ORIGINAL PAPER

Short-term rockburst risk prediction using ensemble learning methods

Weizhang Liang[1,2] · **Asli Sari**[2] · **Guoyan Zhao**[1] 🟢 · **Stephen D. McKinnon**[2] · **Hao Wu**[3]

Abstract

Short-term rockburst risk prediction plays a crucial role in ensuring the safety of workers. However, it is a challenging task in deep rock engineering as it depends on many factors. More recently, machine learning approaches have started to be used to predict rockbursts. In this paper, ensemble learning methods including random forest (RF), adaptive boosting, gradient boosted decision tree (GBDT), extreme gradient boosting and light gradient boosting machine were adopted to predict short-term rockburst risk using microseismic data from the tunnels of Jinping-II hydropower project in China. First, labeled rockburst data with six indicators based on microseismic monitoring were collected. Then, the original rockburst data were randomly divided into training and test sets with a 70/30 sampling strategy. The hyperparameters of the ensemble learning methods were tuned with fivefold cross-validation during training. Finally, the predictive performance of each model was evaluated using classification accuracy, Cohen's Kappa, precision, recall and F-measure metrics on the test set. The results showed that RF and GBDT possessed better overall performance. RF obtained the highest average accuracy of 0.8000 for all cases, whereas GBDT achieved the highest value for high (moderate and intense) risk cases with an accuracy of 0.9167. The proposed methodology can provide effective guidance for short-term rockburst risk management in deep underground projects.

Keywords Rockburst · Short-term risk · Ensemble learning · Prediction · Microseismic monitoring

✉ Guoyan Zhao
  gyzhao@csu.edu.cn

✉ Hao Wu
  hoekwu@csu.edu.cn

1   School of Resources and Safety Engineering, Central South University, Changsha 410083, China

2   The Robert M. Buchan Department of Mining, Queen's University, Kingston K7L 3N6, Canada

3   School of Mines, China University of Mining and Technology, Xuzhou 221116, China

1 Introduction

Rockbursting is a rock mass instability phenomenon accompanied by spalling, cracking, splitting or even ejection due to the sudden release of strain energy (Gong et al. 2019). For deep mines and tunnels, rockbursting is becoming an increasingly prominent problem. It has been reported that many countries have encountered the rockburst hazard, making it a worldwide challenge (Keneti and Sainsbury 2018; Hu et al. 2019). For example, seismicity and rockbursting 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). An intense rockburst can cause casualties and huge economic losses. Durrheim (2010) reported the largest rockburst with a local magnitude of 5.2 in the Klerksdorp district of South Africa. It caused two deaths, fifty-eight injuries and severe damage to the surface buildings. Zhang et al. (2012) described four extremely intense rockburst cases in the tunnels of Jinping II Hydropower Station. One of them occurred at the depth of 2330 m and caused seven deaths, one injury and destruction of a tunnel boring machine. Because of its serious consequences, predicting the risk of rockbursting efficiently is crucial and plays a significant role in disaster prevention.

Rockburst risk prediction can be classified into two types: long-term and short-term (Zhou et al. 2012). Long-term rockburst prediction is generally conducted in the design or early excavation stages of projects. The intrinsic rock mechanics parameters, such as strength, stiffness, brittleness and energy storage capacity, and the in situ or mining-induced stress are used to establish the predictive model (Liang et al. 2019). The likelihood of rockbursting for various rock types in different stress conditions is estimated. The results are then used to provide guidance for the subsequent construction stages. Short-term rockburst prediction aims to predict the risk of rockbursting in the near future based on in situ measurement techniques. Among these techniques, microseismic monitoring is one of the most widely used methods (Zhang et al. 2018). By monitoring and analyzing the microseismic waves released by the rock fracture, some premonitory characteristics of rockbursts are determined, which can be used to predict the risk of rockbursting. The most commonly used microseismic indicators for rockburst prediction include the number of events (Srinivasan et al. 1997), b value (Ma et al. 2018), energy index (Liu et al. 2013), apparent volume (Ma et al. 2018) and others (Cai et al. 2018; Liang et al. 2020) and their changing rate over a time window. For instance, Brady and Leighton (1977) found that the seismic activity increased rapidly and then decreased distinctly before a rockburst; Ma et al. (2018) revealed that an elbow point existed in the b-value time curve prior to the mainshocks; Liu et al. (2013) found that the apparent volume and spatial correlation length increased, whereas energy index, fractal dimension and b value dropped before the occurrence of large events; Xue et al. (2020) used the number of daily events N and b value to predict rockburst risk and found that rockbursting was more likely to occur when $(\lg N)/b$ was larger than 1. However, the precursory characteristics for rockburst prediction are not always consistent in different geological environments; a general early warning threshold has not yet been determined. Although numerous promising results have been achieved in various aspects of rockburst research, short-term rockburst prediction still remains problematic. Currently, there is no uniform approach or consensus in the engineering community.

Rockbursting is affected by many complex factors that include the properties of the rock mass, geological structures, stress regime and excavation activities (Wu et al. 2019). Some of these factors are uncertain, correlated, and cannot be determined quantitatively (Liang et al. 2019). In addition, there is a highly nonlinear relationship between rockburst risk

level and influencing factors, which makes it very difficult to determine the contribution of each rockburst factor. Thus, the prediction accuracy for rockbursting is greatly restricted. In addition, different mechanisms can cause rockbursts. Rockbursts can be classified into three categories, namely strainbursts, pillar bursts and fault-slip bursts (Kaiser and Cai 2012). Strainbursts and pillar bursts are generally produced due to high strain energy concentration and release from the rock mass itself, whereas fault-slip bursting is induced by a remote source such as slippage of faults or other geological structures. Strainbursts are the most common type of rockbursts experienced in tunnels. In strainbursts, the location of the seismic event is the same as where damage occurs, providing the opportunity to predict the rockburst risk using microseismic indicators in tunnels.

More recently, machine learning (ML) approaches have started to be used to predict rockbursts because of their ability to handle nonlinear and complex problems. Rockbursting is generally treated as a classification problem according to the risk level. In this paper, we propose using ensemble learning methods for predicting short-term rockburst risk in tunnels based on microseismic monitoring. Ensemble learning is a branch of machine learning that incorporates multiple models to achieve particular tasks where the errors of an individual model is likely compensated by others (Sagi and Rokach 2018). These models are generally decision trees. Decision trees fit the short-term rockburst prediction task well because they do not need any underlying distribution assumption, can handle multiple types of data and work especially well with small datasets. However, decision trees also have some disadvantages, such as being sensitive to small variations of the input data and being prone to overfitting, which means the model fits a given data too closely and cannot predict the unknown behavior effectively, particularly if the number of samples is small (Kotsiantis 2013; Rokach 2016). Ensemble learning techniques overcome this drawback by combining multiple models, thus reducing the bias (Woźniak et al. 2014). Therefore, the overall predictive performance of the ensemble classifier would be better than that of a single model (Dev and Eden 2019; Krawczyk et al. 2017). The typical ensemble learning methods include random forest (RF) (Breiman 2001), adaptive boosting (AdaBoost) (Freund and Schapire 1997), gradient boosted decision tree (GBDT) (Friedman 2001), extreme gradient boosting (XGBoost) (Chen and Guestrin 2016) and light gradient boosting machine (LightGBM) (Ke et al. 2017). Recently, these algorithms have received a lot of attention and have proven to have excellent prediction performance in many engineering geology fields, such as formation lithology classification (Dev and Eden 2019), landslide assessment (Dou et al. 2020) and rock core image processing (Chauhan et al. 2016). To the best of our knowledge, a comparative analysis of these algorithms for short-term rockburst risk prediction has not been conducted.

The original contribution of this paper is to develop a methodology that uses ensemble learning methods to predict short-term rockburst risk. First, the rockburst case data based on microseismic monitoring are collected. Then, five ensemble learning algorithms are used to predict the short-term rockburst risk. Finally, the predictive performance of each model is comprehensively analyzed and evaluated using five metrics.

2 Literature review

In the last decade, many machine learning (ML) approaches have been used to predict the risk of rockbursts, of which only a small sub-set targeted short-term rockburst prediction. A summary of approaches that use ML for rockburst prediction is provided in

Table 1 Summary of ML algorithms on rockburst prediction

Authors (year)	Types	Algorithms
Feng et al. (1994)	Long-term	ANN
Su et al. (2009	Long-term	GP
Zhou et al. (2012)	Long-term	GA-SVM and PSO-SVM
Dong et al. (2013)	Long-term	RF
Gao (2015)	Long-term	ACCA
Jiang et al. (2016)	Long-term	GA-SMOTE-C4.5 DT
Zhou et al. (2016)	Long-term	LDA, QDA, PLSDA, NB, KNN, MLPNN, DT, SVM, RF and GBM
Zhou et al. (2016)	Short-term	SGB
Li et al. (2017)	Long-term	BNs
Li et al. (2017)	Long-term	GA-ELM
Li et al. (2018)	Long-term	RS-CM, NB, KNN and RF
Li and Jimenez (2018)	Long-term	LR
Afraei et al. (2018)	Long-term	RMs
Pu et al. (2018)	Long-term	SVM
Pu et al. (2018)	Long-term	DT
Faradonbeh and Taheri (2019)	Long-term	ENN, GEP and C4.5 DT
Feng et al. (2019)	Short-term	MIVA-MFA-PNN
Pu et al. (2019)	Long-term	FOA-GRNN
Wu et al. (2019)	Long-term	Copula-LSSVM
Pu et al. (2019)	Long-term	K-means-SVM
Liu and Hou (2019)	Long-term	PSO-BPNN, PSO-PNN and PSO-SVM
Faradonbeh et al. (2019)	Long-term	SOM and FCM
Pu et al. (2019)	Long-term	SVM and GP
Zheng et al. (2019)	Long-term	Entropy weight integrated with grey relational BPNN
Ghasemi et al. (2020)	Long-term	C5.0 DT
Zhou et al. (2020)	Long-term	FA-ANN
Xue et al. (2020)	Long-term	PSO-ELM
Ji et al. (2020)	Short-term	GA-SVM

Table 1. These approaches can be classified into three categories as supervised learning approaches, comparative decision strategies and unsupervised learning approaches. Artificial neural networks (ANN) (Feng and Wang 1994), Gaussian process (GP) (Su et al. 2009), RF (Dong et al. 2013), stochastic gradient boosting (SGB) (Zhou et al. 2016), Bayesian networks (BNs) (Li et al. 2017), logistic regression (LR) (Li and Jimenez 2018), regression models (RMs) (Afraei et al. 2018) and DT (Pu et al. 2018; Ghasemi et al. 2020) have been used as supervised learning attempts of long-term rockburst prediction. The majority of these approaches are not very suitable for the typical size and structure of rockburst datasets because ANN typically need large datasets for training and tend to overfit; GP assumes each observation is normally distributed which is not appropriate in many applications and it uses the whole sample and features information to perform the prediction; BNs cannot obtain good prediction performances when the correlation between features is large because it assumes feature independence; LR

tends to underperform when multiple or nonlinear decision boundaries exist; RMs are sensitive to outliers; and DT is unstable and prone to overfitting.

Some more advanced supervised learning techniques were adapted through combining different algorithms and other performance improvement methods. Zhou et al. (2012) used the genetic algorithm (GA) and particle swarm optimization (PSO) to optimize the hyperparameters of support vector machines (SVM); Jiang et al. (2016) adopted the GA-based synthetic minority over-sampling technique (SMOTE) and C4.5 DT algorithms for the imbalance data classification of rockbursts; Li et al. (2017) used the extreme learning machine (ELM) optimized by GA to improve the accuracy of rockburst prediction; Pu et al. (2019) combined the fruit fly optimization (FOA) and generalized regression neural networks (GRNN) to predict rockbursts; Wu et al. (2019) used Copula theory and least-squares support vector machine (LSSVM) to establish rockburst prediction probability models; Liu and Hou (2019) introduced PSO to optimize back propagation neural network (BPNN), probabilistic neural network (PNN) and SVM simultaneously for rockburst prediction; Zheng et al. (2019) developed a entropy weight integrated with grey relational BPNN model to analyze rockburst risk; Zhou et al. (2020) integrated the firefly algorithm (FA) and ANN for rockburst risk prediction; and Xue et al. (2020) established the PSO-ELM model to determine rockburst risk. In these papers, the prediction performances of supervised learning algorithms are improved through hyperparameter optimization, sampling technique and data preprocessing.

Comparative analysis of multiple ML algorithms was performed by several researchers for long-term rockburst prediction. For instance, Zhou et al. (2016) compared the predictive performance using ten supervised learning methods, which included linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), partial least-squares discriminant analysis (PLSDA), naive Bayes (NB), k-nearest neighbor (KNN), multilayer perceptron neural network (MLPNN), DT, SVM, RF and gradient boosting machine (GBM); Lin et al. (2018) made a performance comparison of the rough set (RS)—cloud model (CM), NB, KNN and RF for rockburst prediction; Faradonbeh and Taheri (2019) predict rockburst hazard based on emotional neural network (ENN), gene expression programming (GEP) and C4.5 DT; and Pu et al. (2019) used a generative model (GP) and a discriminative approach (SVM) to evaluate rockburst liability, respectively. When compared to only using one algorithm, these comparative analyses yield more stable and robust results. However, except for DT and the ensemble approaches, these algorithms are not suitable for solving the prediction issues with small datasets.

The last category of approaches considers the difficulty of obtaining the actual rockburst levels in some instances. Therefore, they adopt an unsupervised learning approach, where the labels of the data do not need to be known in advance. The classification is performed based on clustering and taking the distance from the center of the cluster. Gao (2015) used an abstraction ant colony clustering algorithm (ACCA) for the rockburst prediction to improve the computational efficiency and accuracy of the traditional ACCA; Pu et al. (2019) adopted the k-means method to relabel the original data and then used SVM to predict the rockburst risk for kimberlite pipes; Faradonbeh et al. (2019) advanced the two clustering approaches, namely, self-organizing map (SOM) and fuzzy c-mean (FCM), for rockburst prediction in deep underground projects. However, because the actual rockburst levels can be obtained in most situations, supervised learning methods received the most research attention for rockburst prediction.

As opposed to the extensive literature on long-term rockburst prediction, only three papers focus on short-term rockburst prediction. Zhou et al. (2016) predicted the field rockburst damage using the SGB model; Feng et al. (2019) combined the mean impact value

algorithm (MIVA), modified firefly algorithm (MFA) and PNN model to predict the short-term rockburst risk based on the monitored microseismicity; and Ji et al. (2020) used GA-SVM model to predict rockbursts (high energy tremors) in coal mines. Short-term rockburst prediction is very important as it can be used as a mechanism for real-time warning to prevent accidents. This paper addresses the need in the literature for machine learning methods of short-term rockburst prediction.

3 Data and variables

From the work of Feng et al. (2013), a total of 93 rockburst cases for short-term risk prediction are obtained from the tunnels of the Jinping-II hydropower project in China. This hydropower station includes seven parallel tunnels with an average length of 16.67 km and a maximum depth of 2525 m. The maximum, intermediate and minimum principal stresses are found by back analysis to be approximately 63 MPa, 34 MPa and 26 MPa, respectively (Feng et al. 2013). The main lithology is marble, which has a saturated uniaxial compressive strength of 30–114 MPa, elastic modulus of 25–40 GPa and tensile strength of 3–6 MPa, and the surrounding rock masses are hard and intact (Feng et al. 2019; Ma et al. 2015). Under these geological conditions, the rockburst risk is very high. The rockburst intensity is divided into four levels based on radiated energy and damage phenomena, namely none, slight, moderate and intense. The classification method of rockburst risk level is described in Table 2 (Feng et al. 2013; Chen et al. 2015). In this paper, these four levels acting as the model output are labeled as 0, 1, 2 and 3, respectively. The distribution of the rockburst risk levels in the overall dataset is given in Fig. 1. The complete rockburst case data are presented in "Online Appendix A."

This dataset is obtained from the microseismic monitoring system, which provides many parameters that can be used to quantify seismic events. Details of the data acquisition process are discussed in Feng et al. (2015). In general, the number of events, seismic energy and apparent volume are used to represent seismic activity. Among them, the number of events reflects the frequency of seismicity; seismic energy indicates the energy released by seismic events in the form of waves and characterizes the strength of seismicity. The seismic energy of P or S waves is calculated by (Glazer 2018):

$$E = 8\pi\rho c \int_0^\infty v^2(f)df, \tag{1}$$

where ρ indicates the density of source material; c is the velocity of P or S wave; $v^2(f)$ represents the velocity power spectrum; and f denotes the frequency.

The apparent volume represents the volume of inelastic deformation zone of seismic source, which can be determined by (Liu et al. 2013):

$$V_A = \mu P^2/E, \tag{2}$$

where μ indicates the shear modulus of rock; and P is the seismic potency.

Since seismic events are induced by dynamic rock fractures, the parameters of seismic events can reflect the characteristics of rock damage to some extent. In this study, six indicators are selected to predict the short-term rockburst risk, and their statistics are shown in Table 3. According to the definition of these seismic event parameters, it can be known that C_1, C_2 and C_3 signify the number, strength and size of rock mass fractures during the

**Table 2** Rockburst risk level classification method (Feng et al. 2013; Chen et al. 2015)

Rockburst levels	Radiated energy/J	Phenomena
None	(0, 1]	No obvious surface failure; nearly no cracking sound; no damaged support system
Slight	(1, 10^2]	Slight spalling or slabbing; 10–30 cm ejected rock fragment size; slight cracking sound; failure depth less than 0.5 m; no damaged support system
Moderate	(10^2, 10^4]	Severe spalling and slabbing; 30–80 cm ejected rock fragment size with; cracking sound like a detonator blasting; 0.5–1.0 m failure depth: damaged shotcrete lining between rock bolts
Intense	(10^4, 10^7]	Large volume of rock mass ejected; 80–150 cm ejected rock fragment size; fresh fracture planes in the failure zone; lasting sound like an explosive with an impact wave; 1.0–3.0 m failure depth; support system destroyed

Fig. 1 Distribution of the rock-burst risk levels
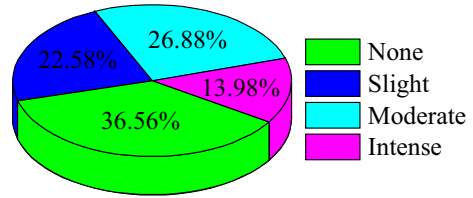
Table 3 Statistics of each indicator

Indicator	Min	Max	Mean	Standard deviation
Cumulative number of events C_1 (unit)	1	70	13.011	13.690
Logarithm of the cumulative release energy C_2 (J)	0.780	7.094	4.389	1.441
Logarithm of the cumulative apparent volume C_3 (m^3)	2.511	5.168	4.150	0.660
Event rate C_4 (unit/day)	0.111	12.250	1.735	1.738
Logarithm of the energy rate C_5 (J/day)	0.178	5.890	3.562	1.332
Logarithm of the apparent volume rate C_6 (m^3/day)	1.666	4.393	3.334	0.588
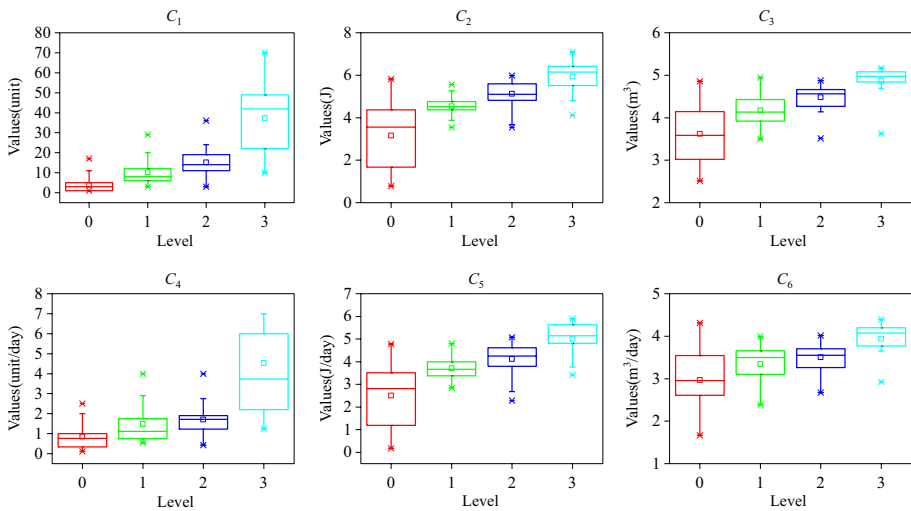
Fig. 2 Box plot of each indicator in different levels

rockburst development process, respectively. In addition, C_4, C_5 and C_6 are adopted to indicate the effect of time. Therefore, these six variables can be introduced to illustrate the state of the rock mass fractures (Feng et al. 2015). To make calculation convenient, the values of C_2, C_3, C_5 and C_6 are used in their logarithmic form. After taking the logarithm, it does not change the correlation of the data, but compresses the scale of the variables and reduces the absolute value of the data.

The box plot of each indicator for the four rockburst levels is shown in Fig. 2. Overall, the rockburst level is positively correlated with each indicator. The larger the

indicators values, the higher the rockburst level. However, some outliers exist in all indicators under each level, which indicates the complexity of the rockburst formation mechanism. In addition, for the same indicator in different levels, the distance between upper and lower quartiles (the height of the box) varies, and the range of indicator values also has some overlapping parts. Therefore, the effect of all indicators needs to be incorporated for better accuracy.

4 Methodology

In this paper, RF, AdaBoost, GBDT, XGBoost and LightGBM are used to predict the short-term rockburst risk and their predictive performances are comprehensively compared from multiple perspectives. The structure of the proposed methodology is indicated in Fig. 3. First, the raw rockburst data is randomly separated into training set (70% of the data) and test set (30% of the data). Of note, the proportion of rockburst samples with different risk levels in training and test sets is kept consistent during data partitioning. Second, a fivefold cross-validation (CV) method is applied on the training set to obtain the optimal hyperparameter values of five ensemble learning models (RF, Ada-Boost, GBDT, XGBoost and LightGBM). Third, each model with tuned hyperparameters is then fitted by the training set. Fourth, the test set is adopted to evaluate the performance of models using classification accuracy, Cohen's kappa, precision, recall and F-measure. Finally, the optimal model can be obtained based on these performance evaluation metrics. If its predictive performance is acceptable, it can be used for engineering applications. The detailed descriptions of ensemble learning models, hyperparameter optimization and model evaluation metrics are given in this section.
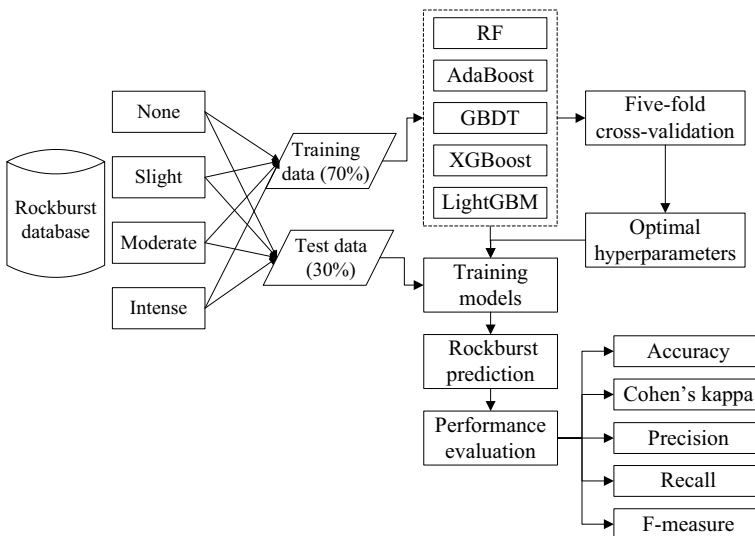

Fig. 3 Structure of the proposed methodology

## 4.1 Ensemble learning models

According to the dependencies between the base learners, the ensemble learning methods can be mainly divided into two types: bagging (Breiman 1996) and boosting (Schapire 1990). In bagging methods, the base learners are usually weak and have no dependencies, which allow them to be implemented in a parallel manner. The diagram of bagging ensemble learning is shown in Fig. 4. First, a bootstrap sampling technique is adopted to generate sample sets from the original dataset. Then, the base learners are independently trained using each sample set. Finally, the predicted result is obtained based on the integration rules. A voting method is usually used for classification problems. RF proposed by Breiman (2001) is a typical representative of the bagging approach. Nevertheless, RF differs from a bagging approach while selecting the features. The features are randomly selected from original features in the RF, which is favorable for improving the generalization ability of the model.

In boosting methods, dependencies exist between the weak base learners, and the subsequent base learner increases the emphasis on the misclassified instances obtained by the previous base learners. The diagram of boosting ensemble learning is shown in Fig. 5. First, the initial sample set is used to train the first base learner. Then, the input of samples for the next base learner is adjusted based on the classification error of the first base learner. This process is repeated with the subsequent base learners and is terminated upon reaching the specified number of iterations or meeting the convergence conditions. Lastly, the final result is obtained by combining the results from all base learners. AdaBoost, GBDT, XGBoost and LightGBM are boosting approaches that use decision trees as base learners (Kotsiantis 2014). In AdaBoost, the weighted samples are taken as the input for the next base learner. The samples' weights are adjusted based on the principle that misclassified samples obtain the larger weights. In addition, the final result is obtained using the weighted results from all base learners. The weight of the base learners is attained according to their classification error rate, and the base learner with a smaller error rate is provided a larger weight (Sagi and Rokach 2018). In GBDT, the residual of the previous

Fig. 4 Diagram of bagging ensemble learning
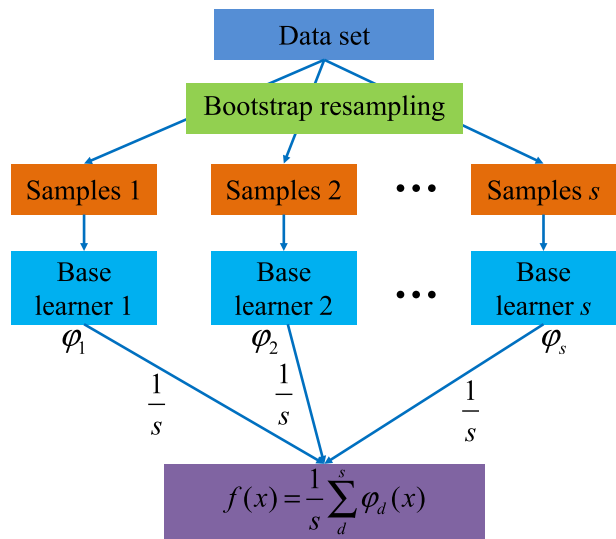
Fig. 5 Diagram of boosting
ensemble learning
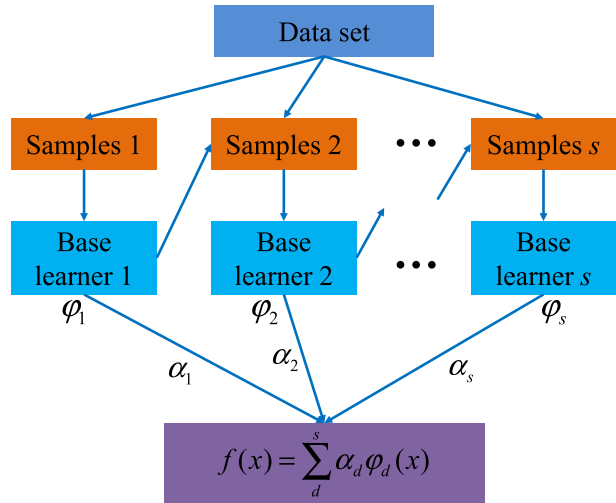
base learner is selected as the input for the next base learner. Therefore, the goal of the subsequent base learners is to reduce the residual. The final result is obtained by adding the results from all base learners. GBDT has a good classification performance because the loss decreases along the gradient direction in each iteration (Deng et al. 2019). XGBoost and LightGBM are proposed as improvements to GBDT. In the XGBoost algorithm, a regularization term is added in the objective function to prevent overfitting, and a second-order Taylor expansion of the loss function instead of a first-order derivative in GBDT is used (Chen and Guestrin 2016). In LightGBM, a histogram-based algorithm is utilized to increase the training speed and reduce memory consumption, and leaf-wise growth strategy of the trees with a maximum depth limit is adopted (Ke et al. 2017).

The strengths and weaknesses of the ensemble learning methods used in this paper are summarized in Table 4. These algorithms were implemented in Python using the scikit-learn library (Pedregosa et al. 2011) for RF, GBDT and AdaBoost, and XGBoost and LightGBM were carried out using the xgboost and lightgbm libraries, respectively.

4.2 Hyperparameter optimization

Hyperparameters are the parameters within the ML algorithm itself that need to be adjusted based on the dataset. To improve the performance of the model, the hyperparameters should be optimized rather than specified manually. In this study, some key hyperparameters in the ensemble learning models are selected for optimization. For the RF algorithm, the number of trees in the forest (n_estimators) and the maximum depth of the tree (max_depth) are selected. For ensemble learning algorithms belonging to the boosting type, the same hyperparameters are selected to make the comparison more meaningful, which include the maximum number of trees (n_estimators) and the shrinkage coefficient of each tree (learning_rate).

K-fold CV is commonly used for hyperparameter configuration (Zhou et al. 2016). In general, the value of K between five to ten is selected (Jung 2018). After considering the number of samples, fivefold CV is used in this study. The training set is randomly split into fivefolds. Among them, fourfolds make up the training sub-set, and the remaining

Table 4 Strengths and weaknesses of each ensemble learning method

Algorithm	Strengths	Weaknesses
RF	Can handle high-dimensional and small datasets well. Due to the introduction of randomness, it is not sensitive to outliers and has a good anti-noise ability	Difficult to interpret, can be computationally and memory-wise costly
AdaBoost	Flexible and not easy to overfit. Under this frame, the weak learners can be built using various classification models	Sensitive to outliers. Because the samples that are difficult to classify would get more attentions, it is susceptible to noise
GBDT	Not sensitive to outliers because it uses some robust loss functions. In addition, it has the flexibility to process various types of data, including continuous and discrete values	Difficult to train data in parallel due to the dependencies between weak learners
XGBoost	Not easy to overfit because the regularization term is added to the objective function	Extremely time-consuming for a large amount of data, because the greedy method is used to determine the optimal feature segmentation point
LightGBM	Not easy to overfit because the trees leaf-wise growth strategy is adopted. Furthermore, it has a fast training speed and does not require a high computational memory, and the complexity of data segmentation is low	Although it is more suitable for processing a large amount of data, it does not show particular advantages for the small datasets

one is used as the validation set. The training sub-set is utilized for fitting the models, and the validation set is utilized for assessing the model's performance. This process is repeated five times by changing permutation of the folds until each sample in the training set is predicted. Then, the average accuracy of the validation set is taken into account to determine the hyperparameter values.

4.3 Model evaluation metrics

For the rockburst risk prediction problems, the predictive performance for both total samples and the samples belonging to a certain risk level should be taken into account. Therefore, five metrics including two global metrics (classification accuracy and Cohen's Kappa) and three within-class metrics (precision, recall and F-measure) are adopted to evaluate the performance of each model (Kumar 2019). For a certain model, these metrics can be calculated based on the confusion matrix. Suppose a confusion matrix is:

$$
M = \begin{bmatrix}
k_{11} & k_{12} & \cdots & k_{1n} \\
k_{21} & k_{22} & \cdots & k_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
k_{n1} & k_{n2} & \cdots & k_{nn}
\end{bmatrix},
\tag{3}
$$

where n indicates the number of levels, k_{ii} indicates the number of samples of level i that are correctly predicted, and k_{ij} indicates the number of samples of level i that are predicted to level j.

The accuracy represents the ratio of each level that is classified correctly, which can be calculated by:

$$
\text{Accuracy} = \frac{1}{m} \sum_{i=1}^{n} k_{ii},
\tag{4}
$$

where m indicates the number of samples in the dataset.

The Cohen's Kappa is a robust criterion that is generally utilized to evaluate the interrater reliability for categorical variables. The range of its value is from -1 to 1, with larger values indicating better agreement. A value greater than or equal to 0.4 indicates a good agreement (McHugh 2012). The Cohen's Kappa can be calculated by:

$$
\text{Kappa} = \frac{m \sum_{i=1}^{n} k_{ii} - \sum \left(\sum_{i=1}^{n} k_{ij} \times \sum_{j=1}^{n} k_{ij} \right)}{m^2 - \sum \left(\sum_{i=1}^{n} k_{ij} \times \sum_{j=1}^{n} k_{ij} \right)}.
\tag{5}
$$

The precision is the proportion of the number of samples that are correctly predicted to the total number of samples that are predicted, which can be calculated by:

$$
\text{Precision} = \frac{k_{ii}}{\sum_{i=1}^{n} k_{ij}}.
\tag{6}
$$

The recall is the proportion of the number of samples that are correctly predicted to the actual total number of samples, which can be calculated by:

$$\text{Recall} = \frac{k_{ii}}{\sum_{j=1}^{n} k_{ij}}. \tag{7}$$

The higher the precision and the recall, the better the results are. However, these two metrics are contradictory in some cases. Then, the F-measure is proposed by comprehensively considering their values, which can be calculated by:

$$F\text{-measure} = \frac{(\beta^2 + 1) \times \text{Precision} \times \text{Recall}}{\beta^2 \times (\text{Precision} + \text{Recall})}, \tag{8}$$

where β indicates the weighting factor, and in general, when $\beta=1$, then F-measure is called F_1.

5 Results and analysis

5.1 Cross-validation results

The hyperparameter optimization results for each model based on the training sets are shown in Table 5. First, the range and interval of values for each hyperparameter were specified, which were identical for the same hyperparameters in different models to make the comparison more reasonable. The search range for the n_estimators was between 10 and 100 with an interval of 10, the search range for the max_depth was between 1 and 10 with an interval of 1, and the search range for the learning_rate was between 0.01 and 0.2 with a an interval of 0.01.

The average accuracy of models for each group of hyperparameter values is shown in Fig. 6. According to Fig. 6, the overall performance of each model can be determined. Comparing with other models, the results of GBDT for different hyperparameter values are more stable. With the increase in hyperparameter values, the average accuracy of Ada-Boost decreases. However, with the increase in hyperparameter values, the average accuracy of XGBoost and LightGBM increases. For the RF, there are several peaks, but its overall accuracy is higher. Based on the best average accuracy of fivefold CV, the optimal hyperparameter values of each model are obtained (see Table 5).

Table 5 Hyperparameter optimization results

Model	Hyperparameter	Range of values	Interval of values	Optimal values
RF	n_estimators	10, 100	10	30
	max_depth	1, 10	1	3
AdaBoost	n_estimators	10, 100	10	50
	learning_rate	0.01, 0.2	0.01	0.01
GBDT	n_estimators	10, 100	10	70
	learning_rate	0.01, 0.2	0.01	0.02
XGBoost	n_estimators	10, 100	10	60
	learning_rate	0.01, 0.2	0.01	0.14
LightGBM	n_estimators	10, 100	10	100
	learning_rate	0.01, 0.2	0.01	0.19
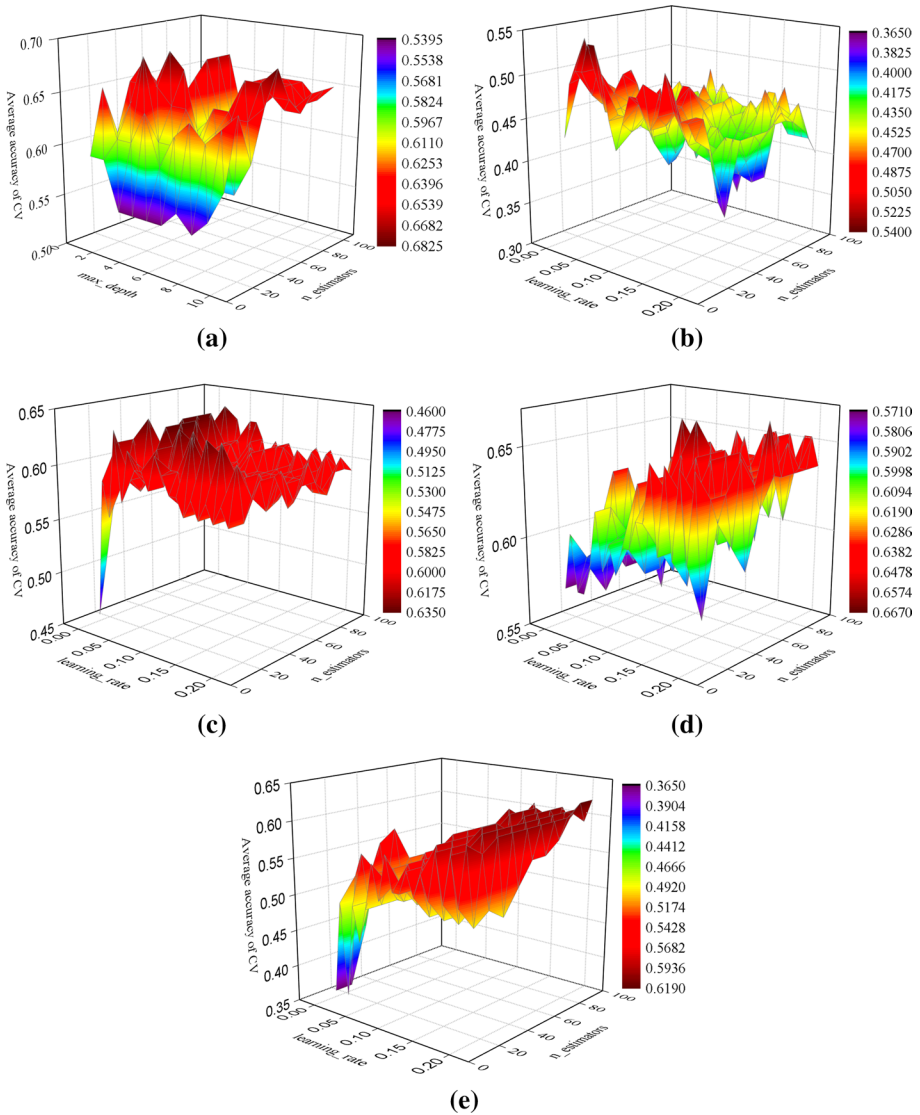
Fig. 6 Average accuracy of CV for different hyperparameters. **a** RF; **b** AdaBoost; **c** GBDT; **d** XGBoost; **e** LightGBM

5.2 Prediction results of the test set

The models with the optimal hyperparameter values were used to predict short-term rock-burst risk on the test set. The prediction results of each model are shown in Table 6. The results of each model form a confusion matrix as defined in Sect. 4.3. The values on the main diagonal represent the number of cases that classified correctly, and others are that misclassified. The values in the first row and fourth column and the fourth row and first column are particularly important. The value in the first row and fourth column indicates

Table 6 Prediction results of each model

		Predicted																		
		RF				AdaBoost				GBDT				XGBoost				LightGBM			
		0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
True	0	11	0	0	0	9	0	2	0	10	0	0	1	9	0	2	0	10	1	0	0
	1	2	3	2	0	2	2	3	0	3	2	2	0	2	2	3	0	3	2	2	0
	2	0	1	7	0	0	1	7	0	0	0	8	0	0	1	7	0	0	2	6	0
	3	0	0	1	3	0	0	2	2	0	0	1	3	0	0	2	2	0	0	2	2

the number of samples with risk level 0 that were predicted as level 3. In this case, it would cause unnecessary panic and loss of prevention costs. According to Table 6, only one such case exists in GBDT, and there are no such cases in other models. The value in the fourth row and first column represents the number of samples with risk level 3 that were predicted as level 0. In this condition, not predicting a major seismic event could produce very serious accidents and even casualties. There is not this kind of case in these five models, which indicates that the ensemble learning models have a certain reliability for this situation.

The accuracy and Cohen's Kappa of each model are shown in Fig. 7. It can be observed that RF performs the best with a highest accuracy of 0.8, followed by GBDT and XGBoost with an accuracy of 0.7667 and 0.7333, respectively. AdaBoost and Light-GBM perform the worst with an accuracy of 0.6667. However, the Cohen's Kappa value of AdaBoost is larger than that of LightGBM, which indicates that the performance of AdaBoost is better than LightGBM. In addition, the rank of the RF, GBDT and XGBoost in term of Cohen's Kappa remains the top three performers. Therefore, after considering both accuracy and Cohen's Kappa, it can be concluded that the performance rank is RF > GBDT > XGBoost > AdaBoost > LightGBM.
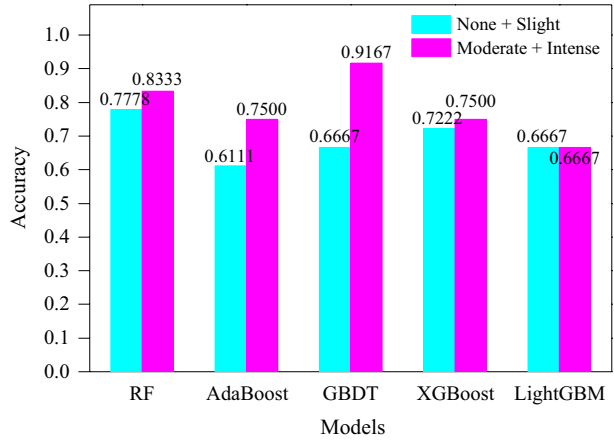
However, the moderate and intense rockbursts should receive more attention in engineering, because they can cause more serious consequences. Therefore, the none and slight risk rockburst cases are combined into a group (low risk rockburst), and the

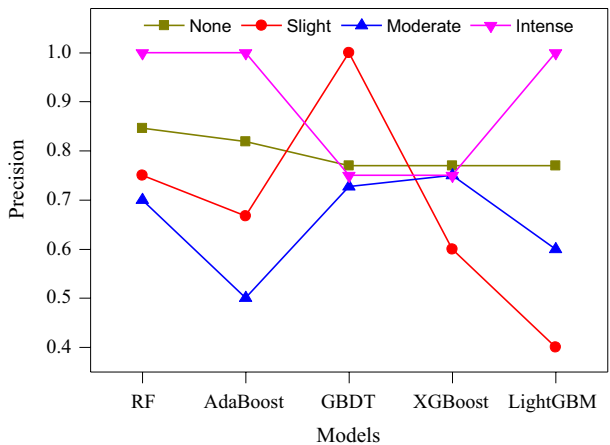**Fig. 7** Accuracy and Cohen's Kappa values of each model
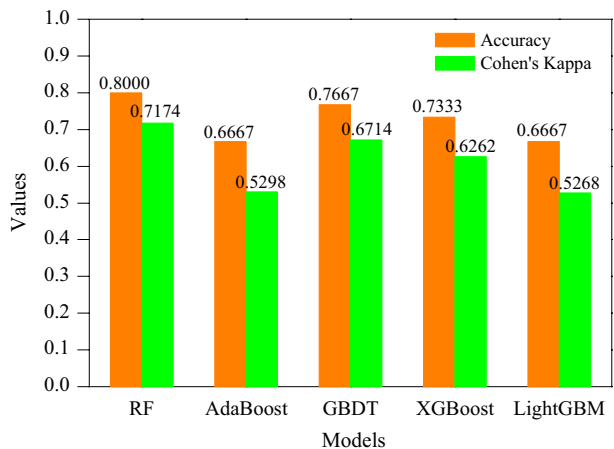
**Fig. 8** Accuracy for low and high risk rockburst


Fig. 9 Precision of each model for each level

moderate and intense risk rockburst cases are combined into another group (high risk rockburst). The accuracy of these two groups is displayed in Fig. 8. In term of high risk rockburst, GBDT possessed the highest value with an accuracy of 0.9167, followed by RF with an accuracy of 0.8333. LightGBM possesses the lowest value with an accuracy of 0.6667. XGBoost and AdaBoost had the same accuracy on high risk rockburst, whereas XGBoost performed better than AdaBoost when predicting low risk rockburst. Therefore, if the center of focus is high risk rockbursts, then the performance rank is GBDT > RF > XGBoost > AdaBoost > LightGBM.

To compare the classification performance of each model for each level, precision, recall and F_1 values are calculated, respectively. Among them, classification precision represents the ability to predict samples correctly. The precision value of each model for each level is shown in Fig. 9. It can be seen that the precision of different models for each level is different. RF, AdaBoost and LightGBM perform the best for the risk of intense rockbursting with a precision of 1.0000, XGBoost performs the best for the risk of moderate rockbursting with a precision of 0.7500, GBDT performs the best for the risk of slight rockbursting with a precision of 1.0000, and RF performs the best for the risk of none rockbursting with

Fig. 10 Recall of each model for each level
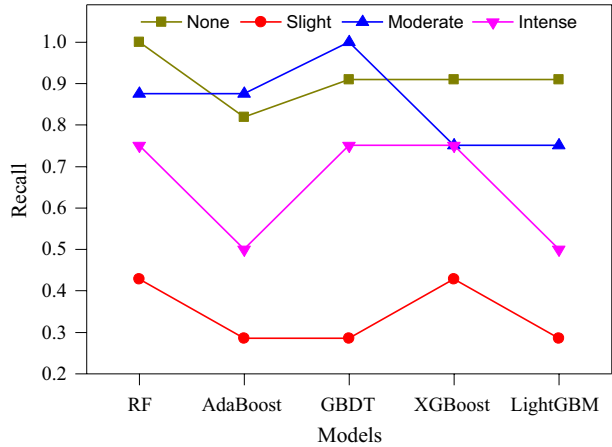
Fig. 11 F_1 value of each model for each level
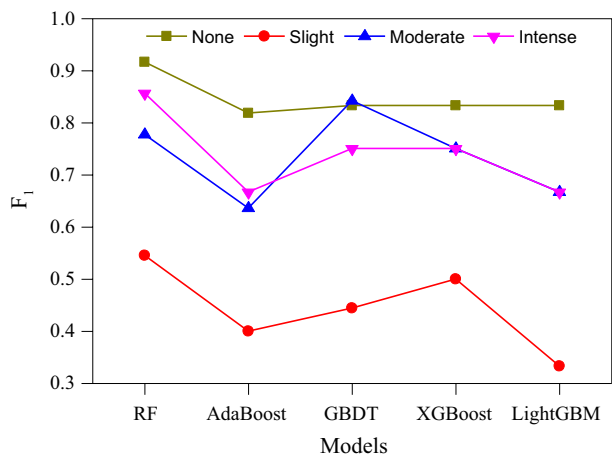
a precision of 0.8462. Overall, the precision of the intense and none risk is higher than the moderate and slight risk. The reason may be that the discrimination boundaries of the risk of moderate and slight rockbursting are more uncertain.

Recall indicates the ability to correctly predict as many events as possible in the actual samples. The recall value of each model for each level is shown in Fig. 10. Comparing with precision values, the variance of recall values among five models is smaller. RF, GBDT and XGBoost achieved the highest recall value (0.7500) for the risk of intense rockbursting, GBDT achieved the highest recall value (1.0000) for the risk of moderate rockbursting, RF and XGBoost achieve the highest recall value (0.4286) for the risk of slight rockbursting, and RF achieves the highest recall value (1.0000) for the risk of none rockbursting. Comparing with other risk levels, the recall value of the risk of slight rockbursting is lowest.

In general, high precision and recall values indicate a good result. However, their values are not positively correlated and may sometimes even be negatively correlated. F_1 is an effective index that measures the accuracy of both precision and recall. The F_1 value of each model for each level is shown in Fig. 11. RF possesses the highest F_1 values for the risk of intense, slight and none rockbursting. The values are 0.8571, 0.5455 and

0.9167, respectively. For the moderate risk of rockbursting, GBDT possesses the highest F_1 value (0.8421), but RF also ranks second with a F_1 value of 0.7778. Therefore, RF has a best performance for rockburst risk prediction from this view. In addition, according to Fig. 11, all of these models have a lower performance for the prediction of the risk of slight rockbursting.

5.3 Overfitting analysis

Overfitting is an important indicator to evaluate the performance of machine learning. A good algorithm should not only perform well in training data, but also have the ability to predict the future data reliably. The prediction accuracy of training and test sets for these five ensemble learning methods is shown in Fig. 12. It can be seen that the GBDT, XGBoost and LightGBM are overfitting to the training set while performing poorly on the test set and AdaBoost performs worst in both training and test sets, whereas RF has a better generalization ability. From this point of view, this study identifies RF as a first-choice algorithm for the short-term rockburst risk prediction for this dataset.

6 Discussion

The database used in this study is the same as that in Feng et al. (2019), but the results obtained are not directly comparable because their test set was manually specified and only contained 10 samples. However, in our work, the training and test sets were randomly split, and the test set included 30 samples to minimize the effects of overfitting and to obtain a more precise prediction of the model performance. In this study, RF obtained the highest accuracy for the overall rockburst samples, whereas GBDT obtained the best accuracy for the risk of moderate and intense rockburst cases. Combined with the precision, recall and F-measure metrics, the optimal model corresponding to different rockburst levels was not always the same. Nevertheless, RF and GBDT had better comprehensive performance metrics. The reason may be that the dataset includes many outliers and these two methods are not sensitive to outliers.

Fig. 12 Prediction accuracy of training and test sets
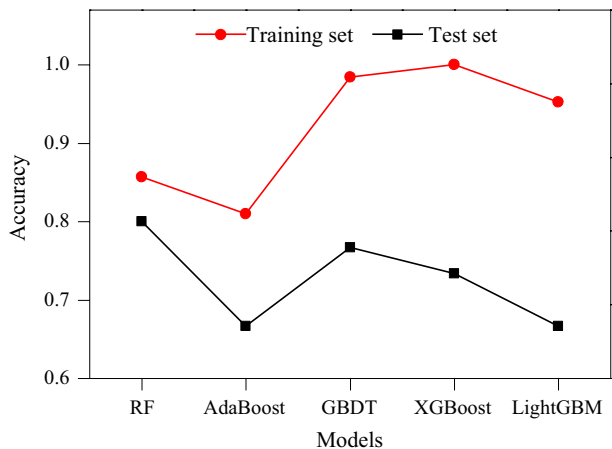
For the performance evaluation of models, most researchers only used the overall accuracy (the ratio of correctly predictive number to the total number) (Wu et al. 2019; Feng et al. 2019; Pu et al. 2019a, b; Liu and Hou 2019). However, this measure does not indicate classification performance of each risk level. Using other metrics of performance, we observed the prediction performance of slight rockbursting was lower than for other risk levels, which is less important than the correct classification of moderate and intense rockburst in engineering applications. There might be three reasons for this misclassification. The first is that the discrimination boundary based on radiated energy is too tight. For example, from Table 2, it can be seen if the radiated energy is equal to 100 J, rockburst risk is slight, whereas if the radiated energy is equal to 101 J, rockburst risk is moderate. The second one is that the discrimination interval of slight level using radiated energy is smaller than moderate and intense levels. The third one is that there is a certain subjectivity to distinguishing the rockburst risk level according to the observed rockburst damage. In addition, the most serious consequence occurs where an intense rockburst takes place when no risk is predicted. On the other hand, economic losses incur when an intense rockburst is predicted where in reality there is no risk. Therefore, using other types of performance measurements than classification accuracy helps understand the prediction results better. In this study, we adopted accuracy, Cohen's Kappa, precision, recall and F-measure to assess the predictive performance of models. In addition, the stability of models for different hyperparameter values, the accuracy of moderate and intense rockbursting, and the misjudgment between the risk of none and intense rockbursting are analyzed.

The proposed methodology is theoretically more scalable for dynamic data, which is important given that the number of seismic events in a database increases with time. The combination of multiple models provides better robustness and can solve such problems more effectively. In actual engineering projects, the optimal methods can vary from dataset to dataset. However, following our proposed methodology, the best fitting ensemble method can always be found in other datasets by comparing the performance of different methods. If the optimal approach obtains a reliable predictive performance, it can be used to predict rockburst risk in new areas. In this study, the highest predictive accuracy is 0.8000 for all cases and 0.9167 for high risk cases, which indicates the proposed methodology can be applied and provide effective guidance for short-term rockburst risk management.

Although the overall prediction results are satisfactory using the proposed methodology, there are also some limitations.

1. The dataset of rockburst samples is relatively small. ML algorithms rely heavily on the quality of the dataset. In general, model overfitting occurs when the training dataset is small, which would decrease the model's generalization and reliability. Although numerous rockburst cases have been reported from all over the world, a public dataset based on microseismic monitoring is rare. If a more comprehensive rockburst database can be established, the short-term rockburst prediction problem could be solved more efficiently using ensemble learning methods.

2. Only the microseismic indices are selected for the short-term rockburst prediction. The rockburst is affected by various factors, such as the intrinsic nature of rock, stress conditions, geologic structure and external disturbance. It is still uncertain whether the microseismic information can completely reflect the comprehensive influence of these factors. Although the selected six microseismic indicators can describe the state of the rock mass fractures, it is meaningful to explore the prediction results by combining the microseismic indicators and other influencing factors.

3. The proposed risk prediction method is only suitable for this kind of rockbursting, in which the source of the seismic event is at the same location as damage in tunnels. This is commonly referred to as strainbursting. There is also another type of rockburst that is induced by the impact of seismic waves due to fault slip or blasting, which is called impact-induced bursting (He et al. 2012). The mechanisms of these two types of rockbursts are different. This study only focuses on the short-term risk prediction for strainbursts.

7 Conclusions

To ensure the safety of workers and the smooth implementation of the projects, short-term rockburst risk prediction using algorithms with high performance is desired in deep level rock engineering. This study compared the comprehensive performance of five ensemble learning methods including RF, AdaBoost, GBDT, XGBoost and LightGBM for short-term rockburst prediction. The hyperparameter values of these five models were optimized based on a training set using fivefold CV. GBDT obtained the most stable results when assigning different hyperparameter values. The model performances were evaluated using the test set after tuning the hyperparameters. Through the analysis of model performance metrics, the RF and GBDT methods yielded the best prediction results. In addition, the predictive performances for different rockburst levels using these models were different. The prediction performance for the risk of none, moderate and intense rockbursting were better than that for the risk of slight rockbursting. Overall, the predictions obtained provide a valuable reference for engineers to base decisions on.

In the future, a rockburst database with a larger number and a higher quality should be established through joint efforts worldwide. In addition, the rockburst data acquisition framework including the indicators and risk levels will be unified. Considering that rockbursting is affected by numerous factors, the effect of different indicators on the prediction results is worth investigating.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Afraei S, Shahriar K, Madani SH (2018) Statistical assessment of rock burst potential and contributions of considered predictor variables in the task. Tunn Undergr Space Technol 72:250–271

Brady BT, Leighton FW (1977) Seismicity anomaly prior to a moderate rock burst: a case study. Int J Rock Mech Min 14:127–132

Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140

Breiman L (2001) Random forests. Mach Learn 45(1):5–32

Cai W, Dou LM, Zhang M, Cao WZ, Shi JQ, Feng LF (2018) A fuzzy comprehensive evaluation methodology for rock burst forecasting using microseismic monitoring. Tunn Undergr Space Technol 80:232–245

Chauhan S, Rühaak W, Khan F, Enzmann F, Mielke P, Kersten M, Sass I (2016) Processing of rock core microtomography images: using seven different machine learning algorithms. Comput Geosci 86:120–128

Chen T, Guestrin C (2016) Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIG-KDD international conference on knowledge discovery and data mining, San Francisco, pp 785–794

Chen BR, Feng XT, Li QP, Luo RZ, Li SJ (2015) Rock burst intensity classification based on the radiated energy with damage intensity at Jinping II hydropower station, China. Rock Mech Rock Eng 48(1):289–303

Deng SK, Wang CG, Wang MY, Sun Z (2019) A gradient boosting decision tree approach for insider trading identification: an empirical model evaluation of China stock market. Appl Soft Comput 83:105652. https://doi.org/10.1016/j.asoc.2019.105652

Dev VA, Eden MR (2019) Formation lithology classification using scalable gradient boosted decision trees. Comput Chem Eng 128:392–404

Dong LJ, Li XB, Peng K (2013) Prediction of rockburst classification using random forest. Trans Nonferrous Met Soc 23(2):472–477

Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu ZF, Chen CW, Han Z, Pham BT (2020) Improved landslide assessment using support vector machine with bagging, boosting, and stacking ensemble machine learning framework in a mountainous watershed, Japan. Landslides 17(3):641–658

Durrheim RJ (2010) Mitigating the risk of rockbursts in the deep hard rock mines of South Africa: 100 years of research. In: Brune J (ed) Extracting the science: a century of mining research. Society for Mining, Metallurgy, and Exploration, Littleton, pp 156–171

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

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. https://doi.org/10.1007/s00521-019-04353-z

Feng XT, Wang LN (1994) Rockburst prediction based on neural networks. Trans Nonferrous Met Soc 4(1):7–14

Feng XT, Chen BR, Zhang CQ, Li SJ, Wu SY (2013) Mechanism, warning and dynamic control of rockburst development processes. Science Press, Beijing **(in Chinese)**

Feng GL, Feng XT, Chen BR, Xiao YX, Yu Y (2015) A microseismic method for dynamic warning of rockburst development processes in tunnels. Rock Mech Rock Eng 48(5):2061–2076

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

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

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

Glazer SN (2018) Mine seismology: data analysis and interpretation. Springer, Dordrecht

Gong FQ, Yan JY, Li XB, Luo S (2019) A peak-strength strain energy storage index for bursting proneness of rock materials. Int J Rock Mech Min 117:76–89

He MC, Xia HM, Jia XN, Gong WL, Zhao F, Liang KY (2012) Studies on classification, criteria and control of rockbursts. J Rock Mech Geotech Eng 4(2):97–114

Hu XC, Su GS, Chen K, Li TB, Jiang Q (2019) Strainburst characteristics under bolt support conditions: an experimental study. Nat Hazards 97(2):913–933

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

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

Jung Y (2018) Multiple predicting K-fold cross-validation for model selection. J Nonparameter Stat 30(1):197–215

Kaiser PK, Cai M (2012) Design of rock support system under rockburst condition. J Rock Mech Geotech Eng 4(3):215–227

Ke GL, Meng Q, Finley T, Wang TF, Chen W, Ma WD, Ye QW, Liu TY (2017) LightGBM: a highly efficient gradient boosting decision tree. In: 31st Annual conference on neural information processing systems, Long Beach, pp 3146–3154

Keneti A, Sainsbury BA (2018) Review of published rockburst events and their contributing factors. Eng Geol 246:361–373

Kotsiantis SB (2013) Decision trees: a recent overview. Artif Intell Rev 39(4):261–283

Kotsiantis SB (2014) Bagging and boosting variants for handling classifications problems: a survey. Knowl Eng Rev 29(1):78–100

Krawczyk B, Minku LL, Gama J, Stefanowski J, Woźniak M (2017) Ensemble learning for data stream analysis: a survey. Inform Fusion 37:132–156

Kumar P (2019) Machine learning quick reference. Packt Publishing Ltd., Birmingham

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 N, Feng XD, Jimenez R (2017a) Predicting rock burst hazard with incomplete data using Bayesian networks. Tunn Undergr Space Technol 61:61–70

Li TZ, Li YX, Yang XL (2017b) Rock burst prediction based on genetic algorithms and extreme learning machine. J Cent S Univ 24(9):2105–2113

Liang WZ, Zhao GY, Wu H, Dai B (2019a) Risk assessment of rockburst via an extended MABAC method under fuzzy environment. Tunn Undergr Space Technol 83:533–544

Liang WZ, Zhao GY, Wang X, Zhao J, Ma CD (2019b) Assessing the rockburst risk for deep shafts via distance-based multi-criteria decision making approaches with hesitant fuzzy information. Eng Geol 260:105211

Liang WZ, Dai B, Zhao GY, Wu H (2020) A scientometric review on rockburst in hard rock: two decades of review from 2000 to 2019. Geofluids 2020:1–17

Lin Y, Zhou KP, Li JL (2018) Application of cloud model in rock burst prediction and performance comparison with three machine learning algorithms. IEEE Access 6:30958–30968

Liu YR, Hou SK (2019) Rockburst prediction based on particle swarm optimization and machine learning algorithm. In: International conference on information technology in geo-engineering, Cham, pp 292–303

Liu JP, Feng XT, Li YH, Sheng Y (2013) Studies on temporal and spatial variation of microseismic activities in a deep metal mine. Int J Rock Mech Min 60:171–179

Ma TH, Tang CA, Tang LX, Zhang WD, Wang L (2015) Rockburst characteristics and microseismic monitoring of deep-buried tunnels for Jinping II Hydropower Station. Tunn Undergr Space Technol 49:345–368

Ma X, Westman E, Slaker B, Thibodeau D, Counter D (2018a) The b-value evolution of mining-induced seismicity and mainshock occurrences at hard-rock mines. Int J Rock Mech Min 104:64–70

Ma TH, Tang CA, Tang SB, Kuang L, Yu Q, Kong DQ, Zhu X (2018b) Rockburst mechanism and prediction based on microseismic monitoring. Int J Rock Mech Min 110:177–188

McHugh ML (2012) Interrater reliability: the kappa statistic. Biochem Medica 22(3):276–282

Ontario Ministry of Labour (2015) Mining health, safety and prevention review. Toronto: Government of Ontario. https://www.labour.gov.on.ca/english/hs/pubs/miningfinal/

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

Pu Y, Apel DB, Wang C, Wilson B (2018a) Evaluation of burst liability in kimberlite using support vector machine. Acta Geophys 66(5):973–982

Pu Y, Apel DB, Lingga B (2018b) Rockburst prediction in kimberlite using decision tree with incomplete data. J Sustain Min 17(3):158–165

Pu Y, Apel DB, Pourrahimian Y, Chen J (2019a) Evaluation of rockburst potential in kimberlite using fruit fly optimization algorithm and generalized regression neural networks. Arch Min Sci 64(2):279–296

Pu Y, Apel DB, Xu H (2019b) Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier. Tunn Undergr Space Technol 90:12–18

Pu Y, Apel DB, Wei C (2019c) Applying machine learning approaches to evaluating rockburst liability: a comparison of generative and discriminative models. Pure appl Geophys 176:4503–4517

Rokach L (2016) Decision forest: twenty years of research. Inform Fusion 27:111–125

Sagi O, Rokach L (2018) Ensemble learning: a survey. Wires Data Min Knowl 8(4):e1249

Schapire RE (1990) The strength of weak learnability. Mach Learn 5(2):197–227

Srinivasan C, Arora SK, Yaji RK (1997) Use of mining and seismological parameters as premonitors of rockbursts. Int J Rock Mech Min 34(6):1001–1008

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

Woźniak M, Graña M, Corchado E (2014) A survey of multiple classifier systems as hybrid systems. Inform Fusion 16:3–17

Wu SC, Wu ZG, Zhang CX (2019) Rock burst prediction probability model based on case analysis. Tunn Undergr Space Technol 93:103069. https://doi.org/10.1016/j.tust.2019.103069

Xue RX, Liang ZZ, Xu NW, Dong LL (2020a) Rockburst prediction and stability analysis of the access tunnel in the main powerhouse of a hydropower station based on microseismic monitoring. Int J Rock Mech Min 126:104174. https://doi.org/10.1016/j.ijrmms.2019.104174

Xue YG, Bai CH, Qiu DH, Kong FM, Li ZQ (2020b) Predicting rockburst with database using particle swarm optimization and extreme learning machine. Tunn Undergr Space Technol 98:103287

Website of the XGBoost library. https://xgboost.readthedocs.io/en/latest/

Website of the LightGBM library. https://lightgbm.readthedocs.io/en/latest/

Zhang CQ, Feng XT, Zhou H, Qiu SL, Wu WP (2012) Case histories of four extremely intense rockbursts in deep tunnels. Rock Mech Rock Eng 45(3):275–288

Zhang MW, Liu SD, Shimada H (2018) Regional hazard prediction of rock bursts using microseismic energy attenuation tomography in deep mining. Nat Hazards 93(3):1359–1378

Zheng YC, Zhong H, Fang Y, Zhang WS, Liu K, Fang J (2019) Rockburst prediction model based on entropy weight integrated with grey relational BP neural network. Adv Civ Eng. https://doi.org/10.1155/2019/3453614

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

Zhou J, Li XB, Mitri HS (2016a) Classification of rockburst in underground projects: comparison of ten supervised learning methods. J Comput Civil Eng 30(5):04016003

Zhou J, Shi XZ, Huang RD, Qiu XY, Chen C (2016b) Feasibility of stochastic gradient boosting approach for predicting rockburst damage in burst-prone mines. Trans Nonferrous Met Soc 26(7):1938–1945

Zhou J, Guo HQ, Koopialipoor M, Armaghani DJ, Tahir MM (2020) Investigating the effective parameters on the risk levels of rockburst phenomena by developing a hybrid heuristic algorithm. Eng Comput. https://doi.org/10.1007/s00366-019-00908-9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.