



# Assessment of rockburst risk using multivariate adaptive regression splines and deep forest model

Deping Guo<sup>4</sup> · Hemao Chen<sup>1,2,3</sup> · Libin Tang<sup>1,2,3</sup> · Zhixiong Chen<sup>1,2,3</sup> · Pijush Samui<sup>5,6</sup>

Received: 19 March 2021 / Accepted: 28 June 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

Rockburst is a major instability issue faced by underground excavation projects, which is induced by the instantaneous release of a large amount of strain energy stored in rock mass. Because of its disastrous damage to infrastructures and facilities, more and more studies have been focused on rockburst prediction. However, due to highly nonlinear relationships between the occurrence of rockburst and potential triggering factors, traditional mechanism-based prediction methods have great difficulties in providing the reliable results. In this study, a multivariate adaptive regression splines (MARS) model and a novel deep forest algorithm were applied to predict and classify rockburst intensity of a database including 344 rockburst cases collected worldwide. The *t*-distributed stochastic neighbor embedding method (*t*-SNE) was utilized for nonlinear dimensionality reduction and visualization of the original input features. After that, the Gaussian mixture model was adopted to relabel original data to determine relative intensity of these rockburst cases. Then, the MARS model and deep forest model were constructed with these newly labeled data. Their performances were compared with some widely used machine learning methods, such as random forest, extreme gradient boost, and ANN model. The results clearly proved the capability of the proposed models to assess and forecast rockburst risk. It also proved that these approaches should be used as cross-validation against each other. The Shapley additive explanations method was adopted to investigate the relative importance of input features of the developed MARS model. The result shows that  $\sigma_\theta$  and  $\sigma_c$  are the most important features for rockburst intensity prediction, where  $\sigma_\theta$  is the tangential stress around underground opening and  $\sigma_c$  refers to uniaxial compressive strength of the rock.

**Keywords** Adaptive regression splines model · Deep forest model · Gaussian mixture model · Machine learning · Rockburst prediction · Random forest · XGBoost

## 1 Introduction

Rock bursting is a catastrophic rock mass failure, which usually occurs in deep burial and high geo-stress areas surrounded by hard and brittle rock masses during

excavation and unloading. In congested urban regions, more and more underground projects are continually developing deeply into the earth. These deep constructions include coal and metal mining, transportation engineering, energy storage, and high-level radioactive waste disposal.

✉ Zhixiong Chen  
chenzhixiong@cqu.edu.cn

<sup>1</sup> Key Laboratory of New Technology for Construction of Cities in Mountain Area, Chongqing University, Ministry of Education, Chongqing 400045, China

<sup>2</sup> School of Civil Engineering, Chongqing University, Chongqing 400045, China

<sup>3</sup> National Joint Engineering Research Center of Geohazards Prevention in the Reservoir Areas, Chongqing University, Chongqing 400045, China

<sup>4</sup> Sichuan Railway Investment Group Co., Ltd, Tianfu 1st St., High-Tech Zone of Chengdu Sichuan Province, Chengdu 100036, China

<sup>5</sup> Geographic Information Science Research Group, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

<sup>6</sup> Faculty of Environment and Labour Safety, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

Their excavation of cavities or tunnels in rock masses causes high stress release around the opening. The sudden emitted elastic deformation may cause ejection or launching of rock, and even earthquake. Rock particles can be ejected with a velocity up to 50 m/s.

Rock bursts usually cause considerable damage to infrastructures and facilities and may even cause fatal injuries. For example, in the USA, from 1936 to 1993, 172 rockburst cases resulted in more than 78 fatalities and 158 injuries [28]. Germany recorded more than 40 cases involving injuries and deaths by rockbursts between 1983 and 2007 [3]. In South Africa, rockburst occurred only seven times in 1908 while the number of rockbursts rose to 680 in 1975. Almost all gold mines have suffered from rockburst when gold mines gradually shifted from shallow mining to deep mining [76]. Rockbursting and seismicity were deemed as the biggest threat to workers' safety in Ontario's underground mines based on an investigation by the Ontario Ministry of Labour in Canada (Ontario Ministry of Labour 2015). From 1933 on, rockbursts have occurred in over 100 mines in China [49]. The most catastrophic rock burst happened in Jinping II Hydropower Station in 2009. The maximum buried depth of the drainage tunnel is about 2525 m with an in situ stress expected to be 72 MPa. On November 28, the support systems in the drainage tunnel were damaged by a rock outburst, which led to the destruction of a tunnel boring machines cost at 120 million RMB and seven deaths [77]. It is gradually acknowledged that rockbursts happen from time to time not only in mining industry, but also in hydropower, transportation, and other deep underground engineering.

Therefore, rockburst prediction is of crucial importance to the design and construction of many underground projects. A large number of rockburst cases show that two basic conditions are necessary for rockburst initialization: (1) The rock is hard, brittle, and relatively intact, capable of storing a large amount of elastic strain energy; (2) high stress environment (including in situ stress and induced stress). The former is the internal cause of rockburst, and the latter is the external cause. After the extensive research of more than seven decades, a number of methods have been developed to evaluate the rockburst potential at the construction site. The accessible assessment methodologies can be categorized into three types: (1) empirical index methods, (2) on-site monitoring methods, and (3) machine learning models.

The empirical index models assume that the rockburst probability can be expressed by some index calculated from: stress condition (for example, Turchaninov criterion, Russenes criterion, and Hoek criterion [25]), rock brittle index (for example, Hucka criterion [27], Singh criterion [50]), energy index (for example, elastic energy  $W_{et}$  criterion [33]), brittleness index modified (BIM) [1], rock

stiffness, instability, fractal fragmentation of rock, and so on. The empirical index models are generally developed from a certain project and cannot be universally applied to other sites because the rock discontinuity structural plane, dimension of the underground cavity, initial geo-stress and excavation-induced stresses, and construction processes vary significantly in different areas.

The on-site monitoring methods monitor the changes of a certain physical quantity (stress and strain, resistance, microseismic, acoustic emission, electromagnetic radiation, etc.) to indirectly reflect the changes of physical and mechanical properties of rock mass during excavation, so as to establish a relationship between the monitoring information and the state of rock mass, and then use it as the precursor information to predict the occurrence of rockburst. The examples of on-site monitoring methods mainly include microgravimetry method, electromagnetic radiation methods, acoustic emission method, electrical resistance method, computer tomography method, and microseismic monitoring method. But the relationship between rockburst intensity and rockburst impact factors is highly nonlinear, which cannot be expressed explicitly [5]. These two reasons make traditional, mechanism-based prediction methods unable to create a precise forecast.

In the last decade, novel and robust machine learning approaches have been developed to solve geotechnical problems [2, 6–10, 22, 54, 56, 59, 67, 68, 70], including prediction on both the rockburst occurrence and its intensity. These approaches can be classified into two categories as supervised learning methods and unsupervised learning methods (Table 1). The supervised learning methods use training samples with predefined labels. To train a supervised learning model for rockburst prediction, training samples of real rockburst cases are fed into the model with labels corresponding to rockburst intensity rankings defined by on-site judgment, engineering experience, engineering specifications or calculation. Ensemble models use stacking technique, which can combine advantages of several single models to form a stronger one [12]. But there are two main shortcomings making the labels (rockburst intensity) of training samples in supervised learning methods not consistent. First, the data were collected from all over the world and the rockburst classification criterion of each project may not be the same. Two data records with the same rockburst ranking using different criterion systems may represent different rockburst intensities. For example, Russia uses criterion has three or five rankings while the criterion in China and some other countries have four rankings. For the Russian criteria, the 3rd ranking means the most critical rockburst, but for Chinese criteria, 3rd ranking is not the most severe one. The machine learning models for rockburst prediction trained by inconsistent labels may not give reliable outcomes. Hence,

**Table 1** Review of machine learning approaches for rockburst prediction

Supervised learning methods		Unsupervised learning methods
Single model	Ensemble models	
ANN, Feng et al. [17]	GA-SVM and PSO-SVM, Zhou et al. [76]	ACCA, Gao [20]
GP, Su et al. [52]	GA-SMOTE-C4.5 DT, Jiang et al. [31]	K-means-SVM, Pu et al. [45, 46]
RF, Dong et al. [13]	GA-ELM, Li et al. [37]	SOM and FCM, Faradonbeh et al. [15, 16]
SGB, Zhou et al. [78]	MIVA-MFA-PNN, Feng et al. [18]	
LDA, QDA, PLSDA, NB, KNN, MLPNN, DT, SVM, RF, GB, and SGB, Zhou et al. [75]	FOA-GRNN, Pu et al. [44]	
BNs, NB, KNN, and RF, Li et al. (2017) [34]	PSO-BPNN, PSO-PNN and PSO-SVM, Liu and Hou [38]	
LR, Li and Jimenez [35]	FA-ANN, Zhou et al. [74]	
RMs, SVM, DT, Pu et al. [45]	PSO-ELM, Xue et al. [60]	
ENN, GEP, and C4.5 DT, Faradonbeh and Taheri [16]	GA-SVM, Ji et al. [29]	
C5.0 DT, Ghasemi et al. [21]	KNN-RNN, SVM-RNN, DNN-RNN and KNN-SVM-DNN-RNN, Xin Yin et al. [32]	

the rockburst intensity rankings (labels in the training data set) should not be determined merely by case records.

Because of the deficiency in supervised learning approaches, the unsupervised learning methods are adopted by some researchers to relabel the intensity rankings based on clustering. Clustering is a technique used to divide original data into different classes or clusters according to a certain standard (such as distance criterion), so that the similarity of data in one same cluster is as large as possible. By contrast, the difference is as wide as possible for the data that are not in the same cluster at the same time. In other words, after clustering, data of the same type are gathered together as much as possible, and different data are separated as much as possible. Gao [20] used an abstraction ant colony clustering algorithm (ACCA) to enhance the performance of the model for rockburst prediction. Pu et al. employed the k-means model to relabel the original data and then used support vector machine method to predict the rockburst potential of kimberlite pipes [45], 28. Faradonbeh et al. [16], 30 developed the self-organizing map (SOM) and fuzzy c-mean (FCM) clustering approaches for rockburst prediction in deep excavation.

Although most of the recent machine learning models (for example, ANN and GP) can provide acceptable accuracy, they are known as ‘black-box’ models whose decision processes are not so straightforward to understand easily. Many of the models are based on deep neural networks (DNN) structures. The training speed is quite slow because there are many parameters that need to be determined

before training, and the generalization ability of the model is greatly affected by these parameters, so the process to find the optimal combination of parameters will cost plenty of time. However, the explicit traditional linear regression model tends to underperform when multiple or nonlinear decision boundaries exist. Fortunately, the multivariate adaptive regression splines (MARS) model can give the explicit expression of highly nonlinear surfaces, which is much easier to interpret and has the ability to handle highly nonlinear problems, such as geotechnical engineering problems [23, 24, 55, 57, 64–66, 71]. The deep forest model proposed by Zhou and Feng opens a door for non-NN style deep learning, exhibiting the possibility of constructing deep models without using back propagation [73]. It has shown comparable or superior performance than DNN models on a variety of complex classification tasks with faster training speed and better parameter robustness. In addition, this method can adapt to training data sets of different size, enhancing its applicability and feasibility.

In order to overcome the above-mentioned deficiency problems for rockburst prediction, a database consisted of 344 groups of real rockburst cases data was reconstructed without their original rockburst intensity labels. Then, the Gaussian mixture model was applied to perform clustering of this database and relabel the rockburst intensity. After that, these relabeled data were fed into a multivariate adaptive regression splines (MARS) model and a deep forest model to train and make the prediction, respectively. Then, their performances were compared with some other well-known methods, such as random forest, extreme

gradient boost (XGBoost), and ANN model. Finally, the Shapley additive explanations method (SHAP) was used to investigate the relative importance of input features of the constructed MARS model.

## 2 Machine learning methodology

### 2.1 MARS methodology

The MARS model was developed by an American physicist J. Friedman [19] to fit and predict complex multi-variable functions. MARS draws more and more attention in the field of mega data because it has several main advantages such as: (1) The MARS model can give explicit expression of high dimension data, having better interpretability than other ‘black-box’ type methods, i.e., ANN and SVM; (2) the outcome of the MARS model only contains important input features, and the redundant variables are determined and deleted automatically; and (3) the MARS model has fewer parameters to tune.

In general, a MARS model consists of three main components: basis function (BF), knots, and spline function. In the simplest case, a basis function is in the linear form, and each spline function only contains one basis function. As clearly shown in Fig. 1a, the data can be approximated using four linear lines with different slopes in four intervals; therefore, a simple MARS expression can be:

$$y = 58.9 - 4.61 \times \text{BF}_1 + 3.18 \times \text{BF}_2 - 6.29 \times \text{BF}_3 + 1.99 \times \text{BF}_4 \quad (1)$$

in which  $\text{BF}_1 = \max(0, 16 - x)$ ,  $\text{BF}_2 = \max(0, x - 11)$ ,  $\text{BF}_3 = \max(0, x - 6.5)$ , and  $\text{BF}_4 = \max(0, 6.5 - x)$ . The three points of  $x = 6.5, 11$ , and  $16$  are the places where the behavior of the function changes significantly, which are also called as knots. The simplest basis function is the piecewise linear functions with a knot occurring at value  $c$  as:

$$\max(0, x - c) = \begin{cases} x - c, & \text{if } x \geq c \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

For a more general case, the spline function is the product of different basis functions. The spline function can be raised to different power with different combinations of basis function so as to manipulate the degree of smoothness of the resultant function estimate. The training data are divided into continuous regions and are fitted in each region with separate splines. The number of basis functions, the product degree, and knot locations are automatically determined like recursive partitioning.

The fundamental idea of MARS is to use the combination of basis functions to approximate the dependence of a

response variable  $y$  on input variable  $x = x_1, x_2, \dots, x_i$  with  $i$  features given  $n$  observed samples  $\{Y_k, X_k\}_{k=1}^n$ . The expression of the response variable  $y$  can be modeled as:

$$y = f(x_1, \dots, x_n) + e = f(x) + e \quad (3)$$

where  $e$  is the fitting error of  $f$  in the domain of  $(X_1, X_2, \dots, X_n)$ . The  $f(x)$  is a MARS model is constructed as a linear combination of BFs as:

$$f(X) = \alpha_0 + \sum_{m=1}^M \alpha_m \text{SP}_m(x) \quad (4)$$

where each  $\text{SP}_m(x)$  is a spline function. It can be a single basic function, or the product of two or more basic functions. The coefficients  $\alpha$  are constants estimated by the least-squares method.

The process of building a MARS model consists of two phases to fit the model in Eq (3): a forward constructive phase and a backward pruning phase. In the forward constructive phase, the algorithm begins with only the intercept term  $\alpha_0$  and adds spline functions to the model repeatedly in a manner which can reduce the training error most rapidly. The forward phase is executed until the number of spline functions exceeds a certain value or the improvement in model performance is insignificant. The spline function can be added to a current model with  $m$  spline functions is:

$$\hat{\alpha}_{M+1} \text{SP}_m(x) \max(0, x_j - c) + \hat{\alpha}_{M+2} \text{SP}_m(x) \max(0, c - x_j) \quad (5)$$

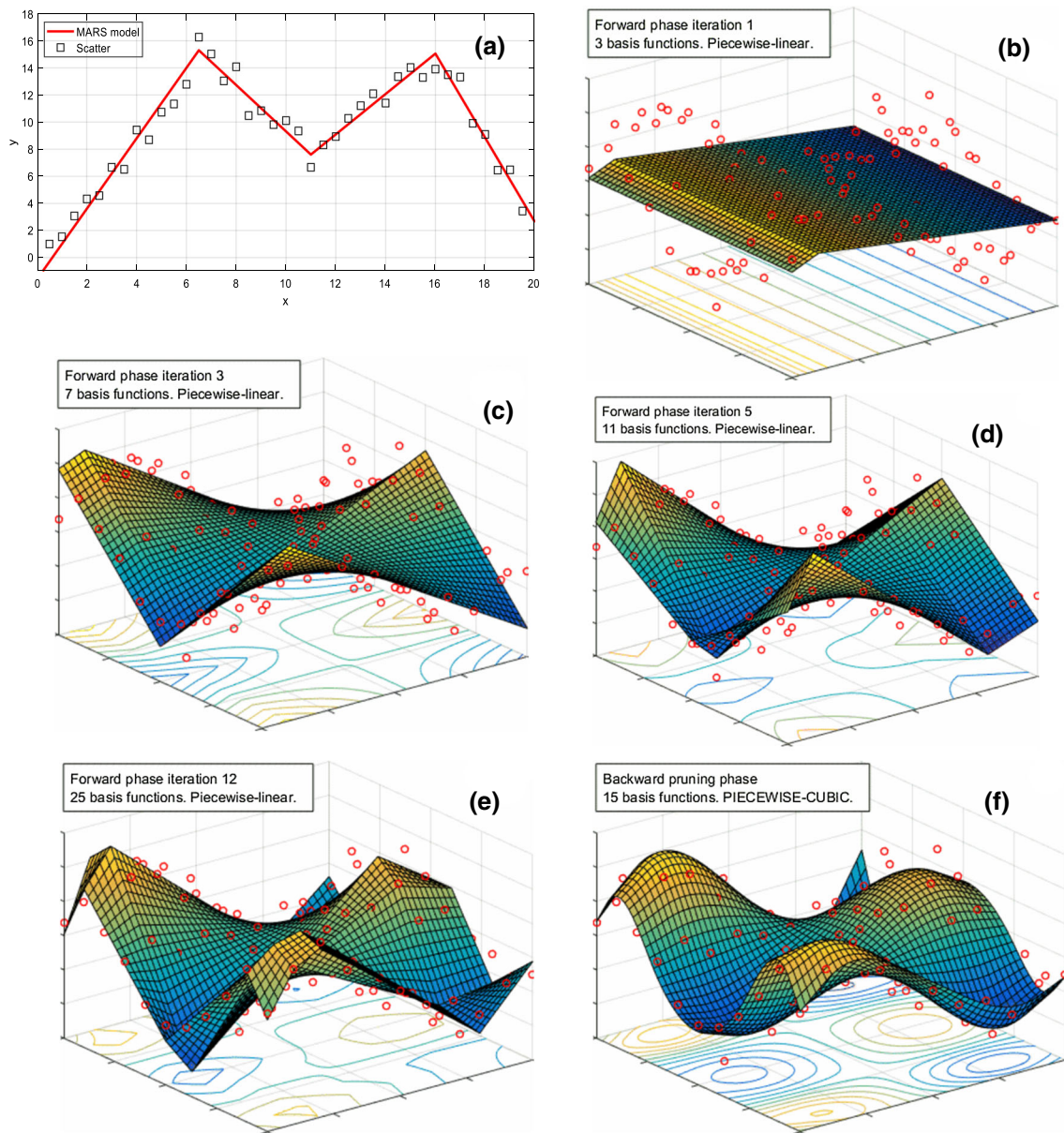
in which the coefficients  $\alpha$  can be obtained through the method of least squares.

Because a large number of spline functions are added over and over again in the forward constructive phase, the model becomes overfitting after all cases have been exhausted. Then in the backward pruning phase, the model is pruned by removing one least important redundant spline function at a time. The spline function that gives the least contribution to prediction is deleted to obtain an optimal submodel. The pruning is repeated to have a series of submodels until the model has only the intercept term. Then from these submodels of different size, the one with the lowest generalized cross-validation (GCV) value is chosen as the final MARS model. GCV can be evaluated as:

$$\text{GCV} = \frac{1}{n} \frac{\sum_{i=1}^N [O_i - f(x_i)]^2}{\left[1 - \frac{C(M)}{n}\right]^2} \quad (6)$$

in which  $O_i$  is the observed values,  $n$  is the number of observations in the training set,  $M$  is the number of spline functions, and  $f(x_i)$  denotes the predicted values of the MARS model. The numerator is the mean squared error





**Fig. 1** **a** Knots and linear splines for a simple MARS example. **b–f** The forward phase of MARS model to simulate the surface of  $z = \sin(0.83 \cdot \pi \cdot x) \cdot \cos(1.25 \cdot \pi \cdot y)$  and the final frame after backward pruning phase and after the model is turned into piecewise cubic

(MSE) of a candidate model. In the denominator,  $C(M) = M + d \times (M - 1)/2$  is the penalty function considering the larger uncertainty with increasing model complexity.  $d$  is the penalizing parameter with a default value of 3. The GCV can reduce the chance of overfitting by penalize both the error of the model, the number of spline functions and number of knots. The removal of an extraneous spline function in each pruning step has the chance to reduce the GCV of the model.

Figure 1b–e shows how the MARS model tries to fit the surface of  $z = \sin(0.83 \cdot \pi \cdot x) \cdot \cos(1.25 \cdot \pi \cdot y)$  in the forward phase by adding more and more basic functions. Figure 1f shows the shape of final MARS model after

backward pruning phase and after the model is turned into piecewise cubic. It can be clearly seen that the MARS model is able to accurately learn the nonlinear relationship.

After the ‘best’ MARS model is determined, the relative importance of the input variables can be assessed using the method called analysis of variance (ANOVA) decomposition [19]. This is implemented by evaluating the variance caused by removing all basic functions corresponding to that particular input feature. An input feature generating larger value of GCV and smaller value of  $R^2$  implies greater importance.

## 2.2 Deep forest methodology

Nowadays, in order to handle complicated learning tasks such as computer vision problems, deep learning methods have been widely used due to their effectiveness and feasibility. Recently, many of the popular machine learning models are based on deep neural networks (DNN) structure, which consist of multiple layers of parameterize differentiable nonlinear modules that can be trained by backpropagation. However, there are some shortages about a DNN model, such as:

1. It needs a larger number of training data sets to obtain acceptable results.
2. The DNN structure is complex and requires a large amount of calculations.
3. DNN has a lot of hyper-parameters, and the training performance heavily depends on the tuning process of such parameters, which makes the training process very time-consuming and difficult.

In order to solve these problems, Zhou and Feng developed a deep forest model [73], which is also named as multi-grained cascade forest (gcForest). It has the following advantages:

1. Performance is highly competitive to deep neural networks in a broad range of tasks.
2. It has fewer parameters, simpler training process, and adaptive forest level.
3. Training speed is much faster.
4. It can perform well on small-scale data sets.
5. The tree structure is easier to analyze and understand theoretically.

The deep forest algorithm is a supervised ensemble machine learning algorithm based on the random forest (RF) algorithm. The random forest regression (RF), proposed by Breiman [4], is an ensemble method which gives the prediction based on the average of a series of decision trees trained from the supervised data. RF uses random sampling and random construction of tree nodes to improve generalization ability and avoid overfitting; therefore, it has been widely used as a powerful regression and classification tool in the field of geotechnical engineering [63, 69, 72].

Deep forest employs a cascade structure, where each level of cascade receives feature information processed by its preceding level, and outputs processed results to the next level. This cascade structure is inspired by the deep neural networks which relies on the layer-by-layer processing of raw features. Each layer of the cascade structure is an ensemble of decision tree forests. During the cascading process, each time before building a new layer, the

performance of the entire cascade structure will be evaluated on the cross-validation data set. If there is no significant effect enhancement, the training process is terminated, so the number of layers is determined automatically. This automated network construction makes it very suitable for discrete or tabular data.

As a prediction algorithm based on decision trees with a certain depth, the deep forest algorithm divides the prediction process into two stages: the multi-grained scanning stage and the cascade forest stage.

### 1. The multi-grained scanning stage

The multi-grained scanning process uses sliding windows of different sizes to extract features from the raw input, which can generate new features with different dimensions, thereby improving the model's learning ability. It is similar to the convolutional neural network, which is to group multiple adjacent features to yield new features by taking the relationship between the features into account.

The multi-grained scanning stage is divided into two processes, feature scanning and feature transformation, as shown in Fig. 2. Assume that  $n$  raw features are entered, with the size of the sliding window,  $m$ , and the sliding step size,  $l$ . The sliding window scans the original input feature to extract feature information, then  $p$  new features will be generated as:

$$p = \frac{(n - m)}{k} + 1 \quad (7)$$

Then this  $p$  new features from the original raw feature are fed into a complete random forest and a random forest to generate corresponding class probability vectors. Each of the random forest and complete random forest will give a  $q$ -dimensional class probability vector as prediction output, and then, these class probability vectors are connected into an  $r$ -dimensional transformed feature vector, where  $L$  is given by Eq. (2).

$$r = 2 \times \left[ \frac{(n - m)}{k} + 1 \right] \times q \quad (8)$$

The dimension of transformed feature vector obtained by multi-grained scanning is much higher than the original raw feature; therefore, more information can be extracted and enhance the learning ability of the model.

### 2. The cascade forest stage

Figure 3 gives the illustration of the cascade forest structure. The cascade forest structure is composed of multiple layers of forests. This layer-by-layer training can enhance the ability to characterize the feature information. Each layer is an ensemble of different types of decision tree forests to encourage the diversity. For example, assume

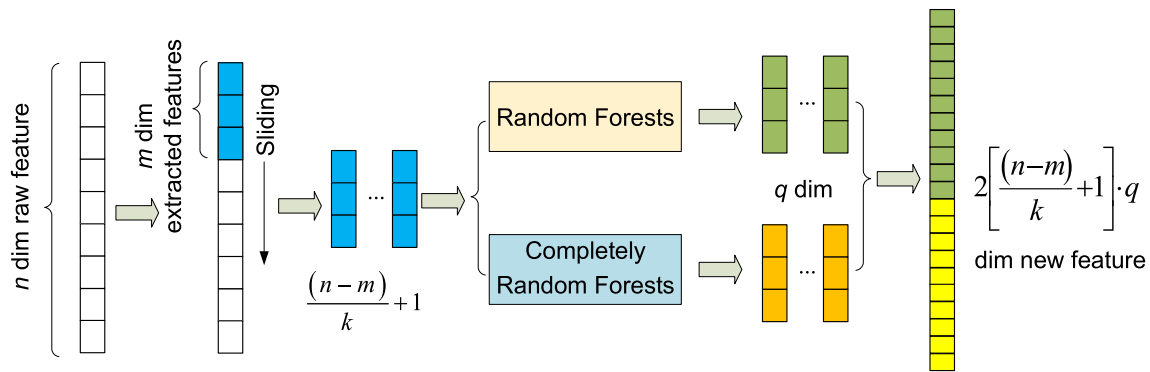


Fig. 2 Illustration of the multi-grained scanning stage

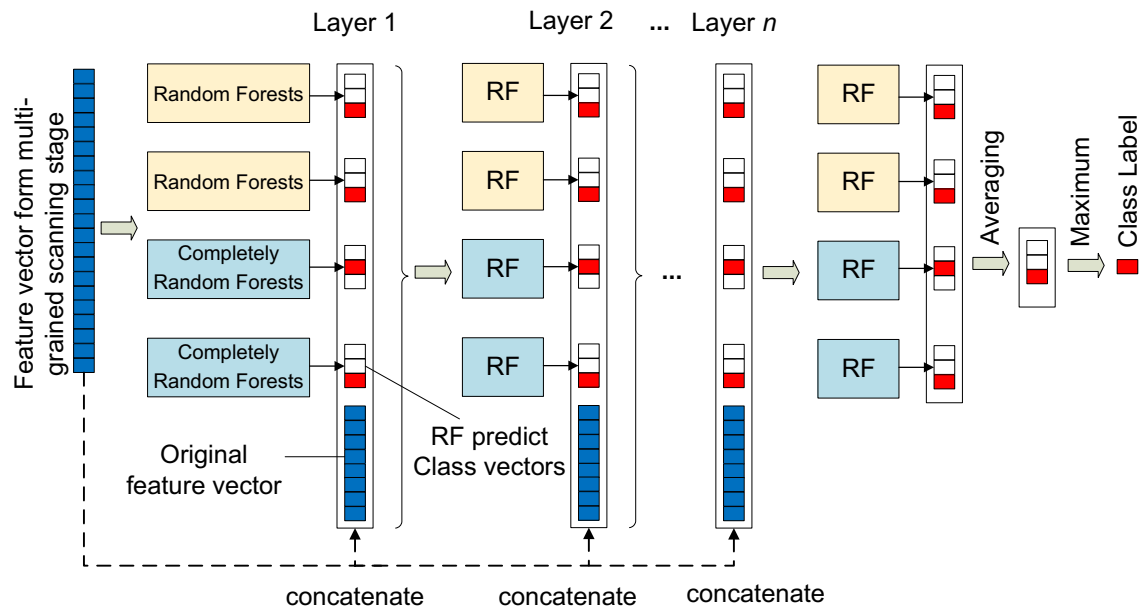


Fig. 3 The structure of the deep forest

that each layer of the cascade consists of two random forests (black) and two completely random tree forests (blue). The use of different types of decision tree forests is to introduce the diversity, which is crucial for ensemble construction.

The input feature vector in the first layer of the cascade forest is the feature vector generated in the multi-grained scanning stage. Suppose there are three classes to predict; thus, each forest in the first layer in the cascade forest will output a three-dimensional class vector, which is then concatenated for re-representation of the original input and fed into the next level of cascade. After training of each layer, the generated class vector and the original feature vector are concatenated as the input of next layer. After training layer by layer, the final prediction will be obtained by aggregating the four three-dimensional class vectors at the last level, and taking the class with the maximum aggregated value.

To reduce the risk of overfitting, class vector produced by each forest is generated by  $k$ -fold cross-validation. If there is no significant performance improvement after adding a certain number of cascade forest layers, the training process is terminated, so the number of layers is automatically determined. This adaptability of the layers can make the model suitable for samples of different sizes. Compared with DNN model, the deep forest cascade structure has obvious advantages. First of all, the parameters of deep forest mainly rely on the parameters of basic component classifier such as random forest or completely random number forest, which are much less than the hundreds of tuning parameters of deep neural networks. Second, the number of cascade forest layer is adaptively completed without additional manual intervention. Finally, deep forest provides a cascade framework that provides an important reference for the deep learning to be realized with non-differentiable modules [73].

### 3 Data collection and statistical analysis

#### 3.1 Description of the database.

As shown in Table 2, the rockburst prediction database used in this study contains 344 data record from previous research [60]. The index properties of rocks play a crucial role in rockburst intensity prediction of underground projects. A criterion for susceptibility to brittle spalling was suggested by Diederichs [11], which relates the rheology of the rock to the potential for brittle behavior versus a more ductile shear behavior. Where spalling failure potential is

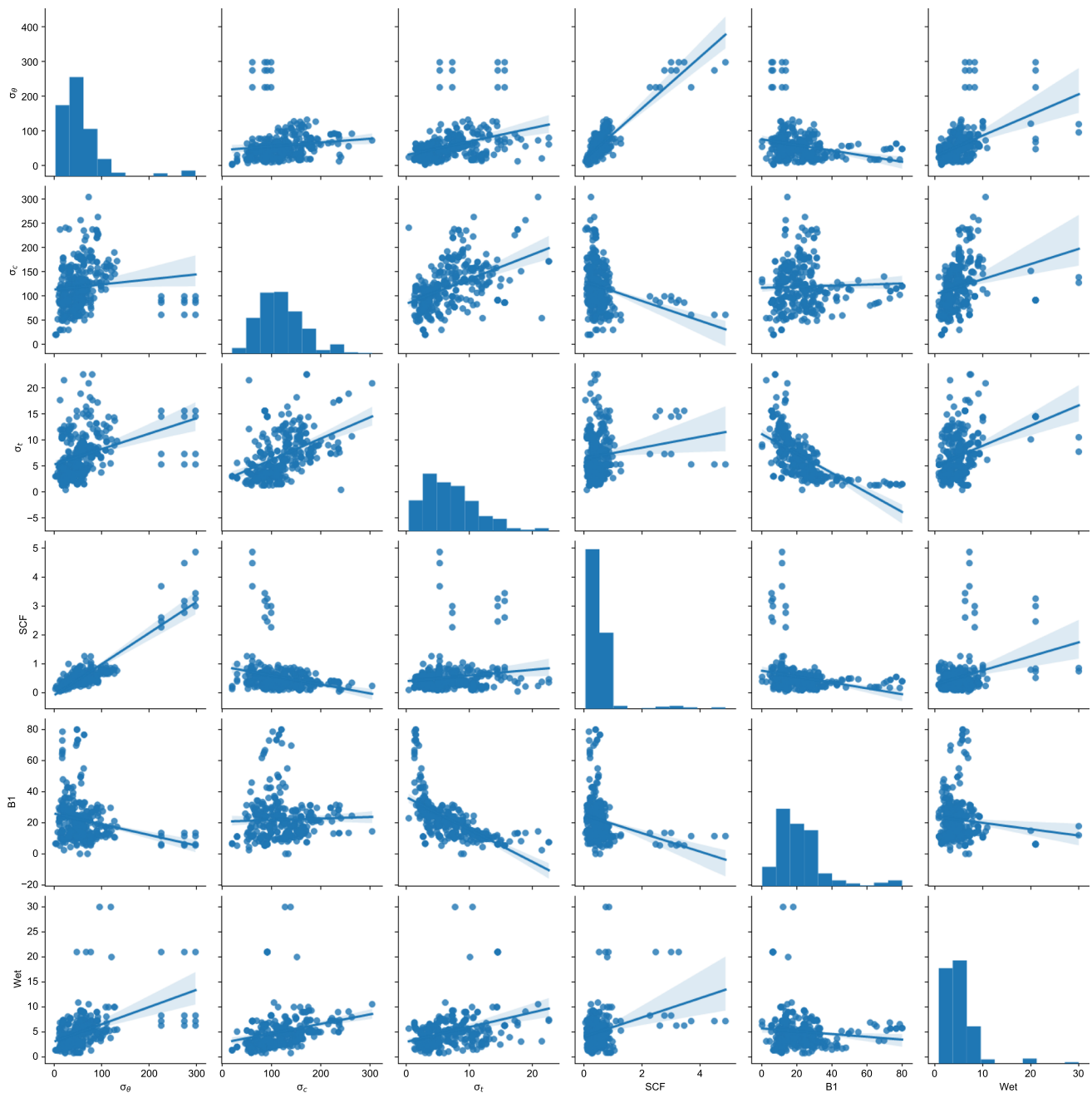
**Table 3** Statistical summary of input variables in the database

	$\sigma_\theta$	$\sigma_c$	$\sigma_t$	SCF	$B_1$	$W_{et}$
Mean	57.73	119.34	7.00	0.54	22.05	5.12
Std	48.07	46.91	4.20	0.58	14.00	3.66
Min	2.60	20.00	0.40	0.05	0.15	0.81
25%	29.88	86.30	3.60	0.28	12.58	2.96
50%	49.08	114.00	6.24	0.42	20.40	4.75
75%	67.20	150.00	9.40	0.61	26.82	6.30
Max	297.80	304.20	22.60	4.87	80.00	30.00
Skewness	2.99	0.66	0.99	4.45	1.97	3.44
Kurtosis	11.29	0.62	0.96	23.46	5.03	17.14

**Table 2** Summary of the original rockburst data

Cited from	Record range	Example no.	Rock type	$\sigma_\theta$ / Mpa	$\sigma_c$ / Mpa	$\sigma_t$ / MPa	SCF	$B_1$	$W_{et}$	Rockburst intensity
Zhou et al. [75]	1–246	1	Granodiorite	90	170	11.3	0.53	15.04	9	Moderate
		2	Syenite	90	220	7.4	0.41	29.73	7.3	Light
		10	Lead and Zinc ore	108.4	140	8	0.77	17.5	5	Strong
		246	Granite porphyry	57.97	70.68	4.19	0.6	25.51	2.87	Light
Xue et al. [61]	247–266	247	Biotite granite	98.02	148.52	6.66	0.66	22.3	3.23	Moderate
		248	Biotite granite	116.88	162.33	12.3	0.72	13.2	5.23	Strong
		249	Biotite granite	43.21	116.78	3.93	0.37	29.73	3.52	Light
		265	Biotite granite	35.34	95.5	2.26	0.37	42.3	2.75	None
		266	Biotite granite	39.11	105.7	2.83	0.37	37.35	3.08	Light
Pu et al. [46]	267–278	267	Kimberlite	18.17	49.1	1.56	0.37	31.4	3.3	Moderate
		268	Kimberlite	21	60	3.17	0.35	18.9	1.7	Moderate
		274	Kimberlite	35.16	79.9	3.12	0.44	25.6	2.5	Light
		278	Kimberlite	29.12	57.1	3.34	0.51	17.1	2.2	Moderate
Liu et al. [39]	279–294	279	N/A	30.9	238	7.6	0.13	31.2	7.4	None
		280	N/A	64.5	215	9	0.3	24	6.6	Moderate
		281	N/A	75.5	151	18.2	0.5	8.3	3.1	Light
		294	N/A	132.4	172	9.8	0.77	17.5	5.5	Strong
Jia et al. [30]	295–300	295	Fused tuff	14.49	70.21	2.04	0.21	34.42	3.11	Light
		296	Sandstone	17.79	81.8	2.45	0.22	33.39	3.6	Light
		298	Micro-weathered tuff	25.62	99.6	3.42	0.26	29.12	4.6	Moderate
		300	Weak-weathered tuff	26.68	46.8	1.8	0.57	26	2.7	Moderate
Du et al. [14]	301–307	301	Griotte	30.9	82.56	6.5	0.37	12.7	3.2	Light
		302	Siltstone	61	171.5	22.6	0.36	7.59	7.5	Light
		303	Skarn	89	128.6	13.2	0.69	9.74	4.9	Strong
		307	Syenite	91.3	225.6	17.2	0.4	13.12	7.3	Strong
Wu et al. [58]	308–314	308	Riassic marble	67.73	112.5	4.1	0.6	27.44	6.02	Moderate
		309	Riassic marble	64.43	112.5	3.4	0.57	33.09	5.8	Light
		311	Riassic marble	54.51	99.59	4.1	0.55	24.29	3.26	Strong
		314	Riassic marble	68.98	112.5	4.1	0.61	27.44	5.32	Moderate
Li et al. [36]	315–344	315	Mixed gneiss	48.7	180	8.3	0.27	21.69	5	Strong
		321	Mixed gneiss	59.9	200	9.8	0.3	20.41	5	Moderate
		326	Mixed gneiss	18.7	179	5.7	0.1	31.4	7.4	Light
		344	Mixed gneiss	62.5	165	9.4	0.38	17.55	9	Moderate

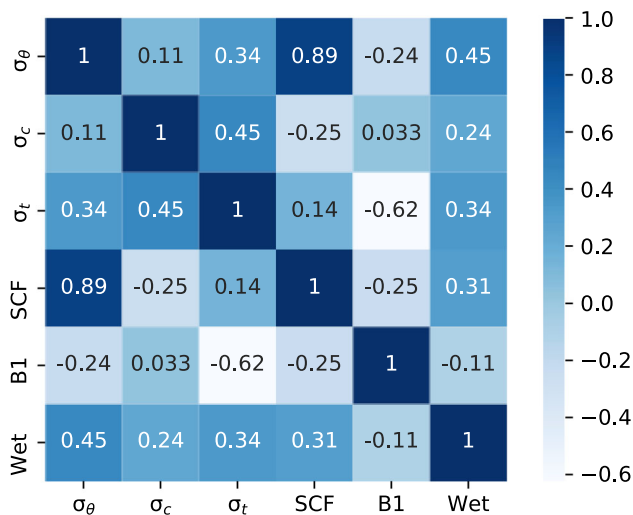




**Fig. 4** The relationship of the input features and the histogram of the frequency distribution of the input features

low, it is likely that shear processes will dominate, resulting in squeezing behavior rather than bursting at depth. Spalling can occur in lower strength rocks without strain bursting (low potential energy). Strainburst potential can be dictated by  $\sigma_c$  as stronger rocks allow for a larger build-up of strain energy. The elastic strain energy storage index  $W_{et}$  was defined by Kidybinskia (1981) as the proportion of strain energy retained to that dissipated during a single loading–unloading cycle in a uniaxial compression test [32]. The rock with high  $W_{et}$  value fails in a violent and

dynamic mode and has high rockburst intensity. Russenes (1974) devised a classification graph of rockburst intensity by taking the ratio of maximum tangential stress around the excavation ( $\sigma_\theta$ ) and the strength of the rock which can be expressed by the uniaxial compressive strength  $\sigma_c$  [47]. Martin et al. (1999) defined this index as stress concentration factor ( $SCF = \sigma_\theta / \sigma_c$ ). Peng et al. (1996) proposed a rock brittleness index as  $B_1$  ( $B_1 = \sigma_c / \sigma_t$ ) [43]. Case histories reveal that rockburst in brittle rock may occur when the SCF or  $B_1$  reaches certain values [43]. The maximum



**Fig. 5** Spearman rank correlation coefficient matrix heatmap of the input variables

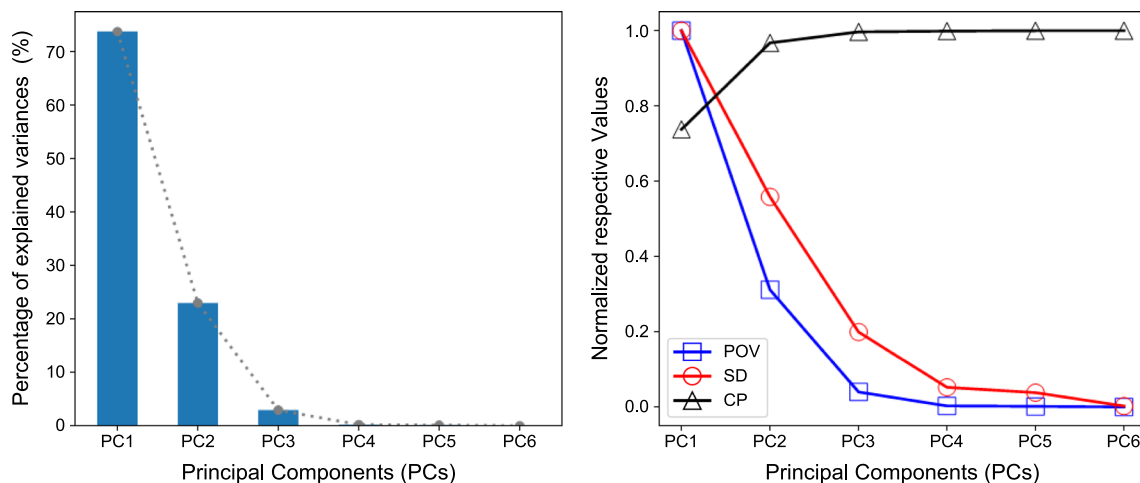
tangential stress around the excavation ( $\sigma_\theta$ ) is a key factor that is affected by the rock stress, groundwater, shape, and diameter of excavation.

Although there are more than twenty rockburst indicators in the literature, due to the limited ability of this study and the attempt to collect more records into the database,

only the six most influential and most commonly appeared input features, i.e., the maximum tangential stress  $\sigma_\theta$ , the uniaxial compressive strength  $\sigma_c$ , the uniaxial tensile strength  $\sigma_t$ , the elastic energy index  $W_{et}$ , and the rock brittleness index  $B_1$  were considered as input parameters.

The four rockburst intensity labels are none, light, moderate, and high, as listed in the last column of Table 2. Table 3 shows the statistical summary of input variables in the database. Figure 4 shows the relationship of the input features and the distribution of the input features. It can be seen that SCF has the highest skewness, and the distribution of  $\sigma_c$  is relatively close to the normal distribution.

Although the training samples are collected from reliable studies, there are some potential deficiencies making the rockburst intensity labels inconsistent. First of all, as shown in the first column in Table 2, these data are collected by a wide variety of researchers within a long time range. (For example, the data collected by Zhou et al. [75] were mainly published between 1991 and 2013. The other 98 data sets recording recent studies were mainly reported between 2006 and 2019). Therefore, it is unknown which classification criterion was adopted by relevant projects. Second, different numbers of ranks are adopted under different classification criteria. For example, Russiauses criterion has three or five rankings while the criteria in



**Fig. 6** The relative importance of the principal components

**Table 4** Principal component analysis result

Particulars	PC1	PC2	PC3	PC4	PC5	PC6
Standard deviation	45.81	25.58	9.13	2.39	1.74	0.11
Proportion of variance	73.77%	22.99%	2.93%	0.20%	0.11%	0.00%
Cumulative proportion	73.77%	96.76%	99.69%	99.89%	100.00%	100.00%
Considered in analysis	Y	Y	Y	N	N	N

China and some other countries have four rankings. Lastly, even the rankings of two records are the same, but they may correspond to different real rockburst intensity under different criteria. For the Russian criteria, the 3rd ranking corresponds to the most critical rockburst, but for Chinese criteria, 3rd ranking is not the most severe. The machine learning models for rockburst prediction trained by input training samples with inconsistent labels may not give reliable outcomes. In order to overcome this, the original rockburst intensity labels were ignored and relabeled using the clustering method of Gaussian mixture model.

Figure 5 shows the Spearman rank's correlation coefficient matrix heatmap of the input variables. It can be obviously observed that the correlation coefficient between  $\sigma_\theta$  and SCF is highest as 0.89, which indicates a very strong positive correlation. (Generally speaking,  $|R|=0$  implies an uncorrelated relationship;  $|R|<0.2$  implies a very weak correlation;  $0.2<|R|<0.4$  stands for a weak correlation;  $0.4<|R|<0.6$  stands for a moderate correlation;  $0.6<|R|<0.8$  implies a strong correlation;  $0.8<|R|<1$  implies a very strong correlation;  $|R|=1$  implies fully correlated.) Also, it can be found that  $B_1$  and  $\sigma_t$  have a medium negative correlation as their correlation coefficient is  $-0.62$ . When the correlation between the feature variables is strong, there will be a so-called multicollinearity problem which will make the ordinary least-squares type algorithm create inaccurate estimates of the regression coefficients. Figure 6 and Table 4 show the importance of the six principal components using the principal component analysis (PCA) method. As shown, the first three principal components (PC1, PC2 and PC3) play more than 99% of the role in the six PCs. Therefore, the information from the original six input features can be expressed by the first three PCs. What's more, this three-dimensional embedded space can guarantee a straightforward visualization and retain the original data characteristics as many as possible.

### 3.2 Dimensionality reduction and clustering of the data

Dimensionality reduction is a popular technique widely used in data mining nowadays. It is capable of distilling the origin information into a reduced number of new variables by a combination of closely statistically related variables. After dimensionality reduction, computational complexity can be reduced, while efficiency and accuracy of the model can be improved. There are several dimensionality reduction methods, which are praised and accepted by current data analysts, such as principal component analysis (PCA) [26], classical multidimensional scaling (MDS) [53], linear discriminant analysis (LDS) and the  $t$ -distributed stochastic neighbor embedding ( $t$ -SNE) method [75]. The  $t$ -SNE method was adopted in this study because it can give better results for visualization in order to understand the structure of the data intuitively. It can reveal characteristics of data more effectively, such as how many data sets are roughly clustered in each cluster, which data clusters into one cluster and which clusters are relatively far apart, etc. As mentioned, the relationships between rockburst intensity and input features are highly nonlinear,  $t$ -SNE has better performance than some other dimensionality reduction methods because it converts affinities of original data points to probabilities instead of extracting representative elements by linear transformation.

After the dimensionality reduction in the data, the Gaussian mixed model (GMM) was used to perform clustering and relabeling of the rockburst intensity. GMM clustering assumes that the data can be seen as generated from several Gaussian distributions, and the data points from the same distribution are classified into the same category [51]. The result of the GMM algorithm returns the probability that a data point belongs to different categories. GMM clustering are more flexible than the commonly used k-means clustering, for example:

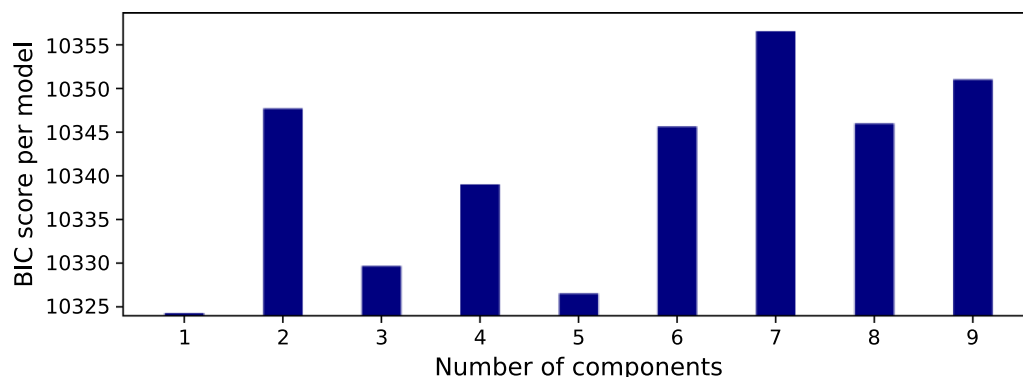
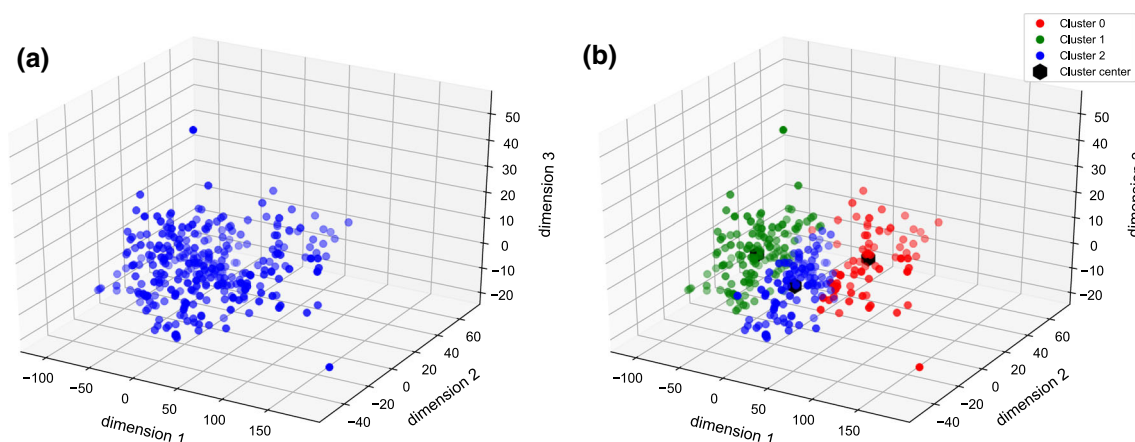
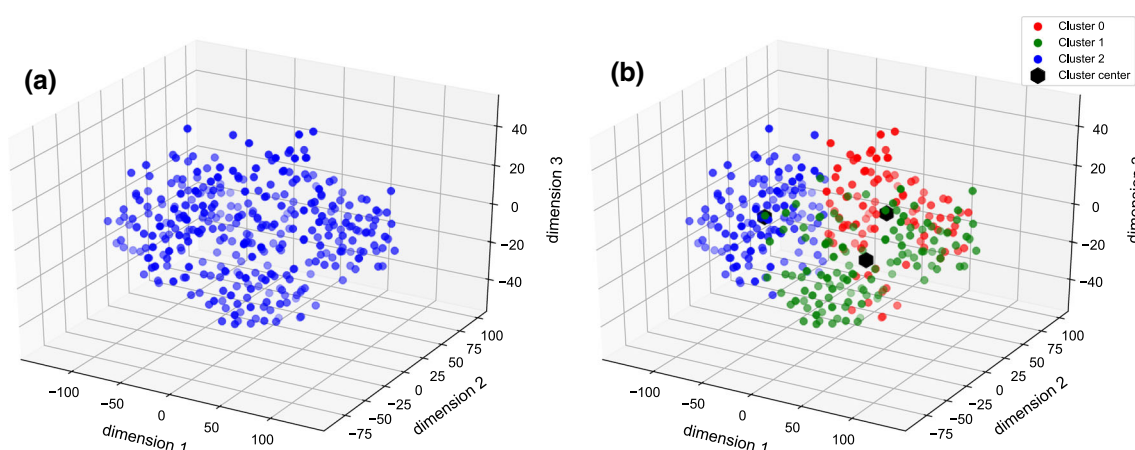


Fig. 7 BIC score for models with different number of components



**Fig. 8** **a** Data visualization after dimensionality reduction with PCA and **b** after clustering with the Gaussian mixture model



**Fig. 9** **a** Data visualization after dimensionality reduction with *t*-SNE and **b** after clustering with the Gaussian mixture model

**Table 5** New rockburst intensity labels after clustering

Number of rockburst case	Previous rockburst intensity label	Label after Clustering
1	M	1
2	L	0
3	L	0
4	M	0
5	M	1
...	...	...
342	H	2
343	H	2
344	M	0

1. GMMs uses the mean and standard deviation of the data, and the clusters can be elliptical rather than limited to circles. k-means is a special case of GMMs, because for k-means when the variance is close to 0 in all dimensions, the clusters will be circular. For the

rockburst problems, the interactions between input features are nonlinear, which will make the k-means algorithm ineffective.

2. k-means clustering is not good for clustering imbalanced sample categories, while the GMM takes the weight for each category into account.
3. The interpretability of k-means clustering is poor because it directly divides the sample space to find the optimal surface. GMM starts from the distribution of the sample itself and calculates the joint probability distribution to obtain the classification result. The interpretability of GMM is better than that of k-means.
4. k-means is a 'hard' classification which return the result of 0 or 1, while GMM is a 'soft' classification give the probability of a point belongs to different categories.

The main task of clustering is to group a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups. The cluster number  $k$  (number of components in GMM model) should

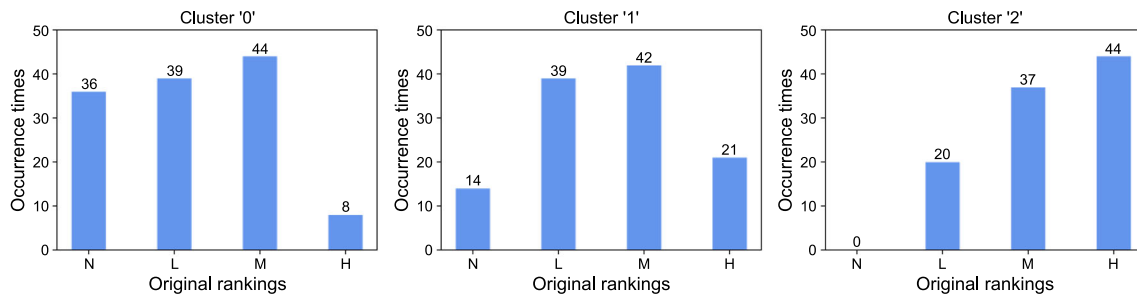


Fig. 10 Distribution of original rockburst ranking in the three clusters

be determined before clustering. In this study, the Bayesian information criterion (BIC) was used to help the selection of the number of components [75]. Given any two estimated models, the model with the lower value of BIC is the one to be preferred. BIC can be regarded as a metric to select the number of components, which is given in Eq. (9)

$$\text{BIC} = -2 \cdot \ln(L) + k \cdot \ln(n) \quad (9)$$

In which  $n$  is the number of observations, or equivalently, the sample size;  $k$  is the number of free parameters to be estimated.  $L$  is the maximized value of the likelihood function for the estimated model.

Figure 7 shows the BIC score for models with a different number of components from 1 to 9. The number of components equals to one means, there is only one cluster and no classification at all. Therefore, this option is unsuitable. In the existing rockburst ranking criteria, the maximum rank equals to 5. Because most of the criteria use ranks less than 5, the number of components equals to and greater than 5 are not considered here. The component number equals to 2 has the lowest BIC score among component's number equal to 2, 3 and 4. Generally speaking, 4 ranks corresponds to rockburst intensity level of 'High,' 'Moderate,' 'Low,' and 'No.' But the rockburst intensity of 'No' and 'Low' may be ambiguous among different criteria. Considering the above factors and the results of previous research [46], the number of components (clusters) was set to be 3 in this study.

Figure 8 shows the data visualization after dimensionality reduction with PCA and clustering with the GMM model. It can be found that some data points are far way from the cluster centers, and the three clusters are too close to each other. In contrast, from the results using  $t$ -SNE method (Fig. 9), it can be observed that the data points are distributed more evenly, the centers of each cluster are far apart, and the three clusters are separated in a more satisfactory way.

Table 5 shows the previous rockburst intensity labels and the new labels using the clustering result. The new labels are expressed using number '0,' '1,' '2,' denoting which cluster they belong to. For original rankings, 'N'

means there were no rockburst recorded, or there were only a few insignificant rockburst. 'L' indicates that the rockburst intensity was low. 'M' denotes the recorded rockburst intensity was 'Moderate,' while 'H' stands for high and severe rockburst. As mentioned before, the original robust rankings were probably defined using different criteria or codes, which may generate inconsistent results. For this reason, the GMM clustering method was employed to obtain consistent new labels based on original input features.

Figure 10 shows the distribution of original rockburst ranking in the three new clusters generated by GMM method. For totally 344 rockburst cases, the number of cases being classified to cluster '0,' cluster '1' and cluster '2' are 127, 116, and 101, respectively. For cluster '0,' the percentage of data with original label 'No rockburst,' 'Low rockburst,' and 'Moderate rockburst' are 28.3%, 30.7%, and 34.6%. Meanwhile, only 6.3% cases belong to 'High rockburst.' It can be noticed that the majority of records in cluster '0' belongs to relatively minor rockburst intensity. In contrast, over 80% of the data in cluster '2' belongs to relatively stronger rockburst intensity: there are 43.6% for 'High rockburst,' 36.6% for 'Moderate rockburst,' while there are only 19.8% for 'Low rockburst' and none for 'No rockburst.' The distribution in cluster '1' is between cluster '0' and cluster '2': the frequencies are 12.1%, 33.6%, 36.2%, and 18.1% for 'N,' 'L,' 'M,' and 'H,' respectively.

It should be noticed that 'No rockburst' always represents the lowest intensity and 'High rockburst' stands for the strongest intensity for common practice in evaluation criteria. What's more, the rockburst intensity of 'No' and 'Low' may be difficult to distinguish precisely. Based on the above considerations, cluster '2' definitely should be defined as the most severe rockburst intensity group, because 44% cases belongs to the primary 'high rockburst' ranking. The rockburst ranking of cluster '0' should be lower than cluster '1.' On the one hand, 28% cases in cluster '0' are 'No rockburst,' which is higher than the number of is 12% in cluster '1.' On the other hand, for 'High rockburst,' cluster '0' has '6%' cases, which are less than '18%' of cluster '1.' This means cluster '0' has more



percentage of data in low rockburst intensity and less data in strong rockburst intensity comparing to cluster '1.' Finally, the relative rockburst ranking for these three clusters are defined as: cluster '0' represents 'Low' rockburst intensity, cluster '1' refers to 'Moderate' and cluster '2' stands for the 'High' rockburst intensity.

## 4 Development of the MARS and deep forest rockburst prediction model

### 4.1 Classifier performance measure indicator

Rockburst prediction can be considered as a multi-classification problem, and based on the input feature values, the rockburst intensity is evaluated as 'No'/'Low,' 'Moderate,' or 'High.' For classification problems, many evaluation metrics have been proposed to assess the model performance in previous studies [42]. In this work, the commonly used accuracy, sensitivity, precision, and F1-score were used as classifier performance metrics.

Accuracy (ACC) is the most frequently used criteria and it is easy to understand, revealing the percentage of correctly predicted cases in the total cases. Generally, the higher the ACC, the better the classifier. Sensitivity (SNS) represents the proportion of correctly predicted samples in all samples actually belongs to a certain class, which measures the classifier's ability to recognize a certain class. Precision (PRC) is a measure of accuracy in a certain class instead in the total data set, representing the proportion of correct predictions among all samples predicted as an ascertain class. In short, for a certain class, SNS evaluates how many correct results were given by the model, while PRC evaluates how much of the result given by the model is correct. Because SNS and PRC are suitable for different problems, F1-score ( $F_{1score}$ ) is adopted as a harmonic mean of the SNS and PRC. Since  $F_{1score}$  is always close to the smaller value of SNS and PRC, whenever SNS or PRC is low, the  $F_{1score}$  score can send a reminder.

The above-mentioned evaluation metrics were calculated by the confusion matrix. The confusion matrix is a table describing the performance of the classification model. It contains information about the actual and predicted classification performed by the classifier, and this information is used to evaluate the performance of the classifier. The original confusion matrix is only suitable for binary classification problems, but the rockburst prediction is a three-classification problem in this work. The multi-classification was transformed into a two-classification problem using the one-vs-rest strategy, that is, a certain class was taken as positive sample, and the rest of the class was considered as negative samples. For example, 'High Rockburst' can be considered as one class, while 'No

Rockburst,' 'Low Rockburst,' and 'Moderate Rockburst' can be put together as another class.

The explanation of binary confusion matrix is shown in Fig. 11. The confusion matrix is composed of four elements, which are denoted by TP (true positive), FP (false positive), FN (false negative), and TN (true negative). Take 'High Rockburst' as an instance, TP means both the actual and predicted label is 'High Rockburst.' FP means that the predicted label is 'High Rockburst' but the actual label is not. FN means the actual label is 'High Rockburst' but the predicted label is not. TN means both the actual and predicted label is not 'High Rockburst.'

After counting the number of TP, FP, TN, and FN, the four classifier performance metrics can be calculated as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$SNS = \frac{TP}{TP + FN} \quad (11)$$

$$PRC = \frac{TP}{TP + FP} \quad (12)$$

$$F_1 = 2 \times \frac{SNS \times PRC}{SNS + PRC} \quad (13)$$

ACC is a global metric, which can be calculated using the following expression:

$$ACC = \frac{\text{total correct prediction}}{\text{total correct prediction} + \text{total wrong prediction}} \quad (14)$$

		Predicted label	
		High rockburst	Low/Moderate rockburst
Actual label	High rockburst	TP (True Positive)	FN (False Negative)
	Low/Moderate rockburst	FP (False Positive)	TN (True Negative)

Fig. 11 Binary confusion matrix (taking high rockburst as an example)

## 4.2 Performance of the constructed MARS and DF model

In this study, the tenfold cross-validation method was used to test model performance and optimize the parameters. For independent test validation, the 344 data sets from the database were randomly shuffled and divided into 75% (258) training set and 25% (86) test set. For the MARS model, the most important parameter is  $\text{Max}_{\text{terms}}$ , which controls the maximum number of terms generated by the forward pass. Figure 12 gives the Accuracy, SNS, PRC, and  $F_{\text{score}}$  change curves of MARS model with parameter  $\text{Max}_{\text{terms}}$ . It can be seen that all these values increased with the increasing  $\text{Max}_{\text{terms}}$ , but the accuracy in the testing set reached the peak when  $\text{Max}_{\text{terms}}$  equals 60. Therefore, the optimal  $\text{Max}_{\text{terms}}$  was chosen as 60. The other main parameters for developing the MARS model are listed in Table 6.

It can be found that in terms of the rockburst database used in this study, the DF model was much more convenient for use, because there were fewer parameters to optimize and the default value generally gave the satisfactory results. The major parameter influenced model performance was  $n_{\text{trees}}$ , which is the number of trees in each estimator. Figure 13 gives the Accuracy, SNS, PRC, and  $F_{\text{score}}$  change curves of MARS model with parameter  $n_{\text{trees}}$ . It can be seen that these values slightly increased with the increasing  $n_{\text{trees}}$ , but the Accuracy in the testing set reached a maximum value when  $n_{\text{trees}} = 80$ . Therefore the optimal  $n_{\text{trees}}$  was adopted as 80. The other main parameters for the developed DF model are listed in Table 6.

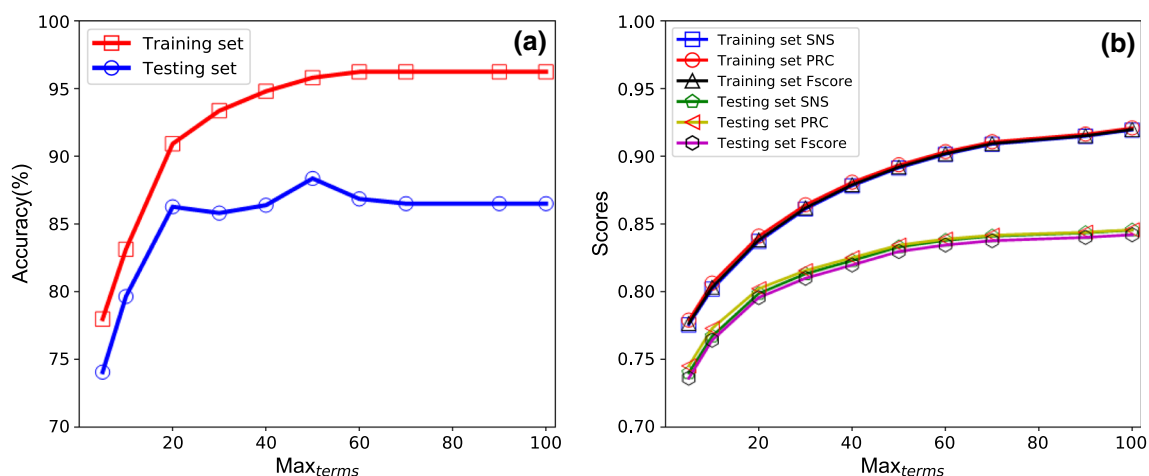
By comparing Figs. 10 and 11, it can be observed that the highest accuracy is sooner achieved by the deep forest model than the MARS model. First of all, this phenomenon should be attributed to the powerful 'ensemble' nature of

DF model. The two cores in ensemble learning are accuracy and diversity of the base classifier, while DF model is good at both. It is generally considered that random forest is the best classifier in a classification problem, and the DF uses the combination of different types of random forest model. At the same time, the diversity between the structures of the decision tree is also very significant, so their integration can achieve satisfactory results. Another reason is that the number of DF layers is adaptively determined. Each layer of DF model uses supervision information. When the next layer is constructed, the output of the previous layer has been determined. Therefore, once the performance improvement is not good, the process of building new layers is terminated. Consequently, DF has better robustness and lower parameter sensitivity, resulting in a faster training speed.

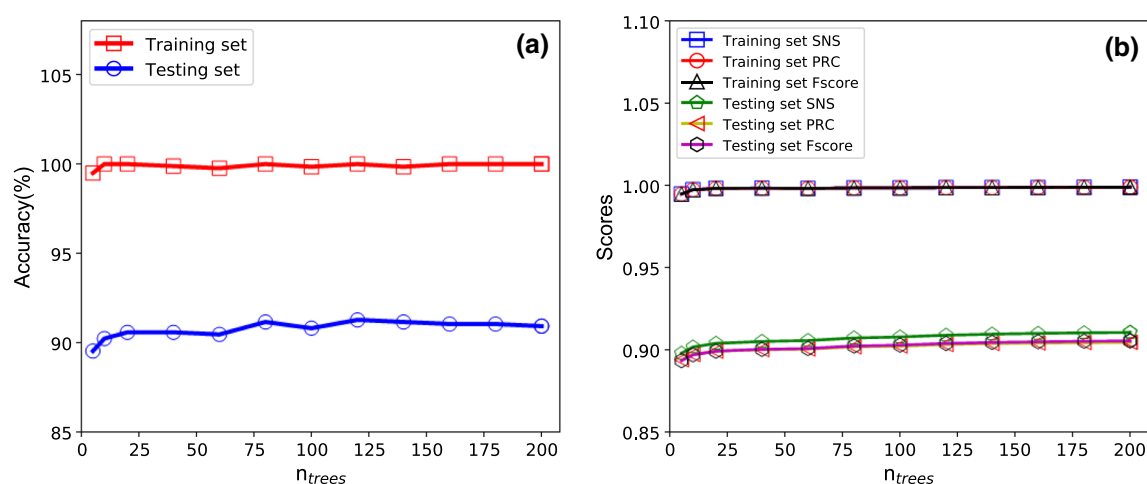
The classifier performance metrics including ACC, SNS, PRC, and  $F_1$  were calculated for the developed MARS model and DF model in both the training set and testing set, as shown in Tables 7 and 8, to test their applicability and practicability. The performances of MARS model and DF model were compared with the

**Table 6** The main parameters for the developed MARS and DF model

Model	Parameters	Value after optimization
MARS	$\text{Max}_{\text{terms}}$	50
	max_degree	10
	penalty	3
	minspan_alpha	0.5
DF	$n_{\text{trees}}$	80
	max_layers	20
	$n_{\text{estimators}}$	2



**Fig. 12** Classifier performance indicator change curve of MARS model with parameter  $\text{Max}_{\text{terms}}$



**Fig. 13** Classifier performance indicator change curve of deep forest model with parameter  $n_{trees}$

**Table 7** Evaluation metrics for the five classification models in training set

Model	ACC (%)	SNS			PRC			$F_1$		
		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
MARS	95.6	0.971	0.947	0.946	0.985	0.926	0.956	0.978	0.936	0.951
DF	99.8	1.000	1.000	0.995	0.998	0.998	1.000	0.999	0.999	0.997
RF	100.0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
XGB	99.9	0.995	0.999	1.000	1.000	1.000	0.992	0.997	0.999	0.996
ANN	90.0	0.918	0.894	0.902	0.932	0.889	0.890	0.925	0.891	0.896

**Table 8** Evaluation metrics for the five classification models in testing set

Model	ACC (%)	SNS			PRC			$F_1$		
		Low	Moderate	High	Low	Moderate	High	Low	Moderate	High
MARS	88.1	0.915	0.846	0.860	0.931	0.834	0.853	0.922	0.835	0.853
DF	90.8	0.918	0.896	0.919	0.945	0.878	0.892	0.930	0.885	0.903
RF	90.2	0.904	0.906	0.903	0.957	0.864	0.878	0.929	0.883	0.888
XGB	89.7	0.885	0.910	0.900	0.962	0.867	0.852	0.921	0.887	0.873
ANN	90.1	0.918	0.893	0.901	0.933	0.889	0.891	0.925	0.892	0.896

popular random forest model, extreme gradient boost (XGBoost) and ANN model. For a given class, the different combinations of precision (PRC) and sensitivity (SNS) give the information about classifier performance:

1. High precision + high sensitivity: the model can detect this class well;
2. High precision + low sensitivity: the model cannot detect this class well, but when it detects this class, the result is highly credible;
3. Low precision + high sensitivity: the model can detect this class well, but the detection results also contain points of other classes;
4. Low precision + low sensitivity: the model cannot detect this class well.

It can be seen from Tables 7 and 8 that both the MARS model and DF model have satisfactory precision and sensitivity score for all the ‘Low,’ ‘Moderate,’ and ‘High’ rockburst intensity, which means they can classify all the three class satisfactorily. The DF model has the highest accuracy of 90.8% in the testing set, which means that the DF model has the best generalization ability, and it is not prone to overfitting. In contrast, although the RF and XGB model have higher accuracy than the DF model in training set, they have lower accuracy than DF model in the testing set. The developed MARS model also has

**Table 9** Expressions of basis functions (BFs) for the MARS model

BF	Expression
BF <sub>1</sub>	$BF_1 = \max(0, 95.67 - \sigma_\theta)$
BF <sub>2</sub>	$BF_2 = \max(0, \sigma_c - 127.37)$
BF <sub>3</sub>	$BF_3 = \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>4</sub>	$BF_4 = \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>5</sub>	$BF_5 = \max(0, 45.92 - \sigma_\theta) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>6</sub>	$BF_6 = \max(0, \sigma_c - 194) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>7</sub>	$BF_7 = \max(0, 194 - \sigma_c) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>8</sub>	$BF_8 = \max(0, \sigma_c - 125.37) \times \max(0, 194 - \sigma_c) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>9</sub>	$BF_9 = \max(0, 125.37 - \sigma_c) \times \max(0, 194 - \sigma_c) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>10</sub>	$BF_{10} = \max(0, 124.15 - \sigma_c) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>11</sub>	$BF_{11} = \max(0, \sigma_\theta - 46.2) \times \max(0, 124.15 - \sigma_c) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>12</sub>	$BF_{12} = \max(0, 46.2 - \sigma_\theta) \times \max(0, 124.15 - \sigma_c) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>13</sub>	$BF_{13} = \max(0, \sigma_\theta - 90.99) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>14</sub>	$BF_{14} = \max(0, 0.9 - SCF) \times \max(0, 125.37 - \sigma_c) \times \max(0, 194 - \sigma_c) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>15</sub>	$BF_{15} = \max(0, \sigma_c - 106.8) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>16</sub>	$BF_{16} = \max(0, 8.6 - \sigma_t) \times \max(0, 90.99 - \sigma_\theta) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>17</sub>	$BF_{17} = \max(0, \sigma_\theta - 16.87) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>18</sub>	$BF_{18} = W_{et} \times B_1 \times \max(0, 0.9 - SCF) \times \max(0, 125.37 - \sigma_c) \times \max(0, 194 - \sigma_c) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>19</sub>	$BF_{19} = \sigma_\theta \times W_{et} \times B_1 \times \max(0, 0.9 - SCF) \times \max(0, 125.37 - \sigma_c) \times \max(0, 194 - \sigma_c) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>20</sub>	$BF_{20} = \sigma_\theta \times \max(0, 8.6 - \sigma_t) \times \max(0, 90.99 - \sigma_\theta) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>21</sub>	$BF_{21} = \max(0, 46.38 - \sigma_\theta) \times \max(0, 6.38 - W_{et}) \times \max(0, 90.99 - \sigma_\theta) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>22</sub>	$BF_{22} = \max(0, \sigma_\theta - 58.2) \times \max(0, 8.6 - \sigma_t) \times \max(0, 90.99 - \sigma_\theta) \times \max(0, \sigma_\theta - 45.92) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>23</sub>	$BF_{23} = \sigma_c \times \max(0, 58.2 - \sigma_\theta) \times \max(0, 124.15 - \sigma_c) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>24</sub>	$BF_{24} = \sigma_c \times \max(0, \sigma_\theta - 58.2) \times \max(0, 124.15 - \sigma_c) \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$
BF <sub>25</sub>	$BF_{25} = \sigma_c \times \max(0, \sigma_c - 69.2) \times \max(0, 95.67 - \sigma_\theta)$

acceptable performance, with the ACC = 95.6% in training set and ACC = 88.1% in testing set. The most important thing is that the MARS model can give the explicit expression comparing to the other 'black-box' type model and being mostly spreadsheet friendly.

Table 9 lists the basis functions (BFs) and their corresponding equation for the developed MARS model. It is observed from Table 9 that significant interactions have occurred between BFs (e.g., BF<sub>3</sub> and BF<sub>1</sub>, BF<sub>6</sub> and BF<sub>5</sub>). The presence of interactions suggests that the built MARS model is not simply additive and that interactions play a significant role in building an accurate model for predictions. The equation of MARS model is given by Eq. (15).

The process to predict the rockburst intensity using the proposed MARS model is as following:

(1) Calculate a rockburst intensity related parameter  $L_{MARS}$  using Eq (15):

$$\begin{aligned}
 L_{MARS} = & 2.0419 - 0.038396 \times BF_1 - 0.011074 \\
 & \times BF_2 + 0.004928 \times BF_3 - 0.00062402 \\
 & \times BF_4 + 0.00041579 \times BF_5 + 1.7307e - 05 \\
 & \times BF_6 + 2.0813e - 05 \times BF_7 - 7.8661e - 07 \\
 & \times BF_8 - 4.8075e - 07 \times BF_9 - 5.2951e - 05 \\
 & \times BF_{10} + 1.129e - 05 \times BF_{11} - 1.4546e - 06 \\
 & \times BF_{12} + 0.0067865 \times BF_{13} - 2.8687e - 06 \\
 & \times BF_{14} + 4.2365e - 05 \times BF_{15} + 0.00011755 \\
 & \times BF_{16} + 3.4187e - 06 \times BF_{17} - 1.7648e - 07 \\
 & \times BF_{18} + 2.7167e - 09 \times BF_{19} - 1.9574e - 06 \\
 & \times BF_{20} + 0.0045943 \times BF_{21} + 1.8573e - 06 \\
 & \times BF_{22} + 1.8204e - 08 \times BF_{23} - 1.5521e - 07 \\
 & \times BF_{24} - 4.3686e - 05 \times BF_{25}
 \end{aligned} \quad (15)$$

(2) Calculate a  $3 \times 1$  matrix  $\phi$  with three element  $\phi_1, \phi_2, \phi_3$  which are parameters related to the rockburst probabilities as:

$$\begin{aligned}\phi &= [\phi_1 \quad \phi_2 \quad \phi_3] \\ &= L_{\text{MARS}} \times [-12.0458 \quad -0.1264 \quad 12.1723] \\ &\quad + [10.6661 \quad 3.70110 \quad -14.367]\end{aligned}\quad (16)$$

Then the rockburst probability of a sample belong to 'Low,' 'Moderate,' and 'High' can be expressed as  $p_1, p_2$ , and  $p_3$ , which can be calculated from the following formula:

$$\begin{aligned}\text{Probability} &= \begin{bmatrix} \text{Low} & \text{Moderate} & \text{High} \\ p_1 & p_2 & p_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{e^{\phi_1}}{e^{\phi_1} + e^{\phi_2} + e^{\phi_3}} & \frac{e^{\phi_2}}{e^{\phi_1} + e^{\phi_2} + e^{\phi_3}} & \frac{e^{\phi_3}}{e^{\phi_1} + e^{\phi_2} + e^{\phi_3}} \end{bmatrix}\end{aligned}\quad (17)$$

Rockburst intensity

$$= \begin{cases} \text{Low} & \text{if } \max(p_1, p_2, p_3) = p_1 \\ \text{Moderate} & \text{if } \max(p_1, p_2, p_3) = p_2 \\ \text{High} & \text{if } \max(p_1, p_2, p_3) = p_3 \end{cases}\quad (18)$$

Take the first line of data in the database as example,  $\sigma_\theta = 90$ ,  $\sigma_c = 170$ ,  $\sigma_t = 11.3$ ,  $\text{SCF} = 0.53$ ,  $B_1 = 15.04$ ,  $W_{\text{et}} = 9$ , from Eq.(15), we can compute  $L_{\text{MARS}} = 1.355$ .

Then from Eq. (16):

$$\begin{aligned}\phi &= 1.355 \times [-12.0458 \quad -0.1264 \quad 12.1723] \\ &\quad + [10.6661 \quad 3.70110 \quad -14.367] \\ &= [-5.662 \quad 3.529 \quad 2.133]\end{aligned}$$

And from Eq. (17):

$$\begin{aligned}\text{Probability} &= \begin{bmatrix} \frac{e^{-5.662}}{e^{-5.662} + e^{3.529} + e^{2.133}} & \frac{e^{3.529}}{e^{-5.662} + e^{3.529} + e^{2.133}} & \frac{e^{2.133}}{e^{-5.662} + e^{3.529} + e^{2.133}} \end{bmatrix} \\ &= [0.000082 \quad 0.802 \quad 0.198]\end{aligned}$$

Finally from Eq. (18), when  $p_1 = 0.000082$ ,  $p_2 = 0.802$  and  $p_3 = 0.198$ , the rockburst intensity predicted by the MARS model can be referred as 'Moderate.'

To investigate which input feature has more important influence on predicting the rockburst intensity, the Shapley additive explanations method (SHAP) [42] was used in this study. SHAP computes Shapley values from coalitional

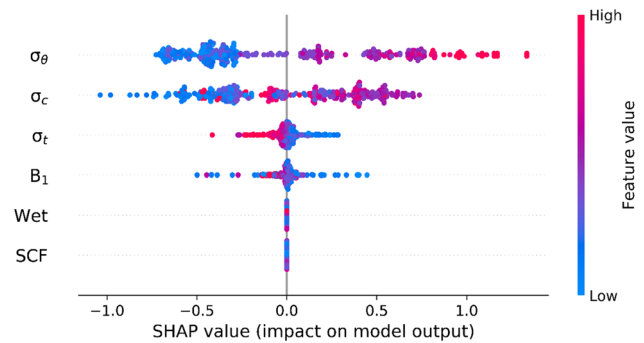


Fig. 14 SHAP summary plot showing the impacts of the input features for the developed MARS model

game theory. The feature values of a data instance act as players in a coalition. Shapley values tell us how to fairly distribute the 'payout' (equal to the prediction) among the features. In general, the features with a high positive SHAP value have more significant impact on the final prediction and tend to make the prediction with higher accuracy.

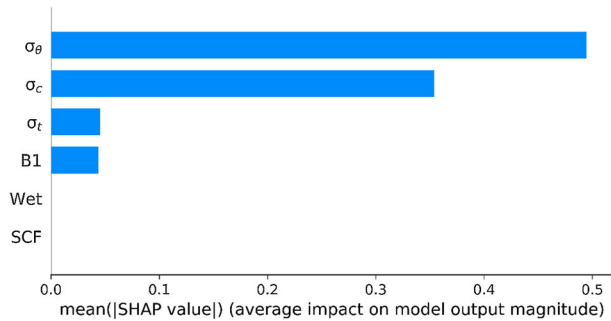
Figure 14 shows the summary plot about impact of the six input features  $\sigma_\theta$ ,  $\sigma_c$ ,  $B_1$ ,  $\sigma_t$ ,  $\text{SCF}$ ,  $W_{\text{et}}$  on the rockburst prediction of MARS model using the SHAP value. The scattered points on the graph represent all the samples. The arrangement of these six feature from bottom to top is based on the importance of features according to their SHAP values. A wide area indicates a large number of samples gathered around this range of value. The color on the right indicates the value of the feature: red indicates that the feature value is high, while blue indicates that the feature value is low. It can be found that the  $\sigma_\theta$  has the highest impact on the rockburst prediction, following by  $\sigma_c$  and  $\sigma_t$ . Also it can be noticed that for  $\sigma_\theta$ , a larger proportion of data points gathered around the zone with negative SHAP value about  $-0.5$  to  $-0.75$ , and these points are mainly with blue color. This means for  $\sigma_\theta$ , a lower  $\sigma_\theta$  value will make a lower rockburst prediction intensity. At the

same time, for  $\sigma_\theta$ , a few data points with bright red color distribute in the zone with high SHAP value about  $0.75$  to  $1.2$ . This suggests that the higher value of  $\sigma_\theta$  tends to yield a prediction of high rockburst intensity. The same trend can be observed for the input feature of  $\sigma_c$ . But for  $\sigma_t$ , the data points distributed in negative SHAP value zone are all red

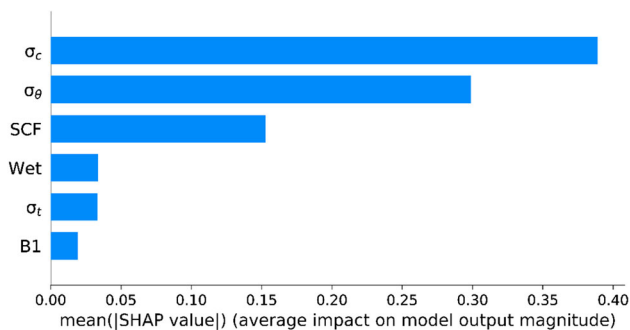


color, which implies that the higher tension strength of the rock will decrease the rockburst prediction intensity.

Figures 15 and 16 show the calculated relative importance of the input features for the developed MARS model



**Fig. 15** The relative importance of the input features for the developed MARS model

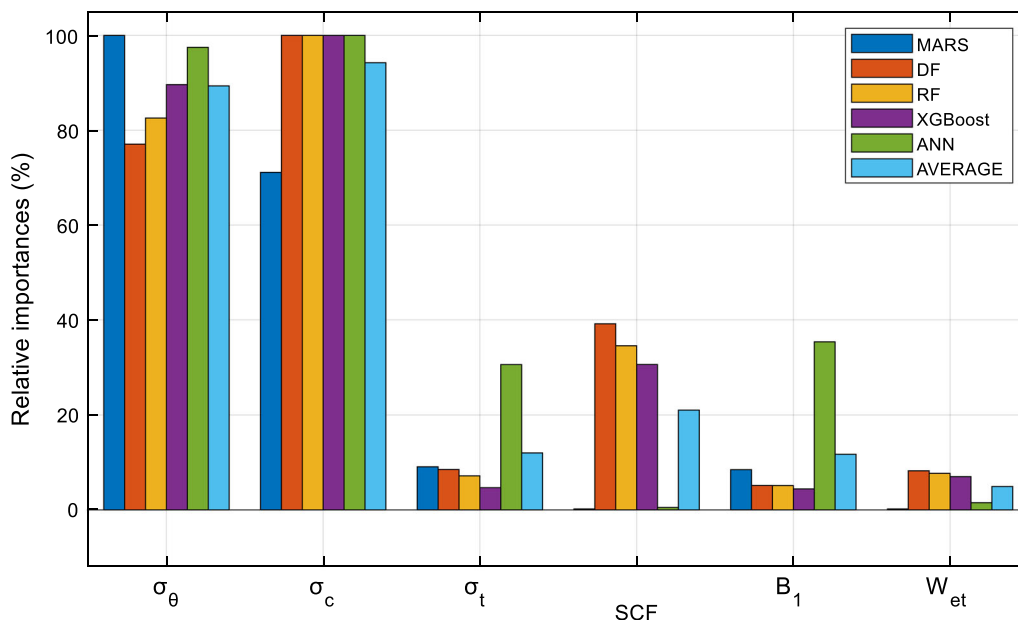


**Fig. 16** The relative importance of the input features for the developed DF model

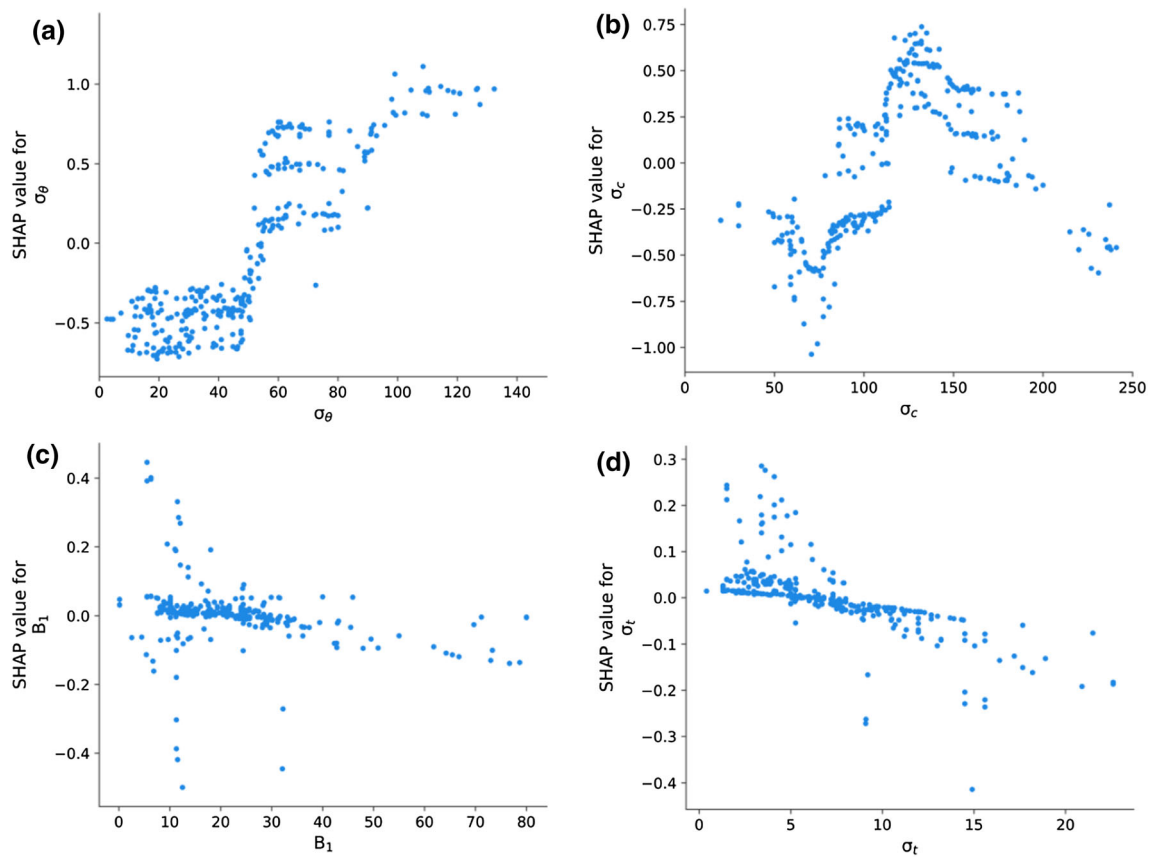
and DF model using SHAP method. Although the results from MARS and DF model are not completely the same,  $\sigma_\theta$  and  $\sigma_c$  are definitely among the number one and number two input features with the highest importance. The sequences of importance for the four remaining input features are not easy to determine. Therefore, the results from MARS, DF, RF, XGBoost and ANN model are compared and averaged, as shown in Fig. 17. Without any doubt,  $\sigma_\theta$  and  $\sigma_c$  are the most important features.

Figure 18 indicates the influence of a single input feature on the rockburst intensity prediction for the developed MARS model base on SHAP value. For example, in Fig. 18a, the x-axis represents the  $\sigma_\theta$  values, and for the input data points the  $\sigma_\theta$  values range from 0 to 140. Meanwhile, the y-axis represents the corresponding SHAP values for each of these data points which ranges from  $-1.0$  to  $1.0$ . It can be clearly found that the SHAP value increases with the  $\sigma_\theta$  almost linearly, which means that a higher  $\sigma_\theta$  value will lead to a stronger rockburst intensity. It coincides with the previous studies, because  $\sigma_\theta$  is the tangential stress around underground opening extracted from project site, and stronger rockburst tends to happen in high stress region.

Figure 18b shows the impact of  $\sigma_c$  on the rockburst intensity. It is obvious that the change of SHAP value with  $\sigma_c$  can be divided into two stages. When the value of  $\sigma_c$  is smaller than 150 MPa, the SHAP value increases with the  $\sigma_c$ , then SHAP value decreases with increasing  $\sigma_c$  when  $\sigma_c$  is higher than 150 MPa. It implies that when the uniaxial compressive strength  $\sigma_c$  of the rock is under 150 MPa, an increasing  $\sigma_c$  means a harder rock which can accumulate



**Fig. 17** Comparison of the relative importance of the input variables using MARS, DF, RF, XGBoost and ANN model

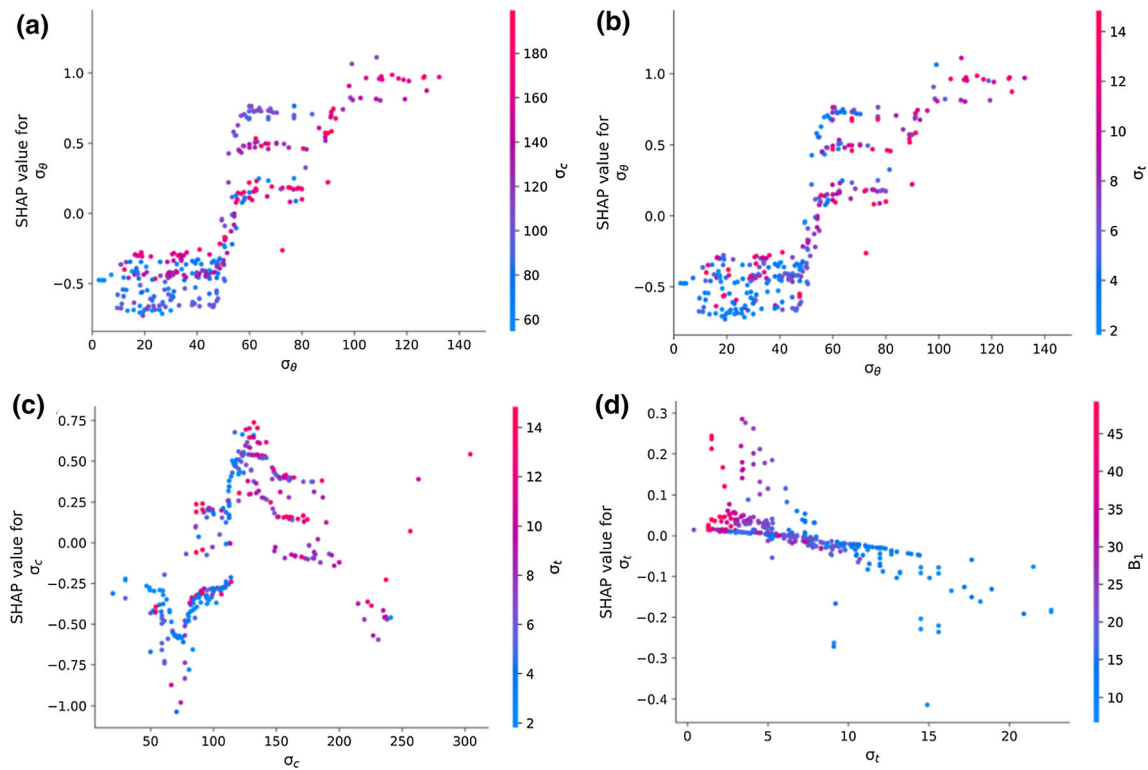


**Fig. 18** The influence of a single input feature on the rockburst intensity prediction for the developed MARS model base on SHAP value: **a**  $\sigma_\theta$ , **b**  $\sigma_c$ , **c**  $\sigma_t$ , and **d**  $B_1$

more energy through deformation and leads to stronger rockburst. But the integrity and good quality of the rock will mitigate the rockburst intensity when the compressive strength  $\sigma_c$  of the rock is higher than 150 MPa. This trend also can be found in Fig. 18c.  $B_1 = \sigma_c/\sigma_t$  presents the ratio of the uniaxial compressive to the tensile strength of the rock, with a certain value of  $\sigma_t$ , the increase of  $B_1$  means a higher compressive strength  $\sigma_c$  and a better rock quality to prevent strong rockburst. From Fig. 18d it can be observed that the SHAP value decreases with the increasing  $\sigma_t$  value. This is consistent with the previous experimental research results. Fracture and buckling are the two stages before the occurrence of rockburst, and buckling is the prerequisite for the occurrence of rockburst. Increasing the tensile strength effectively increases the rock deformation before fracture and buckling and thus decreases the severity of rockbursts.

Figure 19 illustrates how the interaction of two input features effects the rockburst intensity prediction for the developed model. Take Fig. 19a as example, the  $x$ -axis represents the  $\sigma_\theta$  value, and the left  $y$ -axis represents the

SHAP values for corresponding  $\sigma_\theta$  value. The right  $y$ -axis represents different  $\sigma_c$  value using colors from blue to red, for which the blue color means a low  $\sigma_c$  value and the red color stands for a high  $\sigma_c$  value. It can be seen that in different  $\sigma_\theta$  intervals, the SHAP values brought by different  $\sigma_c$  are different. For the  $\sigma_\theta$  with small value from 0 to 50, small  $\sigma_c$  will give a lower SHAP value (blue points) and high  $\sigma_c$  will give a high SHAP value (red points). This implies that when the tangential stress around underground opening is less than 50 MPa, the rock with low compressive strength will have weak rockburst intensity. But for the data points with  $\sigma_\theta$  values in the range of 50 to 85 MPa, it can be found that low value of  $\sigma_c$  will give a higher SHAP value. This means when the  $\sigma_\theta$  increase from 50 to 80 MPa, a lower compressive strength will weaken the rock's resistance toward fracture and buckling, resulting in stronger rockburst. The same pattern can be found in Fig. 19b. As plotted in Fig. 19c, when the compressive strength  $\sigma_c$  is less than 150 MPa, a lower tensile strength cause lower rockburst intensity. Generally, a high  $B_1$



**Fig. 19** The interaction of two input features on the rockburst intensity prediction for the developed MARS model base on SHAP value: **a**  $\sigma_\theta$  versus  $\sigma_c$ , **b**  $\sigma_\theta$  versus  $\sigma_t$ , **c**  $\sigma_c$  versus  $\sigma_t$ , **d**  $\sigma_t$  versus  $B_1$

corresponds to strong rockburst when the tensile strength is less than 10 MPa, as shown in Fig. 19d.

## 5 Conclusions

In this study, a multivariate adaptive regression splines (MARS) model and a novel deep forest model (DF) were developed for prediction of rockburst intensity. More than three hundred rockburst cases collected worldwide were used for model construction. The *t*-SNE method was utilized for nonlinear dimensionality reduction of the original input features. The Gaussian mixture model was adopted as the clustering method to relabel original rockburst intensity. Based on what we have gained so far, the following conclusions can be drawn:

Compared to the ‘black-box’ nature of ANN models, the MARS model shows an advantage of outputting explicit expressions. The deep forest model has shown superior performance than DNN on rockburst prediction problem, having faster training speed and better parameter robustness, and additionally, it can adapt to training sets of different size.

The Shapley additive explanations method (SHAP) was used to investigate the relative importance of input features. The result shows that  $\sigma_\theta$  and  $\sigma_c$  are two most

important features for rockburst intensity prediction, and there are some complicated interactions hidden between input features. It can be found that a higher  $\sigma_c$ , lower  $\sigma_t$  and  $B_1$  value will cause stronger rockburst intensity. When  $\sigma_c$  is under 150 Mpa, an increasing  $\sigma_c$  will lead to stronger rockburst. But it will mitigate the rockburst intensity when the increasing  $\sigma_c$  is beyond 150 MPa.

A main issue limiting the precision and generalization of developed models in this study is the quantity and quality of training data. A larger sets of standardized data are needed to solve overfitting problems and take geological variations into account. The 344 data sets collected in this work should be considered as a relatively small number of samples, compared with other common machine learning tasks.

Currently, only six most influential and most commonly used input features are considered in this study. The parameters representing the rock rheology and geological condition need to be added in the future to enhance the forecasting ability for rockburst prediction.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (Grant No. 51778092), Natural Science Foundation (cstc2017jcyjAX0073), Chongqing.

## References

- Aubertin M, Gill DE, Simon R (1994) On the use of the brittleness index modified (BIM) to estimate the post-peak behavior of rocks. In: Aubertin M, Gill DE, Simon R (eds) 1st North American Rock Mechanics Symposium: models and measurements challenges from industry. A.A. Balkema Austin, Texas, pp 945–952
- Bai XD, Cheng WC, Ong D, Li G (2021). Evaluation of geological conditions and clogging of tunneling using machine learning. *Geomech Eng* (in press)
- Baltz R, Huckle A (2008) Rockburst prevention in the German coal industry. In: Proceedings of the 27th international conference on ground control in mining. West Virginia University, Morgantown, WV, pp 46–50
- Breiman L (2001) Random forests. *Mach Learn* 45:5–32
- Cai W, Dou LM, Si GY, Cao AY, He J, Liu S (2016) A principal component analysis/fuzzy comprehensive evaluation model for coal burst liability assessment. *Int J Rock Mech Min Sci* 81:62–69
- Cheng WC, Bai XD, Sheil BB, Li G, Fei W (2020) Identifying characteristics of pipejacking parameters to assess geological conditions using optimisation algorithm-based support vector machines. *Tunn Undergr Space Technol* 106(1):103592
- Cheng WC, Li G, Liu N, Xu J, Horpibulsuk S (2020) Recent massive incidents for subway construction in soft alluvial deposits of taiwan: a review. *Tunn Undergr Space Technol* 96:103178.1–103178.18
- Cheng WC, Ni JC, Arulrajah A, Huang HW (2018) A simple approach for characterising tunnel bore conditions based upon pipe-jacking data. *Tunn Undergr Space Technol* 71(1):494–504
- Cheng WC, Ni JC, Huang HW, Shen JS (2019) The use of tunnelling parameters and spoil characteristics to assess soil types: a case study from alluvial deposits at a pipejacking project site. *Bull Eng Geol Env* 78(4):2933–2942
- Cheng WC, Wang L, Xue ZF, Ni JC, Rahman MM, Arulrajah A (2019) Lubrication performance of pipejacking in soft alluvial deposits. *Tunn Undergr Space Technol* 91:102991
- Diederichs MS (2007) Mechanistic interpretation and practical application of damage and spalling prediction criteria for deep tunnelling. *Can Geotech J* 44(9):1082–1116
- Dietterich TG (2000) Ensemble methods in machine learning. In: International workshop on multiple classifier systems, pp1–15
- Dong LJ, Li XB, Peng K (2013) Prediction of rockburst classification using random forest. *Trans Nonferrous Met Soc* 23(2):472–477
- Du ZJ, Xu MG, Liu ZP, Wu X (2006) Laboratory integrated evaluation method for engineering wall rock rock-burst. *Gold* 27(11):26–30 ((in Chinese))
- Faradonbeh RS, Haghshenas SS, Taheri A, Mikaeil R (2019) Application of self-organizing map and fuzzy c-mean techniques for rockburst clustering in deep underground projects. *Neural Comput Appl* 32:8545–8559
- Faradonbeh RS, Taheri A (2019) Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques. *Eng Comput* 35(2):659–675
- Feng XT, Wang LN (1994) Rockburst prediction based on neural networks. *Trans Nonferrous Met Soc* 4(1):7–14
- Feng GL, Xia GQ, Chen BR, Xiao YX, Zhou RC (2019) A method for rockburst prediction in the deep tunnels of hydropower stations based on the monitored microseismicity and an optimized probabilistic neural network model. *Sustainability* 11(11):3212
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 19:1–67
- Gao W (2015) Forecasting of rockbursts in deep underground engineering based on abstraction ant colony clustering algorithm. *Nat Hazards* 76(3):1625–1649
- Ghasemi E, Gholizadeh H, Adoko AC (2020) Evaluation of rockburst occurrence and intensity in underground structures using decision tree approach. *Eng Comput* 36(1):213–225
- Goh ATC, Zhang RH, Wang W, Wang L, Liu H, Zhang WG (2020) Numerical study of the effects of groundwater drawdown on ground settlement for excavation in residual soils. *Acta Geotech* 15:1259–1272
- Goh ATC, Zhang WG, Zhang YM, Xiao Y, Xiang YZ (2018) Determination of EPB tunnel-related maximum surface settlement: A Multivariate adaptive regression splines approach. *Bull Eng Geol Env* 77:489–500
- Goh ATC, Zhang Y, Zhang R, Zhang W, Xiao Y (2017) Evaluating stability of underground entry-type excavations using multivariate adaptive regression splines and logistic regression. *Tunn Undergr Space Technol* 70:148–154
- Hoek E, Brown ET (1980) Underground excavations in rock. The Institution of Mining and Metallurgy, London
- Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *J Educ Psychol* 24(6):417
- Hucka V, Das B (1974) Brittleness determination of rocks by different methods. *Int J Rock Mech Min* 11:389–392
- Iannacchione AT, Zelanko JC (1993) Occurrence and remediation of coal mine bumps: a historical review
- Ji B, Xie F, Wang XP, He SQ, Song DZ (2020) Investigate contribution of multi-microseismic data to rockburst risk prediction using support vector machine with genetic algorithm. *IEEE Access* 8:58817–58828
- Jia QJ, Wu L, Li B, Chen CH, Peng YX (2019) The comprehensive prediction model of rockburst tendency in tunnel based on optimized unascertained measure theory. *Geotech Geol Eng* 37(4):3399–3411
- Jiang K, Lu J, Xia KL (2016) A novel algorithm for imbalance data classification based on genetic algorithm improved SMOTE. *Arab J Sci Eng* 41(8):3255–3266
- Kidybinski A (1981) Bursting liability indices of coal. *Int J Rock Mech Min Sci Geomech Abstr* 18(4):295–304
- Kwasniewski M, Szutkowski I, Wang JA (1994) Study of ability of coal from seam 510 for storing elastic energy in the aspect of assessment of hazard in Porabka–Klimontow colliery. *Sci. Rept. Silesian Technical University*
- Li N, Feng XD, Jimenez R (2017) Predicting rock burst hazard with incomplete data using Bayesian networks. *Tunn Undergr Space Technol* 61:61–70
- Li N, Jimenez R (2018) A logistic regression classifier for long-term probabilistic prediction of rock burst hazard. *Nat Hazards* 90(1):197–215
- Li TZ, Li YX, Yang XL (2017) Rock burst prediction based on genetic algorithms and extreme learning machine. *J Cent South Univ* 24(9):2105–2113
- Li TZ, Li YX, Yang XL (2017) Rock burst prediction based on genetic algorithms and extreme learning machine. *J Cent S Univ* 24(9):2105–2113
- Liu YR, Hou SK (2019) Rockburst prediction based on particle swarm optimization and machine learning algorithm. In: International conference on information technology in geoenvironment, Cham, pp 292–303
- Liu R, Ye YC, Hu NY, Chen H, Wang XH (2018) Classified prediction model of rockburst using rough sets-normal cloud. *Neural Comput Appl* 31:1–9
- Maaten LVD, Hinton G (2008) Visualizing data using *t*-SNE. *J Mach Learn Res* 9(November):2579–2605

41. Martin C, Kaiser P, McCreath D (1999) Hoek-Brown parameters for predicting the depth of brittle failure around tunnels. *Can Geotech J* 36(1):136–151
42. McLachlan G, Peel D (2004) *Finite mixture models*. Wiley
43. Peng Z, Wang YH, Li TJ (1996) Griffith theory and rock burst of criterion. *Chinese J Rock Mech Eng* 15:491–495
44. Pu Y, Apel DB, Pourrahimian Y, Chen J (2019) Evaluation of rockburst potential in kimberlite using fruit fly optimization algorithm and generalized regression neural networks. *Arch Min Sci* 64(2):279–296
45. Pu Y, Apel DB, Wang C, Wilson B (2018) Evaluation of burst liability in kimberlite using support vector machine. *Acta Geophys* 66(5):973–998
46. Pu Y, Apel DB, Xu H (2019) Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier. *Tunn Undergr Space Technol* 90:12–18
47. Russenes BF (1974) Analysis of rock spalling for tunnels in steep valley sides. M.Sc. thesis, Norwegian Institute of Technology, Trondheim, Norway, 247 (in Norwegian)
48. Shapley LS (1953) A value for  $n$ -persons games. *Ann Math Stud* 28(7):307–318
49. Shi XZ, Zhou J, Dong L, Hu HY, Wang HY, Chen SR (2010) Application of unascertained measurement model to prediction of classification of rockburst intensity. *Chin J Rock Mech Eng* 29(1):2720–2726
50. Singh SP (1987) The influence of rock properties on the occurrence and control of rockbursts. *Min Sci Technol* 5:11–18
51. Steele RJ, Raftery AE (2010) Performance of Bayesian model selection criteria for Gaussian mixture models 1. *Dept Stat*.
52. Su GS, Zhang KS, Chen Z (2009) Rockburst prediction using Gaussian process machine learning. In: 2009 international conference on computational intelligence and software engineering, Wuhan, pp 1–4
53. Torgerson WS (1952) Multidimensional scaling: I. Theory. *Method Psychometrika* 17(4):401–419
54. Wang L, Chongzhi Wu, Tang L, Zhang W, Lacasse S, Liu H, Gao L (2020) Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. *Acta Geotech* 15(11):3135–3150
55. Wang L, Wu C, Gu X, Liu H, Mei G, Zhang W (2020) Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. *Bull Eng Geol Environ* 79(6):2763–2775
56. Wengang Z, Li Y, Chongzhi WU, Hongrui Li, Goh A, Zhang R (2020) Prediction of lining response for twin-tunnel construction in anisotropic clays using machine learning techniques. *Undergr Space*. <https://doi.org/10.1016/j.undsp.2020.02.007>
57. Wengang Z, Runhong Z, Wei W, Fan Z, Chee AGT (2019) A Multivariate Adaptive Regression Splines model for determining horizontal wall deflection envelope for braced excavations in clays. *Tunn Undergr Space Technol* 84:461–471
58. Wu SC, Wu ZG, Zhang CX (2019) Rock burst prediction probability model based on case analysis. *Tunn Undergr Space Technol* 93:103069
59. Xdb A, Wcca B, Bbs C, Ge LA (2021). Pipejacking clogging detection in soft alluvial deposits using machine learning algorithms. *Tunn Undergr Space Technol* (in press)
60. Xue YG, Bai CH, Qiu DH, Kong FM, Li ZQ (2020) Predicting rockburst with database using particle swarm optimization and extreme learning machine. *Tunn Undergr Space Technol* 98:103287.
61. Xue YG, Zhang XL, Li SC, Qiu DH, Su MX, Li LP, Li ZQ, Tao YF (2018) Analysis of factors influencing tunnel deformation in loess deposits by data mining: a deformation prediction model. *Eng Geol* 232:94–103
62. Yin X, Liu Q, Pan Y et al (2021) Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: comparison of eight single and ensemble models. *Nat Resour Res* 30:1795–1815
63. Zhang W, Chongzhi Wu, Li Y, Wang L, Samui P (2019) Assessment of pile drivability using random forest regression and multivariate adaptive regression splines. *Georisk*. <https://doi.org/10.1080/17499518.2019.1674340>
64. Zhang WG, Goh ATC (2013) Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Comput Geotech* 48:82–95
65. Zhang WG, Goh ATC (2016) Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci Front* 7:45–52
66. Zhang WG, Goh ATC (2016) Evaluating seismic liquefaction potential using multivariate adaptive regression splines and logistic regression. *Geomech Eng* 10(3):269–284
67. Zhang W, Li H, Li Y, Liu H, Chen Y, Ding X (2021) Application of deep learning algorithms in geotechnical engineering: a short critical review. *Artif Intell Rev*. <https://doi.org/10.1007/s10462-021-09967-1>
68. Zhang WG, Li HR, Wu CZ, Li YQ, Liu ZQ, Liu HL (2020) Soft computing approach for prediction of surface settlement induced by earth pressure balance shield tunneling. *Undergr Space*
69. Zhang W, Wu C, Zhong H, Li Y, Wang L (2021) Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci Front* 12:469–477
70. Zhang W, Zhang R, Chongzhi Wu, Goh ATC, Lacasse S, Liu Z, Liu H (2020) State-of-the-art review of soft computing applications in underground excavations. *Geosci Front* 11:1095–1106
71. Zhang WG, Zhang YM, Goh ATC (2017) Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation. *Tunn Undergr Space Technol* 64:24–33
72. Zhang W, Zhang R, Wu C et al (2020) Assessment of basal heave stability for braced excavations in anisotropic clay using extreme gradient boosting and random forest regression. *Undergr Space*
73. Zhou ZH, Feng J (2017) Deep forest: towards an alternative to deep neural networks
74. Zhou J, Guo H, Koopialipour M et al (2020) Investigating the effective parameters on the risk levels of rockburst phenomena by developing a hybrid heuristic algorithm. *Eng Comput* 1:1–16
75. Zhou J, Li XB, Mitri HS (2016) Classification of rockburst in underground projects: comparison of ten supervised learning methods. *J Comput Civil Eng* 30(5):04016003
76. Zhou J, Li XB, Shi XZ (2012) Long-term prediction model of rockburst in underground openings using heuristic algorithms and support vector machines. *Saf Sci* 50(4):629–644
77. Zhou KP, Lin Y, Deng HW, Li JL, Liu CJ (2016) Prediction of rock burst classification using cloud model with entropy weight. *Trans Nonferrous Metals Soc China* 26(7):1995–2002
78. Zhou J, Shi XZ, Huang RD, Qiu XY, Chen C (2016) Feasibility of stochastic gradient boosting approach for predicting rockburst damage in burst-prone mines. *Trans Nonferrous Met Soc* 26(7):1938–1945