



Applying machine learning approach in predicting short-term rockburst risks using microseismic information: a comparison of parametric and non-parametric models

Prabhat Man Singh Basnet¹ · Aibing Jin¹ · Shakil Mahtab²

Received: 28 January 2023 / Accepted: 12 July 2024
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Abstract

Microseismic (MS) information is often utilised in deep underground engineering projects for the early warning of short-term rockburst hazards. Due to the complex nature of rockburst occurrence, predicting short-term rockburst is always challenging. Recently, machine learning (ML) methods are often employed in different geotechnical engineering applications. Parametric and non-parametric ML methods are two different kinds of approaches, each with distinct characteristics. However, the current applications in short-term rockburst prediction are focused on non-parametric methods. Therefore, this paper proposes and studies the feasibility of a parametric model over the non-parametric model, adopting two fundamental parametric and non-parametric ML models, including logistic regression and support vector machine, to predict short-term rockburst using MS information based on two types of normally and non-normally distributed datasets. After modelling, precision, recall, F1 score, and receiving operating curve are considered to evaluate the model's strength in predicting tasks. The results indicate that the parametric model, which obtained an average F1 score and AUC score of 0.72 and 0.91 on a normally distributed dataset achieved more remarkable output in evaluating short-term rockburst risk. Limited data availability is always a challenge in short-term rockburst prediction. In such cases, parametric models can accurately classify the rockburst risk levels due to their characteristics of assuming the predefined function, simplifying the learning processes independent of the data size. However, normally distributed data is beneficial for them that allows a perfect fit. The presented work effectively identifies the rockburst risk in deep underground excavation projects regardless of data size.

Keywords Short-term rockburst · Parametric and non-parametric models · Logistic regression · Support vector machine · Intelligent model

✉ Aibing Jin
jab47U@hotmail.com

¹ Key Laboratory of Ministry of Education for Efficient Mining and Safety of Metal Mine, University of Science and Technology Beijing, Beijing 100083, China

² Department of Geotechnical Engineering, College of Civil Engineering, Tongji University, Shanghai 200092, China

1 Introduction

With increasing depletion of near surface deposits as well as underground space development, underground engineering operations are driving deeper and deeper, causing problems related to rock mechanics. This has led to an increase in underground stress causing rockburst hazard which can be seriously damaging to an excavation (Bruning 2018). Rockburst is a phenomenon which occurs in an underground opening from the violent failure of rock mass due to extreme stress concentrations (Hoek 2000; Song et al. 2024). Rockburst particularly threatens the safety of workers and causes destruction to underground works, especially in deep level mining in which there are tunnels and shafts with hard brittle rock (Nussbaumer 2000). Since the beginning of human exploration towards minerals and underground structures, rockburst has become a common instability phenomenon hindering tremendous projects. The earliest recorded instance of rockburst, the Altenberg tin mine in Germany during 1640, is an example of such event that compelled a halt to mining operations for several decades (Ortlepp 2005). Thereafter, various countries around the world such as Australia, China, South Africa, the USA, etc., have accounted for cases of rockburst in mining and tunnelling works (Głowacka 1993; Wang et al. 2021; Ling et al. 2023; Zhang et al. 2024a, b). Due to the sudden and severe consequences led by the rockburst hazard, researchers around the world have been urged to identify prediction methods and measures for control. Generally, rockburst prediction is investigated on the basis of two main approaches: long-term prediction and short-term prediction (Man Singh Basnet et al. 2023). Long-term rockburst typically focuses on predicting the tendency of different levels of intensity at different sections of mines and tunnels, relying on different rock mechanical parameters during the design stage of the project. Nevertheless, long-term prediction does not tell the time of occurrence or location of events (Liang and Zhao 2022). Unlikely in long-term prediction, short-term prediction employs field monitoring to estimate early warnings of rockburst during the excavation phase (Feng 2017; Zhang et al. 2024a, b). In order to predict the rockburst in real time, microseismic monitoring is commonly used in calculating three-dimensional source location and the different levels of damage severity depending on the monitored information (Song et al. 2024). Hence, scholars have utilised various kinds of MS information to evaluate the occurrence of rockburst. For instance, Feng et al., with the help of six MS parameters, established a formula for dynamic warning of rockburst (Feng et al. 2014). Feng et al. studied the fractal behaviour of energy distribution associated with MS events during the rockburst development process (Feng et al. 2016). Chen et al. based on radiated MS energy, developed a quantitative intensity classification method (Chen et al. 2015). Additionally, Ma et al., applying MS monitoring, investigated the relation between rockburst and its influencing factors (Ma et al. 2015). Further, Liu et al. examined the MS parameters in the Hongtoushan copper mine and concluded that all multiparameter demonstrated precursory behaviour before the ground pressure hazard (Liu et al. 2013). Zhang et al. investigated the occurrence processes of the fault slip rockburst utilising in-situ failure analysis, geological surveys, and MS information to understand the development and occurrence mechanisms. The results demonstrated that fault-slip rockbursts could be effectively monitored and analysed using MS technology, highlighting the importance of understanding geological and stress conditions to mitigate such events (Zhang et al. 2022). Tang et al. studied the mechanism of rockburst in water-rich (WR) areas during tunnel excavation using microseismic monitoring and experimental analysis, comparing

WR and water-poor (WP) sections. The results showed that rockbursts are influenced by the classification of surrounding rock and excavation methods, with high in-situ stress conditions accelerating rockburst occurrence in WR areas (Tang et al. 2023).

Although different scholars experimentally studied rockburst occurrence utilising various MS multi parameters, due to inconsistencies aroused in establishing the exact relation between MS information and the intensity warning of rockburst, researchers developed mathematical and intelligent models to predict rockburst risks. Recently, intelligent machine learning methods have gained more concern because they do not need prior understanding related to the input and output variables and instead simply learn the pattern from data to predict the outcome (Pu et al. 2019; Man Singh Basnet et al. 2023). Taking advantage of ML application, attempts have been made to predicting short-term rockburst. Feng et al. developed the probabilistic neural network (PNN) model, feeding 93 rockburst cases obtained from Jinping II hydropower. This model is further improved using a mean impact value algorithm and modified firefly algorithm for better performance. After optimizing the PNN model, accuracy increased by almost 25% in comparison to the single PNN model (Feng et al. 2019). Liang et al. employed the ensemble boosting technique to verify its strength in predicting short-term rockburst. The final result depicted random forest and the gradient boosted decision tree achieved better performance in classifying rockburst intensity (Liang et al. 2020). Liang et al. further developed the stacking ensemble technique, embedding six different classifiers as a base learner. The model is evaluated using different performance metrics and the final outcome simply illustrated ensemble classifiers are more powerful than each base learner in performing the task (Liang et al. 2021). Zhao et al. created a decision tree (DT) model and also investigated the relationship between MS features and rockburst. The model was successfully employed to predict case histories and has only misclassified two sample (Zhao et al. 2021). Additionally, Yin et al. took 1500 MS events and proposed a tree-based model for real time prediction. In order to establish precursory MS sequences, the dimensionality technique is used and data is labelled to form 300 precursory MS sequences via the grouping rule. Finally, two types of precursor trees with and without pruning were used to validate the result (Yin et al. 2021). Jin et al. built a nonlinear support vector machine (nonlinear-SVM) and tested 22 samples. Initially, radial basis function (rbf) was identified as the best performing kernel which is utilised to build a final model by optimizing the hyperparameters, proving remarkable

Table 1 Characteristics of four rockburst intensity levels (Chen et al. 2015)

Rockburst intensity classes	Characteristics
None	No distinct failures can be observed on the surface of rock mass. Construction and support system have no any effect
Slight/Weak	Slight spalling and slabbing appear as a main failure type on the surface of rock mass; slight rock mass ejection occurs with the fragments having size of around 10–30 cm; failure depth reach <0.5 m and slight cracking sound can be heard
Moderate	Severe slabbing and spallings;30–80 cm ejected rock fragment size with; cracking sound that resembles detonator blasting; the size of failure depth reaches 0.5–1.0 m; shotcrete lining between rock bolts gets damaged
Intense	Massive volume of rock mass ejected;80–150 cm ejected rock fragment size; failure zone contains fresh fracture planes; a lasting sound that resembles thunder and explosives; and the failure depth reaches 1.0–3.0 m

in predicting samples with few misclassifications (Jin et al. 2022). Feng et al. employed knowledge of clustering analysis for establishing the rockburst intensity warning using MS monitoring parameters. The main advantage of the proposed work is that it only needs few samples as an input. The predicted cases of Jinping II hydropower China from his work corresponded with the actual situation (Feng et al. 2020). Finally, Basnet et al. developed an explainable risk prediction model and predicted short-term rockburst risks. The results showed that the model not only predicts the risks but also interprets the decisions made by the model (Basnet et al. 2024).

All aforementioned works are examples of the non-parametric model approach. In fact, recent trends of ML employment in short-term rockburst belong to non-parametric models. As other different types of ML models, parametric models such as logistic regression (LR), naïve bayes (NB), linear discriminant analysis (LDA), etc. have unique theory bases and processes of modelling in contrast to non-parametric models. Nonetheless, parametric models have been infrequently employed in short-term rockburst risk prediction, which is why it is tempting to explore its significance now. The characteristic of a parametric model is that it learns using a pre-defined functional form with a fixed number of parameters. While parametric models are generally less sensitive to the size of the data than non-parametric models, and it basically assumes the data to follow a normal distribution. On the contrary, non-parametric models do not rely on shape of function and the distribution of data but instead generalise well when the number of data is large enough. As a result, to check the feasibility of the parametric model over the non-parametric model, two types of datasets are prepared that are normally (normally transformed) and non-normally (original) distributed data using a small numbers of samples and inputs.

Rockburst prediction work is often complex and complicated process, and there is no study of parametric models in short-term rockburst prediction. Hence, this work compares the result from a parametric and non-parametric model typified by LR and SVM, respectively. The remainder of this paper is designed as follows: the dataset description section describes the source of data and provides the statistical information of two different datasets; the preliminaries section provides a brief of what are parametric and non-parametric models followed by an elaboration of LR and SVM. In the model building section, both LR and SVM are constructed using two different datasets which are compared and evaluated.

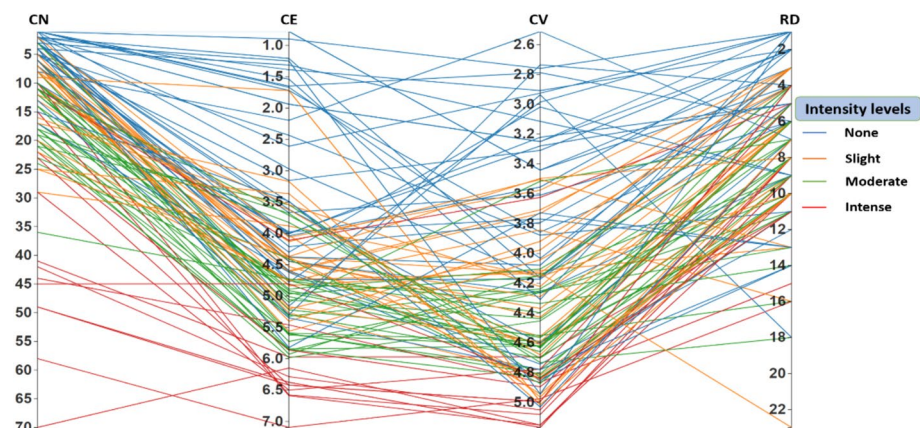


Fig. 1 Parallel plot of distribution of four input indices with respect to intensity levels

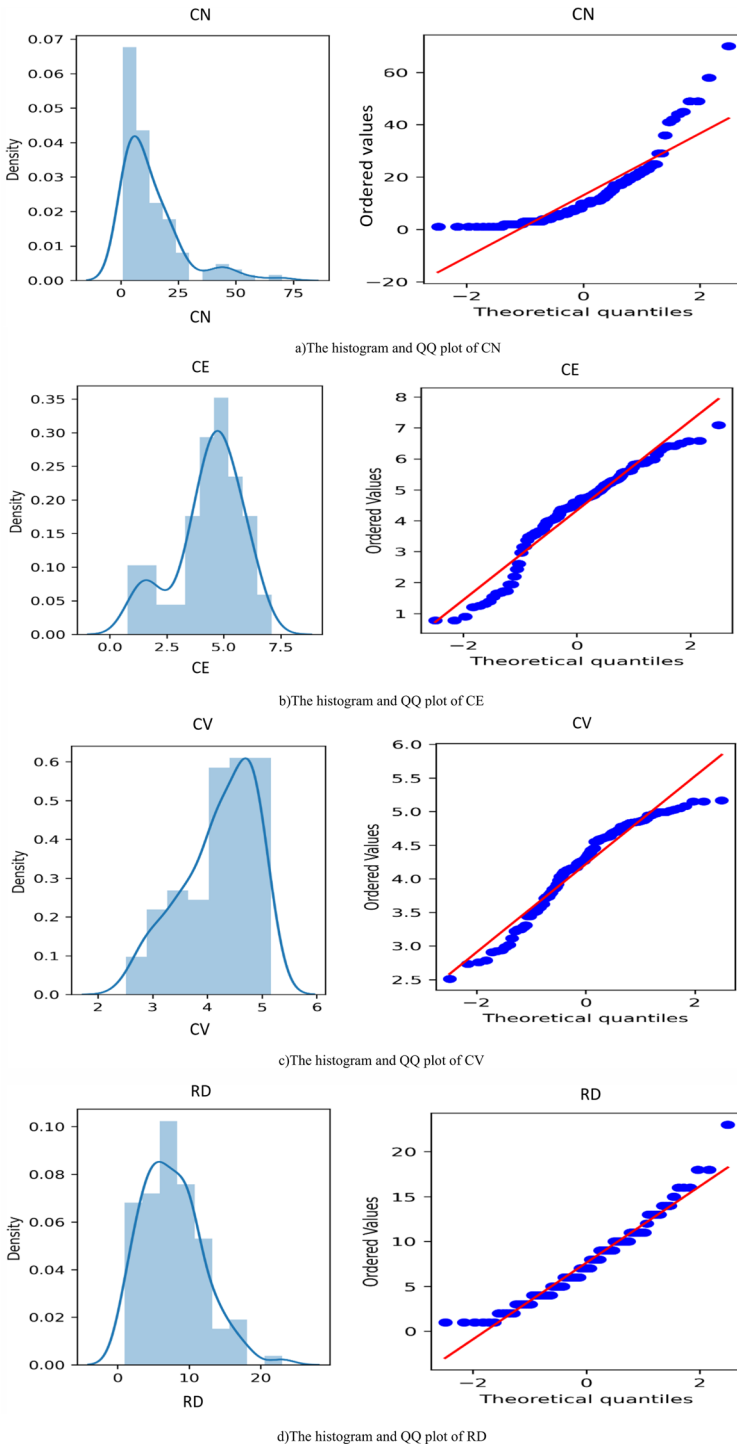


Fig. 2 Histogram and QQ plot of features in Dataset I. **a** The histogram and QQ plot of CN. **b** The histogram and QQ plot of CE. **c** The histogram and QQ plot of CV. **d** The histogram and QQ plot of RD

2 Dataset description

The fundamental step for building ML models simply requires data samples to train and test the model. Therefore, a dataset has been collected by gathering information from the field work and various articles on short-term rockburst based on microseismic parameters. 108 case records were gathered from different sources (Feng et al. 2019, 2013; Liu et al. 2021; Zhao et al. 2021) which contains four features, such as cumulative no. of events (CN), cumulative released energy (CE), cumulative apparent volume (CV) and rockburst development day (RD). CN accounts for the number of microfractures during the events, CE tells the amount of radiated energy that measures the fracturing strength of rock mass, CV is the volume of the inelastic deformation which also describes the amount of damage in rock mass and RD refers to an incubation period of incident. In this study two predictors CE and CV are taken in logarithmic form to ensure the constant correlation and make calculation convenient. All of the mentioned input indices are used to predict the target variable “rockburst intensity”. Basically, there are four types of rockburst intensity levels which are categorised relying on the degree of damages. The characteristics of the four intensity levels are shown in Table 1.

The microseismic features contain the information that is utilised to build a model in order to predict the target (rockburst intensity). The characteristics and patterns between rockburst feature and target can be seen in Fig. 1 which is a parallel coordinate plot that describes the relation between inputs and outputs. Parametric models generally rely on assumption of normality for better performance; therefore, two datasets have been prepared in order to compare the models in both datasets, one with original data that is directly collected from the field and literature and another with transformed data with normal distribution. The first one is named “Dataset I” whereas the second is named “Dataset II”. Since both datasets contain target variables in the form of categorical variables and ML models are unable to handle such type of labels, categorical labels are transformed into numeric form by assigning numeric value to each class; this is represented by 0, 1, 2, 3 for None, Slight, Moderate and Intense rockburst, respectively. Both datasets contain 38 None cases, 27 Slight cases, 29 Moderate cases and 14 Intense cases.

2.1 Dataset I visualization

Dataset I contains the original information collected from the field and literature in its original form; for parametric ML models, normality is considered as a basic assumption because they rely on a known form of mapping function to make a prediction which aids in better fitting of a line or hyperplane. Therefore, there are two ways of checking normality using graphical and numerical methods (Park 2008). These two methods often employ the knowledge of histogram, Q-Q plot, skewness and kurtosis values to determine whether feature variables are normally distributed (Field 2013). Basically, normally distributed data should have a histogram plot resembling a bell shape curve and Q-Q plot having all data points falling on the straight line (Thode 2002). Similarly, skewness and kurtosis measure the symmetry and heaviness of distribution tails respectively. Observing Fig. 2, on the histogram and Q-Q plot, each feature is not much normally distributed; CN and RD are right skewed and kurtotic whereas CE and CV are left skewed. In terms of Q-Q plot, all of the

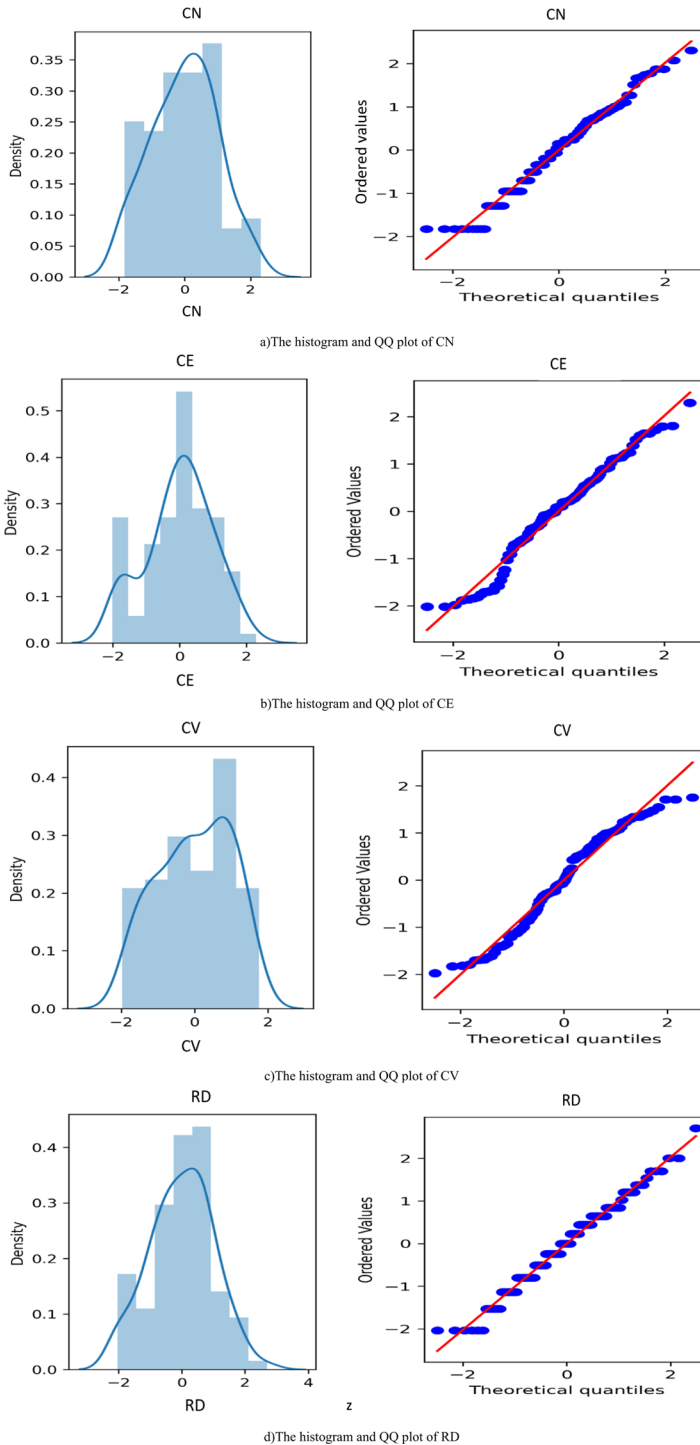
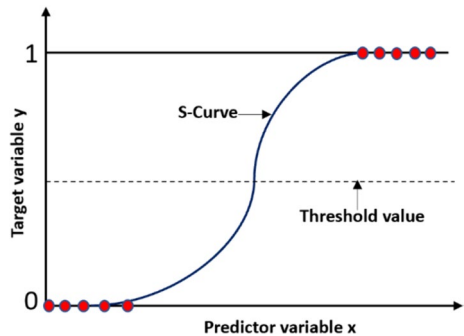


Fig. 3 Histogram and QQ plot of each feature. **a** The histogram and QQ plot of CN. **b** The histogram and QQ plot of CE. **c** The histogram and QQ plot of CV. **d** The histogram and QQ plot of RD

Fig. 4 Working principle of LR



data points do not fall on the straight line which simply means Dataset I is not much normally distributed. The publicly available Dataset can be found in Appendix.

2.2 Dataset II visualisation

From a statistical point of view, Dataset I does not meet the requirement of assumption. Therefore, Dataset II has been prepared applying Box-Cox transformation, which transforms the non-normally distributed data into Gaussian or Gaussian like distribution (Box and Cox 1964). Drawing on Fig. 3, it is obvious that all features almost follow the Gaussian distribution. On the histogram plot, the distribution of independent variables more or less resembles the bell shape curve. In comparison to Dataset I, the Q-Q plot for Dataset II depicts that, for every feature, almost all of the data points fall on the straight line.

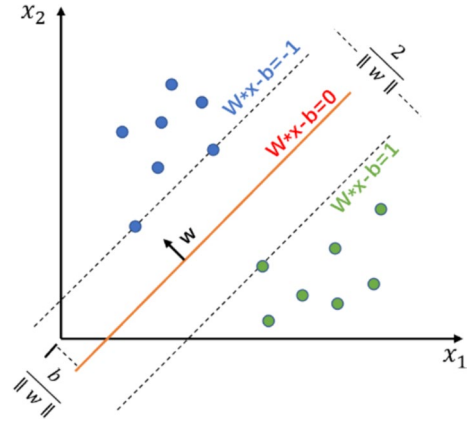
3 Preliminaries

3.1 Parametric and non-parametric model

Parametric models generally assume the form of function (f) which portrays the relation between input(X) with output(Y), represented as $Y=f(x)$. The algorithm utilises training records to learn the target mapping function. Parametric models presume the population can be accurately represented by a probability distribution with a defined set of parameters (James et al. 2013). This is acquainted with whether the population is normal, unless then it may be easily estimated using a normal distribution that is possible by invoking the central limit theorem. Parametric approaches are often model based and rely on assumption regarding the shape of the function to be inferred. The suitable model is then considered to estimate the collection of parameters.

Non-parametric models do not make assumptions regarding the structure of the function to be estimated. These techniques work by estimating the unknown function (f), which may have any form and does not require any assumption about the population's parameters (James et al. 2013; Russell 2010). In fact, the non-parametric method does not require data to be normally distributed; it simply tends towards additional precision which attempts to

Fig. 5 SVM general principle for hyperplane maximisation (Jin et al. 2022)



seek the best fit for the observations. Generally, a huge amount of data is necessary to approximate the undermined the function (f) exactly.

Both parametric and non-parametric models have their own form of strength and shortcomings depending on the specific problems. To explore the significance in short-term rockburst prediction with fewer data and features, Logistic Regression and Support Vector Machine classifiers are adopted as a parametric and non-parametric model, respectively.

3.2 Logistic regression

Logistic regression can be regarded as a transformed form of linear regression applying sigmoid function. Logistic regression utilises sigmoid function to map the input variables and output probabilities. Say, $X=(X_1, \dots, X_n)$ is an input vector containing the information of the features, then the conditional probability of output variable Y is estimated relying on a set of observations on X, represented as $P(Y=y|X=x)$ (Cox 1958). General graphical representation of the probability curve of logistic regression can be seen in Fig. 4 in which the vertical axis is the probability result for a given classification and whereas the horizontal axis stands for the value of X.

Since, the distribution of $y|x$ is assumed to be a Bernoulli distribution, the conditional probability is written as:

$$p(y|x) = f(x)^y(1 - f(x))^{1-y} \tag{1}$$

$f(x)$ denotes the parameter of the Bernoulli distribution which is a function of the input data,

$$f(x) = p(y = 1|x) = E(y|x).$$

Further, $f(x)$ can be calculated utilising the logistic function and the linear transformation of the input variable is given as:

$$f(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_{1x})}} \tag{2}$$

$\beta_0 + \beta_{1x}$ is analogous to the linear form $y = ax + b$. The sigmoid function attempts to confine the Y value between 0 and 1.

LR experimentally uses several parameters; the “C” parameter refers to an inverse of regularisation degree in which the greater the value of C, the lesser the regularisation. The “fit_intercept” parameter defines whether a constant number is necessary for the decision function. Likewise, “solver” aids in the optimisation problem. Newton-cg, lbfgs, liblinear, sag and saga are the common types of solvers.

3.3 Support vector machine

Support vector machine is a non-parametric model that is built for classification and regression relying on the statistical learning framework (Cortes and Vapnik 1995). SVM maximises the width of the margin between two observations by mapping a collection of training samples to points in space. Then, new instances are mapped into the same space and fit into the category to which they belong. SVM can perform linear as well as nonlinear classification problems; if data is nonlinear, it simply uses a kernel trick to classify the nonlinear data by representing it into the higher dimensional feature spaces.

For training samples in a feature space having n data points, it can be written as; $(x_1, y_1), \dots, (x_n, y_n)$, where $y_i \in \{+1, -1\}$, x_i denotes the i th feature vector and y_i estimates the label of x_i . SVM aims to maximise the margin between the set of points x_i as given in Fig. 5. For the set of points x , any hyperplane can be represented with the following equation:

$$w^T x - b = 0 \quad (3)$$

w signifies a normal vector to the hyperplane.

If samples are linearly separable, two parallel hyperplanes that separate the two classes of data are selected so as to ensure that the maximum width existed between them. The area covered by these two parallel hyperplanes is renowned as margin and their respective equations can be defined as

$$w^T x - b = 1 \quad (4)$$

Data points that lie on or above this boundary are indicated with label 1

$$w^T x - b = -1 \quad (5)$$

On the contrary, data points that are situated on or below this boundary are labelled with -1. Geometrically, the distance between these hyperplanes can be denoted as $\frac{2}{\|w\|}$, where $\|w\|$ must be as minimum as possible to provide the largest margin between two planes. The distance is calculated applying the distance from a point to plane equation. In order to prevent the data points falling inside the margin, constraints are added. For each i , constraints are expressed by

$$w^T x_i - b \geq 1, \text{ for } y_i = 1 \quad (6)$$

Or

$$w^T x_i - b \geq -1, \text{ for } y_i = -1 \quad (7)$$

Solving (6) and (7), the following terms can be obtained

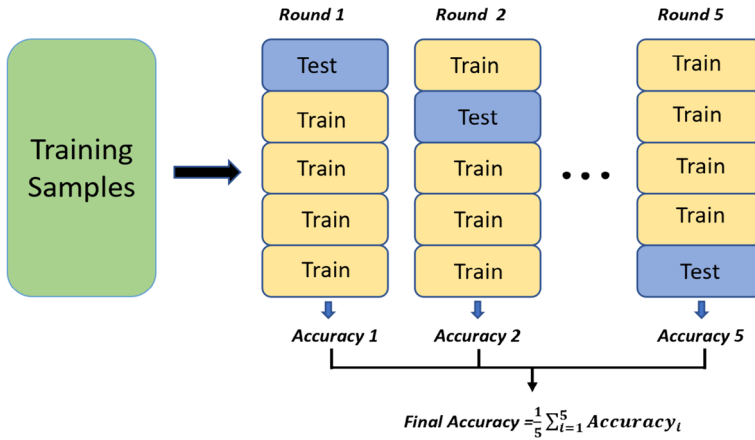


Fig. 6 Schematic diagram of fivefold cross-validation

$$y_i(w^T x_i - b) \geq 1, \text{ for all } 1 \leq i \leq n \tag{8}$$

By minimising $\| w \|$ subject to Eq. (8), values for w and b can be obtained that determine our classifier. $x \mapsto \text{sgn}(w^T x - b)$, $\text{sgn}(\cdot)$ which denotes the sign function. The maximum margin hyperplane is totally decided by the x_i which is situated nearest to it. Such x_i are known as support vectors.

On the other hand, for the non-linearly separable data, a hinge function is utilised.

$$\text{Max}(0, 1 - y_i(w^T x_i - b)) \tag{9}$$

Simply, y_i denotes i th target and $w^T x_i - b$ is the i th output. The optimisation mainly focuses on minimising the given function

$$C \| w \|^2 + \left[\frac{1}{n} \sum_{i=1}^n \text{Max}(0, 1 - y_i(w^T x_i - b)) \right] \tag{10}$$

C is a non-negative parameter which penalises for misclassification. Minimising Eq. (10) is regarded as a constrained optimisation problem having a differential objective function which then expressed for every data point by introducing a new variable ξ_i in order to rewrite the Eq. (8) in the following constraint condition

$$y_i(w^T x_i - b) \geq 1 - \xi_i \tag{11}$$

Finally, the problem of optimisation is given by the following order

$$\min \frac{1}{n} \sum_{i=1}^n \xi_i + C \| w \|^2 \tag{12}$$

Subject to $y_i(w^T x_i - b) \geq 1 - \xi_i$ and $\xi_i \geq 0 \ i=1, 2, \dots, n$

Applying the Lagrange function, Eq. (12) turns the simplest problem.

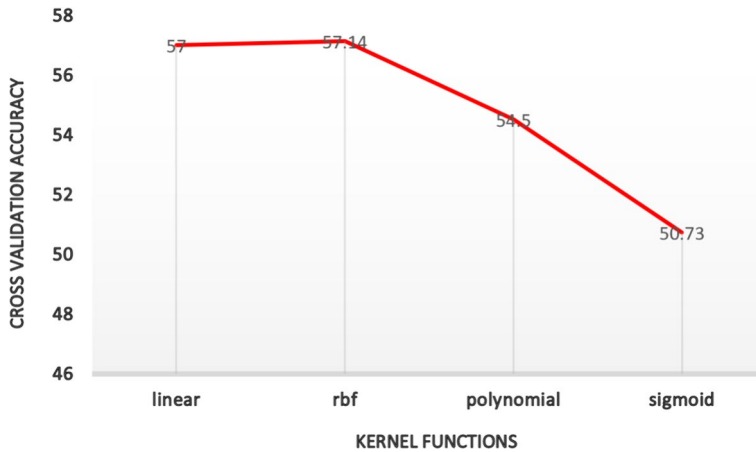


Fig. 7 Best performing SVM kernel on dataset I

$$Maximize f(c_1 \dots c_n) = \sum_{i=1}^n c_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i c_i (x_i^T x_j) y_j c_j \tag{13}$$

Since this belongs to the dual maximisation problem and the c_i is denoted by the given expression

$$w = \sum_{i=1}^n c_i y_i x_i \tag{14}$$

Once the optimal result is obtained after computing Eq. (13), the hyperplane and decision-making function is decided.

In the case of nonlinear SVM, kernel function k which satisfies the $k(x_i, x_j) = \varphi(x_i)\varphi(x_j)$ is introduced for the transformed data points. Now, in terms of optimisation problem, SVM can be expressed as;

$$\begin{aligned}
 Maximize(c_1 \dots c_n) &= \sum_{i=1}^n c_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i c_i k(x_i, x_j) y_j c_j \\
 \text{subject to } \sum_{i=1}^n c_i y_i &= 0 \text{ and } \leq c_i \leq \frac{1}{2nC} \text{ for}
 \end{aligned}
 \tag{15}$$

The final value of c_i can be gained by resolving the Eq. (15) and $b - \left[\sum_{i=1}^n c_j y_j k(x_i, x_j) \right] - y_i$. After all, decision making non-linear SVM is given by

$$b - \left[\sum_{i=1}^n c_j y_j k(x_i, x_j) \right] - y_i$$

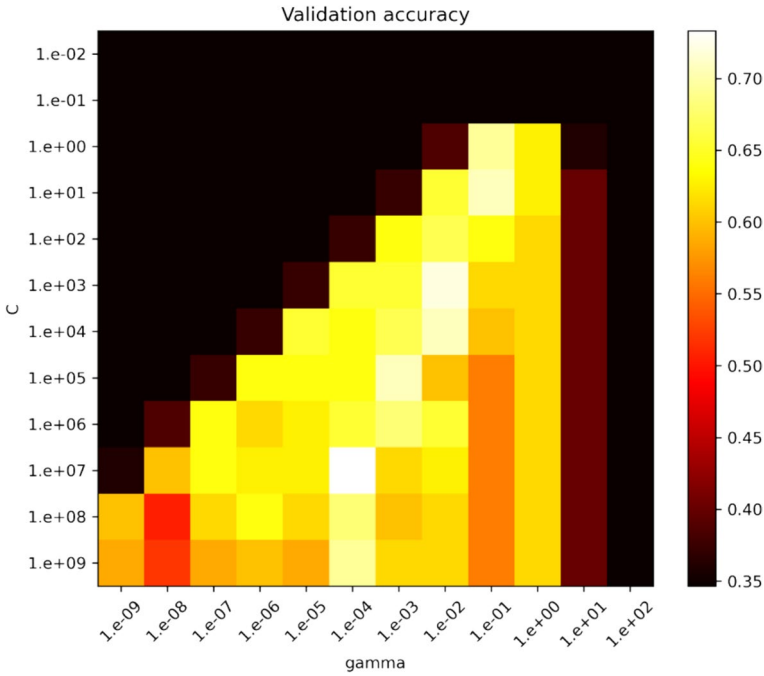


Fig. 8 Hyperparameter tuning for SVM_I

4 Model building

4.1 Model building with dataset I

Models being trained and tested on dataset I have been termed LR_I and SVM_I . The modelling process is conducted in the python scikit-learn library (Pedregosa et al. 2012). In the initial phase, after importing the dataset, it is randomly partitioned into the training and testing sample in the ratio of 70:30, meaning that 30% of the observations separated from the dataset did not participate during the training phase. They are placed apart to validate the model’s performance. Further, the training and testing dataset is scaled using standardisation to avoid the negative influence caused by the differing magnitude of the features. Standardisation is adopted for feature scaling, which scales down the range of every feature with zero mean along with a standard deviation of 1 using the given formula.

$$x' = \frac{A - \mu}{\sigma} \tag{16}$$

where A is the original value of the variable, x' indicates the standardised feature, μ and σ denote the mean and standard deviation of features, respectively. During the modelling process, concerning the size of our training set, an underfitting and overfitting issue could cause a problem; therefore, to address this matter, fivefold cross-validation is adopted while training the model. The schematic diagram of fivefold cross-validation is shown in Fig. 6. The idea behind cross-validation is that the training set is randomly broken down into five subsets with approximately equal dimensions. Then, training is conducted in five rounds so

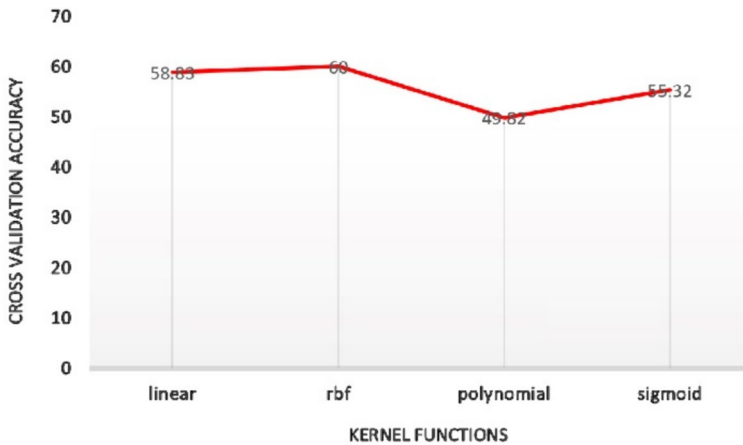


Fig. 9 Best performing SVM kernel on dataset II

that four subsets are combined in each round, forming a training set, whereas the left one is tested as a validation set. The obtained results from five rounds are then averaged to produce an overall accuracy. This strategy often helps to generalise the model's ability while treating new data to yield an unbiased result.

4.2 Modelling process for LR_i

In general, LR is a binary classifier that performs binary classification; hence it is extended to deal for multiclass classification problems (Greene 2003). While training the model, two hyperparameters, C and $penalty$, are optimised applying the Gridsearch method embedding fivefold cross-validation. C is an inverse of regularisation strength that adds penalty to increase the magnitude of the parameter values for reducing the overfitting. Likewise, $penalty$ is a form of regularisation and there are basically three types of penalty parameter $l1$, $l2$ and elasticnet. $l1$ penalises the sum of absolute values of weights and $l2$ penalises the sum of squares of the weights. In elasticnet, both $l1$, $l2$ terms are added. In this study, the C parameter value is chosen between $1e+8$ to $1e-3$ and four attributes for the penalty parameter are considered, including $l1$, $l2$, elasticnet and none. Nevertheless, the model achieved the best accuracy score of 66.67% on a test set when the optimal value for $C=10$ and $l2$ as penalty parameter is obtained.

4.3 Modelling process for SVM_i

Similar to the multiclass LR_i model, for modelling SVM_i , a multiclass approach of one vs one (OVO) (Wu et al. 2004) is adopted by creating a base binary classifier represented as C_{pq} (p and q denotes the p th and q th categories present in the training samples) for every pairwise category available in training records. Altogether, $m(m-1)/2$ base classifiers are created when there are m categories in the training records. For training records T , if the base classifier placed T in category p then vote is given to category p ,

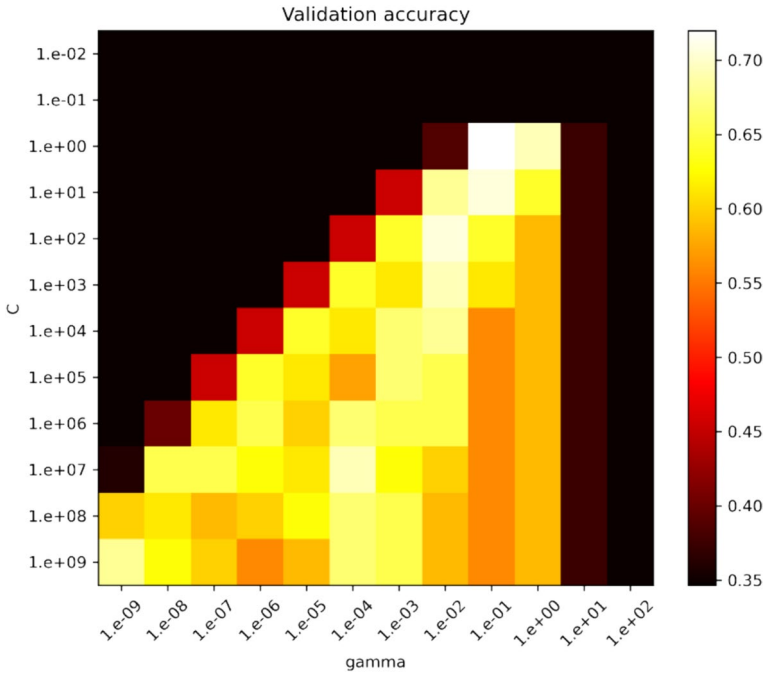


Fig. 10 Hyperparameter tuning for SVM_{II}

otherwise it is voted to category q . Training sample T was categorised into the category with the most votes up until all of the based classifiers had been voted on.

As is already indicated in Sect. 3.3, SVM can handle both linear and nonlinear data, therefore there are basically four types of kernel functions: linear, radial basis function(rbf), polynomial and sigmoid. These kernel functions work differently depending upon the nature of the data. In order to estimate the relevant kernel function that performs best with given data, kernel function is first identified using gridsearch method with five-fold cross validation with default parameter settings. The obtained result is shown in Fig. 7. Out of four kernels, rbf has the highest cross-validation accuracy of 57.14, as we can also see that the accuracy between linear kernel and rbf kernel has only a slight difference and this represents the data is slightly non-linear. Hence the model is further evaluated based on rbf kernel optimising hyperparameters for SVM.

As rbf has the best performance among other kernels, while building the classification model, two hyperparameters C and $gamma$ for rbf are further optimised through the grid-search approach embedding fivefold cross validation. C is regarded as a penalty parameter that estimates the number of errors classifiers make during classification. Similarly, the $gamma$ parameter estimates the data distribution which is mapped into the new dimension. While tuning the hyperparameters, the search range for C parameter is happened between $1e+09$ to $1e-2$ and $1e-09$ to $1e+02$ for the $gamma$ parameter in grid form, respectively. The pair that provides the best accuracy is selected as an optimal hyperparameter values. The heatmap in Fig. 8 shows the accuracy for different pairs of C and $gamma$ and the different colour indicates different accuracies. When the optimal C and $gamma$ parameter is obtained as $1e+07$ and $1e-04$, the model achieved the prediction accuracy of 60.61 for the test sample.

4.4 Model building with dataset II

The models that are being trained and tested on Dataset II are termed LR_{II} and SVM_{II} respectively. As in Sect. 4.1, in order to train the model, the dataset is still partitioned into the same ratio of 70:30. The same sample numbers have been used for training like in Dataset I, and the same rule also applies for testing samples. Feature scaling has not been done for this dataset because, after the box-cox transformation, all features are in a similar range due to auto standardization. However, a similar approach of fivefold cross validation is adopted like in Sect. 4.1 to overcome the problem of overfitting and underfitting.

4.5 Modelling process for LR_{II}

While creating a model on dataset II, the same consistency is maintained. The multi-class classification approach is adopted and, whilst optimising hyperparameters, C and $penalty$ parameters are optimised via embedding fivefold cross validation. The hyperparameter value for the C parameter is still considered between $1e+8$ to $1e-3$ and the four attributes $l1$, $l2$, elasticnet and none are selected as a $penalty$ parameter. After completing optimisation, LR_{II} obtained an accuracy of 72.73 for predicting the test set when C and the $penalty$ parameter equals 10 and $l2$ respectively. In comparison to Dataset I, the logistic regression model has better performance on Dataset II, this simply exaggerate parametric model is more reliable at prediction if data is relatively normal in distribution.

4.6 Modelling process for SVM_{II}

During the modelling process, SVM_{II} follows the same standard of SVM_I . Depending on the characteristics of different data, each kernel functions may have different results. Thus, the most suitable kernel is first estimated by applying gridsearch with five-fold cross-validation using default parameter settings. The result for the most appropriate kernel is shown in Fig. 9. The kernel rbf still has the best accuracy result for Dataset II, similar to Dataset I. Among the four kernels, rbf has the highest score of 60 whereas polynomial has the lowest score of 49.82.

Further, SVM_{II} is built with the rbf kernel by optimising the C and $gamma$ parameter with gridsearch and applying fivefold cross validation. The parameter tuning range is adjusted the same as with SVM_I . The hyperparameter tuning process is shown in Fig. 10; when the optimal value of C and $gamma$ reached 1 and 1, the model has the best prediction accuracy of 63.64 for the test sample.

5 Performance measurement and evaluation

In Sect. 4, prediction models were built based on two types of datasets after hyperparameter tuning for each model. Even though classification accuracy is regarded as a straight forward measure to evaluate performance, nevertheless, it cannot be reliable when data imbalance is a problem. As mentioned, in Sect. 2, the classes distribution proportion is not equal

and, in such case, precision, recall and F1 measure metrics aid in ensuring how robust the classifier is at classifying each class correctly. In rockburst prediction, intense and moderate levels are considered as “high risk” rockburst whereas none and slight are regarded as “low risk” rockburst. Generally, these two kinds of risks should be paid more attention because classifying high risk as a low risk spread false alarms that might cause unexpected casualties and could bring serious consequences to the project. Similarly, falsely predicting low risk rockburst as high risk rockburst leads to large amounts of money being spent on controlling rockburst even if it is not happening. For these reasons, these two problems should be addressed properly. A confusion matrix for high-risk and low risk rockburst is presented in Table 2.

In Table 2, True Positive (TP) indicates the positively predicted sample which is actually positive, False Negative (FN) represents a negatively predicted observation which in reality is positive, False Positive (FP) denotes a predicted positive that is actually negative and True Negative (TN) is a predicted negative which is actually negative. For any model, precision, recall and F1 score is calculated based on this matrix.

Precision(P) is the total number of positively predicted samples upon the sum of positively predicted samples and the number of false positives (Goutte and Gaussier 2005). Simply given by formula:

$$P = \frac{TP}{TP + FP} \tag{17}$$

Similarly, Recall (R) is the ratio of true positive upon the sum of true positive and false negatives (Goutte and Gaussier 2005) represented by the formula below.

$$R = \frac{TP}{TP + FN} \tag{18}$$

A good classifier should have higher precision and recall value but in real case there is always a trade-off between them. Thus, F1 score estimates the quality of model by calculating the harmonic mean between precision and recall (Goutte and Gaussier 2005) using given formula:

$$F1 = 2 \frac{P \times R}{P + R} \tag{19}$$

From Table 2, if only the classification of low risk is considered a major concern because falsely predicting low risk as high risk rockburst leads to unnecessary costs for controlling the risk, we would expect a model which has a low false positive value because the higher value of the false positive is disadvantageous when the model predicts too many cases of low risk as high risk. A situation in which a false positive is most dangerous is regarded as a type I error. To minimise type I errors, a high precision model is more effective for accurate classification. Similarly, if high-risk prediction is the primary concern, models that have less false negative value are advantageous because once the models predict high-risk as low-risk, user may think there are no consequences when in actuality

Table 2 Confusion matrix for high-risk and low-risk rockburst

		Predicted	
		High-risk	Low-risk
Actual	High-risk	True positive	False negative
	Low-risk	False Positive	True negative

situation there is a high risk rockburst might occur and be very harmful in the end. In this case, a false negative is more dangerous, which is a type II error. For type II error, models that have a higher recall value are more beneficial. The final result for the models trained on two different datasets can be found in the confusion matrix plot in Table 3.

From the table we can see that parametric models have achieved better accuracy results than non-parametric models in both types of datasets, however, LR_{II} has better results which simply illustrates that parametric models perform better when features are transformed into normal distribution. In the case of SVM_{II} , it is also less benefited by Dataset II but accuracy has not differed much in comparison to SVM_I . Viewing the confusion matrix for each model, the samples in the diagonal entries denotes the true predictions made by the model with zero misclassification rate. All the samples that do not fall inside the diagonal matrix are false predictions made by the models whose misclassification rates are not zero. The matrix is asymmetric for each classifier because the classification results for each intensity class varies. The precision for models on two different datasets is shown in Fig. 11.

As already discussed above, when a type I error is more dangerous, higher precision is given more priority. Therefore, for low-risk cases SVM_I (1.0) and SVM_{II} (0.90) have better precision for none cases followed by LR_I (0.90) and LR_{II} (0.84). LR_{II} (0.4) outperforms LR_I (0.27), SVM_I (0.26) and SVM_{II} (0.2) for slight cases. In contrast, we can also check the precision for high-risk cases, all four models also have a precision score of one for intense rockburst, which means all of them can accurately classify the intense cases correctly. For moderate cases, LR_{II} has the higher precision score of (0.81) among others. However, LR_I has slightly better precision in comparison to SVM_I and SVM_{II} .

When a type II error is more dangerous, high recall models are beneficial. Thus, from Fig. 12, for high-risk cases, LR_I (0.75) shows higher recall value for intense rockburst with SVM_{II} (0.75) and LR_{II} (0.75). For moderate rockburst, LR_I achieved a greater recall value followed by LR_{II} (0.5) and SVM_{II} (0.5). Nevertheless, SVM_I shows a least recall score of 0.41. On the contrary, we can also see the recall score for low risk rockburst. In terms of no rockburst LR_I (0.83) and LR_{II} (0.91) are better than SVM_I (0.66) and SVM_{II} (0.83). LR_{II} (0.8) and SVM_I (0.8) shows equal recall value for slight rockburst whereas SVM_{II} (0.40) has the lowest recall score.

Regardless of whether high or low risk rockburst is more important, for real scenarios in terms of controlling and safety purposes, both cases of high risk and low risk should be paid equal importance and predicted accurately. Thus, both precision and recall should have higher value for both high-risk and low-risk cases. Generally, for an optimal model, a higher value of precision and recall are desirable but trade-offs always exist between them which means as one increases another decrease. In rockburst prediction, high-risk and

Table 3 Confusion matrix of models trained on dataset I and dataset II for the testing set

		Predicted															
		LR_I			SVM_I			LR_{II}			SVM_{II}						
True	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3	
	0	10	2	0	0	8	4	0	0	11	1	0	0	9	3	0	0
	1	1	3	1	0	0	4	1	0	1	4	0	0	0	5	0	0
	2	0	6	6	0	0	7	5	0	0	6	6	0	0	7	5	0
	3	0	0	1	3	0	0	1	3	0	0	1	3	0	0	2	2
Accuracy		66.67			60.61			72.73			63.64						

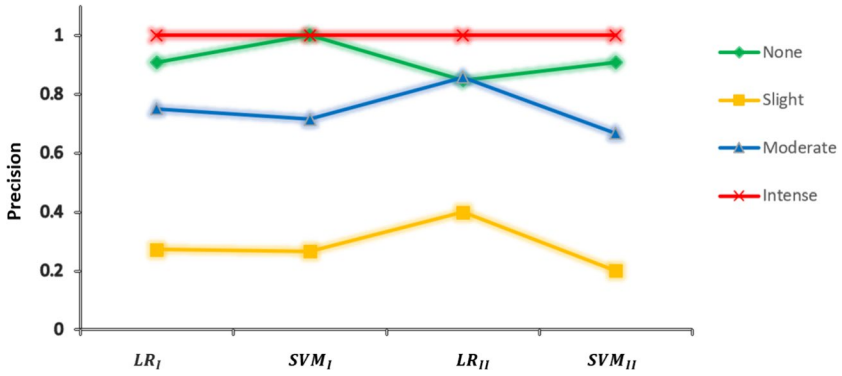


Fig. 11 Precision for models on two different datasets

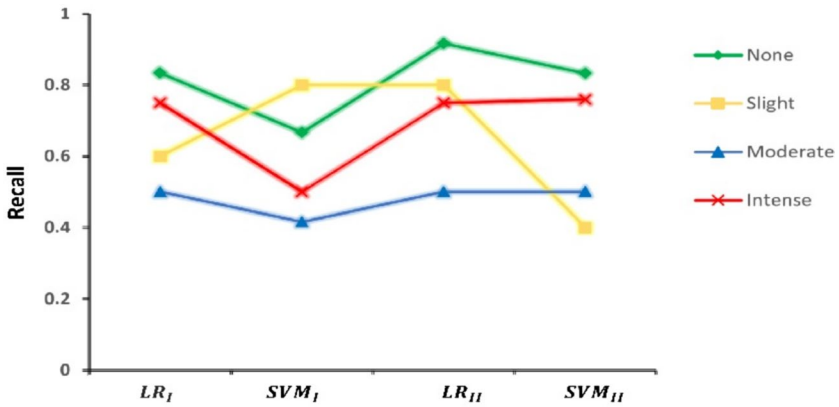


Fig. 12 Recall for models on two different datasets

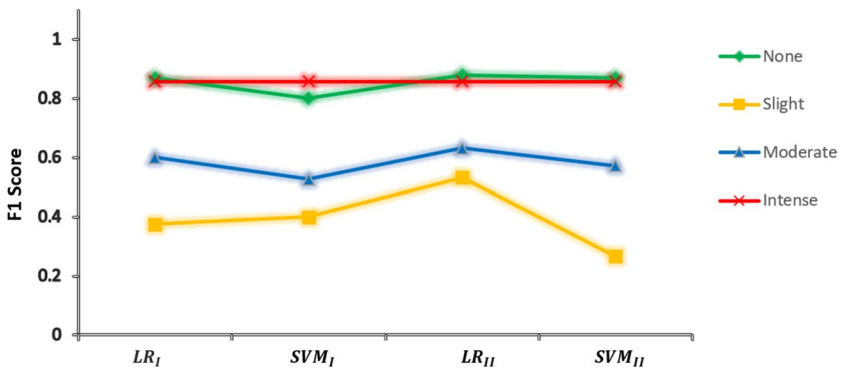


Fig. 13 F1 score for models on two different datasets

low-risk cases should be treated equally in same importance. As a consequence, we cannot simply rely on single precision and recall score to measure the robustness of each classifier.

Hence, when reliability of the model is not interpretable using single precision and recall score, the F1 score metric that evaluates the classifier's performance can be further computed by calculating the harmonic mean of precision and recall. The F1 score for each intensity class is given in Fig. 13. Based on the F1 score metric, best predicting models can be ranked as $LR_{II} > LR_I > SVM_I > SVM_{II}$ which obtained the average F1 score of 0.7254, 0.6754, 0.6458, 0.6412 respectively. Overall, LR outperforms SVM which has a better F1 score for each intensity level in both type of datasets.

For further evaluating the performance of parametric and non-parametric models, a ROC (receiving operating curve) is used for measurement. Initially, ROC is used for binary classification as a performance metric (Spackman 1989). Here, we have extended it for a multiclass classification approach by drawing ROC for each intensity class. Later averaging ROC for both models on two different datasets has been drawn. Multiclass ROC features Y-axis with true positive rate and X-axis with false positive rate. In general, the best classifier is the one which has larger area under the curve (AUC). In this experiment, the AUC for averaging-ROC estimates the robustness of classifiers.

In general, a different value of AUC determines the different characteristic function for the classifier's performance. Most commonly suggested values can be specified as 0.5, 0.7–0.8, 0.8–0.9 and above 0.9. 0.5 denotes no discrimination at all, whereas, 0.7–0.8 and 0.8–0.9 suggests acceptable discrimination and excellent discrimination, respectively. Similarly, the most outstanding model should have an AUC greater than 0.9 (Hosmer Jr et al. 2013). From Fig. 14a and b if we see two models LR_I and SVM_I trained on dataset I, LR_I has the AUC score of 0.9 compared to SVM_I which has the AUC score of only 0.86. This simply illustrates that when the availability of data is comparatively small, a parametric model can still reliably predict short-term rockburst. In the same way, from Fig. 14c and d, if we observe LR_{II} , SVM_{II} trained on Dataset II, LR_{II} has the highest AUC score of 0.91 even when compared to three other models. Further, we can notice that non-parametric SVM_I and SVM_{II} trained on two different datasets have similar AUC scores, however in terms of average F1 score, SVM_I is slightly ahead of SVM_{II} which means a class imbalance problem non-parametric model trained on original dataset can achieve better result.

6 Results and discussion

Parametric and non-parametric ML models are two distinct types of methods with unique theory and modelling processes. However, in short term rockburst risk evaluation, application of the parametric model approach is infrequent as these kinds of models do not rely on data size and normally distributed features are beneficial for them compared to non-parametric models. Therefore, parametric (LR) and non-parametric (SVM) is employed and studied by preparing two different datasets: the original data (Dataset I) which is not much normally distributed and another (Dataset II) that is prepared by applying box-cox transformation which is approximately normally distributed. At first, the model is built on Dataset I by partitioning it into the ratio of 70:30, during the modelling process in Dataset I standardisation technique is preferred for feature scaling to bring all the features in the similar range for both LR and SVM. Once the hyperparameter optimisation is done LR achieved an accuracy of 66.67 whereas SVM obtained only 60.61 for the testing set in Dataset I. Similarly, the models have been trained on Dataset II maintaining the same consistency and configuration as in Dataset I, with the only difference being feature scaling that is not performed for Dataset II because, after box cox transformation, all feature values

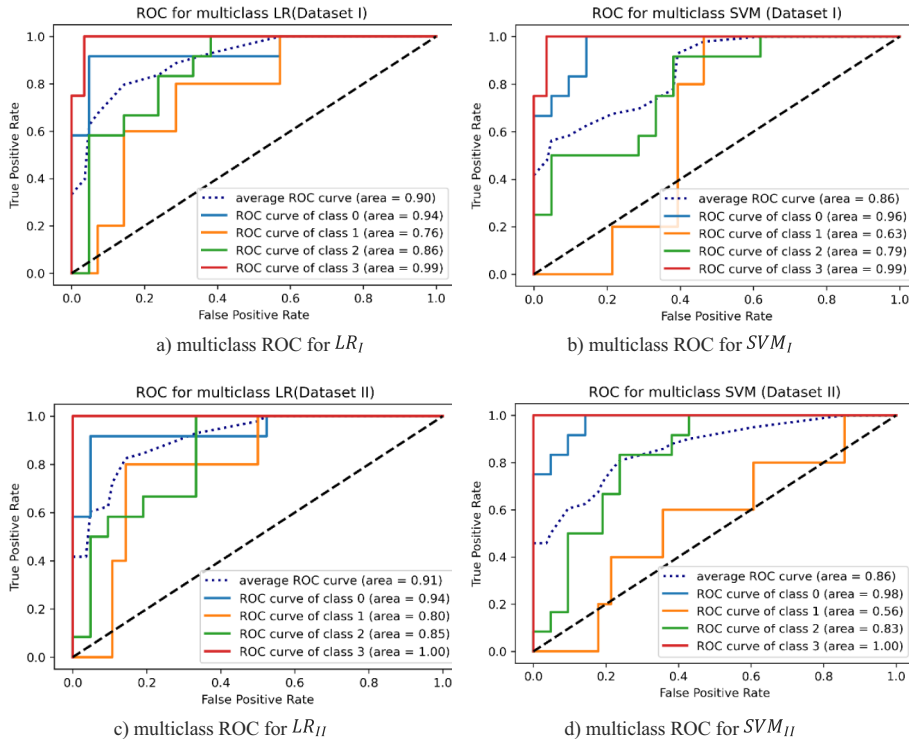


Fig. 14 Multiclass ROC for $LR_I, LR_{II}, SVM_I, SVM_{II}$. **a** multiclass ROC for LR_I . **b** multiclass ROC for SVM_I . **c** multiclass ROC for LR_{II} . **d** multiclass ROC for SVM_{II}

are set into auto scaling. The LR model still obtained the best accuracy of 72.73 whereas SVM achieved an accuracy of 63.64 for testing data. From the accuracy result, parametric LR scored higher accuracy in Dataset II.

Although accuracy is often regarded as a straightforward measure for model performance, it cannot interpret the whole performance. Therefore, unlikely in previous studies, three other evaluation metrics, precision, recall and F1 score, are further calculated and average AUC is computed for each model to identify their robustness. In the perspective of predicting high risk and low risk rockburst cases, the final F1 score says that LR trained and tested on Dataset II is more suitable than a model trained on Dataset I. Similarly, SVM trained on Dataset I is slightly ahead of that trained on Dataset II. As we can see, SVM_{II} has a comparatively higher accuracy than SVM_I but from the F1 score point of view, SVM_I is more appropriate to classify both high risk and low risk in imbalanced classes as it can lead to correct guidance. Moving forward, AUC is calculated for each classifier trained on both datasets to see their robustness. Overall, LR_{II} has the best AUC score of 0.91 followed by LR_I (0.90); likewise, SVM_I and SVM_{II} scored equal AUC in terms of performance which simply exaggerates that, for short-term rockburst risk prediction, parametric models can still effectively identify risks even in small datasets with fewer features. These models are even more capable of predicting accurate outcomes if non-normally distributed data are transformed into a normal distribution.

Short-term rockburst prediction using microseismic monitoring during the rockburst development process relies on different factors. Depending upon the complex geological conditions during the data acquisition process, acquiring sufficient records and features for training the model is always challenging. The parametric model such as LR is often more powerful in defining the true relationship between input and output variables using a known form of functions that can also predict the outcome without being independent of the size of the data. This is because once the functional form is determined, the parametric model estimates the coefficients from training data to give a better predictive model. The amount of data will not impact the training processes. However, if the independent variables are more or less normally distributed, they can find the best fit for accurate prediction. The work of Feng et al. (Feng 2017) states that the degree of microseismicity is positively correlated with the frequency and intensity of rockburst, which means MS multi-parameters and rockburst risk levels also have higher level of correlation. Parametric models are often more effective when dealing with inputs and outputs that have a strong relationship because assuming a specific functional form, such as a line or hyperplane, simplifies the training process. Therefore, a parametric model like logistic regression gave better results regardless of the data size compared to non-parametric models. The primary reason for the poor performance of non-parametric SVM on both datasets is that even though non-parametric models do not depend on the assumption of mapping function and normality of the variable distribution, in order to perfectly learn from the training samples as well as to generalise on unseen observation, it requires a large amount of supportive data and inputs to learn the pattern from it so that it may define the required functional form to best fit the training instances. However, in short-term rockburst prediction, it is always challenging to acquire the sufficient amount of data required to train the non-parametric models, due to this reason models trained with insufficient records often loses ability to generalise on unseen samples and cannot give the higher prediction result. On the other hand, if we see the accuracy of non-parametric SVM on two datasets, the accuracy has only a little difference. However, based on the other performance metrics, it has almost similar performance in every aspect. This is simply because non-parametric SVM depends on the number of training records for estimating the parameters necessary for making a prediction independent of the distribution of input variables. Hence, unlike the parametric model, only a more enormous amount of data increases the efficiency of the non-parametric model.

Although the overall results are promising, some limitations can be handled in future research in a similar area:

1. The available dataset contains four intensity classes. Among them, the proportion of intense rockburst cases is comparatively less. When an ML model is trained on a dataset where the ratio of some classes is not equal, a model could produce a biased result in such cases. Therefore, a dataset can be further updated in future work by giving equal preferences to each class to yield more accurate outcome; this can be done by adding more cases of minority classes into the dataset. It will further help the model to generalise on observation of each intensity class for real time warning.
2. This paper mainly focuses on developing intelligent predicting models for immediate types of rockburst, such as strain burst and strain-structure slip burst that often occur immediately after excavating in deep engineering projects. There is also a different type of rock bursting known as fault-slip rockburst, which has a different mechanism, but due to the self-similarity of the rockburst development process, monitored MS information can be utilised for the early warning of risk levels (Feng et al. 2017). Future research on this topic using an intelligent model is worthy of exploring.

7 Conclusion

The prediction of rockburst hazard relying on a traditional approach is always challenging. As a result, an intelligent approach that aids in the accurate prediction of risk is necessary. Therefore, this paper proposes and compares two widely adopted parametric and nonparametric methods typified by LR and SVM based on two different datasets. Initially, 108 rockburst data obtained from MS monitoring information that contained four attributes such as cumulative number of events, cumulative radiated energy, cumulative apparent volume and rockburst development day, was collected. Since a parametric model performs better when feature variables are normally distributed, two datasets were prepared to better understand the feasibility of both approaches. Dataset I is created using original data which is not normally distributed whereas normally distributed Dataset II is prepared by applying box-cox transformation to Dataset I. After, the models were built on both datasets by using the same number of samples for training and testing in the splitting ratio of 70:30 for both conditions. While training the model, hyperparameter optimisation is done using the gridsearch method applying five-fold cross validation to find the C and *penalty* parameter for LR as well as the C and *gamma* parameter for the best performing rbf kernel of SVM using the multiclass classification approach. Once the optimal parameter is obtained for each classifier, the result shows that the logistic regression model trained on Dataset II obtained the highest accuracy of 72.73% on the testing sample when compared to other models. For further evaluation, three other performance metrics, precision, recall and F1 score were computed from which we can conclude that LR performed well on both datasets whereas SVM has almost equal performance in both type of datasets. In order to investigate the robustness of classifiers, ROC-AUC is drawn for each class and average AUC has been computed for each classifier. The final output indicates that when the LR model is trained on Dataset II, it achieved the highest AUC score of 0.91. When trained on Dataset I, it has an AUC score of 0.90. However, the SVM has an equal AUC of 0.86 while using both datasets. This simply says that, in terms of short-term rockburst risk prediction, the limited availability data is always a threat. Even though a non-parametric model is good in achieving better output, it still needs more data and inputs to generalise well and yield higher accuracy. Nevertheless, a parametric model performs well because the data size does not influence the model generalisation as it constrains an algorithm to a specified functional form that estimates the coefficients from training data to give a better predictive model. Regardless of data size highly accurate model can still be achieved if the data are relatively normally distributed. Hence, in future research parametric models are worth exploring when the datasets are comparatively small and the number of variables are limited.

Appendix

See Table4

Table 4 Original dataset

CN	CE	CV	RD	Intensity
1	0.78	3.441	3	None
3	4.448	4.261	9	None
3	3.668	3.609	6	None
7	4.3	3.018	9	None
1	2.97	4.164	3	None
5	3.996	3.279	5	None
1	1.65	2.787	6	None
1	0.9	2.759	1	None
4	4.737	4.173	5	None
1	1.67	4.033	3	None
1	1.72	3.857	6	None
2	1.39	2.908	2	None
5	2.435	3.878	11	None
4	1.316	3.114	9	None
1	0.78	3.441	4	None
3	4.211	3.794	4	None
2	1.94	3.25	2	None
11	3.973	3.769	13	None
1	4.78	2.985	1	None
5	4.04	3.555	2	None
5	3.154	3.309	2	None
3	3.616	4.603	7	None
3	4.376	4.079	2	None
1	1.54	4.31	1	None
2	2.61	2.925	1	None
2	5.16	2.936	18	None
3	3.493	4.857	11	None
6	4.594	4.743	12	None
8	2.197	2.511	7	None
4	5.82	3.728	13	None
17	4.619	4.844	14	None
10	4.008	3.221	5	None
6	5.3	2.735	4	None
15	3.486	5.03	7	None
13	5.348	4.78	14	None
3	1.25	4.944	1	None
7	4.834	4.116	13	Slight
10	4.614	4.611	9	Slight
4	4.53	4.557	6	Slight
3	4.61	3.732	3	Slight
10	4.446	4.37	6	Slight
4	4.595	3.708	4	Slight
7	4.381	4.132	4	Slight
13	4.408	4.428	6	Slight
3	4.443	4.291	5	Slight
11	4.029	4.944	9	Slight

Table 4 (continued)

CN	CE	CV	RD	Intensity
22	4.736	4.133	23	Slight
8	4.132	3.504	13	Slight
29	3.882	4.156	10	Slight
20	4.76	3.843	16	Slight
12	3.543	4.223	3	Slight
6	5.561	4.043	6	Slight
6	4.368	3.497	8	Slight
7	5.269	4.817	10	Slight
7	5.4	3.919	4	Slight
8	5.204	3.977	3	Slight
16	3.621	4.681	6	Slight
17	3.172	5.015	10	Slight
2	4.061	3.576	3	Slight
9	1.723	4.993	7	Slight
25	4.381	4.848	10	Slight
25	3.367	4.964	8	Slight
5	5.17	4.594	6	Slight
14	5.841	4.622	9	Moderate
17	4.754	4.397	9	Moderate
18	5.295	4.703	10	Moderate
10	5.322	4.238	7	Moderate
14	4.818	4.266	11	Moderate
17	4.944	4.598	11	Moderate
18	5.602	4.779	10	Moderate
19	5.865	4.263	10	Moderate
20	5.589	4.589	11	Moderate
11	5.926	4.141	9	Moderate
8	5.621	4.62	4	Moderate
12	4.912	4.565	8	Moderate
20	5.982	4.453	8	Moderate
6	5.008	4.627	5	Moderate
11	4.966	4.154	4	Moderate
24	4.748	4.66	6	Moderate
6	5.593	4.809	8	Moderate
11	5.724	4.251	13	Moderate
8	5.219	4.552	16	Moderate
21	3.543	4.732	18	Moderate
3	5.06	4.438	7	Moderate
19	3.68	4.832	10	Moderate
23	4.408	4.873	9	Moderate
12	5.098	3.516	7	Moderate
36	4.729	4.336	14	Moderate
21	5.848	4.834	11	Moderate
10	4.886	4.105	6	Moderate
18	3.825	4.703	6	Moderate

Table 4 (continued)

CN	CE	CV	RD	Intensity
41	5.968	4.694	11	Intense
49	6.419	4.995	4	Intense
58	7.094	4.975	16	Intense
49	6.373	5.168	8	Intense
22	5.859	4.895	10	Intense
10	6.576	5.081	8	Intense
15	6.587	5.152	9	Intense
70	6.147	5.152	10	Intense
44	5.459	4.865	15	Intense
11	4.11	3.624	5	Intense
42	6.284	5.05	7	Intense
45	4.803	4.838	11	Intense

Funding This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declarations

Conflict of interest There are no conflicts of interest to declare.

References

- Basnet PMS, Jin AB, Mahtab S (2024) Developing an explainable rockburst risk prediction method using monitored microseismicity based on interpretable machine learning approach. *Acta Geophys*. <https://doi.org/10.1007/s11600-024-01338-y>
- Box GE, Cox DR (1964) An analysis of transformations. *J Roy Stat Soc B* 26(2):211–243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Bruning TD (2018) A combined experimental and theoretical investigation of the damage process in hard rock with application to rockburst
- 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. <https://doi.org/10.1007/s00603-013-0524-2>
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1007/BF00994018>
- Cox DR (1958) The regression analysis of binary sequences. *J Roy Stat Soc Ser B (methodol)* 20(2):215–242. <https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>
- Feng XT (2017) *Rockburst: Mechanisms, monitoring, warning, and mitigation*. Butterworth-Heinemann, Oxford
- Feng XT, Chen BR, Zhang CQ, Li SJ, Wu SY (2013) Mechanism, warning and dynamic control of rockburst development process. Science Press Beijing, Beijing
- Feng GL, Feng XT, Chen BR, Xiao YX, Yu Y (2014) A microseismic method for dynamic warning of rockburst development processes in tunnels. *Rock Mech Rock Eng* 48(5):2061–2076. <https://doi.org/10.1007/s00603-014-0689-3>
- Feng XT, Yu Y, Feng GL, Xiao YX, Chen BR, Jiang Q (2016) Fractal behaviour of the microseismic energy associated with immediate rockbursts in deep, hard rock tunnels. *Tunn Undergr Space Technol* 51:98–107. <https://doi.org/10.1016/j.tust.2015.10.002>
- Feng XT, Liu JP, Chen BR, Xiao YX, Feng GL, Zhang FP (2017) Monitoring, warning, and control of rockburst in deep metal mines. *Engineering* 3(4):538–545. <https://doi.org/10.1016/J.ENG.2017.04.013>

- 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* 11:3212. <https://doi.org/10.3390/su11113212>
- Feng GL, Lin MQ, Yu Y, Fu Y (2020) A microseismicity-based method of rockburst intensity warning in deep tunnels in the initial period of microseismic monitoring. *Energies* 13(11):2698. <https://doi.org/10.3390/en13112698>
- Field A (2013) *Discovering statistics using ibm spss statistics*. Sage, New York
- GlOwacka E (1993) Probabilistic seismic hazard evaluation in underground mines. *Nat Hazards* 7(2):121–132. <https://doi.org/10.1007/BF00680426>
- Goutte C, Gaussier E. 2005. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. Paper presented at: European conference on information retrieval, Springer, New York
- Greene WH (2003) *Econometric analysis*. Pearson Education India, Hoboken
- Hoek E. 2000. *Rock engineering: course notes*
- Hosmer DW Jr, Lemeshow S, Sturdivant RX (2013) *Applied logistic regression*. John Wiley & Sons, Hoboken
- James G, Witten D, Hastie T, Tibshirani R (2013) *Statistical learning*. In: James G, Witten D, Hastie T, Tibshirani R (eds) *An introduction to statistical learning: with applications in r*. Springer, New York, pp 15–57
- Jin AB, Basnet PMS, Mahtab S (2022) Microseismicity-based short-term rockburst prediction using non-linear support vector machine. *Acta Geophys* 70(4):1717–1736. <https://doi.org/10.1007/s11600-022-00817-4>
- Liang WZ, Zhao GY (2022) A review of long-term and short-term rockburst risk evaluations in deep hard rock. *Chin J Rock Mech Eng* 41(1):19–39
- Liang WZ, Sari A, Zhao GY, McKinnon SD, Wu H (2020) Short-term rockburst risk prediction using ensemble learning methods. *Nat Hazards* 104(2):1923–1946. <https://doi.org/10.1007/s11069-020-04255-7>
- Liang WZ, Sari YA, Zhao GY, McKinnon S, Wu H (2021) Probability estimates of short-term rockburst risk with ensemble classifiers. *Rock Mech Rock Eng* 54(4):1799–1814. <https://doi.org/10.1007/s00603-021-02369-3>
- Ling K, Wang Y LDQ, Guo YP, Zhou Z, Zhang LL, He MC (2023) Experimental study on rockburst and spalling failure in circular openings for deep underground engineering. *Rock Mech Rock Eng* 56(4):2607–2631. <https://doi.org/10.1007/s00603-022-03203-0>
- Liu JP, Feng XT, Li YH, Xu SD, Sheng Y (2013) Studies on temporal and spatial variation of microseismic activities in a deep metal mine. *Int J Rock Mech Min Sci* 60:171–179. <https://doi.org/10.1016/j.ijrmm.2012.12.022>
- Liu GF, Jiang Q, Feng G, Chen DF, Chen BR, Zhao ZN (2021) Microseismicity-based method for the dynamic estimation of the potential rockburst scale during tunnel excavation. *Bull Eng Geol Env* 80:1–24
- Luo S, Gong FQ (2023) Evaluation of energy storage and release potentials of highly stressed rock pillar from rockburst control perspectives. *Int J Rock Mech Min Sci* 163:105324. <https://doi.org/10.1016/j.ijrmm.2022.105324>
- Ma TH, Tang CA, Tang LX, Zhang WD, Wang L (2015) Rockburst characteristics and microseismic monitoring of deep-buried tunnels for jinpings ii hydropower station. *Tunn Undergr Sp Technol* 49:345–368. <https://doi.org/10.1016/j.tust.2015.04.016>
- Man Singh Basnet P, Mahtab S, Jin A (2023) A comprehensive review of intelligent machine learning based predicting methods in long-term and short-term rock burst prediction. *Tunn Undergr Sp Technol* 142:105434
- Nussbaumer MM (2000) *A comprehensive review on rock burst*. Mass Inst Technol
- Ortlepp W (2005) *Rasim comes of age—a review of the contribution to the understanding and control of mine rockbursts*. In: Paper presented at: Proceedings of the 6th International symposium on rockbursts and seismicity in mines
- Park HM 2008 *Univariate analysis and normality test using sas, stata, and spss*
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V et al (2012) Scikit-learn: machine learning in python. *J Mach Learn Res* 12:2825–2830
- Pu YY, Apel DB, Liu V, Mitri H (2019) Machine learning methods for rockburst prediction-state-of-the-art review. *Int J Min Sci Technol* 29(4):565–570. <https://doi.org/10.1016/j.ijmst.2019.06.009>
- Russell SJ (2010) *Artificial intelligence : a modern approach*. Third edition. Upper Saddle River NJ: Prentice Hall, [2010] ©2010.
- Song Y, Wang EY, Yang HZ, Liu CF, Li BL, Chen D (2024) Microseismic source location using deep learning: a coal mine case study in China. *J Rock Mech Geotech Eng*. <https://doi.org/10.1016/j.jrmge.2024.02.014>

- Spackman KA (1989) Signal detection theory: valuable tools for evaluating inductive learning. In: Paper presented at: ML
- Tang S, Li J, Tang L, Zhang L (2023) Microseismic monitoring and experimental study on rockburst in water-rich area of tunnel. *Tunn Undergr Sp Technol* 141:105366. <https://doi.org/10.1016/j.tust.2023.105366>
- Thode H (2002) Testing for normality. CRC Press, Boca Raton
- Wang CL, Cao C, Liu YB, Li CF, Li GY, Lu H (2021) Experimental investigation on synergetic prediction of rockburst using the dominant-frequency entropy of acoustic emission. *Nat Hazards* 108(3):3253–3270. <https://doi.org/10.1007/s11069-021-04822-6>
- Wu TF, Lin CJ, Weng R (2003) Probability estimates for multi-class classification by pairwise coupling. *J Mach Learn Res* 5:975–1005
- Yin X, Liu QS, Pan YC, Huang X (2021) A novel tree-based algorithm for real-time prediction of rockburst risk using field microseismic monitoring. *Environ Earth Sci*. <https://doi.org/10.1007/s12665-021-09802-4>
- Zhang W, Feng XT, Yao ZB, Hu L, Xiao YX, Feng GL, Zhang Y (2022) Development and occurrence mechanisms of fault-slip rockburst in a deep tunnel excavated by drilling and blasting: a case study. *Rock Mech Rock Eng* 55(9):5599–5618. <https://doi.org/10.1007/s00603-022-02927-3>
- Zhang QH, Li WG, Yuan L, Zheng TL, Liang ZW, Wang XR (2024) A review of tunnel rockburst prediction methods based on static and dynamic indicators. *Nat Hazards*. <https://doi.org/10.1007/s11069-024-06657-3>
- Zhang Y, Feng XT, Yao Z, Zhang W, Hu L, Yang CX, Fu LJ (2024) Study on warning method for fault rockburst in deep TBM tunnels. *Rock Mech Rock Eng*. <https://doi.org/10.1007/s00603-024-03830-9>
- Zhao HB, Chen BR, Zhu CX (2021) Decision tree model for rockburst prediction based on microseismic monitoring. *Adv Civ Eng* 2021:8818052

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.