusions can be drawn. In terms of the detailed accuracy of Kappa statistics, neural network>Naive Bayes>decision tree; in TP Rate, neural network>Naive Bayes=decision tree. In terms of FP Rate, decision tree>Naive Bayes>neural network. Precision values of decision tree class 3 and neural network model class 3 reach 1, and the prediction results are the best. The average value of the three models is above 0.7, and the maximum value of the neural network model is 0.899. In term of Recall, neural network>Naive Bayes=decision tree, and the value of the neural network model class 2 is 1, indicating it achieves the best effect. In terms of harmonic average, the neural network achieves the best effect, and neural network>decision tree>Naive Bayes; In terms of MCC value, neural network>decision tree>Naive Bayes, the value of Bayes model is 0.569. Comprehensive analyses and comparison of each index are performed. Neural network is the best one, and naive Bayes is better than decision tree.

3.2.3 Node error rate

Node error rate is mainly embodied in four models: mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), root relative square error (RRSE), as shown in Fig. 2.

As shown in Fig. 2, the minimum MAE of three classifier models is neural network model with a value of 12.16%. In terms of RMSE, neural network model has the smallest value of 23.30%. In terms of the RAE, Naive Bayes and decision tree model have the larger value, closing to 60%. The RRSE of the Naive Bayes and the decision tree model are as high as about 70%. In general, the error of the neural network model is relatively low, and the prediction effect is the best.

4. Development of floor failure depth prediction system

4.1 Main flow of the software

Based on Java, the prediction system of floor failure depth is developed. The software design flow chart is shown in Fig. 3. The main functions of the system are as follows:

- (1) Switching between ordinary users and engineers.
- (2) Data addition and deletion through data storage interface.
- (3) The calculation of floor failure depth is realized by empirical formula, WEKA prediction formula and theoretical calculation.

The main flow of the software is shown in Fig. 4.

Prediction System for Floor Failure Depth is launched, and the operation interface is shown in Fig. 4. After the user name and password are successfully verified, the main interface is entered in the system. Six main menus are included in this system, namely Login, User Management, Empirical Formula, WEKA Prediction, Theoretical

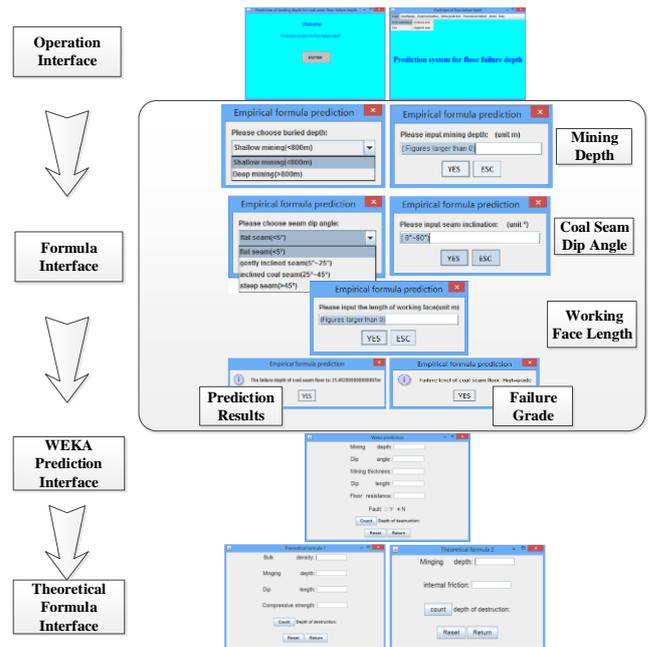


Fig. 4 Flow of the software

Calculation, About and Help. Different functions are implemented in these menus.

4.2 Empirical formula interface

The menu of Empirical Formula includes three sub-menus, namely Empirical Formula 1, Empirical Formula 2 and Empirical Formula 3, which are compiled in accordance with the empirical formulas under three coal mining procedures. The factor of working face length is considered in Empirical Formula 1 and Empirical Formula 2; mining depth, coal seam dip angle and working face length are synthetically examined in Empirical Formula 3. For example, click on Empirical Formula 3 to enter the interface.

In coal mine production practice, according to different mining conditions, including buried depth, inclination angle and so on, the corresponding floor failure depth can be monitored and obtained. That is, different depth or different inclination angle corresponding empirical formula is different, so the corresponding conditions need to be graded.

Firstly, select the mining depth. Then the coal seam mining is divided into Shallow Mining and Deep Mining. Click the bottom of Yes, and input the mining depth in the input box where the input should be greater than 0.

Secondly, select the dip angle. The dip angle of coal seam is divided into Flat Seam ($< 5^\circ$), Gently Inclined Seam ($5^\circ \sim 25^\circ$), Inclined Coal Seam ($25^\circ \sim 45^\circ$) and Steep Seam ($> 45^\circ$). Click Yes, and input the coal seam dip in the input box, which has a prompt of ($0^\circ \sim 90^\circ$).

Thirdly, input working face length, whose unit is meter. which input box with a prompt input greater than 0 number. The selection of working face length is shown in Fig. 4.

Table 8 Analysis of prediction results

Working face	Measured value /m	Calculation of floor failure depth /m				Relative error /%			
		Empirical Formula 1	Empirical Formula 2	Empirical Formula 3	WEKA prediction	Empirical Formula 1	Empirical Formula 2	Empirical Formula 3	WEKA prediction
Magouliang mine 1100	13	13.7	14.0	12.2	10.1	5.04	7.36	6.08	22.14
Huafeng mine 41303	13	13.7	14.0	18.0	9.6	5.04	7.36	38.50	25.94
Dongjiahe mine 507	10.8	13.0	13.4	12.1	7.4	20.44	24.03	11.88	31.36
Xinglongzhuang mine 10302	16	22.3	21.0	22.3	14.3	39.29	31.26	39.68	10.70
Dongpang mine 9103	12.43	8.3	9.1	7.2	3.8	33.55	27.05	42.01	69.36
Bucun mine 9115	10	12.6	13.0	12.1	7.9	25.76	30.18	20.68	20.96
Zhaogu two mine 11050	34.8	20.1	19.3	26.8	19.6	42.16	44.53	23.11	43.68
Bucun mine 9113	7	12.6	13.0	12.8	7.3	79.66	85.98	82.23	4.98
Relative average error						31.36	32.21	33.02	28.64

Click Yes to perform the calculation.

4.3 WEKA prediction interface

Click WEKA Prediction on the main interface, which is shown in Fig. 4. WEKA prediction formula considers the main factors such as mining depth, dip angle, mining thickness, working face length(dip length), with or without faults. The value of floor failure resistance is between 0 and 1. The parameters are input accordingly, and the calculation of floor failure depth is carried out by clicking the Count. Reset is set in the WEKA Prediction Interface to clear the parameters in the input box. The WEKA prediction can be reused for calculations. Click Return to back to the main interface.

4.4 Theoretical formula interface

The main menu of Theoretical Formula includes two sub-menus: Theoretical Formula 1 and Theoretical Formula 2 of Fracture Mechanics and Plastic Mechanics.

Click on Theoretical Formula 1 to enter the left interface shown in Fig. 4. The main parameters include average bulk density of rock (MN/m³), mining depth (m), working face length (m) and compressive strength of rock mass (MPa). Parameters are input into the input frame accordingly.

Click on Theoretical Formula 2 to enter the right calculation interface, as shown in Fig. 4. Two parameters are set in Theoretical Formula 2, namely mining depth and Internal Friction. Input the two parameters in the input box for the calculation.

Click the Count to calculate the floor failure depth of Theoretical Formula 1 or 2. The reset button is set up in the interface of the Theoretical Formula. Click Return to back to the main interface of the system.

5. Application of floor failure depth prediction system

5.1 Comparison of accuracy of different prediction methods

Select the measured data of the floor failure depth, and input the corresponding parameters in the prediction system for the calculation. The measured data are shown in Table 8. After the calculation of the prediction system, the results are summarized and the relative error analysis is performed. The prediction results are shown in Table 8.

As shown in Table 8, the maximum relative average error is 33%, which is calculated by Empirical Formula 3, the minimum is 28.6% calculated by WEKA prediction with the best prediction effect. It shows that the calculation result of this model is closer to the actual error and has higher precision than that of the empirical formula.

5.2 Best use range of WEKA prediction

Import the 35 set of original measurement datas in Table 1 into WEKA to evaluate the prediction results of the system. The X-axes in (a), (b), (c) and (d) in Fig. 5 represent the mining depth (m), dip angle (°), working face length (m) and the failure resistance of floor, respectively. The Y-axes represent the predicted value of the regression model. The cross in the figure represents error at the datum point. The larger the cross, the greater the error.

From the Fig. 5(a), when the mining depth is between 110~550 m, the cross is smaller and the data are more concentrated. When the mining depth is between 550~1000 m, there are more crosses. Therefore, WEKA prediction is more suitable for mining conditions of 110 m~550 m.

From Fig. 5(b), the error between the predicted value and the measured value is smaller when the coal seam dip is between 0°~15°; there are more large crosses when the coal seam dip is between 15°~30°. Therefore, the WEKA prediction model is more suitable for the mining conditions of 0°~15°.

From the Fig. 5(c), when the working face length is 30 ~ 135 m, the cross of data points is smaller. When the working face length is 135~200 m, the prediction accuracy of the predicted value is lower.

From the Fig. 5(d), there is no obvious change in the error in the range of 0.2~0.8 of the failure resistance depth, and the error is not obvious in a certain range. Therefore,

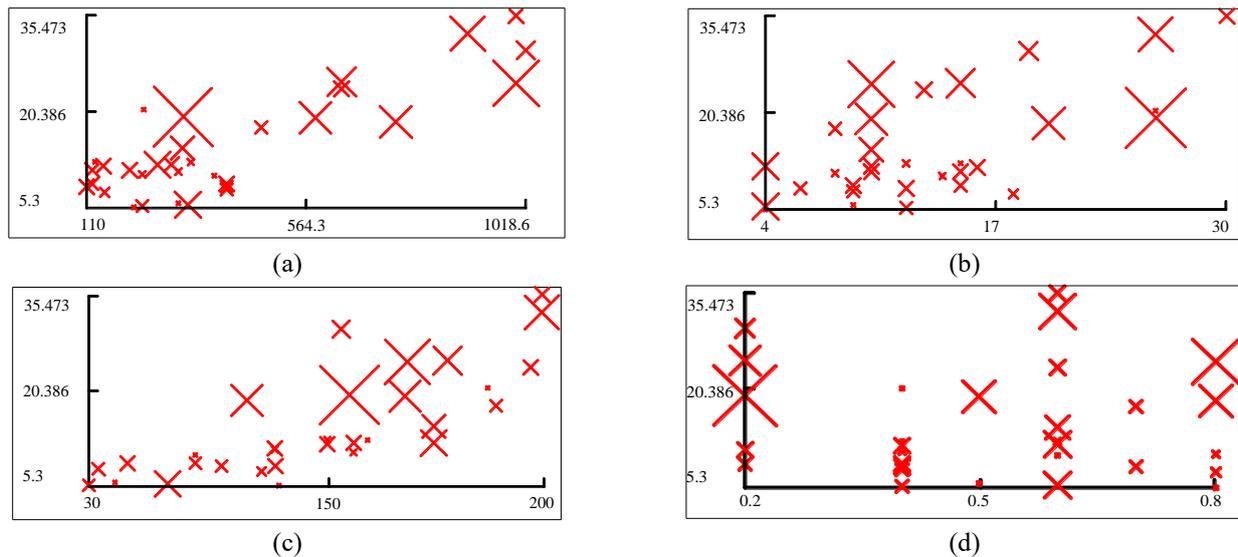


Fig. 5 Analysis of prediction results

6. Conclusions

By de-noising the data of the measured data floor failure depth, the valuable data information is obtained. Through learning training with different models, the optimal model is determined. Based on Java language, the prediction system of floor failure depth is developed and applied.

(1) According to the actual measured data of the floor failure depth in China, the raw data samples of the damage of the base plate of the 44 groups are selected. Since the fault has been removed, the remaining modeling data is 35 sets. After standardization and discretization, 25 groups of valuable data were obtained.

(2) Taking the remaining 25 groups of data as samples, the grey correlation degree of five factors affecting the depth of floor failure in descending order is: mining depth>working face length>floor failure resistance>mining thickness>dip angle of coal seams.

(3) Training samples were trained with naive Bayes, neural network and decision tree, respectively. Besides, the modeling time, confusion matrix, detailed accuracy and node error rate were also trained. Finally, the neural network was verified as the best model.

(4) Based on Java language, the prediction system of floor failure depth is developed, which can realize the switching between ordinary users and engineer users, the increase of data, the deletion and the calculation of floor failure depth under different formulas.

(5) Predictions for 8 sets of measured data were done and analyzed by the system, the prediction results showed that the relative error of WEKA prediction formula was the smallest with the best prediction effect.

(6) The WEKA prediction model had better applicability under the mining conditions of 110 m~550 m coal seam depth, 0° ~ 15° coal seam dip and 30 m~135 m working face length. Because the datas come from the coal mine site in China, the results are more suitable for the coal mine in China, but it also has certain reference value for similar mines abroad.

Acknowledgments

The research described in this paper was financially supported from the funding of the Taishan Scholar Talent Team Support Plan for Advantaged & Unique Discipline Areas and Key research and development plan of Shandong Province (2018GSF117018; 2018GSF120003; 2019GSF111024), Shandong Provincial Natural Science Foundation (ZR2019BEE013) and National Natural Science Foundation of China(51804179; 51604167; 51974173), State Key Laboratory of Mining Disaster Prevention and Control Co-founded by Shandong Province and the Ministry of Science and Technology(MDPC2016ZR01).

References

- Bai, L.Y., Zhao, J.H., Liu, Z.X. and Zhang, Z.X. (2017), "Depth prediction of floor damage based on data mining algorithm", *Coal Eng.*, **49**(6), 92-95. <https://doi.org/10.11799/ce201706027>.
- Blümling, P., Bernier, F., Lebon, P. and Martin, C.D. (2007), "The excavation-damaged zone in clay formations—time-dependent behaviour and influence on performance assessment", *Phys. Chem. Earth*, **32**(8-14), 588-599. <https://doi.org/10.1016/j.pce.2006.04.034>.
- Bukowski, P. (2011), "Water hazard assessment in active shafts in upper Silesian coal basin mines", *Mine Water Environ.*, **30**(4), 302-311. <https://doi.org/10.1007/s10230-011-0148-2>.
- Butrón, C., Gustafson, G., Fransson, Å. and Funehag, J. (2010), "Drip sealing of tunnels in hard rock: A new concept for the design and evaluation of permeation grouting", *Tunn. Undergr. Sp. Tech.*, **25**(2), 114-121. <https://doi.org/10.1016/j.tust.2009.09.008>.
- Camos, A. and Martins, F.F. (2017), "Compressive strength prediction of CFRP confined concrete using data mining techniques", *Comput. Concrete*, **19**(3), 233-241. <http://doi.org/10.12989/cac.2017.19.3.233>.
- Fernandez, G. and Moon, J. (2010), "Excavation-induced hydraulic conductivity reduction around a tunnel - Part 1: Guideline for estimate of ground water inflow rate", *Tunn. Undergr. Sp. Tech.*, **25**(5), 560-566. <https://doi.org/10.1016/j.tust.2010.03.006>.

- Genis, M., Akcin, H., Aydan, O. and Bacak, G. (2018), "Investigation of possible causes of sinkhole incident at the Zonguldak Coal Basin, Turkey", *Geomech. Eng.*, **16**(2), 177-185. <http://doi.org/10.12989/gae.2018.16.2.177>.
- Huang, Z., Jiang, Z., Tang, X., Wu, X., Guo, D. and Yue, Z. (2016), "In situ measurement of hydraulic properties of the fractured zone of coal mines", *Rock Mech. Rock Eng.*, **49**(2), 603-609. <https://doi.org/10.1007/s00603-015-0741-y>.
- Jiang, N., Wang, C., Pan, H., Yin, D. and Ma, J. (2020), "Modeling study on the influence of the strip filling mining sequence on mining-induced failure", *Energy Sci. Eng.*, **8**(6), 2239-2255. <https://doi.org/10.1002/ese3.660>.
- Jiménez, R., Anupol, J., Cajal, B. and Gervilla, E. (2018), "Data mining techniques for drug use research", *Addictive Behaviors Reports*, **8**, 128-135. <https://doi.org/10.1016/j.abrep.2018.09.005>.
- Kaveh, A., Hamzeziabari, S.M. and Bakhshpoori, T. (2018), "Soft computing-based slope stability assessment: A comparative study". *Geomech. Eng.*, **14**(3), 257-269. <http://doi.org/10.12989/gae.2018.14.3.257>.
- Komurlu, E., Kesimal, A. and Hasanpour, R. (2015), "In situ horizontal stress effect on plastic zone around circular underground openings excavated in elastic zones", *Geomech. Eng.*, **8**(6), 783-799. <https://doi.org/10.12989/gae.2015.8.6.783>.
- Li, L.P., Lei, T., Li, S.C., Xu, Z.H., Xue, Y.G. and Shi, S.S. (2015), "Dynamic risk assessment of water inrush in tunnelling and software development", *Geomech. Eng.*, **9**(1), 57-81. <https://doi.org/10.12989/gae.2015.9.1.057>.
- Liu, Q., Chai, J., Chen, S., Zhang, D., Yuan, Q. and Wang, S. (2020), "Monitoring and correction of the stress in an anchor bolt based on pulse prepumped Brillouin optical time domain analysis", *Energy Sci. Eng.*, 1-13. <https://doi.org/10.1002/ese3.644>.
- Liu, S. and Lin, Y. (2010), *Grey System: Theory and its Application*, Springer-Verlag Berlin Heidelberg, 379-401.
- Narain, R., Golas, A. and Lin, M.C. (2010), "Free-flowing granular materials with two-way solid coupling", *Proceedings of the International Conference on Computer Graphics and Interactive Techniques*, Seoul, Korea, December.
- Polak, K., Rozkowski, K. and Czaja, P. (2016), "Causes and effects of uncontrolled water inrush into a decommissioned mine shaft", *Mine Water Environ.*, **35**(2), 128-135. <https://doi.org/10.1007/s10230-015-0360-6>.
- Sato, T., Kikuchi, T. and Sugihara, K. (2000), "In-situ experiments on an excavation disturbed zone induced by mechanical excavation in Neogene sedimentary rock at Tono mine, central Japan", *Develop. Geotech. Eng.*, **84**, 105-116. [https://doi.org/10.1016/S0165-1250\(00\)80010-3](https://doi.org/10.1016/S0165-1250(00)80010-3).
- Schäfer, M. and Teschauer, I. (2001), "Numerical simulation of coupled fluid-solid problems", *Comput. Meth. Appl. Mech. Eng.*, **190**(28), 3645-3667. [https://doi.org/10.1016/S0045-7825\(00\)00290-5](https://doi.org/10.1016/S0045-7825(00)00290-5).
- Si, L., Zhang, H., Wei, J., Li, B. and Han, H. (2021), "Modeling and experiment for effective diffusion coefficient of gas in water-saturated coal", *Fuel*, **284**, 118887. <https://doi.org/10.1016/j.fuel.2020.118887>.
- Wang, C., Shen, B., Chen, J., Tong, W., Jiang, Z., Liu, Y. and Li, Y. (2020), "Compression characteristics of filling gangue and simulation of mining with gangue backfilling: An experimental investigation", *Geomech. Eng.*, **20**(6), 485-495. <https://doi.org/10.12989/gae.2020.20.6.485>.
- Wu, J., Li, S.C., Xu, Z.H., Pan, D.D. and He, S.J. (2018), "Flow characteristics after water inrush from the working face in karst tunneling", *Geomech. Eng.*, **14**(5), 407-419. <https://doi.org/10.12989/gae.2018.14.5.407>.
- Yuan, Y.C., Li, S.C., Zhang, Q.Q., Li, L.P., Shi, S.S. and Zhou, Z.Q. (2016), "Risk assessment of water inrush in karst tunnels based on a modified grey evaluation model: Sample as Shangjiawan Tunnel", *Geomech. Eng.*, **11**(4), 493-513. <https://doi.org/10.12989/gae.2016.11.4.493>.
- Zhang, B. and Meng, Z. (2019), "Experimental study on floor failure of coal mining above confined water", *Arab. J. Geosci.*, **12**(4), 114-123. <https://doi.org/10.1007/s12517-019-4250-2>.
- Zhang, G.H., Jiao, Y.Y., Ma, C.X., Wang, H., Chen, L.B. and Tang, Z.C. (2018), "Alteration characteristics of granite contact zone and treatment measures for inrush hazards during tunnel construction-A case study", *Eng. Geol.*, **235**, 64-80. <https://doi.org/10.1016/j.enggeo.2018.01.022>.
- Zhou, Z.Q., Li, S.C., Li, L.P., Shi, S.S. and Xu, Z.H. (2015), "An optimal classification method for risk assessment of water inrush in karst tunnels based on the grey system", *Geomech. Eng.*, **8**(5), 631-647. <https://doi.org/10.12989/gae.2015.8.5.631>.

GC