Deep Mining 2024 – P Andrieux & D Cumming-Potvin (eds)

doi:10.36487/ACG_repo/2465_52

Harnessing machine learning for seismic event discrimination in deep underground mining: a case study from Western Australia

Roohollah Shirani Faradonbeh [a,*], Jamshid Shakeri [b], Zaniar Ghaderi [c], Peter A Mikula [d], Hyongdoo Jang [a], Abbas Taheri [e]

[a] *Western Australian School of Mines, Curtin University, Australia*

[b] *Department of Mining Engineering, Hamedan University of Technology, Iran*

[c] *Department of Industrial Engineering, Tarbiat Modares University, Iran*

[d] *Mikula Geotechnics Pty Ltd, Australia*

[e] *Department of Mining, Queen's University, Canada*

Abstract

This paper presents a comprehensive study on applying machine learning (ML) techniques to discriminate seismic events in deep underground mining from blast and noise records using data collected from the Vivien gold mine in Western Australia. The dataset/catalogue comprises parameters derived from signals recorded by an ESG microseismic monitoring system, encompassing various parameters such as magnitude, seismic moment, total radiated energy and more, totalling 33,298 records. A rigorous statistical analysis was conducted to address potential multicollinearity issues and identify key input variables. Additionally, the local outlier factor (LOF) method was utilised to remove anomalies, ensuring homogeneity in the dataset for further analysis. The synthetic minority oversampling technique (SMOTE) was then applied to address imbalanced datasets, particularly in classifying seismic record types as seismic events, blasts or noise attributed to rockfall. Eight robust ML algorithms were employed to develop classifiers for predicting record class types. The performance of each model was evaluated using statistical indices, ultimately identifying random forest (RF) as the most accurate method for distinguishing between different record types. Furthermore, a user-friendly graphical user interface (GUI) was also developed to facilitate data analysis based on the proposed RF model, enhancing the interpretation of microseismic monitoring results in practical applications. This study underscores the efficacy of ML approaches in seismic event discrimination to ensure the seismic dataset is clean and reliable for use in geotechnical assessments of seismic hazards and seismic characteristics at the mine.

Keywords: *seismic event, microseismic monitoring, deep underground mining, machine learning, random forest*

1 Introduction

Deep underground mining is usually accompanied by seismic events, especially in hard rock mediums, with the potential to result in casualties, infrastructure damage and significant operational costs (Potvin & Wesseloo 2013). Therefore tracking fracture initiation and propagation induced by stress concentration and redistribution with mining progress is critical to treating or controlling potential hazards like rockburst. This is usually conducted by microseismic monitoring technology and post-processing of signals recorded by geophones and/or accelerometers. However, the analysis and interpretation of seismic events waveforms could be complicated if other types of signals are also recorded, such as blast and noise signals (e.g. the signals induced by the impact of the mining vehicles on tunnel walls, electrical circuits, rockfalls, drilling and orepass operations), as they may produce amplitude and energy values similar to those recorded for the

* Corresponding author. Email address: roohollah.shiranifaradonbeh@curtin.edu.au

seismic events. Inappropriate event labelling can result in identifying fictitious high-stress zones and diverting resources into managing those perceived hazards. Several techniques can be utilised to assess the catalogue of records, including manual techniques (based on waveform spectral analysis), statistical analyses and machine learning (ML)-based approaches. However, as discussed by Faradonbeh et al. (2024), the former two techniques suffer from critical shortfalls such as being time-consuming in evaluating a large database generated by sensors using manual methods, implementing biased user-defined thresholds for discriminating the signals, not being robust enough to extract the highly complex relationship between variables, and being unable to provide reliable and explicit models for users. However, ML-based methods can significantly resolve the foregoing limitations in an intelligent manner, as reported by different researchers (Peng et al. 2019; Rao et al. 2021; Dong et al. 2020; Shirani Faradonbeh et al. 2023).

Most prior studies have only focused on using one or two ML techniques for classifying seismic records based on a limited number of data parameters. Furthermore, in these studies the output parameter was binary (i.e. seismic event and blast signal) and other types of signals mentioned earlier were not considered. Therefore there is no comprehensive study to assess the classification capability of different ML algorithms in dealing with microseismic monitoring data in depth. Through a comprehensive database compiled from the Vivien gold mine, this study aims to propose classifiers to distinguish three types of signals (i.e. seismic events, blasts and noise) using eight robust ML algorithms — artificial neural network (ANN), random forest (RF), decision tree (DT), k-nearest neighbour (KNN), support vector machine (SVM), AdaBoost (AB), Gaussian Naive Bayes (GNB) and logistic regression (LR) — and compare their efficiency. In the following sections, the procedure of data collection and analysis is discussed in more detail.

2 Microseismic monitoring data collection

The database of over 30,000 records was collected over a period of five years using a commercial seismic monitoring system from ESG Solutions (https://www.esgsolutions.com/) installed at Vivien gold mine in Western Australia. The system comprised three triaxial and seven uniaxial sensors, recording the seismic response in the narrow vein gold mine to a depth of about 550 m. Waveform processing was performed and events were classified through the following three stages:

1. automatically by the ESG system software, using settings assigned by ESG
2. manually reviewed and adjusted by ESG personnel
3. all waveforms larger than magnitude -1 were manually inspected and audited for quality control by the site geotechnical engineer, who had over 25 years of experience in waveform processing. There were too many smaller magnitude records to be able to audit them all and, based on the audit of the magnitude waveform, it is certain that some contamination existed in the seismic catalogue.

The catalogue records are characterised by variables including local magnitude, seismic moment, total radiated energy, apparent stress, potency, logarithm of potency, maximum displacement at source, the Es/Ep energy ratio (ratio of energy associated with the S-wave and P-wave), static stress drop, source radius, S-wave corner frequency, apparent volume, residual and number of used sensors. The first 12 variables describe the seismological attributes of the seismic source. The last two parameters describe the performance aspects of the monitoring system:

- ‘Location residual’ describes the extent of mismatch among the sensor waveform picks used to compute a location. A mismatch will always be present. Very low or very high residuals indicate probable errors in waveform interpretation.
- ‘Number of sensors used’ defines how many sensors received an adequate signal from the source to be acceptable for the source parameter calculations.

The dataset is considered highly accurate in terms of the classification of records into three classes: noise (n), seismic event (e) and blast (b). However, the database will contain contaminations, i.e. each waveform class will contain a small number of mislabelled records that should be in a different class. The noise class includes

every record that is not an event or blast. In this mine, most noise was due to the impact of blocks of rock falling from the backs or hanging walls of stopes and striking the lower footwall of the stopes. Minor noise sources arose from the placement of backfill, electrical spikes and some borehole drilling.

3 Preprocessing of database

Table 1 lists the descriptive statistics values of all input parameters for three different classes, however, as noted below, some parameters were identified as being highly correlated and were removed from further analysis. The dataset was examined to assess the relative frequencies of three distinct categories, as depicted in Figure 1. The analysis revealed that the highest proportion within the database is attributed to n instances, accounting for 40.5%, followed by e (31.8%) and b (27.7%) cases, respectively.

Table 1 Descriptive statistics of all variables

Variable (symbol)	Unit	Minimum	Maximum	Mean	Std deviation
Local magnitude (MLOC)	–	–3.067	1.459	–1.018	0.748
Seismic moment (SM)	Nm	25,096	154,170,655,000	658,495,325	2,997,429,109
Total radiated energy (TRE)	J	0.000†	1,981,186	9,598	49,932
Apparent stress (AS)	MPa	0.000	14.098	0.157	0.433
Potency (P)	m^3	0.000	5.139	0.022	0.100
Log potency (L)	m^3	–6.078	0.711	–3.004	1.122
Maximum displacement at source (MDS)	mm	0.043	7.931	0.682	0.703
Es/Ep energy ratio (ESEP)	–	0.023	257.669	11.575	10.655
Static stress drop (SSD)	Pa	0.051	236,453,984	3,136,309	9,093,107
Source radius (SR)	m	0.375	551.971	8.060	8.019
S-wave corner frequency (SWC)	Hz	2.368	3482.490	340.748	311.836
Apparent volume (AV)	m^3	0.452	917,617,444	83,260	5,070,330
Location residual (R)	m	0.000	823.636	13.191	14.536
Number of sensors used (SU)	–	4	10	7.673	1.506

Zero values indicate inclusion in the dataset of entries with very small parameter values

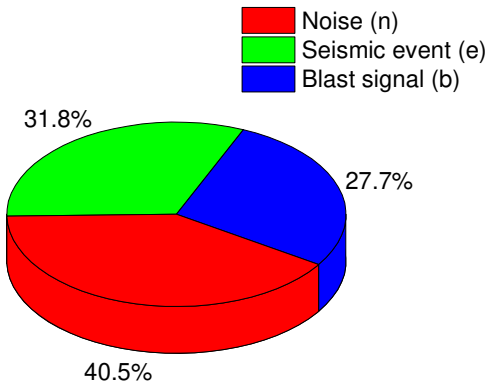


Figure 1 Percentage of microseismic data for three different classes

3.1 Correlation analysis

The presence of multicollinearity, resulting from high correlations among the independent variables, can have a detrimental impact on the reliability of statistical inferences and introduce heightened complexity to the proposed ML-based models (Baghbani et al. 2023; Onyelowe et al. 2021a). To assess this issue, the Pearson correlation coefficient, R (Equation 1), was computed to gauge the relationships between all combinations of input variables. The results were presented as a correlation matrix in Figure 2. Using product-moment analysis it was determined that correlation coefficients falling within the ranges of ±0.0 to ±0.20, ±0.21 to ±0.40, ±0.41 to ±0.60, ±0.61 to ±0.80 and ±0.81 to ±1.0 indicate no correlation, weak, moderate, strong and very strong correlations, respectively, between the pairs of database parameters (Gowida et al. 2021; Sorabi et al. 2024).

$$r = \frac{\sum_{i=1}^{n}(X_i - X_m)(Y_i - Y_m)}{\sqrt{\sum_{i=1}^{n}(X_i - X_m)^2}\sqrt{\sum_{i=1}^{n}(Y_i - Y_m)^2}} \quad (1)$$

where:

n = total number of samples

X_i = values of the variable X

Y_i = values of the variable Y

X_m = average value of the variable X

Y_m = average value of the variable Y.

According to Figure 2 it is evident that the majority of the input data exhibit a weak correlation. However, Figure 2 highlights that the correlation between MLOC-L and SM-P, with a correlation coefficient of 1, is notably robust. Additionally, the correlation coefficients for AS-SSD, MDS-L, and MDS-MLOC are reported as 0.89, 0.91 and 0.91, respectively, signifying exceptionally high correlation levels based on the established criteria. In statistical analysis a high correlation among independent data points is undesirable as it suggests a lack of true independence between the variables, potentially leading to multicollinearity issues in regression analysis. Such high correlation can result in inflated standard errors, complicating the determination of the genuine relationship between variables and affecting the reliability and interpretability of statistical models. Therefore ensuring the independence of variables is crucial to prevent inaccuracies or biases in data analysis. Accordingly, for the highly correlated pairs, the following parameters were removed from the analysis:

- MLOC-L – removed L since both parameters are functions of log (SM), but MLOC is less complex to calculate
- SM-P – removed P since SM is the more fundamental term
- AS-SSD – removed SSD, which is a more model-dependent parameter
- MDS-MLOC – removed MDS, since MLOC is kept in preference over L
- MDS-L – both parameters removed in managing the other correlations.

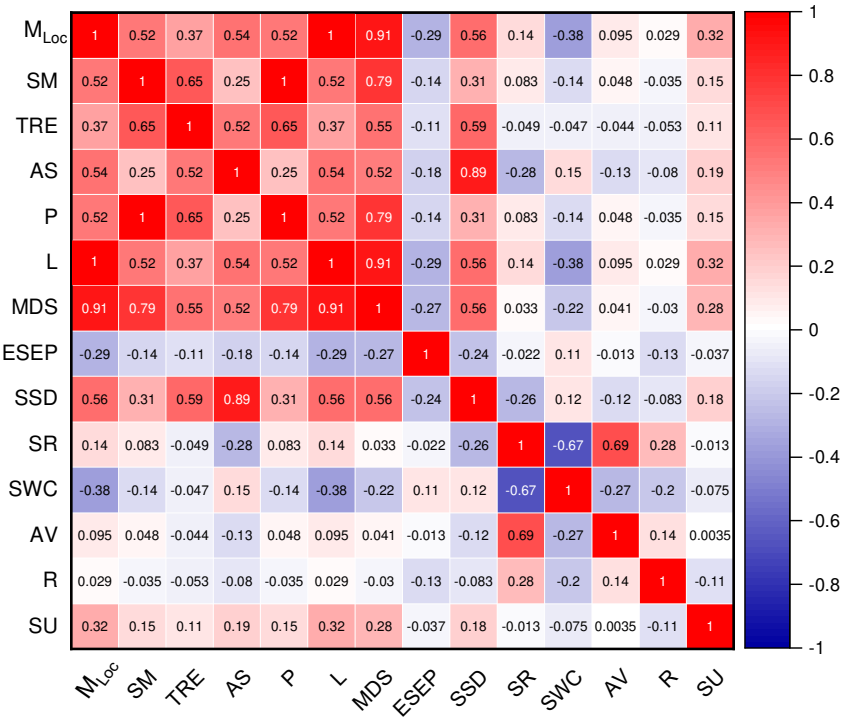


Figure 2 The correlation matrices for the input variables

3.2 Anomaly detection

In the initial data processing stage, assessing the presence of extreme anomalies within the database is imperative. These outliers exhibit distinct behaviour compared to the rest of the datasets, significantly impairing the ability of algorithms to identify the latent patterns and potentially leading to the overfitting problem (Shirani Faradonbeh & Taheri 2019). Various techniques, such as distance-based (e.g. local outlier factor [LOF]), density-based (DB), IQR and matrix factorisation methods (i.e. principal component analysis, PCA), were evaluated to detect outlier data. Among these approaches, DB methods yielded the most favourable outcomes. DB methods involve the computation of the distance between each data point and its nearest neighbours. Data points with significant distance from their neighbouring points are classified as outliers (Breunig et al. 2000). In this study, the LOF method, which falls under the category of DB methods, was employed. Comparative analysis of the results obtained from all the methods indicated that the LOF method outperformed others in accuracy and minimal removal of outlier data (see Table 2). Minimal data removal refers to the method's capacity to detect outliers without unduly eliminating data points that are not genuine outliers. As seen in Table 2, the LOF method discarded only 10% of the data points, thereby preserving most of the dataset. This approach ensures the dataset's integrity and maintains the underlying patterns within the data, which are vital for retaining important information that could influence the performance of subsequent analyses.

Table 2 Evaluation results of outlier data detection models

Method	Data points removed	Ratio (%)
LOF	3,330	10.0
DB	17,626	53.4
IQR	16,009	48.1
PCA	9,502	28.5

The conventional approach to outlier detection entails utilising statistical measures, distance metrics, density estimations and other indicators to quantify the deviation of a sample from the rest of the dataset. The LOF algorithm, an unsupervised method, calculates the local deviation in density for a given data point concerning its neighbourhood, thereby identifying outliers (Breunig et al. 2000). Unlike statistical methods that assume specific probability distributions and clustering algorithms that only make binary judgments regarding outlier status, the LOF algorithm offers a simpler and more intuitive solution (Aranha et al. 2023; Yin et al. 2021). It does not rely on distribution assumptions and quantitatively describes the degree of outliers for each sample. The calculation steps of the LOF algorithm are as follows:

The circle is delineated with sample point O as the central point, and the KNN distance, denoted as $d_K(O)$, is employed as the radius. The region encompassed within this circle is referred to as the K-distance domain.

The K-reachable distance from the sample point P to point O is computed according to the illustration presented in Figure 3, and it follows Equation 2 as stated below:

$$dist_k(P,O) = \max\{d_k(O), d(P,O)\} \tag{2}$$

where d (P, O) is the Euclidean distance between the sample points O and P.

The local reachable density of the sample point P is evaluated using Equation 3:

$$lrd_k(P) = \frac{N_k(P)}{\sum_{o \epsilon N_k(P)} dist_k(P,O)} \tag{3}$$

The LOF of the sample point P is computed using the following equation:

$$LOF_k(P) = \frac{\sum_{O \in N_k(P)} \frac{lrd_k(O)}{lrd_k(P)}}{N_K(P)} \tag{4}$$

where the absolute value of $LOF_k(P)$ approaching 1 signifies that the density of sample point P aligns with that of its neighbouring points, indicating that both P and its neighbours belong to the same cluster.

The small absolute value of $LOF_k(P)$ indicates that the density of point P surpasses that of its neighbourhood points, suggesting that P is densely populated. Conversely, a higher absolute value of $LOF_k(P)$ implies that the density of point P is lower than its neighbouring points, indicating a greater likelihood of P being an outlier. In this study a threshold value of 1.5 was set for $LOF_k(P)$. This means that if the LOF score of data point P exceeds 1.5 the point is considered an outlier. This value was chosen because it is empirically found (Kumaravel et al. 2022; Breunig et al. 2000) to provide a good balance between sensitivity and specificity in outlier detection across a variety of datasets. It is not too high to miss outliers or too low to flag too many normal points as outliers. Finally, the LOF method identified 3,330 data points as outliers. Among these, 160 were linked to b, 188 to e and 2,982 to n. These outliers were then removed from the original database (i.e. 10% of the original database removed).

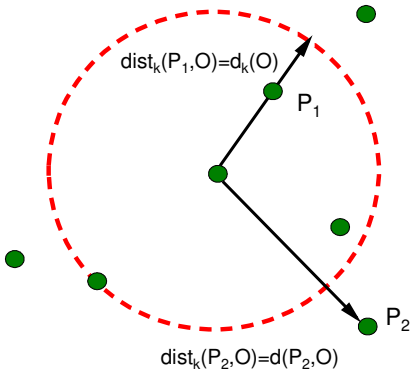


Figure 3 The diagram of reachable distance

3.3 Synthetic minority oversampling technique

In this study, the synthetic minority oversampling technique (SMOTE) method was employed to produce synthetic data. The widely utilised SMOTE algorithm serves the purpose of enhancing imbalanced datasets in the realm of data analysis. Imbalanced datasets pose a challenge where one class is significantly over-represented compared to another, resulting in biases in ML model development and their generalisability (Song et al. 2024; Geetha et al. 2019). To tackle this issue the SMOTE algorithm generates synthetic instances of the minority class by interpolating between existing instances. The process entails randomly selecting a minority instance and identifying its KNN. New instances are then generated by randomly selecting feature values from either the original instance or its KNN (see Figure 4). This procedure effectively balances the classes by introducing new instances into the dataset (Lui et al. 2022). While the SMOTE algorithm is typically employed for handling imbalanced datasets it is also possible to reverse the scenario by creating a dummy class and increasing its volume using this approach. Compared to alternative oversampling methods, the SMOTE algorithm offers several advantages. Firstly, it mitigates the risk of overfitting by avoiding the generation of exact replicas of existing instances. Secondly, it diminishes the likelihood of amplifying noise within the dataset since it generates new instances based on existing ones. Lastly, it demonstrates computational efficiency and can be applied to datasets with a high number of dimensions (Song et al. 2024; Geetha et al. 2019).

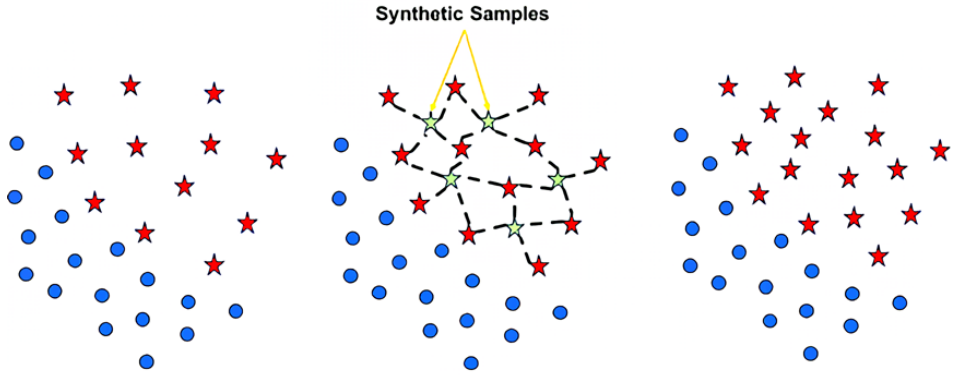


Figure 4 The principle of the synthetic minority oversampling technique

The important steps of performing the SMOTE algorithm are as follows (Chen et al. 2022): Step 1 – For each sample x_i in the minority category, the Euclidean distance is employed as an index to calculate its proximity to all samples in the same class sample set. This process helps identify the KNN. Step 2 – Based on the imbalance ratio of the given dataset, a sampling ratio (M) is defined to determine the proportion of various samples in the new dataset. For each x_i, a few samples are randomly selected from its KNNs, assuming that the chosen nearest neighbour is x_m. From each randomly selected neighbour x_m, a new sample is generated by modifying the original sample using Equation 5, where rand(0,1) represents a random number between 0 and 1.

$$x_{new} = x + rand(0,1) \times |x_i - x|, new \in 1,2,\dots,M \quad (5)$$

Step 3 – These steps are repeated M times to generate M new samples. In other words, if the minority category initially comprises T samples, new MT samples are generated. A balanced dataset is obtained after merging these new samples with the existing few original samples. Given the initial count of 12,194 instances in class n, the SMOTE method was employed to equalise the number of instances in classes e and b. Consequently the count of class e instances increased by 32% whereas the count of class b instances increased by 43%.

3.4 Parameter standardisation

Parameter standardisation is a crucial step in data preprocessing to facilitate the development of intelligent models. This is necessary due to input parameters with diverse ranges of values and units, which can impact

model training performance. Standardising these parameters plays a vital role in expediting the convergence of the model during training. When input features exhibit dissimilar scales or units, it can result in uneven learning rates across different features. This discrepancy can impede the optimisation algorithm's ability to find the optimal solution efficiently and may even lead to convergence in suboptimal regions. A comparable scale is achieved by standardising the input parameters, enabling the optimisation algorithm to converge more effectively. Additionally, standardisation ensures that each feature contributes equally to the learning process of the model (Habib & Okayli 2024). This study implemented the Z-score method (Ali et al. 2023; Wang et al. 2021) to standardise the parameters. The Z-score normalisation approach, also known as the normalisation score, utilises the mean (μ) and standard deviation (σ) values of the given parameter x to transform data points so that they have μ = 0 and σ = 1. Mathematically, the Z-score of a parameter x can be calculated as follows:

$$Z_{score} = \frac{(x-\mu)}{\sigma} \tag{6}$$

4 Machine learning algorithms

ML algorithms are computational models that help computers acquire knowledge and perform tasks such as classification, prediction or decision-making based on available data. These algorithms can be broadly categorised into five main types: supervised, unsupervised, semi-supervised, reinforcement and deep learning algorithms. Each algorithm type serves different purposes and can be applied in various domains, such as regression analysis, classification, clustering and image processing (Shirani Faradonbeh et al. 2019; Shirani Faradonbeh et al. 2023). In the present research, a comprehensive investigation was conducted using eight robust supervised learning algorithms: ANN, RF, DT, KNN, SVM, AB, GNB and LR. The primary objective was to examine the interplay between the independent and dependent variables through an intelligent approach.

4.1 Artificial neural network

ANNs are computational models inspired by the human brain's information processing capabilities. They have evolved since the 1940s into a powerful tool across technological and scientific domains. An ANN's strength lies in its ability to predict, classify and recognise patterns, making it indispensable for tackling intricate problems. The core of ANN operation is a highly interconnected system of simple processing units known as neurons, which can learn about complex dependencies between variables through organisation in layers. The most prevalent ANN architecture, the multilayer perceptron (MLP), comprises an input layer, one or more hidden layers and an output layer (Shakeri et al. 2022a, 2022b; Khoshalan et al. 2021). The classification prowess of ANNs is particularly beneficial as they empower the network to segment input data into distinct categories or groups. By training the ANN on labelled data, the network learns to discern patterns and make accurate predictions or classifications on unseen data. The success of an ANN in classification tasks is mainly due to its capacity to learn and generalise from training examples using an iterative learning process called backpropagation to minimise the difference between predicted and actual outputs. Factors such as the network's architecture, including the number of layers and neurons, the quality and quantity of training data, and the choice of activation functions and optimisation algorithms, significantly influence the network's performance. Despite the challenges, an ANN's remarkable classification abilities, spanning various fields, indicate its power and versatility (Onyelowe & Shakeri 2021; Onyelowe et al. 2021b).

4.2 Random forest

The RF algorithm is widely used in ML for classification and regression tasks. It operates by constructing multiple DTs, known as classification and regression trees (CART), through a process of bootstrap sampling and aggregation. In regression problems the input variables are split into multiple points to calculate the sum of square error (SSE) for predicted and actual values, with each node choosing the minimum SSE value. RF employs an ensemble of independent DT generated through random sampling, where new training sets are created from the original set using bootstrap sampling. DT are then built for each new bootstrap training set,

and the out-of-bag dataset is created to evaluate the performance of each DT. Random feature selection is performed at each step of building trees using cross-validation, helping to decorrelate the input trees and improve the model's accuracy (Li et al. 2022; Matin et al. 2018). The RF algorithm is particularly effective in classification tasks due to its robustness in handling large feature sets and interactions among predictor variables. It can also deal with unstable learners, delivering excellent performance when implemented with high-quality and free implementations. The algorithm determines the importance of each feature in the dataset, aiding in model inference. By using different bootstrap samples and subsets of features to train individual trees, RF achieves faster training and better predictive performance by striking a favourable variance-bias trade-off. Overall, RF is an ensemble-based ML model with numerous DT constructed on diverse bootstrapped datasets. It effectively reduces variance and enhances the predictive capability of DT, making it a valuable tool for classification and regression tasks, particularly in civil and mining engineering (Zhang et al. 2019).

4.3 Decision tree

The DT algorithm is widely used in ML and data mining for classification and prediction tasks. It encompasses various subsets but two common ones are CART and chi-square automatic interaction detector (CHAID). CART is particularly suitable for predicting continuous variables and stands out for its white box nature and simple interpretability. This algorithm facilitates understanding of the relationship between input and output parameters, making it superior to other DT algorithms. CART is robust in handling complex samples and large-scale datasets without being affected. It can be applied as a CT or an RT, and the right partition for a database is determined using indices like the Gini criterion, entropy or the Twoing criterion. By not assuming variable distribution, CART's non-parametric nature makes it versatile. During training, prevention criteria such as a minimum number of observations, maximum tree depth and least error values help avoid developing overly complex trees. Pruning the tree using a verification dataset ensures the optimal subtree is obtained while reducing overtraining (Yari et al. 2023; Ghasemi et al. 2020; Faradonbeh et al. 2022a). In classification tasks, the DT algorithm, especially CART, demonstrates its ability to classify data effectively. Its simplicity and interpretability make it a valuable tool for understanding the relationships between input and output variables. CART can efficiently classify instances into different classes or categories by partitioning the data based on specific indices. With the flexibility to handle complex samples and large datasets, it remains reliable even in challenging scenarios. Using prevention criteria during training helps avoid creating overly complex trees, ensuring a balanced and optimised classification model. The DT algorithm, particularly CART, thus proves to be a powerful and versatile technique for classification tasks in supervised learning (Pu et al. 2018; Ghasemi et al. 2020).

4.4 K-nearest neighbour

The KNN algorithm is a non-parametric method that identifies K instances most closely related to an unknown sample within a given dataset. This algorithm is particularly effective in large datasets and limited dimensionality scenarios, where relative neighbour techniques emerge as efficient local procedures. These techniques are applicable across a spectrum of ML domains, particularly in multi-label classification, regression and semi-supervised learning contexts. KNN is recognised for its ease of implementation, practicality, and user-friendliness, exhibiting similarities with other algorithms such as ANN and RF in its classification and regression analysis applications. At its essence, KNN operates by pinpointing K samples adjacent to unknown samples in the calibration dataset, commonly using distance measures. Optimisation can be achieved by identifying models with congruent attributes (Khan et al. 2022; Gomah et al. 2022). Furthermore, KNN assigns categories to unknown samples by averaging pertinent variables and comparing the outcomes against those of the K known samples. The efficacy of KNN hinges heavily on the selection of K. One of the benefits of KNN is its simplicity in interpretation and execution. Additionally, KNN performs adeptly in environments with non-linear decision boundaries, making it valuable for various classification and regression tasks. Adjusting the value of K provides considerable adaptability in setting thresholds. Unlike other ML frameworks, KNN does not require a distinct training phase, thus simplifying the process of hyperparameter tuning (Mahmoodzadeh et al. 2021; Wang et al. 2024; Ahmad et al. 2021).

4.5 Support vector machine

The SVM algorithm is a powerful and versatile supervised ML technique for classification and regression tasks. Unlike other methods such as ANN, SVM follows the principles of structural risk minimisation to minimise model error. This makes it particularly beneficial when dealing with datasets where the number of features exceeds the number of data points. During the training phase, SVM carefully analyses the input data and constructs a hyperplane that separates the data into distinct classes or regimes. This hyperplane is optimised by maximising the margin between the closest data points belonging to different classes. These crucial data points, known as support vectors, are pivotal in establishing the decision boundary. SVM is highly regarded for its ability to handle complex decision boundaries and efficiently process high-dimensional data. By finding the hyperplane that not only separates the classes but also maximises the margin, SVM achieves good generalisation performance, reducing the risk of overfitting. Furthermore, SVM can handle both linearly separable and non-linearly separable data by utilising kernel functions that transform the data into higher-dimensional feature spaces. This allows SVM to find a hyperplane that effectively separates the classes, even in cases where the data is not linearly separable (Baghbani et al. 2023; Xu et al. 2022, Wang et al. 2024).

4.6 AdaBoost algorithm

The AB algorithm, or adaptive boosting, is an ensemble method widely used in ML that leverages DT as its primary classifier. The core principle behind AB is to adaptively assign weights to each instance, with higher weights assigned to instances that were incorrectly classified. The AB algorithm, developed by Freund & Schapire (1995), is the most commonly employed variant of the boosting algorithm, optimising classifier accuracy. This learning approach is relatively straightforward, constructing a strong classifier from a small number of weak classifiers that are individually efficient but lack predictive power. The objective is to combine these weak classifiers to enhance their overall performance. Consequently the resulting robust classifier can predict the class of new observations within a given dataset. AB enhances the classification efficacy of a simple learning algorithm by aggregating sets of weak classifiers to build a more resilient classifier. In boosting algorithms, the simple learning algorithm is referred to as a weak learner, which selects a limited yet effective set of potentially relevant features. A weak learner cannot accurately categorise the training data even when employing the best classification function. A series of learning challenges must be addressed to strengthen the weak learner. Following each learning cycle, the instances are reweighted to emphasise those misclassified by the preceding weak classifier. The final robust classifier then employs a weighted combination of the weak classifiers to determine each feature's optimal threshold classification function (Ahmad et al. 2022; Pu et al. 2018).

4.7 Gaussian Naive Bayes

The GNB algorithm is an ML classification technique that assumes the dataset's features adhere to a Gaussian distribution. This algorithm is considered a variant of the Naive Bayes classifier, which assumes independence among features. This assumption simplifies the computational process and often yields satisfactory results despite its limited applicability to real-world datasets. To classify a test observation, the algorithm calculates the posterior probability for each class by multiplying the likelihood of the observation given each class (which is assumed to follow a Gaussian distribution) by the prior probability of each class. The observation is then assigned to the class with the highest posterior probability. The likelihood function is determined using the Gaussian distribution, which characterises the probability density function of a symmetrically distributed random variable. One notable advantage of the GNB algorithm is its ability to handle continuous input variables without requiring complex transformations or scaling procedures. This property makes it particularly advantageous when working with datasets that exhibit a normal feature distribution. Furthermore, the algorithm demonstrates computational efficiency. During the training phase the algorithm calculates the mean and standard deviation for each feature within each class, a task that can be swiftly accomplished even with large datasets. During prediction, the algorithm utilises these precomputed statistics to calculate posterior probabilities, resulting in prompt predictions (Dong et al. 2016; Ibrahim et al. 2023; Pu et al. 2020).

4.8 Logistic regression

LR is a supervised ML algorithm renowned for its prowess in tackling binary classification problems. It serves as a potent instrument for scrutinising and categorising variables of a categorical nature, finding applications across domains such as engineering and medicine. The crux of LR lies in the sigmoid function, which imparts a transformative effect on linear regression, morphing it into the logit function and thereby rendering it amenable to dealing with categorical outcomes. LR not only furnishes an assessment of the significance of predictors but also furnishes cues about the direction of the relationship between predictors and the outcome. This attribute bestows upon it an informative output that engenders sound decision-making. LR is harnessed to estimate the likelihood of an event materialising, thereby streamlining data classification into discrete classes. The regression coefficients, gleaned via maximum likelihood estimation, yield valuable insights into the magnitude and direction of the link between explanatory variables and the categorical outcome. This conveys an enhanced comprehension of the underlying determinants that govern the classification decision. Consequently, LR models assume a position of eminence when scrutinising and categorising categorical variables. Their versatility and efficacy have rendered them a favoured choice across various domains (Xu et al. 2023; Dong et al. 2016; Faradonbeh et al. 2022b).

5 Metrics for performance evaluation

Classification metrics serve as a yardstick for evaluating the performance of classification algorithms. This study employed various metrics to compare and assess the adjusted unsupervised ML algorithm. Accuracy (ACC) is a metric that considers both normal and faulty samples. The classification ACC assessment in isolation may lead to misinterpretations, particularly in unevenly distributed observations across different classes or in scenarios involving more than two classes. Employing a confusion matrix proves to be a valuable instrument in comprehending the precise performance of a classification model and identifying the specific types of errors it encounters. Comprised of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) components, the confusion matrix provides a structured overview of classification outcomes. ACC can be computed using Equation 7, where n_{total} signifies the total number of samples (Faradonbeh et al. 2024; Zhou et al. 2022; Ma et al. 2024).

$$ACC = \frac{TP+TN}{n_{total}} \tag{7}$$

An illustration of a three-class classification scenario depicted in Figure 5 showcases how the confusion matrix visually represents the correct classification of microseismic records through its diagonal elements, while any misclassifications are evident in the off-diagonal elements. An optimal classifier would exhibit solely diagonal elements, resulting in a perfect overall accuracy score of 1. The integration of classification metrics, including the analysis of confusion matrices, assumes a pivotal role in evaluating the effectiveness and dependability of ML algorithms (Faradonbeh et al. 2024; Zhou et al. 2022; Ma et al. 2024).

Confusion matrix		Predicted class		
		1st class	2nd class	3rd class
Real class	1st class	X_{11}	X_{12}	X_{13}
	2nd class	X_{21}	X_{22}	X_{23}
	3rd class	X_{31}	X_{32}	X_{33}

Figure 5 The configuration of a standard confusion matrix for a three-class classification scenario

Precision (PR) quantifies the TP value relative to the FP. Equation (8) can be utilised to compute PR, with FP representing the count of FPs.

$$PR = \frac{TP}{TP+FP} \tag{8}$$

Recall (REC) indicates the proportion of anomalous data accurately identified among all anomalies. In industrial applications, REC holds significant importance as FNS carry more severe consequences than false alarms or FPs. Equation 9 calculates REC, with FN representing the count of FNs.

$$REC = \frac{TP}{TP+FN} \tag{9}$$

The F1 score is a harmonic mean between REC and PR, and it can be approximated using Equation 10. Utilising F1 as a metric also helps alleviate the impact of an imbalanced database, where the positive data is considerably smaller than the negative data. Accuracy exhibits limited sensitivity in temporal classification scenarios.

$$F1 = \frac{2(PR \times REC)}{PR+REC} \tag{10}$$

Area under the receiver operating characteristic (AUC-ROC) illustrates the relationship between the false positive rate (FPR) and the true positive rate (TPR). A greater AUC indicates a better classifier model, with a perfect classifier represented by an AUC of 1. Generally, higher metric values indicate better predictive performance, a principle that applies across all chosen metrics.

6 Discussion and results

As delineated previously, this research employed a variety of robust classifier ML algorithms to construct models distinguishing between different classes of microseismic recordings. The efficacy of ML algorithms in discerning intricate associations between input and output variables can be bolstered through the provision of ample training data and the utilisation of appropriate training algorithms. 70% of the dataset was allocated for model training purposes, while the remaining unseen data was reserved for validating the trained model. In this investigation, the confusion matrix and associated performance metrics were utilised to assess the efficacy of the proposed multi-class classifiers in categorising catalogue records. The settings that lead to the highest accuracy in predicting microseismic event types are listed in Table 3. Furthermore, the dataset partitioning technique employed in this investigation involved utilising the random division approach. While this methodology may not fully ensure the statistical coherence between the training and test sets, it upholds the impartiality of the experimental outcomes. It is worth mentioning that an identical training set was utilised for model development to enhance the comparative analysis of all hybrid models.

Table 3 The optimum values of controlling parameters for the developed classifiers (continued next page)

Model	Parameter	Default value
Decision tree	max_depth	5 (none)
	min_samples_split	10
	min_samples_leaf	5
	max_features	sqrt'
Random Forest	n_estimators	100
	max_depth	10 (none)
	min_samples_split	5
	min_samples_leaf	2
	max_features	sqrt'
AdaBoost	n_estimators	50
	learning_rate	1
Support vector machine	Kernel function	Radial basic function

Model	Parameter	Default value
	ε	0.1
	tol	0.001
	γ	1/number of features
K-nearest neighbours	n_neighbours	5
	weights	Uniform
	algorithm	Automatic
Logistic regression	penalty	l2
	C	1
	solver	lbfgs
MLP classifier	hidden_layer_sizes	(100)
	activation	Relu
	solver	Adam
	alpha	0.0001
Gaussian Naive Bayes	priors	None
	var smoothing	1e-9

The confusion matrices and overall accuracy of the developed multi-class classifiers during both the training and testing phases are depicted in Table 4. According to this table, all the classification techniques employed in this research exhibit high precision in forecasting signal types recorded by the microseismic monitoring system, consistently achieving accuracies exceeding 90% during both the training and testing phases. A detailed analysis of Table 4 and Figure 6 shows that the SVM, DT and LR models exhibit nearly identical prediction accuracies for the training data, with 95.88, 95.18, and 95.04 values, respectively. These values are slightly higher than AB's accuracy of 94.86, with a marginal difference of about 1%. Similarly, the KNN and MLP models show comparable accuracies of 96.38 and 96.06 for the training data, slightly lower than the RF model's accuracy of 97.25 by approximately 1%. In contrast, the test data results indicate that the MLP, RF, SVM and LR models have prediction accuracies of 95.61, 95.51, 95.46 and 95.02, respectively, surpassing the DT, AB, and KNN models, which achieved 94.60, 94.59 and 92.94, respectively. Notably, the GNB model yielded the lowest prediction accuracy for seismic events, with values of 90.79 and 90.82 for the training and testing data, respectively.

The findings also indicate that most of the intelligent models, except for the GNB model, predicted seismic events with a variance of about 1% in accuracy between the train and test stages. Among these, the RF and MLP models demonstrated relatively higher accuracies. Specifically, the RF model accurately predicted 21,456 out of 22,062 training data cases and 9,031 out of 9,456 testing data cases, while the MLP model correctly predicted 21,193 training data cases and 9,041 testing data cases. Conversely, the GNB model accurately predicted 20,029 training data cases and 8,588 testing data cases, showing the weakest performance among the models studied. Overall, the results demonstrate that all the classification algorithms show high accuracy in identifying blasts. This is obvious from confusion matrices shown in Table 4, where the orange diagonal data represent the number of correctly predicted cases for each class, highlighting the models' proficiency in predicting blasts. This can also be observed in Table 5, where the precision of each model in categorising the data in proper classes is listed. According to this table, the precision of all models in predicting the blasts is higher than in other classes. This can be related to some distinct features of blast records compared to other recordings, which could make their detection easier for the algorithms (some

algorithms like RF are particularly good at capturing the decision boundaries in the feature space when they are clear and strong signals).

Table 4 The confusion matrices of machine learning classifiers for the training and testing stages

Confusion matrix	Model	Predicted class						Model	Predicted class					
	DT	Train			Test			RF	Train			Test		
		n	e	b	n	e	b		n	e	b	n	e	b
Real class	n	7,254	71	22	3,097	48	14	n	7,286	47	14	3,117	30	6
	e	148	6,830	356	50	2,847	275	e	81	6,960	293	53	2,916	203
	b	12	455	6,914	11	114	3,000	b	4	167	7,210	8	125	2,992
	ACC (%)	95.18			94.59			ACC (%)	97.25			95.51		
	Model	Predicted class						Model	Predicted class					
	AB	Train			Test			SVM	Train			Test		
		n	e	b	n	e	b		n	e	b	n	e	b
	n	7,214	114	19	3,107	44	8	n	7,257	70	20	3,120	29	10
	e	160	67,31	443	73	2,880	219	e	121	6,831	382	51	2,923	198
	b	9	390	6,982	8	159	2,958	b	4	313	7,064	8	133	2,984
	ACC (%)	94.86			94.60			ACC (%)	95.88			95.46		
	Model	Predicted class						Model	Predicted class					
	KNN	Train			Test			LR	Train			Test		
		n	e	b	n	e	b		n	e	b	n	e	b
	n	7,268	56	23	3,115	35	9	n	7,232	76	39	3,118	30	11
	e	107	6,873	354	56	2,895	221	e	158	6,743	433	66	2,902	204
	b	6	253	7,122	6	153	2,966	b	11	377	6,993	8	152	2,965
	ACC (%)	96.38			94.92			ACC (%)	95.04			95.02		
	Model	Predicted class						Model	Predicted class					
	GNB	Train			Test			MLP	Train			Test		
		n	e	b	n	e	b		n	e	b	n	e	b
	n	6,,687	515	145	2,895	216	48	n	7,262	67	18	3,117	37	5
	e	92	6,168	1074	29	2,664	479	e	114	6,853	367	55	2,950	167
	b	3	204	7,174	1	95	3029	b	5	298	7,078	6	145	2,974
	ACC (%)	90.79			90.82			ACC (%)	96.06			95.61		

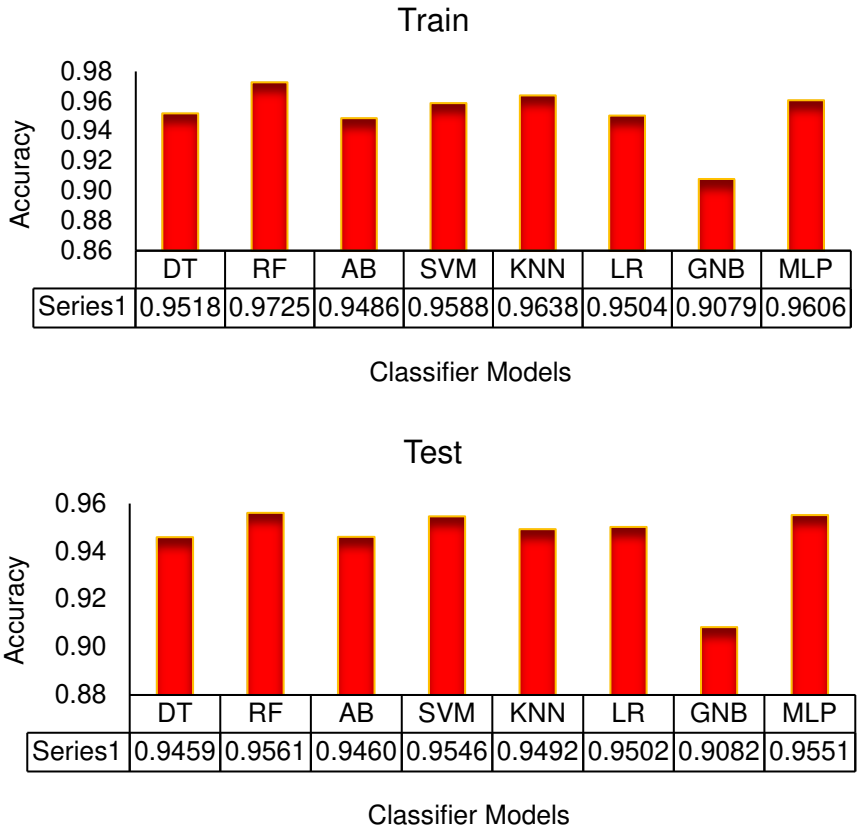


Figure 6 Classification accuracy for all machine learning models

Table 5 The prediction accuracy of class type for different algorithms at the testing stage

Model	b	e	n
DT	0.9807	0.9462	0.9121
RF	0.9808	0.9495	0.9347
AB	0.9746	0.9342	0.9287
SVM	0.9814	0.9475	0.9348
KNN	0.9805	0.9390	0.9280
LR	0.9768	0.9410	0.9324
GNB	0.9897	0.8955	0.8518
MLP	0.9808	0.9419	0.9453

Additional examination of the precision, recall and F1 score metrics for each category in the training and testing phases is outlined in Tables 6 and 7. According to the study conducted by Pu et al. (2020), classification accuracy is frequently used as a performance metric although it may not always be the most suitable measure. In scenarios where users are interested in specific questions, such as the number of identified blasting events out of all events or the accuracy in identifying true microseismic events, classification accuracy alone may not suffice. Therefore the study introduced precision and recall as alternative indicators to assess performance. Furthermore, the researchers employed the F1 score, a metric that combines both recall and precision, to evaluate the effectiveness of the classification algorithms. A higher F1 score indicates a more accurate classification model.

Table 6 Evaluation results of the developed models in the training stage

Method	Train							
	ACC	Score	PR	Score	REC	Score	F1	Score
DT	0.9518	4	0.9517	4	0.9518	4	0.9517	4
RF	0.9725	8	0.9726	8	0.9725	8	0.9725	8
AB	0.9486	2	0.9485	2	0.9486	2	0.9485	2
SVM	0.9588	5	0.9587	5	0.9588	5	0.9587	5
KNN	0.9638	7	0.9638	7	0.9638	7	0.9637	7
LR	0.9504	3	0.9503	3	0.9504	3	0.9503	3
GNB	0.9079	1	0.9120	1	0.9079	1	0.9079	1
MLP	0.9606	6	0.9606	6	0.9606	6	0.9606	6

Table 7 Evaluation results of the developed models in the testing stage and final score

Method	Test								Total	Final
	ACC	Score	PR	Score	REC	Score	F1	Score		
DT	0.9459	2	0.9464	3	0.9459	2	0.9457	2	25	6
RF	0.9551	7	0.9551	7	0.9551	7	0.9549	7	60	1
AB	0.9460	3	0.9459	2	0.9460	3	0.9458	3	19	7
SVM	0.9546	6	0.9546	6	0.9546	6	0.9545	6	44	3
KNN	0.9492	4	0.9492	4	0.9492	4	0.9491	4	44	4
LR	0.9502	5	0.9501	5	0.9502	5	0.9501	5	32	5
GNB	0.9082	1	0.9125	1	0.9082	1	0.9083	1	8	8
MLP	0.9561	8	0.9560	8	0.9561	8	0.9561	8	56	2

In order to identify the most effective ML model for classification during both the training and testing phases, the study implemented a scoring analysis. Each model was assigned a score from 1 to 8 according to how well it performed across various indicators compared to other models. Models were then ranked based on their performance metrics, as outlined in Tables 6 and 7 and Figure 7. The overall score for each model was determined by aggregating its training and testing scores. The RF model emerged as the top performer, demonstrating its proficiency in capturing intricate relationships within the data. Additionally, the study revealed that the MLP, SVM, KNN, LR, DT and AB methods followed RF in the rankings. Conversely, the GNB method ranked lowest in the overall performance evaluation.

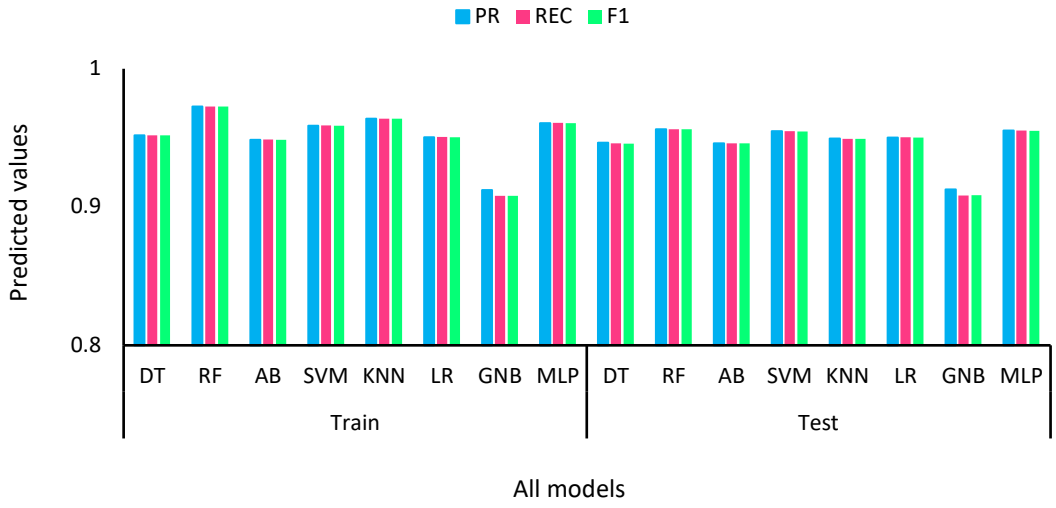
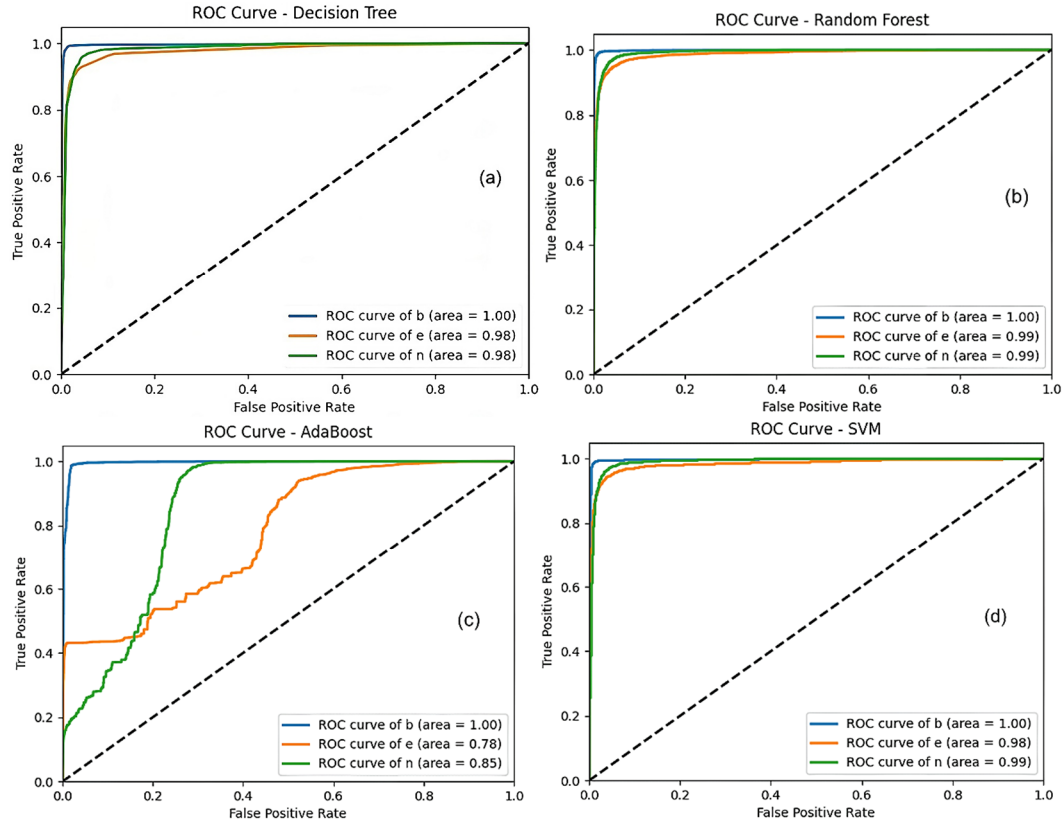


Figure 7 Comparison of precision, recall and F1 scores for all developed models

In addition to utilising a confusion matrix, the ROC curve is a valuable tool for illustrating the level of agreement between predicted and actual values. ROC curve is obtained by plotting the TPR against the FPR. The AUC equal to 0.5 refers to the random classification by the model, while AUC = 1 refers to the perfect classification. The analysis of the ROC curves depicted in Figure 8 reveals that all models excelled in predicting the blast (b) with AUC = 1, which is consistent with the confusion matrix results shown in Table 4. In general, RF and MLP models outperformed SVM, LR, DT, KNN, GNB and AB models by AUC values of 1, 0.99 and 0.99 for classes b, e and n, respectively. This could be related to the higher capability of the RF and MLP algorithms in inferring the latent complex relationships between parameters for this database. On the other hand, the correct prediction of the noise (n) was more difficult, probably because some noise parameter values overlapped more with event parameter values. This correlates with the similar difficulty reported by the geotechnical engineer undertaking regular audits of waveforms on the mine site.



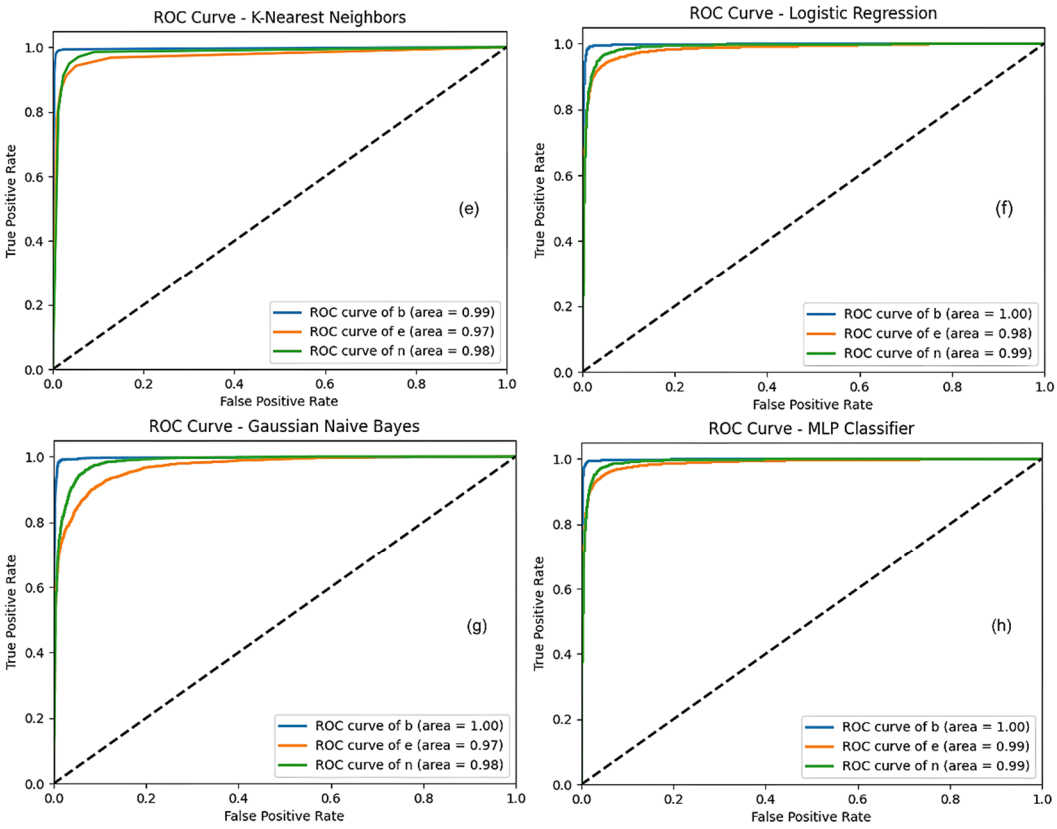


Figure 8 ROC graph for all machine learning models. (a) Decision tree; (b) Random Forest; (c) AdaBoost; (d) Single vector machine; (e) k-nearest neighbour; (f) Logistic regression; (g) Gaussian Naaive Bayes; (h) Multilayer perceptron

7 User interface development

To enhance the usability and accessibility of the intelligent best model formulated in this study for researchers and engineers, a graphical user interface (GUI) was developed, as seen in Figure 9. This GUI empowers users to effortlessly predict the class types recorded by the microseismic monitoring system by inserting values for inputs listed in Table 1. As observed in the above sections, the RF method demonstrated superior accuracy in distinguishing between recorded class types among all the intelligent methods employed in this study. Therefore the GUI described in this section is based on the RF model. It is critical to mention that this GUI has been established based on the data range (minimum-maximum range) specified in Table 1, and any user input data must adhere to this range. In future an expanded database and further research endeavours will enable the optimisation and enhancement of the aforementioned GUI system. Users can access the GUI interface through the following link: https://github.com/Faradonbeh/Microseismic-Monitoring/releases/tag/v1.0.0.

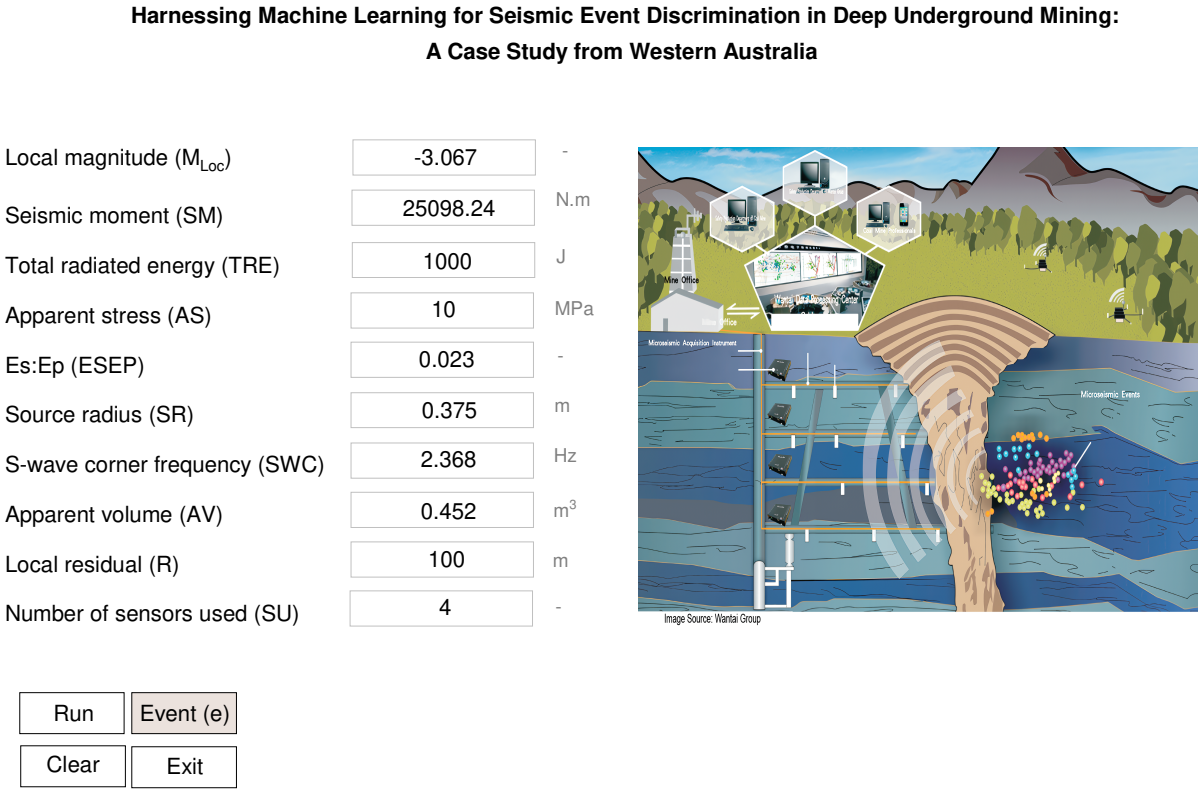


Figure 9 Graphical user interface of the developed random forest-based model

8 Conclusion

This paper has demonstrated the effectiveness of ML techniques in discriminating seismic event classes in deep underground mining using data collected from a gold mine in Western Australia. Potential multicollinearity issues were addressed through a rigorous statistical analysis, and key input variables were identified. The LOF method was utilised to remove anomalies, ensuring homogeneity in the dataset. The SMOTE method was employed to address imbalanced datasets, particularly in classifying seismic event types. Eight ML algorithms were used to develop classifiers. It was found that SVM, DT and LR models exhibited similar prediction accuracies for the training data with accuracy values of 95.88%, 95.18% and 95.04%, respectively, which were slightly higher than the AB model's accuracy (94.86%). KNN and MLP models also showed accuracy values of 96.38% and 96.06%, just below RF's 97.25%. For the testing data, MLP, RF, SVM and LR achieved accuracies of 95.61%, 95.51%, 95.46% and 95.02%, outperforming DT, AB and KNN, which scored 94.60%, 94.59% and 92.94%, respectively. The GNB model revealed the lowest prediction accuracy for both training (90.79%) and testing data (90.82%). Overall, the RF and MLP models demonstrated the highest modelling performance, successfully predicting 21,456 and 21,193 training cases and 9,031 and 9,041 test cases, respectively. In contrast, the GNB model showed the weakest performance by predicting 20,029 training cases and 8,588 testing cases accurately. Following RF and MLP, the SVM, KNN, LR, DT and AB models showed strong performance, whereas GNB ranked the lowest. These results, supported by confusion matrices and precision criteria, highlight that all tested classification algorithms show high accuracy in identifying the blast type.

Furthermore, the ROC curve analysis (Figure 8) showed that all models performed well in predicting the blasts (b), consistent with the confusion matrix results (Table 4). Correct prediction of the noise (n) was more difficult, probably because some noise parameter values overlapped more with event parameter values. However, the RF and MLP models exhibited higher AUC values for all three signal classes (i.e. b, e and n classes) compared to other ML models, demonstrating the higher capability of these techniques in distinguishing different types of recordings. The study also developed a user-friendly GUI based on the

proposed RF model, enhancing the interpretation of microseismic monitoring results in practical applications. Overall, this research underscores the efficacy of ML approaches in seismic event discrimination, offering valuable insights for improved monitoring and safety measures in deep underground mining operations. The findings provide a solid foundation for further advancements in utilising ML to enhance the safety and efficiency of mining operations. Future endeavours could achieve higher accuracy and more generalised models by implementing parallel processing techniques to handle large datasets, performing more comprehensive hyperparameter tuning and including more case studies.

Acknowledgement

The authors gratefully thank ESG Solutions and Ramelius Resources Ltd for making the Vivien gold mine seismic dataset available for analysis and reporting in this paper.

References

Ahmad, M, Kamiński, P, Olczak, P, Alam, M, Iqbal, MJ, Ahmad, F ... & Khan, BJ 2021, 'Development of prediction models for shear strength of rockfill material using machine learning techniques', *Applied Sciences*, vol. 11, no. 13.

Ahmad, M, Katman, HY, Al-Mansob, RA, Ahmad, F, Safdar, M & Alguno, AC 2022, 'Prediction of rockburst intensity grade in deep underground excavation using adaptive boosting classifier', *Complexity*, vol. 2022, no. 1, pp. 1–10.

Ali, M, Zhu, P, Jiang, R, Huolin, M, Ehsan, M, Hussain, W ... & Ullaah, J 2023, 'Reservoir characterization through comprehensive modeling of elastic logs prediction in heterogeneous rocks using unsupervised clustering and class-based ensemble machine learning', *Applied Soft Computing*, vol. 148, no. 110843.

Aranha, PE, Policarpo, NA & Sampaio, MA 2023, 'Unsupervised machine learning model for predicting anomalies in subsurface safety valves and application in offshore wells during oil production', *Journal of Petroleum Exploration and Production Technology*, vol. 14, pp. 567–581.

Baghbani, A, Costa, S, Faradonbeh, RS, Soltani, A & Baghbani, H 2023, 'Modeling the effects of particle shape on damping ratio of dry sand by simple shear testing and artificial intelligence', *Applied Sciences*, vol. 13, no. 7.

Breunig, MM, Kriegel, HP, Ng, RT & Sander, J 2000, 'LOF: identifying density-based local outliers', *ACM SIGMOD Record*, vol. 29, issue 2, pp. 93–104.

Chen, J, Huang, H, Cohn, AG, Zhang, D & Zhou, M 2022, 'Machine learning-based classification of rock discontinuity trace: SMOTE oversampling integrated with GBT ensemble learning', *International Journal of Mining Science and Technology*, vol. 32, no. 2, pp. 309–322.

Dong, L, Tang, Z, Li, XB, Chen, YC & Xue, JC 2020, 'Discrimination of mining microseismic events and blasts using convolutional neural networks and original waveform', *Journal of Central South University*, vol. 27, no. 10, pp. 3078–3089.

Dong, L, Wesseloo, J, Potvin, Y & Li, X 2016, 'Discrimination of mine seismic events and blasts using the fisher classifier, naive bayesian classifier and logistic regression', *Rock Mechanics and Rock Engineering*, vol. 49, pp. 183–211.

Faradonbeh, RS, Ryoza, MG & Sepehri, M 2024, 'Application of artificial intelligence in distinguishing genuine microseismic events from the noise signals in underground mines', *Applications of Artificial Intelligence in Mining, Geotechnical and Geoengineering*, pp. 197–220.

Faradonbeh, RS, Taheri, A & Karakus, M 2022a, 'Fatigue failure characteristics of sandstone under different confining pressures', *Rock Mechanics and Rock Engineering*, vol. 55, no. 12, pp. 1–26.

Faradonbeh, RS, Taheri, A & Karakus, M 2022b, 'The propensity of the over-stressed rock masses to different failure mechanisms based on a hybrid probabilistic approach', *Tunnelling and Underground Space Technology*, vol. 119, no. 6.

Freund, Y & Schapire, RE 1995, 'A desicion-theoretic generalization of on-line learning and an application to boosting', *European Conference on Computational Learning Theory*, Springer, Berlin, pp. 23–37.

Ghasemi, E, Gholizadeh, H & Adoko, AC 2020, 'Evaluation of rockburst occurrence and intensity in underground structures using decision tree approach', *Engineering with Computers*, vol. 36, pp. 213–225.

Geetha, R, Sivasubramanian, S, Kaliappan, M, Vimal, S & Annamalai, S 2019, 'Cervical cancer identification with synthetic minority oversampling technique and PCA analysis using random forest classifier', *Journal of medical systems*, vol. 43, no. 1, pp. 1–19.

Gomah, ME, Li, G, Khan, NM, Sun, C, Xu, J, Omar, AA ... & Zaki, MM 2022, 'Prediction of strength parameters of thermally treated Egyptian granodiorite using multivariate statistics and machine learning techniques', *Mathematics*, vol. 10, no. 23.

Gowida, A, Elkatatny, S & Gamal, H 2021, 'Unconfined compressive strength (UCS) prediction in real-time while drilling using artificial intelligence tools', *Neural Computing and Applications*, vol. 33, pp. 8043–8054.

Habib, M & Okayli, M 2024. 'Evaluating the sensitivity of machine learning models to data preprocessing technique in concrete compressive strength estimation', *Arabian Journal for Science and Engineering*, pp. 1–19.

Ibrahim, AF, Ahmed, A & Elkatatny, S 2023, 'Applications of different classification machine learning techniques to predict formation tops and lithology while drilling', *ACS Omega*, vol. 8, no. 45, pp. 42152–42163.

Khan, K, Ahmad, W, Amin, MN, Aslam, F, Ahmad, A & Al-Faiad, MA 2022, 'Comparison of prediction models based on machine learning for the compressive strength estimation of recycled aggregate concrete', *Materials*, vol. 15, no. 10.

Khoshalan, HA, Shakeri, J, Najmoddini, I & Asadizadeh, M 2021, 'Forecasting copper price by application of robust artificial intelligence techniques', *Resources Policy*, vol. 73.

Kumaravel, VP, Buiatti, M, Parise, E & Farella, E 2000, 'Adaptable and robust EEG bad channel detection using local outlier factor (LOF)', *Sensors*, vol. 22, no. 19.

Li, J, Li, C & Zhang, S 2022, 'Application of Six metaheuristic optimization algorithms and random forest in the uniaxial compressive strength of rock prediction', *Applied Soft Computing*, vol. 131.

Lui, TC, Gregory, DD, Anderson, M, Lee, WS & Cowling, SA 2022, 'Applying machine learning methods to predict geology using soil sample geochemistry', *Applied Computing and Geosciences*, vol. 16.

Ma, J, Ma, C, Li, T, Yan, W, Faradonbeh, RS, Long, H & Dai, K 2024, 'Real-time classification model for tunnel surrounding rocks based on high-resolution neural network and structure–optimizer hyperparameter optimization', *Computers and Geotechnics*, vol. 168.

Mahmoodzadeh, A, Mohammadi, M, Ibrahim, HH, Abdulhamid, SN, Salim, SG, Ali, HFH & Majeed, MK 2021, 'Artificial intelligence forecasting models of uniaxial compressive strength', *Transportation Geotechnics*, vol. 27.

Matin, SS, Farahzadi, L, Makaremi, S, Chelgani, SC & Sattari, GH 2018, 'Variable selection and prediction of uniaxial compressive strength and modulus of elasticity by random forest', *Applied Soft Computing*, vol. 70, pp. 980–987.

Onyelowe, KC, Mahesh, CB, Srikanth, B, Nwa-David, C, Obimba-Wogu, J & Shakeri, J 2021a, 'Support vector machine (SVM) prediction of coefficients of curvature and uniformity of hybrid cement modified unsaturated soil with NQF inclusion', *Cleaner Engineering and Technology*, vol. 5.

Onyelowe, KC & Shakeri, J 2021, 'Intelligent prediction of coefficients of curvature and uniformity of hybrid cement modified unsaturated soil with NQF inclusion', *Cleaner Engineering and Technology*, vol. 4.

Onyelowe, KC, Shakeri, J, Amini-Khoshalann, H, Salahudeen, AB, Arinze, EE & Ugwu, HU 2021b, 'Application of ANFIS hybrids to predict coefficients of curvature and uniformity of treated unsaturated lateritic soil for sustainable earthworks', *Cleaner Materials*, vol. 1.

Pu, Y, Apel, DB & Hall, R 2020, 'Using machine learning approach for microseismic events recognition in underground excavations: Comparison of ten frequently-used models', *Engineering Geology*, vol. 268.

Pu, Y, Apel, DB & Lingga, B 2018, 'Rockburst prediction in kimberlite using decision tree with incomplete data', *Journal of Sustainable Mining*, vol. 17, no. 3, pp. 158–165.

Peng, P, He, Z & Wang, L 2019, 'Automatic classification of microseismic signals based on MFCC and GMM-HMM in underground mines', *Shock and Vibration*, pp. 1–10.

Potvin, Y & Wesseloo, J 2013, 'Towards an understanding of dynamic demand on ground support', *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 113, no. 12, pp. 913–922.

Rao, D, Shi, X, Zhou, J, Yu, Z, Gou, Y, Dong, Z & Zhang, J 2021, 'An expert artificial intelligence model for discriminating microseismic events and mine blasts', *Applied Sciences*, vol. 11, no. 14.

Shakeri, J, Asadizadeh, M & Babanouri, N 2022a, 'The prediction of dynamic energy behavior of a Brazilian disk containing nonpersistent joints subjected to drop hammer test utilizing heuristic approaches', *Neural Computing and Applications*, vol. 34, no. 12, pp. 9777–9792.

Shakeri, J, Amini Khoshalan, H, Dehghani, H, Bascompta, M & Onyelowe, K 2022b, 'Developing new models for flyrock distance assessment in open-pit mines', *Journal of Mining and Environment*, vol. 13, no. 2, pp. 375–389.

Shirani Faradonbeh, R & Taheri, A 2019, 'Long-term prediction of rockburst hazard in deep underground openings using three robust data mining techniques', *Engineering with Computers*, vol. 35, no. 2, pp. 659–675.

Shirani Faradonbeh, R, Ghiffari Ryoza, M, Jang, H & Topal, E 2023, 'Practical Models to distinguish between seismic events and blast signals', *5th International Underground Excavations Symposium and Exhibition*, Istanbul.

Shirani Faradonbeh, R, Vaisey, W, Sharifzadeh, M & Zhou, J 2024, 'Hybridized intelligent multi-class classifiers for rockburst risk assessment in deep underground mines', *Neural Computing and Applications*, vol. 36, no. 4, pp. 1681–1698.

Song, C, Zhao, T, Xu, L & Huang, X 2024, 'Probabilistic prediction of uniaxial compressive strength for rocks from sparse data using Bayesian Gaussian process regression with synthetic minority oversampling technique (SMOTE)', *Computers and Geotechnics*, vol. 165.

Sorabi, P, Ataei, M, Jazi, MRA, Dehghani, H, Shakeri, J & Habibi, MH 2024, 'Utilizing heuristic strategies for predicting the backbreak occurrences in open-pit mines, Gol Gohar Mine, Iran', *Soft Computing*, pp. 1–16.

Wang, J, Liu, P, Ma, L, He, M & Xiong, H 2021, 'A rockburst proneness evaluation method based on multidimensional cloud model improved by control variable method and rockburst database', *Lithosphere*, vol. 2021, special issue 4.

Wang, M, Ye, XW, Jia, JD, Ying, XH, Ding, Y, Zhang, D & Sun, F 2024, 'Confining pressure forecasting of shield tunnel lining based on GRU model and RNN model', *Sensors*, vol. 24, no. 3.

Xu, C, Nait Amar, M, Ghriga, MA, Ouaer, H, Zhang, X & Hasanipanah, M 2022, 'Evolving support vector regression using Grey Wolf optimization; forecasting the geomechanical properties of rock', *Engineering with Computers*, vol. 38, no. 2, pp. 1818–1833.

Xu, Y, Klein, B, Li, G & Gopaluni, B 2023, 'Evaluation of logistic regression and support vector machine approaches for XRF based particle sorting for a copper ore', *Minerals Engineering*, vol. 192.

Yari, M, Armaghani, DJ, Maraveas, C, Ejlali, AN, Mohamad, ET & Asteris, PG 2023, 'Several tree-based solutions for predicting flyrock distance due to mine blasting', *Applied Sciences*, vol. 13, no. 3, https://doi.org/10.3390/app13031345

Yin, X, Liu, Q, Pan, Y, Huang, X, Wu, J & Wang, X 2021, 'Strength of stacking technique of ensemble learning in rockburst prediction with imbalanced data: comparison of eight single and ensemble models', *Natural Resources Research*, vol. 30, pp. 1795–1815.

Zhang, J, Ma, G, Huang, Y, Aslani, F & Nener, B 2019, 'Modelling uniaxial compressive strength of lightweight self-compacting concrete using random forest regression', *Construction and Building Materials*, vol. 210, pp. 713–719.

Zhou, J, Huang, S & Qiu, Y 2022, 'Optimization of random forest through the use of MVO, GWO and MFO in evaluating the stability of underground entry-type excavations', *Tunnelling and Underground Space Technology*, vol. 124.