Proc. of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Jinsong Huang, D.V. Griffiths, Shui-Hua Jiang, Anna Giacomini and Richard Kelly

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5182-7_08-010-cd

ISGSR 2022

Rockburst Intensity Prediction Based on African Vultures Optimization Algorithm-Random Forest Model

Zhong-guang Wu[1] and Qiang zhu[2] and Zong-wei Chen[3] and Shun-chuan Wu[4,5]

[1]Research Center for Standards and Metrology, China Academy of TransportationScience, Beijing 100029, China.
E-mail: kinliwu@163.com
[2]Key Laboratory of Ministry of Education for Efficient Mining and Safety of Metal Mine, University of Science and Technology Beijing, Beijing 100083, China.
E-mail: 121557316@qq.com
[3]Research Center for Standards and Metrology, China Academy of TransportationScience, Beijing 100029, China.
E-mail: chenzw@motcats.ac.cn
[4]Key Laboratory of Ministry of Education for Efficient Mining and Safety of Metal Mine, University of Science and Technology Beijing, Beijing 100083, China.
E-mail: wushunchaun@163.com
[5]Faculty of Land Resources Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650093, China.
E-mail: wushunchaun@163.com
Corresponding Author: CHEN Zong wei

Abstract: Rockburst is a common geological disaster in geotechnical engineering construction, rockburst prediction is closely related to geotechnical engineering and has great significance. Therefore, this paper proposed an African vultures optimization algorithm and random forest combined with AVOA-RF model to achieve the better performance of rockburst prediction. Six key parameters about microseismic, i.e., cumulative event number, event rate, cumulative release energy logarithm, energy rate logarithm, cumulative apparent volume logarithm, and apparent volume rate logarithm are selected to constitute rockburst prediction index system. A data set of 78 rockburst cases is constructed by collecting the monitoring data of rockburst microseismic of Jinping II Hydropower station, and used for train and test the proposed AVOA-RF model. In the process of model building, the average error rate obtained by 10-fold cross validation is used as the fitness value in the African vultures optimization algorithm. The model's optimal parameters were m_{try}=2 and n_{tree}=41. Then, the accuracy, precision, recall, $F1$-score, macro-average, micro-average, and AUC are selected to evaluate the model's prediction performance. The results showed that AVOA-RF model has good performance in rockburst data of test sets and new engineering projects, the accuracy on the test set is 94.4% and the model of AUC is 0.9974. The feature importance obtained by the AVOA-RF model indicated that cumulative release energy logarithm plays the most important role in rockburst. Besides, the proposed model is compared with support vector machine, decision tree, random forest, and probabilistic neural network. The comparison results show that the proposed model has the better performance.

Keywords: Rockburst classification; Machine learning; Random forest; African vultures optimization algorithm.

1 Introduction

Rockburst is a common geological disaster in geotechnical construction, which easily happened in the area of high ground stress and dry and intact hard and brittle rock mas (Guo et al. 2021). At present, the definition of rockburst has been clear: rockburst is an engineering geological disaster caused by rock ejection in surrounding rock with high elastic strain energy accumulated due to excavation disturbance of relevant staff (Wei et al. 2020). Rockburst occurs suddenly and strongly, and rock particles can be ejected at a speed of 8-50 m/s, which seriously threatens the safety of workers and equipment (Xu et al. 1999; Gong and Li 2007; Feng et al. 2012; Liu et al. 2021). Moreover, the prediction of rockburst intensity has been a difficult problem in the field of geotechnical engineering. Therefore, research on the prediction and prevention of rock bursts is very necessary. At present, many scholars have conducted more research on rockburst prediction methods (Sun 2019). Generally, rockburst prediction methods can be roughly divided into three categories: the first category is based on engineering practice and testing to propose new quantitative criteria for rockburst prediction; The second category is based on rockburst impact factors of rockburst comprehensive prediction methods; The third category is based on the data obtained from field monitoring to propose new means of rockburst prediction (Wang 2021). However, the occurrence of rockburst is related to many factors, including geological structure, mining or excavation methods, rock mechanical properties, and in-situ stress (Zhou et al. 2004). Predicting and classifying the intensity of a rockburst is a complex nonlinear process. As a result, current engineering prediction methods have significant limitations. In light of this, some researchers have attempted to predict rockburst using machine learning methods (Pu et al. 2019).

Feng and Zhao (2002) established a rockburst prediction model by analyzing the factors affecting rockburst and then using support vector machine. The results showed that the rockburst prediction method based on support vector machine has high accuracy and the method is scientifically feasible; Sun et al. (2009) used the knowledge of fuzzy mathematics and neural network to build a rockburst prediction model trained by an improved BP algorithm based on typical rockburst data. The model was successfully applied to the impact ground pressure prediction in the Sanhejian coal mine in China, and the results showed that the model was not only accurate but also intelligent; Dong et al. (2013) applied random forest to the rockburst grade determination problem. The results showed that the random forest method has a strong discriminatory ability and low misjudgment rate, which is an effective way to solve the rockburst grade determination; Chen et al. (2016) comprehensively analyzed the main influencing factors of rockburst, selected the rock stress coefficient as the evaluation index, and used the decision tree method for rockburst intensity prediction. The results showed that the decision tree method has the features of simple calculation, accurate and reliability, and high prediction efficiency; Wu et al. (2019) constructed a PCA-PNN model and compared the model with random forest, support vector machine, and artificial neural network models, and the results showed that the rockburst intensity prediction model based on PCA-PNN is reasonable and feasible. In summary, machine learning has achieved many results in rockburst prediction. Although the machine learning methods used in the above studies can be used to predict rockburst, they still have some drawbacks such as long training time and random selection of hyperparameters in the algorithms.

Therefore, in this paper, a rockburst prediction model based on African vulture optimization algorithm-random forest (AVOA-RF) is proposed and compared with prediction models constructed by support vector machine (SVM), decision tree (DT), and probabilistic neural network (PNN) for analysis. With the comparison results of accuracy (ACC), precision (P), recall (R), $F1$-score, and other commonly used model performance evaluation metrics in classification problems, it is shown that the model has better performance and prediction capability.

2 Data sources and description

In this paper, six parameters of microseismic, which are closely related to microfracture activity and can reflect the law of rockburst incubation, are selected as cumulative event number (N), event rate (n), cumulative release energy logarithm (Lg(E)), energy rate logarithm (Lg(e)), cumulative apparent volume logarithm (Lg(V)) and apparent volume rate logarithm (Lg(v)). The data set of rockburst is established by selecting six indexes (Feng et al. 2013). Rockburst intensity is classified into four levels according to the conventional grading method, namely none rockburst, light rockburst, moderate rockburst, and strong rockburst (indicated by the numbers 1 to 4). Figure 1(a) shows the distribution of the rockburst dataset and the proportion of the four rockburst intensity types in the form of a pie chart, which shows that there are 29 none rockburst samples (37%), 17 light rockburst (22%), 21 moderate rockburst (27%), and 11 strong rockburst (14%). Obviously, there is a category imbalance in this dataset. When the model is trained using the unbalanced dataset, many commonly used machine learning algorithms fail to obtain good predictions. The reason for this is that the objective function of these algorithms is usually the overall accuracy, which leads the algorithms to focus too much on samples from the majority class and ignore samples from the minority class (Tang and Xu 2020). To deal with the problems caused by unbalanced data sets for model training, this paper uses a random oversampling technique to increase the number of minority class samples by randomly replicating them until the number of instances of different rockburst classes is balanced. The boxplot of the data set is shown in Figure 1(b). Where◇and□denote outliers and mean values, respectively. The horizontal solid lines in the box refer to the median, and the horizontal lines above and below the boundaries of the box are the third and first quartiles, respectively. The median values of most of the data are not in the center of the box, which indicates that the distribution of most of the data is asymmetric. For none rockburst, except for the N and n, there are no outliers; For light rockburst, except for Lg(v) and Lg(V), there are no outliers; For moderate rockburst, except for N, there are no outliers; For strong rockburst, in addition to n exist outliers, other indicators do not exist. In this paper, these outliers are not treated and are retained in the data set.

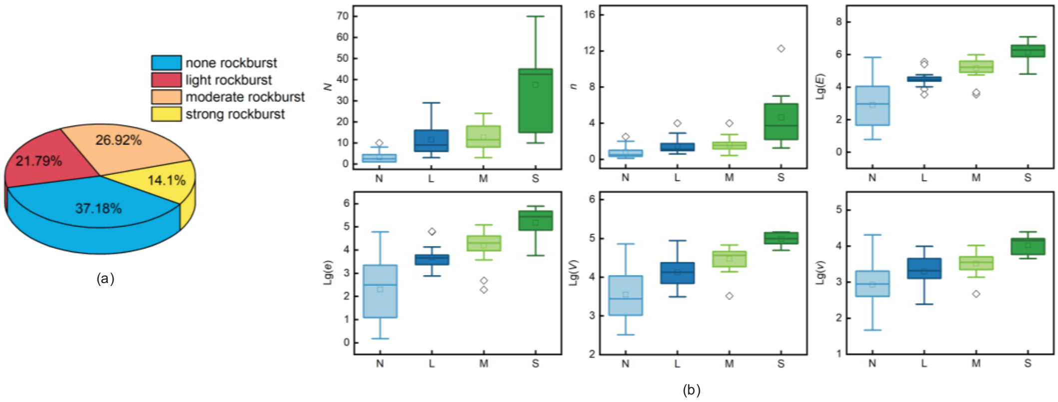

Figure 1. Data visualization: (a) The distribution of rockburst cases; (b) Boxplot of each variable for the four rockburst groups.

3 Methodology

In this paper, a random oversampling technique is applied to categories 2 and 4 within the dataset, and the number of datasets is expanded to 90. Then, the dataset was divided into two groups (i.e., training and testing datasets). In particular, the training set contains 72 cases (80%) and the testing set contains 18 cases (20%). The following content describes the random forest model, the African vulture algorithm, and the performance metrics.

3.1 Random forest

Random forest combines the Bagging idea and the random subspace method (Pavlov 1997). The basic idea of bootstrap is to obtain multiple self-sampling data sets through repeated random sampling of training sets, and construct a classification decision tree for each self-sampling data set. Finally, the output of all trees is gathered for voting to obtain the final result. Random forest algorithm can fit complex nonlinear relations and has the advantages of fast training speed, small generalization error, and difficulty in over-fitting (Ho 1998).

3.2 African vultures optimization algorithm

The inspiration of African vulture optimization algorithm comes from the foraging and navigation behavior of African vulture, which consists of four stages (determining the best vulture in the population, calculating the hunger rate of vulture, exploration stage, and exploitation stage) (Abdollahzadeh et al. 2021). The algorithm uses the hunger rate of vultures (F) to determine whether the vultures are in exploration or development. The structural process of the African vulture optimization algorithm is shown in Figure 2. The following is the calculation equation of F :

$$\begin{cases} t = h \times \left(\sin^{w}\left(\dfrac{\pi}{2} \times \dfrac{i}{n} \right) + \cos\left(\dfrac{\pi}{2} \times \dfrac{i}{n} \right) - 1 \right) \\ \mathrm{F} = (2 \times r_0 + 1) \times z \times \left(1 - \dfrac{i}{n} \right) + t \end{cases} \tag{1}$$

Where, F is the hunger of vultures; i is the current number of iterations, n is the total number of iterations, z is the random number in [-1, 1]; h is the random number in [-2, 2]; r_0 is the random number in (0, 1); w is the determination of parameters for interruptions in the exploration and mining phases, take 2.5.

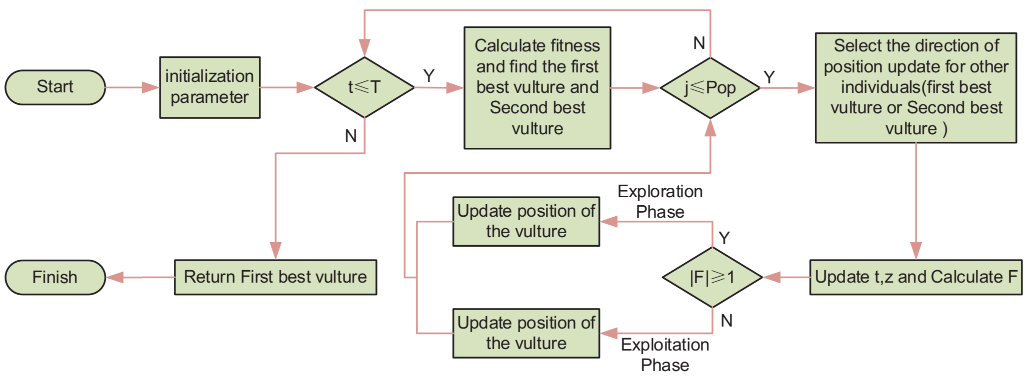

Figure 2. Flowchart of African vultures optimization algorithm.

3.3 Performance evaluation index

In this paper, the commonly used model performance evaluation indexes in classification problems are adopted, including accuracy, precision, recall, $F1$-score, and area under the curve of subject working characteristics (AUC). In addition, Macro-averaging and Micro-averaging indicators related to multi-classification problems are selected (Zhou et al. 2016; Tang and Xu 2020). The calculation formulas of ACC, P, R, and $F1$-score are as follows :

$$\mathrm{ACC}=\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \tag{2}$$

$$\mathrm{P}=\frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{3}$$

$$\mathrm{R}=\frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{4}$$

$$F1\text{-}score=\frac{2\times \mathrm{R}\times \mathrm{P}}{\mathrm{R+P}} \tag{5}$$

Where, TP is the number of positive samples correctly divided into positive samples; TN is the number of negative samples correctly divided into negative samples; FP is the number of negative samples wrongly divided into positive samples; FN is the number of positive samples wrongly divided into negative samples.

4 Model development

Before constructing the model, to reduce the impact of feature size and magnitude on model training and to speed up the convergence, this paper normalizes the existing data and maps them all to the interval of [-1, 1]. Two main parameters need to be determined for the random forest algorithm, which is the number of sample predictors per split node (m_{try}) and the number of classification trees in the random forest (n_{tree}) (Dong et al. 2013). These two main parameters determine the efficiency of the model computation and the accuracy of the results.

In the African vulture optimization algorithm, the population size is 20, the particle dimension is 2, the upper bound of the optimization parameters is [300, 6] and the lower bound is [2, 1], and the maximum number of iterations is 100. The fitness for this optimization is the mean value of the misclassification rate calculated by 10-fold cross validation (CV). The optimal hyperparameters of the algorithm m_{try}=2 and n_{tree}=41 are obtained by the optimization-seeking iteration. Then, the obtained two hyperparameters are substituted into RF for training to obtain the optimal AVOA-RF model. The optimal RF classifier (a combination of RF and AVOA) can be used for further performance evaluation of the test dataset. The full process of model construction and its performance evaluation is shown in Figure 3.

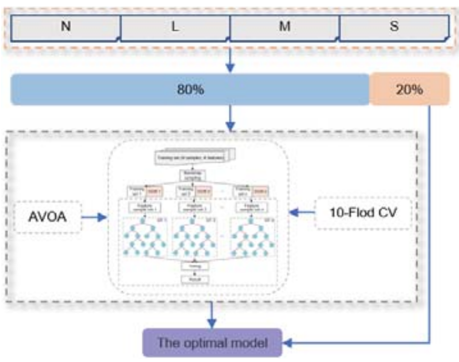

Figure 3. Model building and evaluation.

5 Model validation and analysis

The test set was brought into the optimal AVOA-RF model for prediction, and the confusion matrix of the test set was obtained as shown in Figure 4 (rows indicate the true values and columns indicate the predicted values). Each evaluation index is calculated separately according to the values of the corresponding confusion matrix, as shown in Table 1. The accuracy of the model reached 94.44%, which indicates the excellent performance of the proposed AVOA-RF for rockburst prediction. To further validate the performance of the AVOA-RF model, the results should be compared and analyzed with those of different rockburst prediction models for the same rockburst data. Therefore, SVM, DT, PNN, and RF models were selected for comparison in this paper. These four models are also built based on the training set that has been preprocessed as described above. After constructing the models, the confusion matrix of these four models on the test set was obtained, as shown in Figure 4. The values of the corresponding evaluation metrics are shown in Table 1. The ROC curves of each model, as shown in Figure 5.

Based on the metrics in Table 1, it can be seen that the AVOA-RF model performs better. The accuracy of the proposed AVOA-RF classifier is improved by 5.5%, 22.2%, 11.1%, and 5.5% compared to the SVM, DT, PNN, and RF models, respectively. Figure 5 shows the AUC of AVOA-RF is the largest (AUC=0.9974). Moreover, the AVOA-RF model also shows the optimal performance in the tests of other metrics.

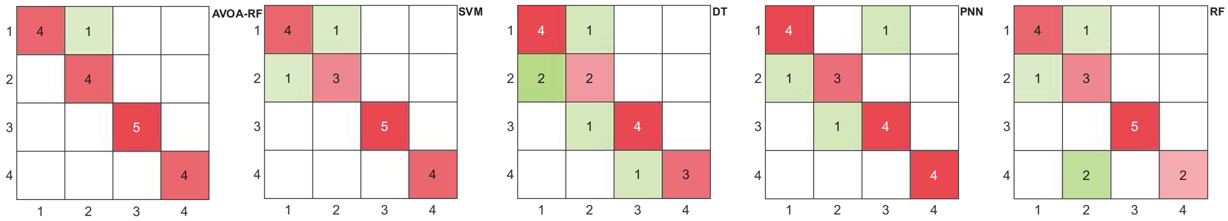

Figure 4. The confusion matrix for each model on the test set.

Table 1. Evaluation indicators of five models.

Model	Evaluation index	Class 1	Class 2	Class 3	Class 4	Macro-averaging	Micro-average	Accuracy
	P	1.00	0.8	1.00	1.00	0.95	0.94	
AVOA-RF	R	0.8	1.00	1.00	1.00	0.95	0.94	94.4%
	$F1$-score	0.89	0.89	1.00	1.00	0.95	0.94	
	P	0.80	0.75	1.00	1.00	0.89	0.89	
SVM	R	0.80	0.75	1.00	1.00	0.89	0.89	88.9%
	$F1$-score	0.80	0.75	1.00	1.00	0.89	0.89	
	P	0.67	0.50	0.80	1.00	0.74	0.72	
DT	R	0.80	0.50	0.80	0.75	0.71	0.72	72.2%
	F1-score	0.73	0.50	0.8	0.86	0.72	0.72	
	P	0.80	0.75	0.80	1.00	0.84	0.83	
PNN	R	0.80	0.75	0.80	1.00	0.84	0.83	83.3%
	$F1$-score	0.80	0.75	0.80	1.00	0.84	0.83	
	P	0.83	0.75	1.00	1.00	0.90	0.89	
RF	R	1.00	0.75	0.80	1.00	0.89	0.89	88.9%
	$F1$-score	0.91	0.75	0.89	1.00	0.89	0.89	

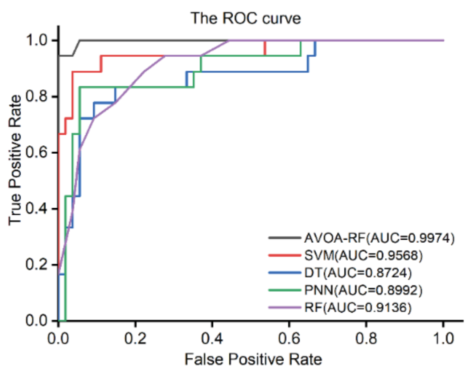

Figure 5. ROC curve for each model.

Feature importance can be able to reflect some extent the degree of influence of features on the prediction model. Feature importance is measured by determining how much each feature contributes to each tree in the random forest (Gini index) and then taking the average value. The Lg(E), Lg(V), and N contribute more to rockburst prediction than the other three metrics. The importance of the six features is shown in Figure 6.

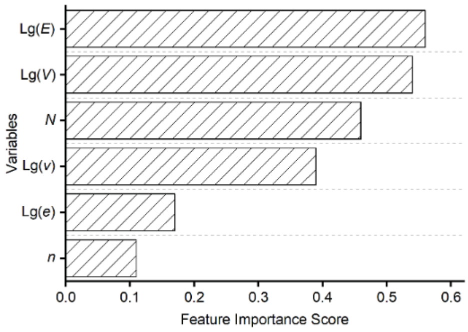

Figure 6. Feature importance of variables.

6 Engineering application

Jinping II hydropower station is located in Sichuan province, China. It has four diversion tunnels with an average length of 16.67km and a diameter of 13m. The overlying rock mass is generally buried 1500-2000m deep, and the maximum depth is about 2525m. The main lithology of the tunnel is marble, which has high brittleness and strength. To further verify the applicability of AVOA-RF rockburst prediction model proposed in this paper, two actual rockburst case data of the 3# diversion tunnel of Jinping Second-level hydropower station were selected for comparative analysis (Feng et al. 2015). The prediction results are shown in Table 2. As can be seen from Table 2, the two rockburst cases of the 3# diversion tunnel were correctly predicted. Therefore, the proposed ensemble classifier is feasible and effective in predicting rockburst risk.

Table 2. Prediction results of rockburst cases.

	N (unit)	n (unit/d)	Lg(E) (J)	Lg(e) (J/d)	Lg(V) (m^3)	Lg(v)(m^3/d)	Actual	Predicted
Case 1	45	4.1	4.803	3.762	4.838	3.796	3	3
Case 2	42	6.0	6.284	5.439	5.050	4.304	3	3

7 Conclusion

In this study, a classification model AVOA-RF combining African vulture optimization algorithm and the random forest is proposed to classify the strength of rockburst in geotechnical engineering. The following conclusions were obtained.

(1) The African vulture optimization algorithm can effectively optimize the hyperparameters of the random forest model. Meanwhile, the AVOA-RF classifier proposed in this paper exhibits a better accuracy (0.944) and an AUC of 0.9974. In addition, the model has been successfully applied to rockburst prediction in engineering.

(2) The results of feature importance show that Lg(E) is the most important variables in rockburst classification.

References

Guo, D., Chen, H., Tang, L., Chen, Z., and Samui, P. (2021). Assessment of rockburst risk using multivariate adaptive regression splines and deep forest model. *Acta Geotechnica*, 1-23.

Wei, X., Chen, T., and Wang, X. et al. (2020). Progress in Research of the Rockburst Hazard. *Modern Tunnelling Technology*, 57(02), 1-12.

Gong, F., and Li X. (2007). A distance discriminant analysis method for prediction of possibility and classification of rockburst and its application. *Chinese Journal of Rock Mechanics and Engineering*, (05), 1012-1018.

Feng, X., Zhang, C., and Chen, B. et al. (2012). Dynamical Control of Rockburst Evolution Process. *Chinese Journal of Rock Mechanics and Engineering*, 31(10), 1983-1997.

Liu, X., Zhang, X., and Yang, W. (2021). Rockburst prediction based on PCA-SVM. China Mining Magazine, 30(07), 176-180.

Xu, L., Wang, L., and Li, T. (1999). Present Situation of Rockburst Research at Home and Abroad. *Journal of Yangtze River Scientific Research Institute*, (04), 25-28.

Sun, C. (2019). A Prediction Model of Rock Burst in Tunnel Based on the Improved MATLAB-BP Neural Network. *Journal of Chongqing Jiaotong University (Natural Sciences)*, 38(10), 41-49.

Wang, K. (2021). Overview of State-of-Art of Rockburst Prediction and Prevention Techniques for Deep-buried Tunnels. *Tunnel Construction*, 41(02), 212-224.

Zhou, P., Wu, J., and Sun, Y. (2004). Forecast of Rockburst and Analysis of the Influence Factors. *Hongshui River*, (01), 71-74.

Pu, Y., Apel, D.B., Liu, V., Mitri, H. (2019). Machine learning methods for rockburst prediction-state-of-the-art review. *International Journal of Mining Science and Technology*, 29(4), 565-570.

Feng, X., and Zhao H. (2002). Prediction of Rockburst Using Support Vector Machine. *Journal of Northeastern University(Natural Science)*, (01), 57-59.

Sun, J., Wang, L., and Zhang, H. et al. (2009). Application of fuzzy neural network in predicting the risk of rock burst. *Procedia Earth and Planetary Science*, 1(1), 536-543.

Dong, L., Li, X., and Peng, K. (2013). Prediction of rockburst classification using Random Forest. *Transactions of Nonferrous Metals Society of China*, 23(02), 472-477.

Chen, S.,Wu, A., and Wang, Y. et al. (2016). Prediction of rockburst intensity based on decision tree model. *Journal of Wuhan University of Science and Technology*, 39(03), 195-199.

WU, S., Zhang, C., and Cheng, Z. (2019). Prediction of intensity classification of rockburst based on PCA-PNN principle. *Journal of China Coal Society*, 44(9), 2767-2776.

Feng, X., Chen, B., and Zhang, C. et al. (2013). Mechanism, warning and dynamic control of rockburst development processes. *Science Press, Beijing*.

Tang, Z., and Xu, Q. (2020). Rockburst prediction based on nine machine learning algorithms. *Chinese Journal of Rock Mechanics and Engineering*, 39 (04), 773-781.

Pavlov, Y.L. (1997) Random Forests. *Karelian Centre Russian Acad.sci.petrozavodsk*, 45(1), 5-32.

Ho, T.K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, 20(8), 832-844.

Abdollahzadeh, B., Gharehchopogh, F.S., and Mirjalili, S. (2021). African vultures optimization algorithm: a new nature-inspired metaheuristic algorithm for global optimization problems. *Computers and Industrial Engineering*, early, 1.

Zhou, J, Li, X, and Hani, M. (2016). Classification of Rockburst in Underground Projects: Comparison of Ten Supervised Learning Methods. *Journal of Computing in Civil Engineering, ASCE*, 30(05), 04016003.

Feng, G., Feng, X., Chen, B., Xiao, Y., and Yu, Y. (2015). A microseismic method for dynamic warning of rockburst development processes in tunnels. *Rock Mech Rock Eng*, 48(5), 2061-2076.