

## Research Article

# The Rock Burst Hazard Evaluation Using Statistical Learning Approaches

Jie Chen,<sup>1</sup> Jingkuan Gao ,<sup>1</sup> Yuanyuan Pu ,<sup>1</sup> Mingzhong Gao,<sup>2</sup> Like Wei ,<sup>3</sup> Chong Wang,<sup>3</sup> Bo Peng,<sup>4</sup> Xusheng Zhao,<sup>5</sup> Guangchao Zhang,<sup>6</sup> and Zhigang Zhang <sup>1</sup>

<sup>1</sup>State Key Laboratory of Coal Mine Disaster Dynamics and Control, Chongqing University, Chongqing 400044, China

<sup>2</sup>College of Water Resources and Hydropower, Sichuan University, Chengdu 610065, China

<sup>3</sup>Information Research Institute, Ministry of Emergency Management, Beijing 100029, China

<sup>4</sup>Sichuan Coal Industry Group Limited Liability Company, Chengdu 610091, China

<sup>5</sup>China Coal Technology and Engineering Group Chongqing Research Institute, Chongqing 400037, China

<sup>6</sup>College of Energy and Mining Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Yuanyuan Pu; [yuanyuanpu@cqu.edu.cn](mailto:yuanyuanpu@cqu.edu.cn)

Received 15 January 2021; Revised 10 February 2021; Accepted 25 March 2021; Published 5 April 2021

Academic Editor: Xiaowei Feng

Copyright © 2021 Jie Chen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The great threat and destructiveness brought by a rock burst make its prediction and prevention crucial in engineering. The rock burst hazard evaluation at project locations is an effective way of preventing rock burst since currently real-time prediction is not available. Since different control factors and discrimination conditions of rock burst were accepted by conventional risk determination methods, the rock burst risk determination in the same area may produce conflicting results. In this study, Naive Bayes statistical learning models based on different model prior distributions representing highly complicated nonlinear relationship between rock burst hazard and impact factors were built to evaluate the rock burst hazards. The results suggested that the Bayes statistical learning model based on a Gaussian prior has the strongest performance over four preset prior distributions. Combining the rock mechanics parameters measured in the laboratory and the stress data collected on the project sites, the proposed model was successfully employed to evaluate the kimberlite rock burst risk of a diamond mine in Canada. The Bayes statistical learning model exhibits its robustness and generalization in rock burst hazard evaluation, which can be generalized for similar engineering cases with enough supported data.

## 1. Introduction

A rock burst is a kind of sudden and severe rock instability, referring to a dynamic geological disaster caused by the sudden release of elastic strain energy accumulated in the rock mass of the underground excavations [1, 2]. Rock burst ejects a large amount of rocks in a short time, which seriously endangers the safety of construction equipment and personnel. With the increase of the mining depth and the tunnel construction, the initial ground stress has seen a drastic increase, which is more likely to induce rock burst [3]. Since the first recorded rock burst that occurred in the United Kingdom in 1738, all mining countries over the world have recorded rock bursts, including China, Canada, the United States, and Australia [4]. The Brunswick lead-zinc

mine in Canada, the Macassa gold mine, the Kalgoorlie gold mine in Australia, and the Idaho lead-zinc-silver mine in the United States have all experienced rock bursts that caused serious fatalities [5–7]. In 2018, 21 workers were killed and 4 were injured in a rock burst that occurred in the connecting lane at 1303 face of Long Yun Coal Mine, He Ze, Shandong Province, China. In addition to mining, rock bursts are also likely to occur in other underground excavations. For example, More than 1000 rock bursts of different levels occurred during the excavation of the No. 2 diversion tunnel in Jinping II Hydropower Station on the Ya Long River, which seriously affected the progress of the project.

The hazard of the rock burst makes the research on it a hot topic in the field of rock mechanics and engineering. Researchers have been committed to predicting the exact

time and location of rock burst based on the mechanism, so as to completely eliminate the threat of rock burst. However, the mechanism of rock burst is complicated, and its control factors are numerous including the mechanical properties of the rock, the field stress environment, construction support parameters, etc. Furthermore, the coupling relationships and interactions among those various influencing factors are complicated. Currently, there is no universal and reliable method for predicting rock burst in real time [8, 9]. Hence, before the construction of a project, the assessment for the rock burst risk in certain high-risk areas has become a major preventive method when rock burst cannot be accurately predicted in real time and space. Adjusting on-site excavation methods and support parameters with reference to the results of rock burst risk assessment is also an important prevention and control measure for rock burst. The evaluation of the rock burst risk is mainly to use the control factors of rock burst to comprehensively evaluate the possibility of the rock burst and the potential rock burst intensity. Conventional rock burst risk assessment methods can be divided into two categories, single index methods and comprehensive index methods. The single index methods use the selected control factors of rock burst to calculate a result value through the preset formula and then compare it with threshold value to determine the risk of a rock burst. The comprehensive index methods mainly use mathematical and statistical models to carry out a weighted mapping of the control factors of rock burst and calculate a comprehensive value to judge the risk of rock burst. The common synthetic index methods to evaluate rock burst risk include principal component analysis [10], fuzzy mathematics [11], and analytic hierarchy process [12]. Researchers have made achievements in evaluating rock burst hazards using single index methods and comprehensive index methods. However, due to some inherent defects of these two approaches, the universality is difficult to meet engineering needs, which is difficult to extend a successful risk assessment case to more engineering sites. For example, for single index methods, the calculated index is not only difficult to comprehensively reflect the mechanical behavior of the rock mass and the instability characteristics under extreme conditions but also requires presetting artificial threshold values for the risk assessment. However, in different rock burst cases, the threshold is determined by the constructor mainly based on experience. For example, both Kidybinski and Singh proposed to use strain energy storage index to judge the risk of coal rock burst. For coal mines in Silesia, Poland, Kidybinski's judgment criteria are that the strain energy storage index less than 2.5 refers to no rock burst hazard; the strain energy storage index greater than 2.5 and less than 3.5 refers to medium rock burst hazard; the energy index greater than 3.5 and less than 5 is regarded as a strong rock burst hazard; the strain energy storage index greater than 5 refers to the violent rock burst hazard [13]. By contrast, Singh puts forward the rock burst hazard criteria for hard rock as follows: value less than 10 is regarded as no rock burst hazard; value greater than 10 and less than 15 is regarded as medium rock burst hazard; and value greater than 15 is regarded as strong rock burst hazard [14]. In terms of the

comprehensive index method, it has to set a threshold value for risk judgment, as well as to assign weights to the selected control factors of rock burst. The aforementioned traditional rock burst risk assessment methods are greatly affected by the subjective judgment of researchers, which is difficult to accurately and objectively judge the risk of rock burst. Hence, the methods accepted in some cases are difficult to generalize to other rock burst cases.

In view of the shortcomings of traditional risk evaluation approaches, it raises a new possibility to use the statistical learning method to judge the risk of rock burst only from case data. Statistical learning is a way of obtaining knowledge from existing data and predicting using new data. Commonly used statistical learning models include support vector machines, feedforward neural networks, logistic regression, and Bayesian methods. Rock burst risk assessment can be regarded as a multiclass supervised-learning task, which means to use the collected rock burst case data (including the values of the control factors of rock bursts and the corresponding intensity level) to train a statistical learning model and then feed new input data control factors for prediction. As an end-to-end data-driven method, statistical learning does not need any prior information about the data as well as considering the complex mapping relationship for the intermediate process in the process of model training and prediction. According to universal approximation theorem, a feedforward neural network with only one hidden layer, which embeds enough neurons, can approximate any continuous function on a compact subset of  $R_n$  with arbitrary precision [15]. Therefore, the statistical learning model is suitable for multiclassification tasks such as rock burst hazard evaluation which is not clear enough, involving many variables and complex mapping relationships.

This study collected rock burst case data to train Naive Bayes classifiers under various prior distributions. By comparing the classification accuracy, the Bayesian classifier based on Gaussian distribution is selected as the rock burst risk assessment model. This model is used to evaluate the rock burst risk of kimberlite in a diamond mine in Canada. Compared with other statistical learning models, the Naive Bayes classifier does not significantly reduce the model accuracy when the training sample size has been reduced, which has strong adaptability to small sample tasks [16]. Rock burst cases are all over the world, and most of them are difficult to access case reports. Due to the lack of control factor data of some cases, the complete and high-quality data that can be collected are limited, which means that the rock burst hazard evaluation task is a small sample training task. Therefore, it is reasonable to use the Naive Bayes classifier to evaluate the rock burst hazard. The remainder of this article is organized as follows: Section 2 discusses the background theory of the Naive Bayes model; Section 3 gives the complete process of model construction, including data collection, model training, and validation; Section 4 uses the trained model to conduct rock burst hazard evaluation in a diamond mine; and finally, the conclusion and discussion are given in Section 5.

## 2. Bayesian Statistical Learning Model

The Naive Bayesian statistical classification method has excellent performance in many practical applications, such as file classification and spam filtering.

Compared with most random processes, the Naive Bayes classification method can extract data features of each dimension more quickly, reducing the difficulty of high-dimensional data calculation. However, the probability calculation of Naive Bayes in practice is significantly different from the true probability, but the impact of this difference on the classification task can be ignored, and the classification effect is well represented. This method assumes that all attributes independently affect the results, given the sample attributes  $X = \{x_1, x_2, \dots, x_n\}$ , and the Bayesian class conditional probability corresponding to the class label  $y$  is

$$P(y|X) = P(y|x_1, \dots, x_n) = \frac{P(y)P(x_1, \dots, x_n|y)}{P(x_1, \dots, x_n)}. \quad (1)$$

According to the independence hypothesis of Naive Bayes:

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i|y). \quad (2)$$

For attribute  $X$ , the probability of class  $Y$  is simplified as

$$P(y|x_1, \dots, x_n) = P(y) \frac{\prod_{i=1}^n P(x_i|y)}{P(x_1, \dots, x_n) = \frac{P(y)}{P(X)} \prod_{i=1}^d P(x_i|y)}, \quad (3)$$

where  $n$  is the number of attributes and  $x_i$  is the value of  $X$  on the  $i$ -th attribute.

Generally, for all categories, the probability  $P(X)$  has the same value, so the Bayesian decision criterion based on the Bayesian optimization classifier is

$$h_{nb}(x) = \arg \max_y P(y) \prod_{i=1}^n P(x_i|y). \quad (4)$$

Obviously, the judgment process of the Naive classifier of expression (4) is to estimate the prior probability  $P(y)$  of the class based on the dataset, and each attribute estimates the conditional probability  $P(x_i|y)$ .

Through the basic formula of the above Bayes algorithm, the basic process is as follows:

- (1) Let  $X = \{x_1, x_2, \dots, x_n\}$  be the sample to be classified, where  $x_i$  is a characteristic attribute of  $X$ .
- (2) The set of categories to be classified is  $C = \{y_1, y_2, \dots, y_n\}$ .
- (3) Calculate the probability of  $P(y_1|X)P(y_2|X), \dots, P(y_n|X)$ , respectively.
- (4) If  $P(y_k|X) = \max\{P(y_1|X), P(y_{k2}|X), \dots, P(y_n|X)\}$ , then  $X$  is considered to be of type  $y_k$ .

According to the prior distribution that the data obey, the Naive Bayes classifier can be divided into Gaussian,

multinomial, complement, Bernoulli, and other models. If the distribution of sample features is continuous, the attribute features can be assumed to obey Gaussian distribution:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right). \quad (5)$$

The parameters  $\sigma_y$  and  $\mu_y$  can be calculated using maximum likelihood estimation (MLE).

Schütze H et al. supposed that when the characteristic attributes are continuous values, which distribution obeys the Bernoulli distribution, the Bernoulli's Naive Bayes classification algorithm can be introduced for calculating posterior distribution [4–6]. The classification decision rule is calculated based on the Bernoulli distribution probability as follows:

$$P(x_i|y) = P(i|y)x_i + (1 - P(i|y))(1 - x_i). \quad (6)$$

Similarly, when the data prior distribution is polynomial distribution or other distribution, the corresponding Bayesian classifier can also be constructed correspondingly.

The control factors of rock burst are numerous and the relationship is complicated. It is difficult to develop the mathematical laws that the control factors obey from the mechanism, which is difficult to predict the predetermined distribution of its characteristics. In this paper, Naive Bayesian models based on Gaussian, polynomial, complement, and Bernoulli distributions are established, respectively, performing statistical learning on the same dataset and finally adopting the prior distribution with the best performance as that prior distribution of the Bayesian classifier.

In general, the Naive Bayes classifier can only conduct binary classification task. The risk assessment of rock burst is a multiclass classification task, which has to construct a Naive Bayes classifier suitable for multiclass classification. This article establishes multiple basic binary classifiers and finally completes the multiclass classification task through a “voting” mechanism. The basic idea is to establish a binary classifier between each pair of all categories, that is to say, there are  $n$  categories to be classified, and  $n(n-1)/2$  classifiers are established. Let  $C_{ij}$  represent the classifier between category  $i$  and category  $j$ . For a single training sample  $X$ , if the classification result belongs to  $i$ , then class  $i$  gets “one vote.” Otherwise, category  $j$  gets “one vote.” Until all  $n(n-1)/2$  classifiers have voted for  $X$ ,  $X$  will belong to the category with the most votes.

## 3. Model Construction and Training

**3.1. Sample Data Collection.** Rock burst is caused by the concentration of on-site stress exceeding the energy storage limit of the rock mass. There are two main control factors of a rock burst, the field stress indexes and the rock mass properties. In order to reflect the nature and characteristics of rock burst comprehensively, this article adopted the tangential stress of surrounding rock  $\sigma_\theta$  (MPa), the ratio of

tangential stress to uniaxial tensile strength  $\sigma_\theta/\sigma_t$ , the ratio of uniaxial compressive strength to uniaxial tensile strength  $\sigma_c/\sigma_t$ , and elastic energy index  $W_{ET}$  as the attributes of the data sample to evaluate the risk of rock burst. According to the commonly accepted standards for ranking a rock burst grades, a rock burst can be divided into four grades [17]. That is, no rock burst, moderate rock burst, strong rock burst, and severe rock burst. These four levels are used as data labels to participate in the construction of statistical learning models by digitizing. 0 means no rock burst risk, 1 means medium rock burst risk, 2 means strong rock burst risk, and 3 means violent rock burst risk.

This article collected 111 rock burst cases from different underground excavations over the world [4, 11, 12, 18–20]. Each data sample comprises four control factors of the rock burst mentioned above in this research as well as the corresponding rock burst grades, of which the dataset recorded a total of 13 samples labelled 0, 29 samples labelled 1, 55 samples labelled 2, and 15 samples labelled 3. The details of the data samples used to build the Bayes models are shown in Table 1.

**3.2. Model Training and Verification.** This article uses the Scikit-Learn statistical learning platform with Python language to build and train the Naive Bayes model. In order to eliminate the dimensional difference of the sample attributes, the sample attributes are standardized according to formula (7). All attributes are scaled between  $[0, 1]$ ,  $x$  is the original value of the attribute, and  $x'$  is the standardized sample attribute value.

$$f: x \longrightarrow x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \quad (7)$$

In order to compare the classification effect of the Naive Bayes models under the four prior distributions, trained models should use unseen samples that have not participated in the training for model verification. In statistical learning, the number of training samples have a great impact on the model's performance. This article collected 111 learning samples, which are typically small. If a part of samples are retained for model validation, it will further reduce the number of samples participating in model training and harm the model performance. Hence, this paper uses a 10-fold cross-validation method to verify the accuracy of the model. That is, all training samples are randomly divided into 10 disjoint subsets. In each round of training, nine subsets are selected for model training, and the remaining one is used for model validation. The final classification accuracy is the average accuracy of 10 round validations. This strategy ensures that all samples participate in model training and also ensures that the validation samples do not participate in, that is, the independence of verification. Figure 1 shows a schematic diagram of the 10-fold cross-validation.

Figure 2 shows the training process of four Bayesian models with different prior distributions.

We can see from Figure 2 that as the training samples enter the model, the accuracy of the Gaussian Bayes classifier is steadily increasing. Until all samples are involved in

training, the model accuracy stabilizes at 0.4. Complement Naive Bayes classifier has the worst accuracy, only 0.25, which is equal to the probability of random guessing. The model obviously did not acquire any knowledge from the training samples. However, the classification accuracy of Gaussian Bayes classifier, Bernoulli Bayes classifier, and polynomial Bayes classifier is not much different, and they are all approximately 0.4. It is difficult to determine the pros and cons of the three solely based on classification accuracy.

**3.3. Unequal Cost Classification Results.** In Section 3.2, the model validation adopted an intuitive metric, that is, the classification accuracy of the model (correctly classified samples/total samples). However, in some special cases, classification accuracy is not a universal measure for classification task. Table 2 gives the confusion matrix of the rock burst classification task in the case of multiclass classification [21]. The elements on the diagonal represent the correct classification, whose misclassification costs are all zero. Other elements represent the corresponding misclassification cost in the case of misclassification. Obviously, this is an asymmetric matrix. For example,  $\text{Cost}_{vn}$  represents that violent rock bursts were misclassified as no rock burst. This may lead to the lack of timely and effective prevention and control measures at the project field, resulting in huge casualties.  $\text{Cost}_{nv}$  represents that it misclassified no rock burst as violent rock burst. This will lead to excessive precautions in areas where there is no danger of rock bursts, resulting in economic and efficiency losses. The first misclassification cost is much higher than the second. Therefore, the risk assessment of the rock burst is a unequal cost classification.

Consider the following case. There are two classifiers A and B; classifier A misclassified all no rock burst cases as violent rock burst cases while classifier B misclassified all violent rock burst cases as no rock burst. The classification accuracy of the two is equal which is zero. However, the misclassification cost for classifier A is lower, and classifier A is better than classifier B. Therefore, the classification accuracy cannot fully reflect the performance of a classifier, especially when used for unequal cost classification tasks.

This article introduces receiver operating characteristic (ROC) curve to measure the performance of Bayesian classifiers [22]. The original intention of ROC curve is designed to serve two classification tasks with unequal costs. Combined with the actual needs of this article, the ROC curve is extended to four classifications. We drew the ROC curve of each category and finally used the average curve to reflect the performance of the entire classifier. According to ROC performance criteria, the classifier with the largest area under the curve (AUC) has the strongest performance. Figure 3 reflects the unequal cost classification performance of the three Bayesian classifiers.

## 4. Engineering Case Analysis

The Bayesian classifier based on Gaussian distribution is found to have the best performance among classifiers with four prior distributions, that is, the strongest rock burst risk

TABLE 1: Sample data.

Number	Case	$\sigma_\theta$	$\sigma_\theta/\sigma_c$	$\sigma_c/\sigma_t$	$W_{ET}$	Risk	Label
1	Yu Zixi Hydropower Station Water Diversion Tunnel, China	90	0.53	15.04	9	Strong	2
2	Yi Chuan No. 2 Tunnel, China	90	0.41	29.73	7.3	Medium	1
3	Cavern Highway Tunnel, Iran	62.6	0.38	17.53	9	Medium	1
4	Takiyami Power Station Headrace tunnel, Japan	55.4	0.32	24.11	9.3	Strong	2
5	Yutoo Hydropower Station, Norway	30	0.34	23.97	6.6	Strong	2
6	Yutoo Hydropower Station, Norway	48.75	0.27	21.69	5	Strong	2
7	Yutoo Hydropower Station, Norway	80	0.44	26.87	5.5	Medium	1
8	Guanyuk Road Tunnel, Japan	89	0.38	28.43	5	Strong	2
9	Jin Ping No. 2 Diversion Tunnel, China	98.6	0.82	18.46	3.8	Strong	2
10	Jin Ping No. 2 Diversion Tunnel, China	35	0.26	14.34	2.9	Medium	1
11	Tasmanian River Hydropower Station, Australia	157.3	0.58	13.18	6.27	Violent	3
12	Tasmanian River Hydropower Station, Australia	148.4	0.45	17.53	5.08	Medium	1
13	Tasmanian River Hydropower Station, Australia	132.1	0.39	20.86	4.63	Strong	2
14	Tasmanian River Hydropower Station, Australia	127.9	0.28	28.9	3.67	Medium	1
15	Tasmanian River Hydropower Station, Australia	107.5	0.2	36.04	2.29	No	0
16	Tasmanian River Hydropower Station, Australia	96.41	0.19	47.93	1.87	No	0
17	Tasmanian River Hydropower Station, Australia	167.2	0.66	13.2	6.83	Violent	3
18	Tasmanian River Hydropower Station, Australia	118.5	0.22	33.75	2.89	Medium	1
19	Hui Ze Lead-Zinc Mine, China	34.15	0.63	4.48	3.17	Medium	1
20	Jin Chuan No. 2 Mine, China	60	0.444	8.976	4.86	Medium	1
21	Jin Chuan No. 2 Mine, China	60	0.902	6.841	2.15	Medium	1
22	Jin Chuan No. 2 Mine, China	60	0.564	9.498	6.11	Medium	1
23	Jin Chuan No. 2 Mine, China	60	0.697	12.05	2.85	Medium	1
24	Jin Chuan No. 2 Mine, China	60	0.402	16.04	3.5	Medium	1
25	Jin Chuan No. 2 Mine, China	60	0.439	13.13	2.12	Medium	1
26	Ma Luping lead-zinc mine, China	63.8	0.58	24.4	6.31	Strong	2
27	Ma Luping lead-zinc mine, China	2.6	0.13	6.67	1.39	No	0
28	Ma Luping lead-zinc mine, China	44.4	0.37	24	5.1	Medium	1
29	Ma Luping lead-zinc mine, China	13.5	0.45	11.2	2.03	Medium	1
30	Ma Luping lead-zinc mine, China	70.4	0.64	24.4	6.31	Strong	2
31	Ma Luping lead-zinc mine, China	3.8	0.19	6.67	1.39	No	0
32	Ma Luping lead-zinc mine, China	57.6	0.48	24	5.1	Strong	2
33	Ma Luping lead-zinc mine, China	19.5	0.65	11.2	2.03	Strong	2
34	Ma Luping lead-zinc mine, China	81.4	0.74	24.4	6.31	Violent	3
35	Ma Luping lead-zinc mine, China	4.6	0.23	6.67	1.39	No	0
36	Ma Luping lead-zinc mine, China	73.2	0.61	24	5.1	Strong	2
37	Ma Luping lead-zinc mine, China	30	1	11.2	2.03	Violent	3
38	Birmingham Iron Mine, America	15.2	0.283	9.68	1.92	No	0
39	Birmingham Iron Mine, America	88.9	0.627	10.7	3.62	Violent	3
40	Birmingham Iron Mine, America	59.82	0.697	11.7	2.78	Strong	2
41	Birmingham Iron Mine, America	32.3	0.479	10.1	1.1	No	0
42	Birmingham Iron Mine, America	30.1	0.34	23.97	6.6	Violent	3
43	Birmingham Iron Mine, America	18.8	0.11	27.22	7	No	0
44	Birmingham Iron Mine, America	34	0.23	25.25	7.6	Medium	1
45	Birmingham Iron Mine, America	38.2	0.72	13.59	1.6	No	0
46	Birmingham Iron Mine, America	11.3	0.13	18.75	3.6	No	0
47	Birmingham Iron Mine, America	92	0.35	24.58	8	Medium	1
48	Birmingham Iron Mine, America	62.4	0.27	24.74	9	Violent	3
49	Birmingham Iron Mine, America	43.4	0.32	18.96	5.6	Violent	3
50	Birmingham Iron Mine, America	11	0.1	21.43	4.7	No	0
51	Cheng Chao Iron Mine, China	18.7	0.23	7.52	1.5	No	0
52	Cheng Chao Iron Mine, China	28.6	0.23	10.22	2.5	Strong	2
53	Cheng Chao Iron Mine, China	29.8	0.23	11.52	4.6	Strong	2
54	Cheng Chao Iron Mine, China	33.6	0.22	14.45	5.2	Strong	2
55	Cheng Chao Iron Mine, China	26.9	0.29	9.8	3.7	Strong	2
56	Cheng Chao Iron Mine, China	55.9	0.44	20.3	8.1	Violent	3
57	Cheng Chao Iron Mine, China	59.9	0.62	8.26	1.8	Medium	1
58	Cheng Chao Iron Mine, China	68	0.64	17.51	7.2	Violent	3
59	Dong Guashan Copper Mine, China	105.5	0.56	9.74	7.27	Strong	2
60	Dong Guashan Copper Mine, China	105.5	0.62	14.05	5.76	Strong	2

TABLE 1: Continued.

Number	Case	$\sigma_\theta$	$\sigma_\theta/\sigma_c$	$\sigma_c/\sigma_t$	$W_{ET}$	Risk	Label
61	Dong Guashan Copper Mine, China	105.5	0.55	11.11	3.97	Strong	2
62	Tong Yu Tunnel K21 + 680, China	47.56	0.81	16.71	5	Medium	1
63	Tong Yu Tunnel K21 + 680, China	43.62	0.56	24.41	6	Medium	1
64	Da Xiangling Tunnel YK55 + 119, China	25.7	0.43	45.9	1.7	No	0
65	Da Xiangling Tunnel ZK55 + 154, China	26.9	0.42	29.9	2.4	Medium	1
66	Da Xiangling Tunnel YK55 + 819, China	40.4	0.56	34.3	1.9	Medium	1
67	Da Xiangling Tunnel ZK55 + 854, China	39.4	0.6	28.3	3.4	Strong	2
68	Da Xiangling Tunnel YK56 + 080, China	38.2	0.53	21	3.6	Strong	2
69	Da Xiangling Tunnel YK56 + 109, China	45.7	0.66	21.5	4.1	Strong	2
70	Da Xiangling Tunnel YK56 + 177, China	35.8	0.52	17.8	4.3	Strong	2
71	Da Xiangling Tunnel YK56 + 343, China	39.4	0.57	25.6	3.8	Strong	2
72	Da Xiangling Tunnel ZK56 + 374, China	40.6	0.61	25.6	3.7	Strong	2
73	Da Xiangling Tunnel YK56 + 421, China	39	0.56	29.2	4.8	Strong	2
74	Da Xiangling Tunnel YK61 + 305, China	57.2	0.71	32.2	5.5	Violent	3
75	Da Xiangling Tunnel YK61 + 382, China	55.6	0.49	49.5	4.7	Strong	2
76	Da Xiangling Tunnel YK61 + 400, China	56.9	0.46	45.5	5.2	Strong	2
77	Da Xiangling Tunnel ZK61 + 440, China	62.1	0.47	55	5	Strong	2
78	Da Xiangling Tunnel YK61 + 445, China	29.7	0.26	42.9	3.7	Medium	1
79	Da Xiangling Tunnel YK61 + 450, China	29.1	0.31	36.1	3.2	Medium	1
80	Da Xiangling Tunnel YK61 + 493, China	27.8	0.31	42.8	1.8	No	0
81	Da Xiangling Tunnel YK61 + 827, China	30.3	0.34	28.3	3	Medium	1
82	Da Xiangling Tunnel YK61 + 382, China	55.6	0.49	49.5	4.7	Strong	2
83	Da Xiangling Tunnel ZK56 + 451, China	41.6	0.61	25	3.7	Strong	2
84	Da Xiangling Tunnel ZK56 + 479, China	40.1	0.55	31.3	4.6	Strong	2
85	Da Xiangling Tunnel ZK61 + 201, China	58.2	0.69	32.1	5.9	Violent	3
86	Da Xiangling Tunnel ZK61 + 352, China	56.8	0.5	50.9	5.2	Strong	2
87	The Hegula Tunnel, Norway	54.4	0.34	21.92	4.9	Strong	2
88	West Code Tunnel, Norway	47.3	0.27	21.34	5	Strong	2
89	Sewage Tunnel, Norway	73.5	0.42	21.6	4.8	Strong	2
90	La Xiwa Hydropower Station Underground Chamber, China	56.3	0.32	23.78	8.9	Strong	2
91	Sheng Qiao Hydropower Station Diversion Tunnel, China	29.7	0.33	25	6.5	Strong	2
92	Er Tan Hydropower Station Branch tunnel, China	88.2	0.41	29.45	6.9	Medium	1
93	Jin Ping Hydropower Station Diversion Tunnel, China	97.2	0.81	18.75	4	Strong	2
94	Yu Zixi Water Hydropower Station Diversion Tunnel, China	87.5	0.53	15.14	8.8	Strong	2
95	A gold mine roadway, South Africa	108	0.5	25.78	7.1	Violent	3
96	Hoist underground cavern, South Africa	93.9	0.47	25.06	6.3	Strong	2
97	Galena Gold Mine, America	52	0.3	25	5.2	Strong	2
98	Sewage Tunnel, Norway	109.4	0.58	31.15	6.9	Strong	2
99	Sewage Tunnel, Norway	85.7	0.46	24.03	5.8	Medium	1
100	Sewage Tunnel, Norway	79.8	0.44	28.57	4.7	Strong	2
101	Long Yangxia Hydropower Station Underground Chamber, China	21.6	0.12	31.03	7.6	No	0
102	Tai Pingyi Hydropower Station Underground Chamber, China	59.4	0.36	17.74	8.9	Medium	1
103	Li Jiaxia Hydropower Station Underground Chamber, China	10.8	0.09	23.53	5.5	No	0
104	Kotoyo Tunnel, Japan	79	0.36	26.83	4.8	Strong	2
105	Vietas Water Power Station Diversion Tunnel, Sweden	81	0.45	27.27	5.1	Medium	1
106	Raibl metal mine tunnel, Italy	103.6	0.74	17.72	5.2	Violent	3
107	Brugge Hydropower Station Underground Chamber, Switzerland	35.6	0.23	29.25	7.9	No	0
108	Waterfall Ditch Hydropower Station Underground Chamber, China	43.4	0.35	20.5	5	Strong	2
109	Qin Ling Tunnel of Xi Kang Railway, China	60.7	0.54	14.2	6.2	Strong	2
110	Bayu Tunnel, China	74.2	0.39	21.3	7.1	Violent	3
111	Kamchik Tunnel, Uzbekistan	59.6	0.4	18.9	6.9	Violent	3

assessment ability. The trained Gaussian classifier was used to evaluate the rock burst hazard of kimberlite in a diamond mine in Northern Canada.

The diamond mine is located in the North Slave Lake region, 250 kilometers north of Yellowknife, the capital of Yukon Province, Canada. The mining area includes three kimberlite pipes with the diameters more than 500 m, all of which are blasted underground. Kimberlite is a kind of

igneous rock that usually contains minerals such as diamonds. The lithology of the kimberlite is hard, which is prone to burst. Figure 4 shows an aerial view of this diamond mine and a mining kimberlite stope.

In order to obtain the corresponding risk assessment attributes, laboratory tests of rock properties and project site investigations are implemented. Twelve locations in two kimberlite pipes are selected to investigate the tangential

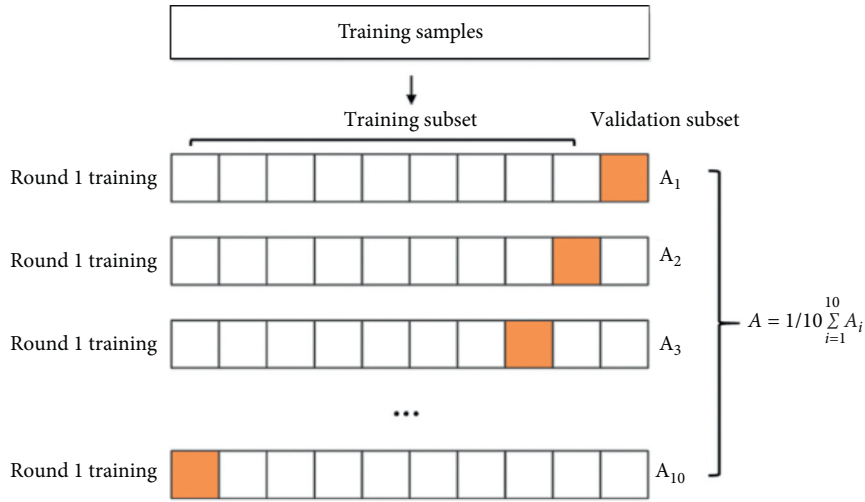


FIGURE 1: Schematic diagram of the 10-fold cross-validation process of the model.

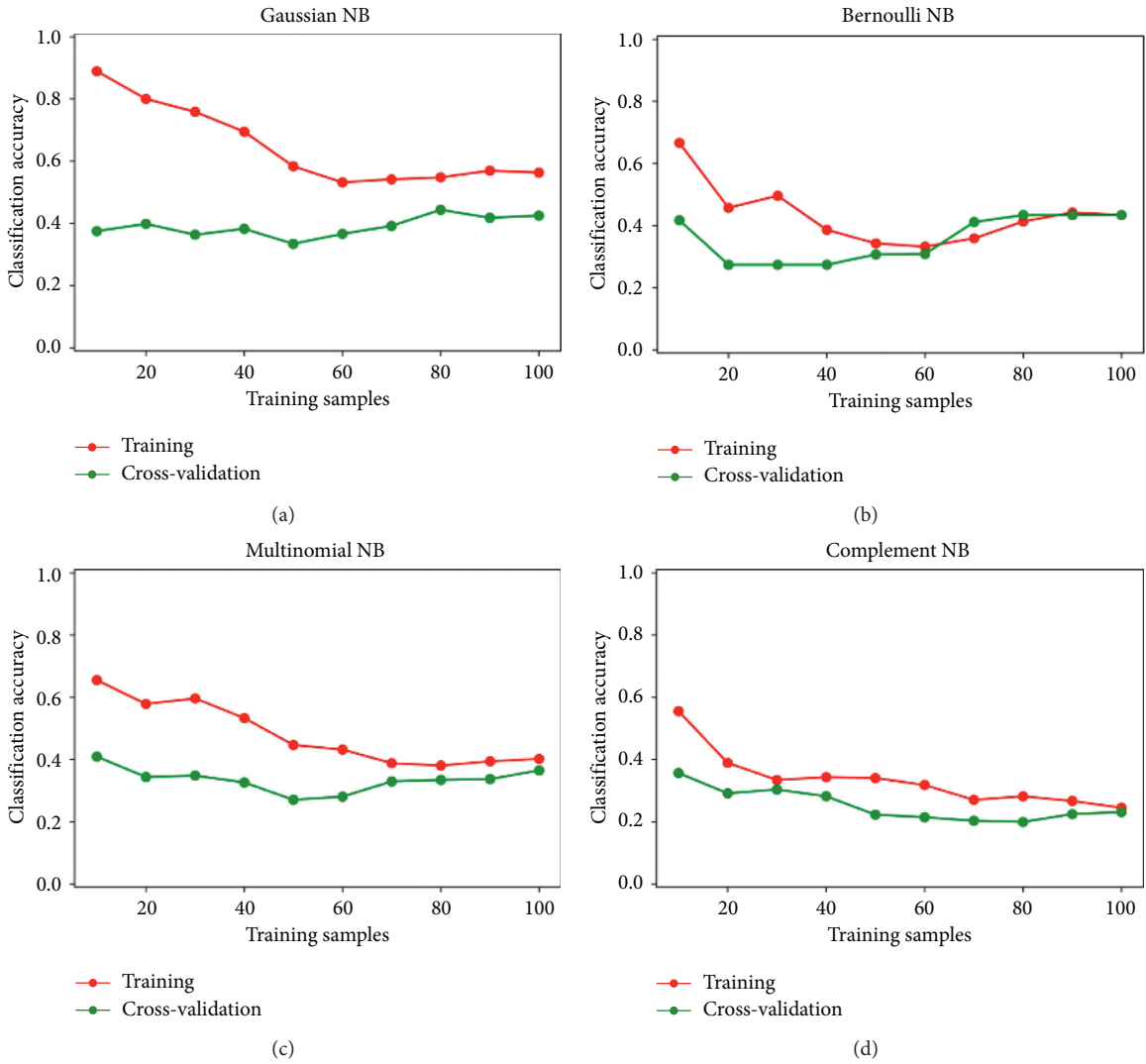


FIGURE 2: Model training process.

TABLE 2: Generalized confusion matrix for risk evaluation task of rock burst.

Misclassification cost		Prediction grade			
		No rock burst	Medium rock burst	Strong rock burst	Violent rock burst
Actual grade	No rock burst	0	$Cost_{NM}$	$Cost_{NS}$	$Cost_{NV}$
	Medium rock burst	$Cost_{MN}$	0	$Cost_{MS}$	$Cost_{MV}$
	Strong rock burst	$Cost_{SN}$	$Cost_{SM}$	0	$Cost_{SV}$
	violent rock burst	$Cost_{VN}$	$Cost_{VM}$	$Cost_{VS}$	0

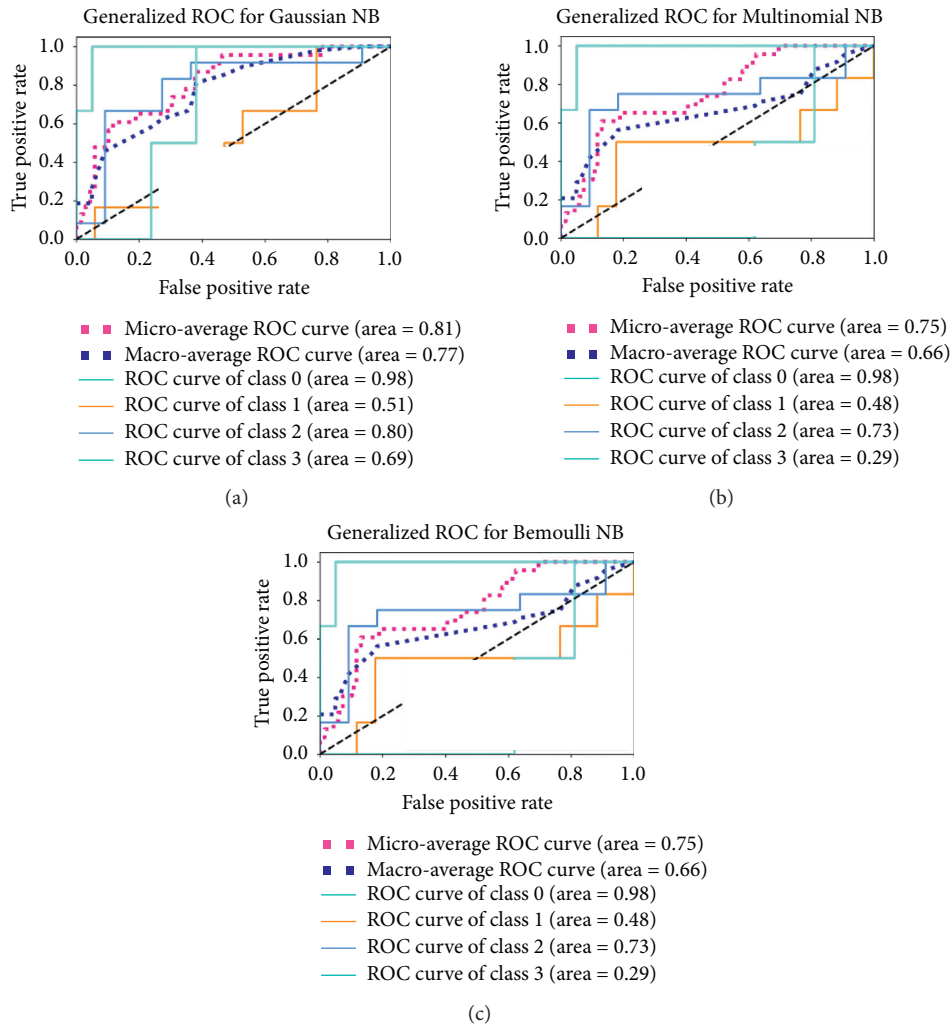


FIGURE 3: ROC curves and corresponding AUCs for three Bayesian models.

stresses of the surrounding rock mass. The rock samples near the survey locations are extracted to test the rock burst attribute parameters. The distribution of the survey area in kimberlite pipes and the schematic diagram of the lab tests are shown in Figure 5.

Table 3 shows the rock burst attributes of 12 kimberlite samples. The Naive Bayes classifier with Gaussian prior distribution is used to evaluate the rock burst hazard, and the corresponding hazard evaluation results are obtained.

The hazard evaluation results show that seven out of the 12 sampling areas had medium rock burst hazard, three had

strong rock burst hazard, and the remaining two areas had no rock burst hazard. The evaluation results matched mining logs which recorded two rock burst cases in this diamond mine in 2017. The first rock burst case occurred in the haul load at the N9750 level of the A154 kimberlite pipe, near the No. 7 sampling location. The second rock burst case occurred in the haul road at the N9850 level of A418 pipe, near the No. 12 sampling location. The mining log recorded the first rock burst case situation in which the rock mass was peeling off from the roof of the roadway and accompanied by a small-scale ejection, damaging the bolt and wire mesh



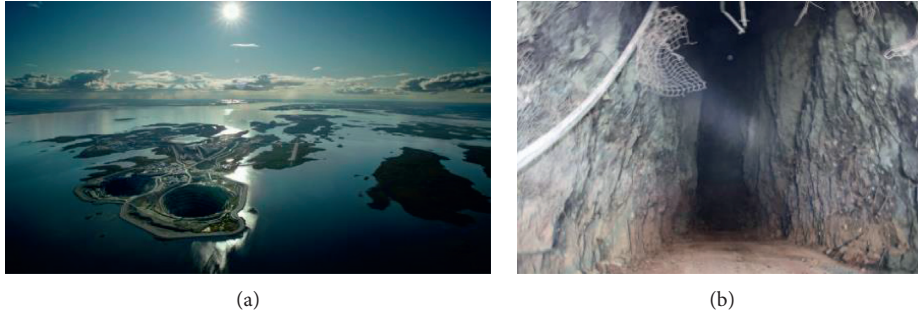


FIGURE 4: An aerial view of the diamond mine and a schematic diagram of mining kimberlite pillars [23].

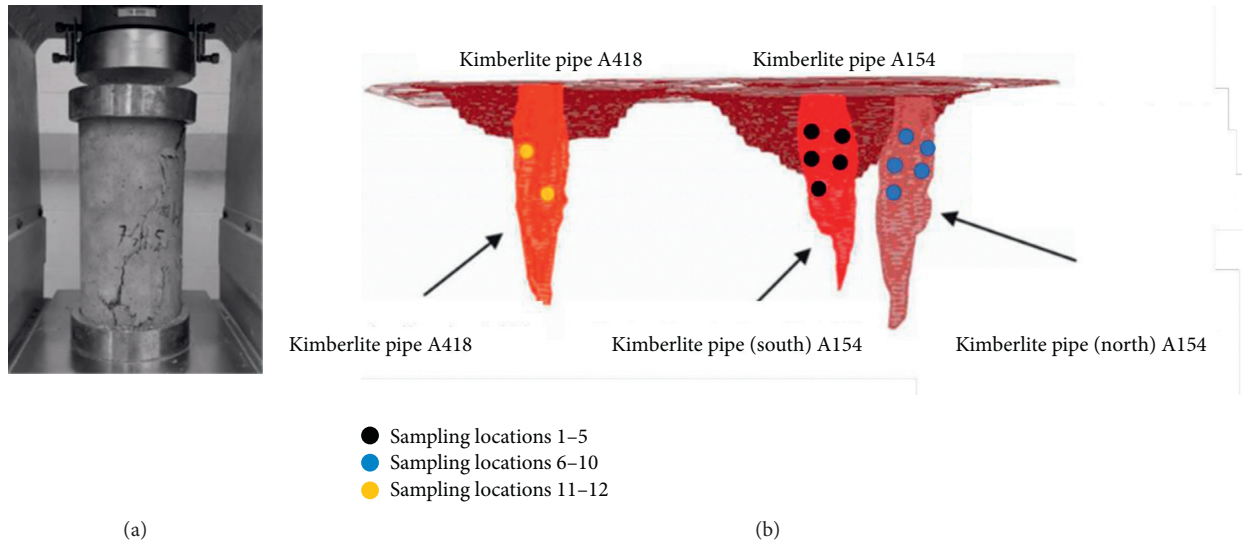


FIGURE 5: Schematic test of mechanical parameters and field sampling points of kimberlite.

TABLE 3: Sample attributes and corresponding rock burst risk assessment results.

Sampling point	$\sigma_\theta$ (MPa)	$\sigma_\theta/\sigma_c$	$\sigma_c/\sigma_t$	$W_{ET}$	Risk assessment results
1	18.167	0.37	31.4	3.3	No
2	21	0.35	18.9	1.7	Medium
3	31.16	0.38	21.2	2.3	Strong
4	46.376	0.62	25.1	3.2	Medium
5	48.64	0.64	18.6	2.5	Medium
6	22.92	0.4	40	1.5	Medium
7	99.088	0.88	30.1	5.2	Medium
8	35.156	0.44	25.6	2.5	Strong
9	15.84	0.32	22.9	2.8	No
10	13.02	0.2	28.5	1.2	Medium
11	21.12	0.4	24.2	2.3	Medium
12	29.121	0.51	17.1	2.2	Strong

support of the roadway. The length of the rock burst area is about 3 m, and no rock burst sound signal was recorded at the location before the occurrence. The second rock burst was even more severe. The rock mass ejected from the roof and the roadway support was completely damaged. The length of the rock burst area reached 12 m. According to the

rock burst evaluation rules, the first rock burst can be determined as a medium rock burst, and the second one can be determined as a strong rock burst. The rock burst cases verify the effectiveness of our proposed Gaussian distribution Bayesian model.

## 5. Conclusion and Discussion

- (1) This paper uses a statistical learning model to obtain information from the collected data of 111 rock burst cases and trains Naive Bayes classifiers based on different prior distributions. By comparing the classification performance of the four classifiers with different prior distributions, we finally determined that the Naive Bayesian model based on Gaussian distribution has the strongest classification ability. The trained model was used to determine the rock burst hazard of kimberlite in a diamond mine in Canada. The evaluation results obtained from the proposed model match the rock burst observations on the site, which verifies the applicability of the model.
- (2) The core part of constructing a Bayesian classifier is to determine the distribution type of training samples.

The four control factors of rock burst selected in this paper represent the two necessary conditions inducing a rock burst. The interactions over these control factors are complicated, which is difficult to determine the data distribution of attributes in advance solely through the rock burst mechanism. This paper compares four commonly used prior distributions through model validation and finds that the Gaussian Bayesian classifier works the best. Compared with the other three prior distributions, the selected rock burst attributes follow the general Gaussian distribution.

- (3) Although the performance of the model is in line with expectations, it can be seen that the performance of the model still needs to be improved. For example, the general classification accuracy of the model in Figure 2 is slightly higher than 0.4, and there is still much room for improvement. Another shortcoming is the “partial discipline” of the model. By analyzing the ROC curve, it can be seen that the model has better performances on the evaluation of no and strong rock burst hazards than on moderate and severe rock burst hazards.
- (4) Taking the characteristics of the proposed Bayesian model into account, the following measures are expected to improve model performance. The straightforward one is to more accurately derive the data distribution of the control factors of the rock burst. Combining the physical characteristics of the rock burst, we can explore the mathematical laws of the control factors, so as to establish the prior distribution of the data more accurately and improve the performance of the Bayesian model. The second is to collect more learning samples. If it is difficult to accurately determine the prior distribution for data, collecting more samples can endow the model the better prediction performance. The first method allows the Bayesian model not to be limited to several commonly used distributions. Building accurate prior distribution will cause a qualitative improvement in model performance. However, when the first method fails to realize, the second one can make up for the complement of model performance. However, due to the fixed prior distribution of the model, the second method can only achieve limited performance improvement of the model.
- (5) This article discards the conventional strategies based on the mechanism for rock burst hazard evaluation. This paper starts from the data, uses the statistical learning model to automatically obtain data information, and matches the knowledge obtained by the new input to complete the task of rock burst risk assessment. The ideas in this article can provide references for the rock burst prevention and prediction of mine dynamic disasters.

## Data Availability

The data used to support the findings of this study are included within the article and are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was funded by the National Key Research and Development Program of China (grant nos. 2017YFC0804201 and 2017YFC0804202).

## References

- [1] X. T. Feng, Y. X. Xiao, G. L. Feng et al., “Study on the development process of rockbursts,” *Chinese Journal of Rock Mechanics and Engineering*, vol. 38, no. 4, pp. 649–673, 2019.
- [2] Q. H. Qian, “Definition, mechanism, classification and quantitative forecast model for rockburst and pressure bump,” *Rock and Soil Mechanics*, vol. 35, no. 1, pp. 1–6, 2014.
- [3] M. C. He, H. P. Xie, S. P. Peng et al., “Study on rock mechanics in deep mining engineering,” *Chinese Journal of Rock Mechanics and Engineering*, vol. 16, pp. 2803–2813, 2005.
- [4] Y. Pu, D. B. Apel, C. Wang, and B. Wilson, “Evaluation of burst liability in kimberlite using support vector machine,” *Acta Geophysica*, vol. 66, no. 5, pp. 973–982, 2018.
- [5] B. Wilson and D. G. F. Hedley, *Rock Bursts: Case Studies from North American Hard-Rock Mines*, Society for Mining, Metallurgy, and Exploration, Littleton, CO, USA, 2009.
- [6] R. Baltz and A. Hucke, “Rock burst prevention in the German coal industry,” in *Proceedings of the 27th International Conference on Ground Control in Mining*, West Virginia University, Morgantown, WV, USA, July 2008.
- [7] Y. Potvin, M. Hudyma, and R. J. Jewell, “Rock burst and seismic activity in underground Australian mines—an introduction to a new research project,” in *Proceedings of the ISRM International Symposium*, Melbourne, Australia, November 2000.
- [8] M. C. He, F. Q. Ren, W. L. Gong et al., “Temperature characteristics during physical simulation test of strain burst,” *Journal of China University of Mining & Technology*, vol. 46, no. 4, pp. 692–698, 2017.
- [9] M. Cai, P. K. Kaiser, H. Morioka et al., “FLAC/PFC coupled numerical simulation of AE in large-scale underground excavations,” *International Journal of Rock Mechanics and Mining Sciences*, vol. 44, no. 4, pp. 550–564, 2007.
- [10] Y. Minami, D. Apel, and H. Xu, “A principal component analysis/fuzzy comprehensive evaluation for rockburst potential in kimberlite,” *Pure and Applied Geophysics*, vol. 175, no. 6, pp. 2141–2151, 2018.
- [11] Y. H. Wang, W. D. Li, Q. G. Li et al., “Comprehensive evaluation method of rockburst based on fuzzy mathematics,” *Chinese Journal of Rock Mechanics and Engineering*, vol. 5, pp. 3–5, 1998.
- [12] J. Gong, N. L. Hu, X. Cui et al., “Rockburst tendency prediction based on AHP-TOPSIS evaluation model,” *Chinese Journal of Rock Mechanics and Engineering*, vol. 33, no. 7, pp. 1442–1448, 2014.
- [13] A. Kidybiński, “Bursting liability indices of coal,” *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 18, no. 4, pp. 295–304, 1981.
- [14] S. P. Singh, “Burst energy release index,” *Rock Mechanics and Rock Engineering*, vol. 21, no. 2, pp. 149–155, 1988.

- [15] J. Park and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, no. 2, pp. 246–257, 1991.
- [16] V. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, Hoboken, NJ, USA, 1998.
- [17] H. B. Zhao, X. T. Feng, and S. D. Yin, "Classification of engineering rock based on support vector machine," *Rock and Soil Mechanics*, vol. 6, pp. 698–701, 2002.
- [18] J. Zhou, X. Li, and X. Shi, "Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines," *Safety Science*, vol. 50, no. 4, pp. 629–644, 2012.
- [19] M. Z. Bai, L. J. Wang, and Z. Y. Xu, "Study on a neural network model and its application in predicting the risk of rock blast," *China Safety Science Journal*, vol. 4, pp. 68–72, 2002.
- [20] Y. F. Bai, J. Deng, L. J. Dong et al., "Fisher discriminant analysis model of rock burst prediction and its application in deep hard rock engineering," *Journal of Central South University (Natural Science)*, vol. 40, no. 5, pp. 1417–1422, 2009.
- [21] J. Chen, J. K. Gao, Y. Y. Pu et al., "Machine learning method for predicting and warning of rock burst," *Journal of Mining and Strata Control Engineering*, vol. 3, no. 1, pp. 57–68, 2021.
- [22] Y. Pu, D. B. Apel, and C. Wei, "Applying machine learning approaches to evaluating rockburst liability: a comparison of generative and discriminative models," *Pure and Applied Geophysics*, vol. 176, no. 10, pp. 4503–4517, 2019.
- [23] Y. Pu, D. B. Apel, and H. Xu, "Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier," *Tunnelling and Underground Space Technology*, vol. 90, pp. 12–18, 2019.