Available online at www.sciencedirect.com

ScienceDirect

Underground Space 17 (2024) 226–245

www.keaipublishing.com/undsp

Research Paper

Imbalanced rock burst assessment using variational autoencoder-enhanced gradient boosting algorithms and explainability

Shan Lin [a], Zenglong Liang [a], Miao Dong [a], Hongwei Guo [a,b,*], Hong Zheng [a]

[a] *Key Laboratory of Urban Security and Disaster Engineering, Ministry of Education, Beijing University of Technology, Beijing 100124, China*
[b] *Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University (PolyU), Hung Hom, Kowloon 999077, Hong Kong, China*

Received 7 October 2023; received in revised form 18 November 2023; accepted 21 November 2023
Available online 14 January 2024

Abstract

We conducted a study to evaluate the potential and robustness of gradient boosting algorithms in rock burst assessment, established a variational autoencoder (VAE) to address the imbalance rock burst dataset, and proposed a multilevel explainable artificial intelligence (XAI) tailored for tree-based ensemble learning. We collected 537 data from real-world rock burst records and selected four critical features contributing to rock burst occurrences. Initially, we employed data visualization to gain insight into the data's structure and performed correlation analysis to explore the data distribution and feature relationships. Then, we set up a VAE model to generate samples for the minority class due to the imbalanced class distribution. In conjunction with the VAE, we compared and evaluated six state-of-the-art ensemble models, including gradient boosting algorithms and the classical logistic regression model, for rock burst prediction. The results indicated that gradient boosting algorithms outperformed the classical single models, and the VAE-classifier outperformed the original classifier, with the VAE-NGBoost model yielding the most favorable results. Compared to other resampling methods combined with NGBoost for imbalanced datasets, such as synthetic minority oversampling technique (SMOTE), SMOTE-edited nearest neighbours (SMOTE-ENN), and SMOTE-tomek links (SMOTE-Tomek), the VAE-NGBoost model yielded the best performance. Finally, we developed a multilevel XAI model using feature sensitivity analysis, Tree Shapley Additive exPlanations (Tree SHAP), and Anchor to provide an in-depth exploration of the decision-making mechanics of VAE-NGBoost, further enhancing the accountability of tree-based ensemble models in predicting rock burst occurrences.

Keywords: Gradient boosting; VAE; Ensemble learning; Explainable artificial intelligence (XAI); Rock burst

1 Introduction

Rock engineering (Wagner, 2019) is an engineering field that studies and applies the principles of rock mechanics and rock engineering mechanics. Rock engineering has a wide range of applications, including underground tunnels, underground space development, dams, rock slopes, rock foundations, mining, and other engineering projects (Lee & Jeong, 2016; Wagner, 2019; Xing et al., 2018).

With the emergence of various new technologies and the growing demand for energy, resources and space, various rock engineering projects are being constructed on a larger scale, at greater depths and in a more complex underground world. The increase in underground construction leads to an increase in geologic hazards, in which rock burst is the most frequent one(Waqar et al., 2023).

Rock burst is a disaster in which the stored elastic strain energy is suddenly released under high geostress conditions after excavation because the energy inside the rock is much larger than the restraining force provided by the surrounding rock (Yao & He, 2008). The earliest rock burst recorded in China can date back to 1933, which occurred

* Corresponding author at: Key Laboratory of Urban Security and Disaster Engineering, Ministry of Education, Beijing University of Technology, Beijing 100124, China.
E-mail addresses: ghway0723@gmail.com, hw-cee.guo@polyu.edu.hk (H. Guo).
Peer review under the responsibility of Tongji University.

https://doi.org/10.1016/j.undsp.2023.11.008

at Fushun Shengli Coal Mine (He & Wang, 2023). Many rock burst events occurred during the establishment of the Jinping-II hydropower station (Feng et al., 2015), including a powerful rock burst event and the maximum depth of destruction of the surrounding rock up to 3 meters, resulting in severe damage (He et al., 2012). Moreover, in the long term, it is estimated that coal will comprise 55% of China's energy consumption by 2023, with nearly 10% of that coming from coal mines enduring rock burst (Pan & Wang, 2023). Therefore, predicting and preventing rock burst holds significant importance for ensuring the safety and sustainable development of deep underground coal mining in China.

From the energy perspective, the cause of rock burst seems to be very simple, but from the practical engineering perspective, because we can't directly measure the energy stored inside the rock and the surrounding rock on its constraints on the energy, we can only predict by a limited number of quantitatively expressed characteristics of the conditions, which makes predicting rock burst very difficult (Wang et al., 2020). In the early research on the prediction of rock burst occurrences, researchers use the fuzzy comprehension evaluation method (Du et al., 2006; Xu et al., 2008) and rough set theory (Zhang et al., 2010) of modeling and other mathematical methods to measure the relationship between the selected features and rock burst occurrences and to achieve the purpose of prediction.

As computer science rapidly evolves, machine learning is increasingly applied in engineering practice (Liu & Lu, 2022; B. Liu et al., 2021, 2022a, 2022b;). Machine learning (ML) (Topuz & Alp, 2023) is a subfield of artificial intelligence (AI) that aims to enable computer systems to learn and improve from data and make decisions accordingly without having to consider the original meaning of the data. Since the data in it does not need to be considered for its actual meaning during computational analysis, ML can be applied to various fields (B. Liu et al., 2023a, 2023b; H. Liu et al., 2023; Xia et al., 2023; Zhu et al., 2023; Zhuang & Zhou, 2019).

In the problem of predicting rock burst, many researchers have obtained better results by building ML models. Based on the analysis of main cause of rock burst, Zhao (2005b) proposed the evolutionary support vector machine (SVM) to predict rock burst. The results showed that this method was feasible and appropriate with significant potential. Ahmad et al. (2022) used a newly developed model to predict rock burst intensity grade using adaptive boosting (AdaBoost) classifier and found that the prediction results are consistent with the actual conditions of the subsequent construction.

Ensemble learning (Dong et al., 2020) is an ML method that improves the performance of a model by combining multiple weak learners into one strong learner. Its basic idea is to get more accurate and stable predictions by combining the predictions of multiple weak learners. Due to the excellent performance of ensemble learning in engineering, many researchers have used ensemble learning methods to predict rock burst and have taken good results. Liang et al. (2021) used an ensemble learning approach for rock burst risk prediction and showed that random forest (RF) and gradient boosting decision tree (GBDT) have better overall performance. Li et al. (2022) constructed an ensemble learning model consisting of multiple classifiers by establishing multiple ML models and integration ideas. The results showed that the robustness of the ensemble model is significantly better than that of the ordinary model, and the practicality of the model is verified in the Sanshandao gold mine. H. Liu et al. (2023) used the histogram gradient boosted tree (HGBT) to model rock burst prediction and verified the feasibility of the model through Sanshandao Gold Mine.

Deep learning (LeCun et al., 2015; Schmidhuber, 2015) is a special method in ML, which uses neural network models to simulate the neural network structure of the human brain, and learns and predicts through the connection and weight adjustment of multi-layer neurons. In predicting rock burst problems, many researchers have obtained better results using neural network models. Li et al. (2005) proposed a neural network model for predicting rock burst, and the prediction results showed that it is feasible and effective to predict rock burst using artificial neural networks. Zhao and Chen (2020) established a data-driven model based on the convolutional neural network (CNN) and compared it with the traditional neural network. The results showed that the data-driven model can effectively tap the complex phenomena and mechanisms of rock bursts occurrences.

Rock burst brings great harm to mines, tunnels, underground engineering, and other rock projects. Solving the rockburst prediction problem can not only ensure the safety of the staff but also improve the stability of the project so that the whole project is more secure and stable. However, rockburst data are highly imbalanced, and in the case of imbalanced data categories, the model may be more inclined to predict the categories with a higher frequency of occurrence and has poorer prediction of the categories with a lower frequency of occurrence, which seriously reduces the performance of the model in dealing with a small number of categories. Therefore, in order to better solve the rockburst prediction problem, solving the data imbalance problem is of primary importance.

Data imbalance is a situation that is encountered in classification problems, where there is a significant difference in the number of samples from different categories, which occurs more frequently in areas such as medical (Khushi et al., 2021; Tasci et al., 2022) and credit card fraud (Mienye & Sun, 2023; Puh & Brkic, 2019). The impact of imbalanced data on ML models is very significant, which may lead to degradation of model performance and bias in prediction results. When the data classification is severely imbalanced, the model tends to be more inclined or even completely to predict the majority class. Therefore, it is crucial to change the imbalanced data into balanced data by generating minority class samples. Generative

modeling in deep learning is an excellent solution to this problem. In deep learning, generative adversarial networks (GAN) (Goodfellow et al., 2014), variational autoEncoder (VAE) (Kingma & Welling, 2013), etc., are the models used to generate new samples with similar characteristics by learning the distribution probabilities of the data. In addition to this, there are some sampling-based methods, such as SMOTE (Chawla et al., 2002) which can also be implemented to generate new samples.

As a generative model, VAE has many applications in text generation, data enhancement, and so on (Liu & Liu, 2019; Y. Wang et al., 2023). In dealing with the problem of data imbalance, many researchers have also addressed this problem by generating a few class samples through VAE. Mirza et al. (2021) compared various methods that rely on generative modeling to deal with imbalanced data, such as GAN and VAE, and their results showed that VAE yields higher Precision and F1-score. Y. Liu et al. (2022) synthesized the advantages of the conditional variational Autoencoder (CVAE) and GAN and established the CVAEGAN-SM model to augment the imbalanced dataset. Zhang and Liu (2022) established a model based on improved conditional variational autoencoder (ICVAE) and borderline synthetic minority (BSM) oversampling techniques for the ICVAE-BSM model. Their results show that the method is more effective in improving the accuracy in the case of data imbalance. Wang et al. (2023) combined the advantages of VAE and GAN to construct the VAE-GAN model, which solved the serious data imbalance problem that occurred in the original system and significantly improved the performance of the original system.

In summary, this paper addresses three main problems: (1) the imbalance problem of rock burst data, (2) the rock burst prediction problem, and (3) the problem that the ensemble learning model is complex to explain. Therefore, the main work of this paper can be divided into four parts, as shown in Fig. 1: (1) constructing the rock burst dataset and analyzing its data structure; (2) using the VAE generative model to generate a few classes of samples to solve the imbalance problem of rock burst data, and building the latest ensemble learning model and the baseline model logistic regression; (3) comprehensively evaluating the established model from the perspectives of both generative data methodology and the classifier; (4) establishing an explainable framework to explain the best-performing model in model evaluation.

2 Data analysis

In this section, we will visualize and analyze the data using Missingno Library, split violin diagrams, heatmap, histogram, and scatter plots. Missingno Library can help us understand the missingness of each feature in the original data, and split violin diagrams can enable us to understand the distribution of each feature in different categories of data. Heatmap, histogram, and scatter plots can help us understand the relationship between features and data distribution in more detail.

2.1 Data source

Five hundred and thirty-seven real data of rock burst occurrences were collected from open literatures (Ahmad et al., 2022; Cai et al., 2001; Du et al., 2006; Faradonbeh et al., 2022; Gong & Li, 2007; Guo et al., 2022; Hao et al., 2016; Jia, 2014; Jiang, 2008; Li et al., 2022; Li et al., 2017; R. Liu et al., 2019; Pu et al., 2019; Shi et al., 2010; Tian, 2021; J. Wang et al., 2022; J. Wang et al., 2023; Xia, 2007; Xu et al., 2008; Xue et al., 2020a, 2020b, 2019; Zhang et al., 2010, 2020, 2011; Zhao, 2005a; Zhao & Chen, 2020; Zhou et al., 2021, 2020, 2016), among which there are eight mainly used characteristics, which are maximum tangential stress of surrounding rock τ_{max}, uniaxial compressive strength σ_C, uniaxial tensile strength σ_t, elastic strain energy index W_{et}, depth of burial D, shear strength coefficient C (the ratio of the maximum tangential stress to the uniaxial compressive strength as Eq. (1)), and two kinds of brittleness indices B_1 (the ratio of the uniaxial compressive strength to the uniaxial tensile strength as Eq. (2)), B_2 (the ratio of the difference between the uniaxial compressive and tensile strengths to the sum of the two as Eq. (3)). These features describe the depth of the formation in which the rock is located, the strength of the rock, and the forces applied, and are widely used to predict rock burst.

$$C = \frac{\tau_{max}}{\sigma_C} \tag{1}$$

$$B_1 = \frac{\sigma_C}{\sigma_t} \tag{2}$$

$$B_2 = \frac{\sigma_C - \sigma_t}{\sigma_C + \sigma_t} \tag{3}$$

Since the data are collected from different literatures recording real-world rock burst cases, a large amount of missing values (about 49%) is observed in data related to feature D as Fig. 2. According to the "80% rule" (Bijlsma et al., 2006) missing values that exist in more than 20% of samples may be removed from the data, not to mention 49% of missing values existing for D. Of course, we can treat the missing values with resort to different data imputation techniques. However, high instances of missing values severely limit the model performance and create high uncertainty around the estimated values replacing the missing data. Therefore, feature D is not retained. On the other hand, features B_1, B_2 and C are computed from τ_{max}, σ_C, and σ_t, but in ML, the features should be as uncorrelated as possible, so they are also not retained. Therefore, we finally use these four features τ_{max}, σ_C, σ_t and W_{et}.

In denoising, we first tried various denoising methods under python outlier detection (PyOD) (Zhao et al., 2019), such as statistically-based denoising, clustering-based local outlier factor (LOF) denoising, and so on. However, the models built using the denoised data are far inferior to the models built before denoising.

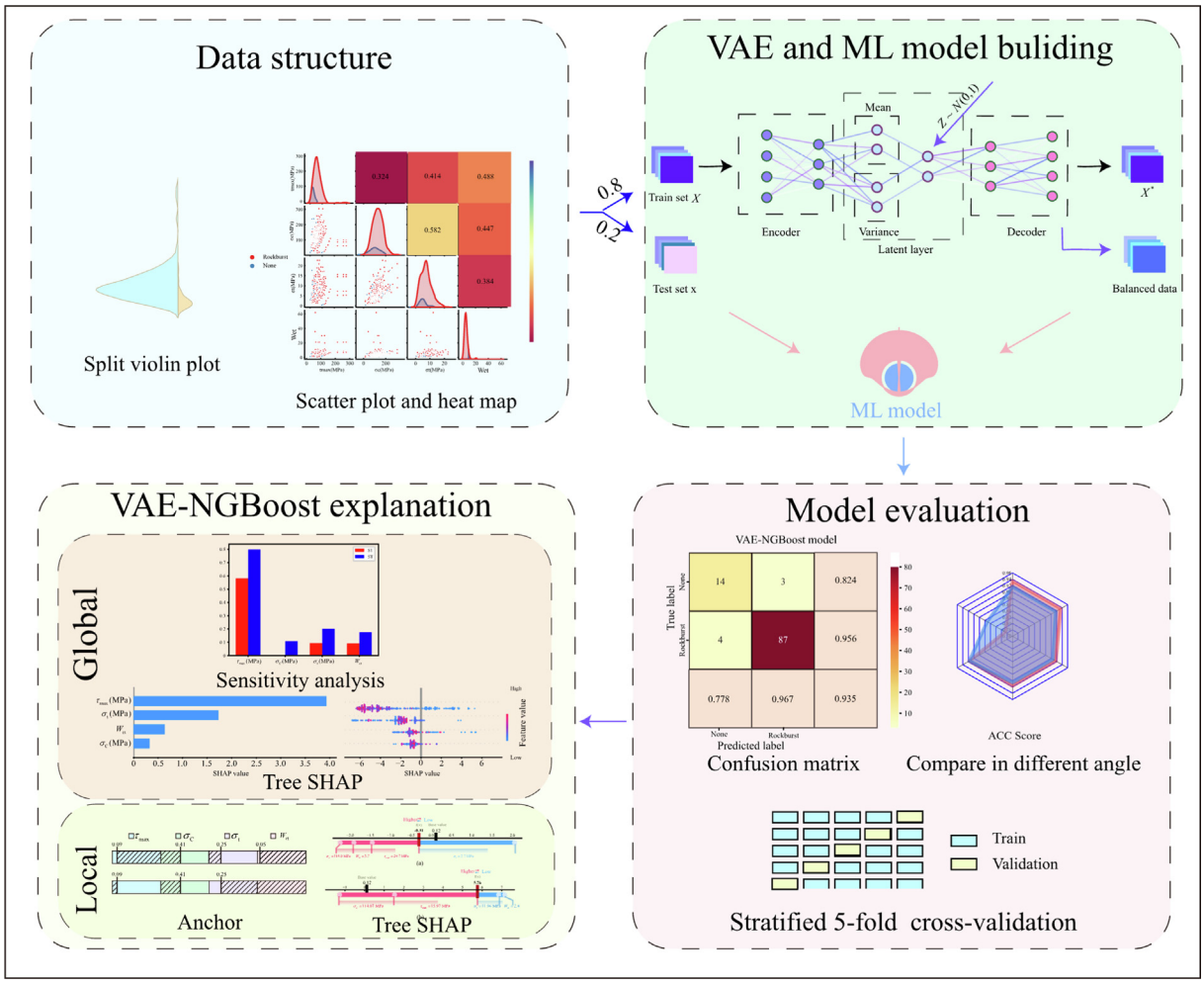

Fig. 1. Schematic graph of data flow in the proposed explainable ensemble learning model.

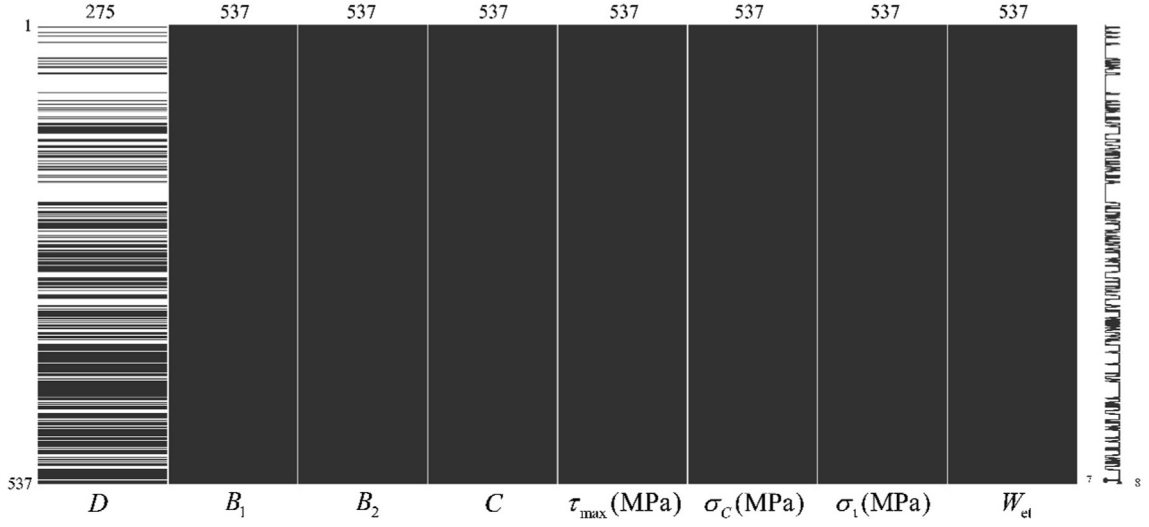

Fig. 2. Visualization of missing values.

This indicates that the real data collected are very reliable, so we do not take denoising measures on the data. In summary, we built a dataset containing 537 four-feature rock burst binary classifications, of which 79 rock burst did not occur and 458 rock burst cases occurred, with a ratio of 1:5.8, which is a seriously imbalanced dataset.

2.2 Data structure

Figure 3 is the violin diagrams of the dataset. This figure clearly demonstrates the distribution of each feature and the imbalance of class. Feature τ_{max} is concentrated from 30.9 to 75.0 MPa and its mean value is 59.16 MPa; σ_C is concentrated from 91.3 to 158.0 MPa, and its mean value is 126.72; σ_t is concentrated from 4.40 to 9.4 and its mean value is 7.39; W_{et} is concentrated from 3.14 to 6.27 and its mean value is 5.17. From this figure, we can also see the imbalance of the data and the differences in the distribution of the various features of the different types of rock burst. The distribution of the features in the non-rockburst data is skewed towards lower values compared to the rockburst data. In particular, the feature τ_{max} is higher than 50 MPa in only a small fraction of the non-rockburst data.

2.3 Feature correlation analysis

Feature correlation analysis is a common method to understand the correlation between features, and the correlation coefficient allows us to evaluate the reasonableness of the selected features. Since the Spearman coefficient is applicable to data with nonlinear relationship and non-normal distribution, it is used in this paper to measure the correlation between features. Figure 4 shows the Spearman coefficient of τ_{max}, σ_C, σ_t and W_{et}. In this figure, the correlations for these four selected features are relatively low, with the highest correlation coefficient only 0.582. Therefore, it can be assumed that the correlation between these features is very weak and meets the requirements for feature selection in ML. Additionally, in the scatter plot in Fig. 4, we further list the data's distribution between each feature, and clearly there is a large variance in distribution and the magnitude can be observed. Therefore, it is necessary to conduct preprocessing data in the initial stage.

3 Methodology

Data imbalance tends to severely degrade the performance of classifiers, and obtaining of real data is expensive, so employing a mathematical approach to data augmentation has become one of the most popular methods in ML. Combining the excellent performance of ensemble learning on geotechnical engineering problems and the effective revelation of uncertainties are often faced by geotechnical engineers and researchers. Therefore, this chapter focuses on the two parts: data augmentation and ensemble learning.

3.1 VAE data augmention

VAE (Kingma & Welling, 2013) is a generative model that utilizes neural network training to obtain two functions: inferential network and generative network. Figure 5 shows the process of VAE training and generating data. The training process is the most important part of VAE, and like unsupervised learning such as AE, and it is important to ensure that the difference between the output data and the original data is relatively small, i.e., the reconstruc-

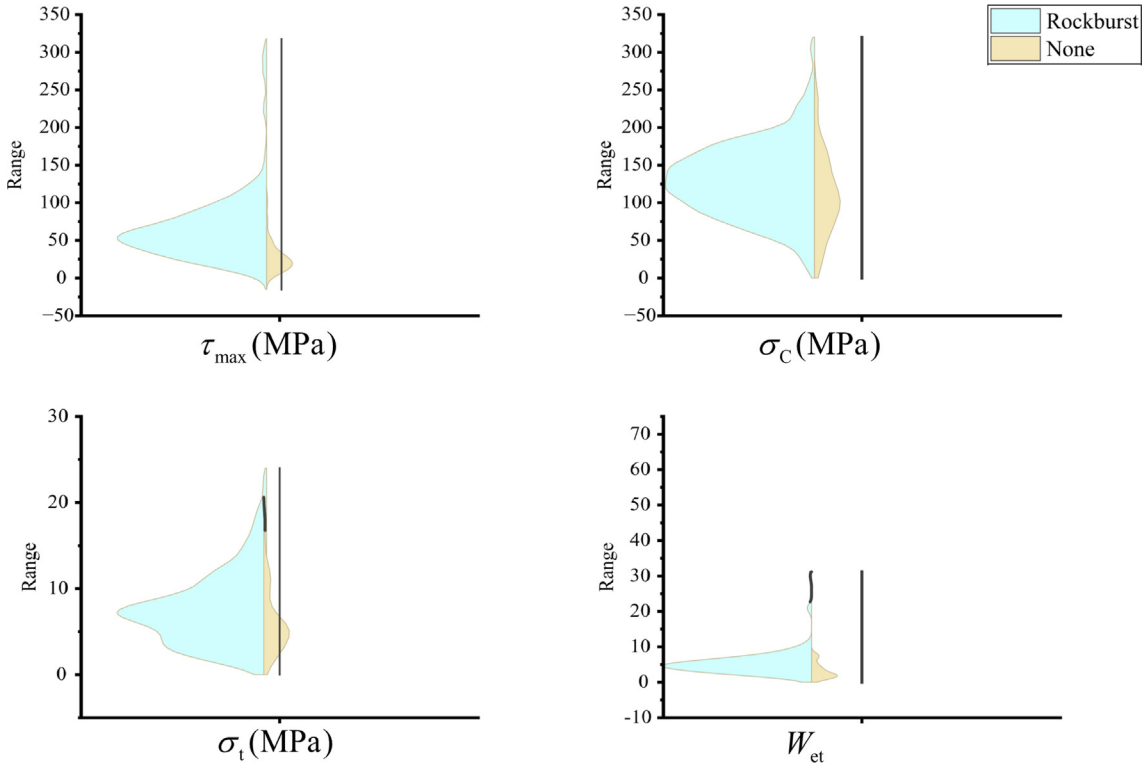

Fig. 3. Split violin plots of the four selected input features.

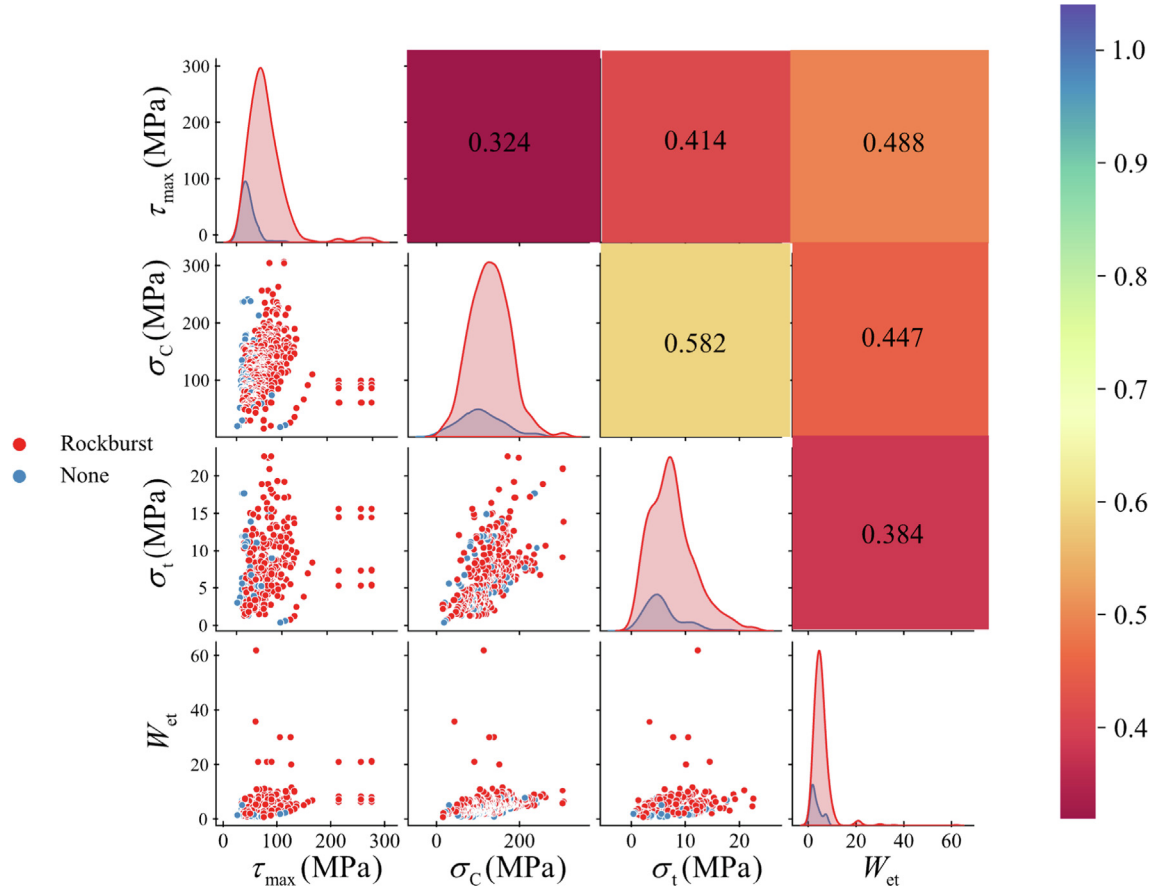

Fig. 4. Scatter plot matrix with coefficient heatmap, histogram, and pairwise scatter plots.

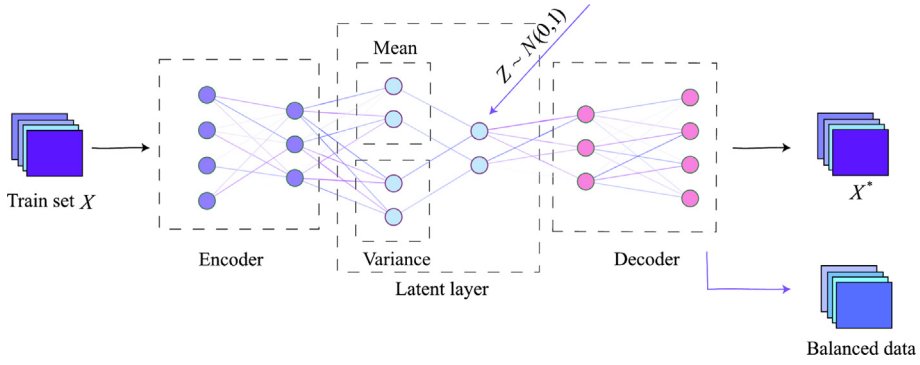

Fig. 5. Workflow of VAE generating minority class data.

tion error is small. In addition to this, unlike AE, the distribution of the latent layer in VAE is made close to $N \sim (0, 1)$ by narrowing the kullback leibler (KL) scatter from the standard normal distribution.

Figure 6 shows the digital training process of VAE, in the order that they are needed, which can be divided into three parts: KL scattering error, reconstruction error, and reparameter process. First, define the following posterior distribution probability of: $p_\theta(z|x)$, encoder: $q_\varphi(z|x)$, decoder: $p_\theta(x|z)$, original data: $\boldsymbol{X} = \left[x^{(1)}, x^{(2)}, \cdots, x^{(n)}\right]$, and scatter between the distributions of A and B: $D_{KL}(A\|B)$.

(1) KL scattering error between $q_\varphi(z|x)$ and $p(z|x)$ can be calculated by Eq. (4):

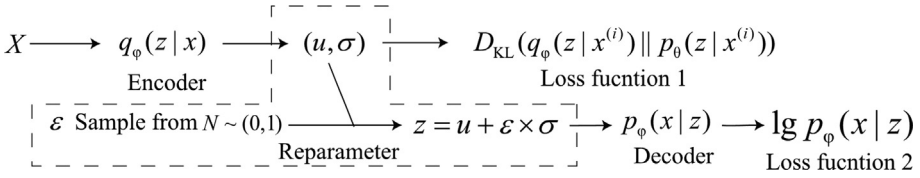

Fig. 6. Digital workflow of training VAE.

$$D_{KL}\left(q_\varphi\left(z|x^{(i)}\right)\|p_\theta\left(z|x^{(i)}\right)\right)$$
$$= E_{z\sim q}\left[\lg\frac{q_\varphi\left(z|x^{(i)}\right)}{p_\theta\left(z|x^{(i)}\right)}\right]$$
$$= E_{z\sim q}\left[\lg q_\varphi\left(z|x^{(i)}\right) - \lg p_\theta\left(z,x^{(i)}\right)\right] + \lg p\left(x^{(i)}\right), \quad (4)$$

where x are known, $\lg p\left(x^{(i)}\right)$ is a constant, therefore D_{KL} gets smaller when $-E_{z\sim q}\left[\lg q_\varphi\left(z|x^{(i)}\right) + \lg p_\theta(z,x)\right]$ gets bigger, where $-E_{z\sim q}\left[\lg q_\varphi\left(z|x^{(i)}\right) + \lg p_\theta(z,x)\right]$ is known as Evidence lower bound (Elbo), $-$Elbo is the first loss function (L_1).

(2) Combining probabilistic modeling ideas, reconstruction error can be expressed in terms of the maximum likelihood estimation (MLE) loss function as Eq. (5):

$$\text{Reconstruction error} = -\frac{1}{T}\sum_{t=1}^{T}\lg p_\theta\left(x^{(i)}|z^{(i,t)}\right), \quad (5)$$

which is the second loss function (L_2), where T is the number of times $z^{(i)}$ sampled.

(3) The reparameterization trick is a solution to the inability of the VAE reconstruction error gradient to backpropagate, and the core principle of which is to get the display mapping (φ) of $y\sim N(u,\sigma^2)$ through $\epsilon\sim N(0,1)$ and Eqs. (6)–(7), so as to compute the expectation of the gradient and backpropagate it.

$$y = u(x) + \sigma(x)\times\epsilon \quad (6)$$
$$\varphi = (u(x),\sigma(x)) \quad (7)$$

In summary, when assuming that $p(z|x)\sim N(0,1)$, the objective function is shown as Eq. (8):

$$L = \alpha L_1 + L_2$$
$$= -\alpha\frac{1}{2}\sum_{J=1}^{J}\left(1+\lg\left(\left(\left(\sigma_j\right)^2 - \left(u_j\right)^2 - \left(\sigma_j\right)^2\right)\right)\right) - \frac{1}{L}$$
$$\times\sum_{l=1}^{L}\lg p_\theta\left(x^{(i)}|z^{(i,l)}\right), \quad (8)$$

where α can be viewed as a regularization factor, and j is the dimensions of z.

From the above training process, L_1 ensures that the hidden layer distribution is close to $N\sim(0,1)$ and L_2 ensures that the reconstructed data are close to the original data. Therefore, generating new data only requires sampling from $N\sim(0,1)$ and putting it into decoding, at which time the new data are most likely to match the original data.

When dealing with imbalanced data, suppose we have a total of M samples, among which there are T minority class samples, the whole processing process can be divided into four steps. (1) Extract the minority class samples from the training set, and then use these to train a VAE network to recognize the features of the minority class samples. (2) Sample $M-2T$ data from $N\sim(0,1)$ as hidden variables. (3) Input the hidden variables into the generative network, so as to generate $M-2T$ minority class samples. (4) Merge with the original dataset and the ratio of samples from the two categories is $M-2T:M-2T$, with category balanced.

3.2 Ensemble machine learning

3.2.1 NGBoost

NGBoost is an ensemble learning method for prediction purposes by approximating the predictive probability distribution, presented by Stanford University researchers (Duan et al., 2019). Figure 7 shows the overall process of NGBoost. NGBoost doesn't use point predictions $E(y|x)$ to solve the problem, but instead uses the conditional probability distribution $P(y|x)$ of the point predictions, which is unique in ML. Besides this, the natural gradient boosting method reduces the magnitude of parameter updates and the risk of the model overfitting to the training data.

NGBoost is flexible, compatible, and easy-to-use, especially on small datasets, which offers competitive performance. In this subsection, we focus on the property scoring rule and the natural gradient.

(1) The proper scoring rule $S(,)$ in natural gradient is a method for assessing how the predicted parameterized distribution H_θ matches the observations y, which needs to fulfill Eq. (9):

$$E_{y\sim Q}[S(Q,y)] \le E_{y\sim H_\theta}[S(H_\theta,y)], \quad (9)$$

where Q is the real distribution of observation y.

Same as VAE, logarithmic score L ($-$MLE) is also used in NGBoost. This score can be expressed as Eq. (10), and a larger value of $S(H_\theta,y)$ indicates a larger difference between H_θ and y.

$$S(H_\theta,y) = -\lg H_\theta(y). \quad (10)$$

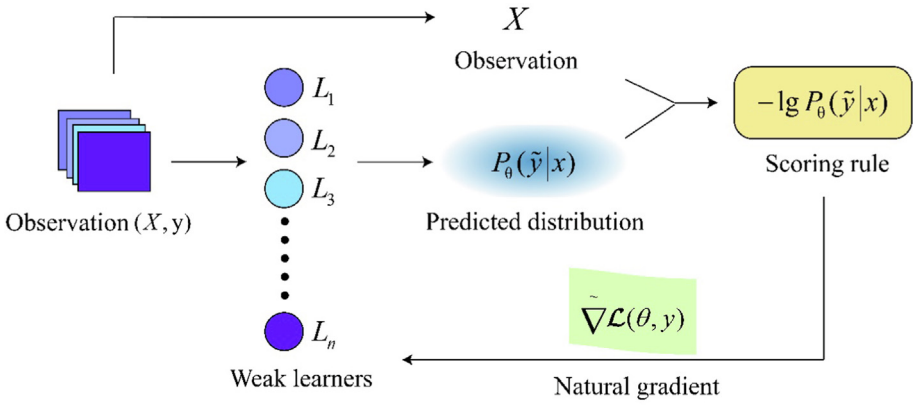

Fig. 7. Diagram of NGBoost algorithm.

The scatter induced by the −MLE can be derived from Eq. (11) as the D_{KL}.

$$S(H_\theta, y) - S(Q, y) = -\lg \frac{Q(y)}{H_\theta(y)} = D_{\mathrm{KL}}(Q(y) \| H_\theta(y)) \quad (11)$$

(2) The generalized natural gradient is the steepest path of ascent in Riemannian space, which addresses the difficulty of dealing with probabilistic predictions in traditional gradient boosting. Combined with the −MLE used in the proper scoring rule, the natural gradient $\widetilde{\nabla}$ can be calculated by Eq. (12):

$$\widetilde{\nabla L}(\theta, y) \propto I_L(\theta, y)^{-1} \nabla L(\theta, y), \quad (12)$$

where $I_L(\theta, y)$ is the Fisher information carried by the observations on $H_\theta(y)$.

3.2.2 GBDT framework classifier

Gradient boosting decision tree (GBDT) is an ensemble learning that trains multiple decision tree models serially and uses the prediction results of the previous tree to improve the prediction of the next tree. Figure 8 shows the framework of the GBDT algorithm, which performs well in dealing with classification and regression problems. Common models using GBDT as a framework include XGBoost, LightGBM, and CatBoost.

(1) XGBoost

XGBoost (Chen et al., 2016) implements the GBDT framework widely employed in engineering, offering improved speed, accuracy, and generalization performance compared to traditional GBDT.

In speed, XGBoost uses parallel computing, which allows multiple GPUs to be used for training simultaneously, thus speeding up the training of the model.

In accuracy, XGBoost improves the original GBDT's objective function by making a second-order approximation to the loss function, which makes the model more accurate in the training process. Its loss function can be represented by Eq. (13):

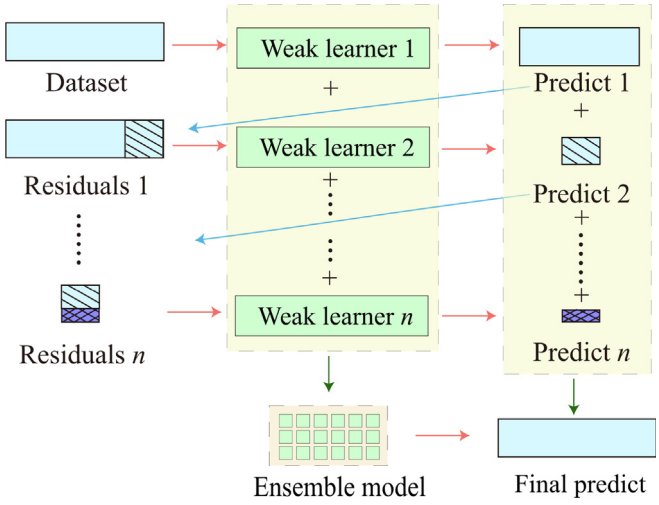

Fig. 8. Diagram of GBDT framework.

$$\mathrm{obj}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right], \quad (13)$$

where $g_i = \partial_{\hat{y}^{(t-1)}} D\left(y_i, \hat{y}^{(t-1)}\right)$, $h_i = \partial^2_{\hat{y}^{(t-1)}} D\left(y_i, \hat{y}^{(t-1)}\right)$, and $D(A, B)$ defines the difference between A and B.

In generalization performance, XGBoost introduces the regularization technique, which usually uses L_1 and L_2 regularization terms to control the complexity of the model and prevent overfitting. Therefore, the final objective function of the model can be represented by Eq. (14):

$$\mathrm{obj}^{(t)} = \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f), \quad (14)$$

where the regularization term $\Omega(f) = \gamma C + \frac{1}{2}\lambda \|\omega\|^2$ is supplemented in loss to reduce the model complexity and prevent overfitting. Here, C represents node numbers, ω is the corresponding node weight, and λ and γ are hyperparameters to balance C and ω.

(2) LightGBM

Although XGBoost demonstrates a significant improvement in computational speed compared to GBDT, it is still relatively slow when dealing with a large number of data, an issue LightGBM addresses. LightGBM employs a histogram-based splitting algorithm, specifically gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) (Ke et al., 2017). This method involves three main steps for selecting the splitting points: first, discretizing the continuous features; next, constructing a histogram based on the discretized values; and finally, choosing the optimal splitting points based on statistical information from the histogram.

GOSS is designed to reduce the utilization of data instances while preserving the accuracy of the learned models. Provided a dataset with n instances, the detailed steps for this algorithm are as follows:

1) Calculate the absolute value of the gradient and sort the data from high to low based on it, and then select one of the top-p $\times 100\%$ to assign to subset P.
2) Set up a dataset Q of size $q \times |P^c|$ from $|P^c|$ by random sampling, where $|P^c|$ is the not selected data in step1.
3) Finally the data is partitioned through $\widetilde{V}_J(d)$, where $\widetilde{V}_J(d)$ is the estimated variance gain of feature j at point d on the dataset $P \bigcup Q$.

$\widetilde{V}_J(d)$ can be calculated by Eq. (15):

$$\widetilde{V}_J(d) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in P_l} g_i + \frac{1-p}{q}\sum_{x_i \in Q_l} g_i\right)^2}{n_l^j(d)} + \frac{\left(\sum_{x_i \in P_r} g_i + \frac{1-p}{q}\sum_{x_i \in Q_r} g_i\right)^2}{n_r^j(d)}\right), \tag{15}$$

where P_l or $Q_l = [x_i \in P \text{ or } Q : x_{ij} \le d]$, P_r or $B_r = [x_i \in P \text{ or } Q : x_{ij} > d]$, and $\frac{1-p}{q}$ is used to normalize the sum of the gradients over Q back to the size of P^c.

The main objective of EFB is to reduce the computation time by reducing the feature dimensionality. EFB is composed of Greedy Bundling (GB) and Merge Exclusive Features (MEF) algorithm, where GB addresses how to determine which features can be merged, and MEF addresses how to merge these features together.

(3) CatBoost

CatBoost is another machine learning algorithm that utilizes the GBDT framework but with the ability to handle categorical features and automatic feature scaling, composed by Categorical and Boosting. Compared to other machine learning algorithms that use the GBDT framework, its most significant advantage is its efficient handling of categorical features and the ability to prevent biased prediction (Zhao et al., 2023).

There are two main core algorithms added to CatBoost, ordered target statistics (Ordered TS) and Ordered boosting, where Ordered TS enhances the model generality on unseen datasets and Ordered boosting helps to prevent the biased prediction in gradient estimation (Zhang et al., 2013). Addressing these critical issues can reduce the overfitting of model, and improve the model performance (Prokhorenkova et al., 2018).

Ordered TS is an optimization for the problem of different distributions for the training and test sets resulting from Greedy TS (as Eq. (16)) (Barreca, 2001). Ordered TS can be expressed by Eq. (17). which uses all the observed history to compute its TS, i.e., take $D_K = [x_j : \sigma_j < \sigma_k]$ in Eq. (16) for a training example and $D_K = D$ a test one. D is the dataset and D_K is all the data already used in the dataset.

$$\hat{x}_k^i = \frac{\sum_{j=1}^{n} 1_{[x_j^i = x_k^i]} \cdot y_j + \alpha p}{\sum_{j=1}^{n} 1_{[x_j^i = x_k^i]} + \alpha}, \tag{16}$$

$$\hat{x}_k^i = \frac{\sum_{x_j \in D_k} 1_{[x_j^i = x_k^i]} y_j + \alpha p}{\sum_{x_j \in D_k} 1_{[x_j^i = x_k^i]} + \alpha}. \tag{17}$$

Ordered boosting on the other hand is an optimization strategy for the biases in gradient estimation that result from gradient boosting. Its core idea is to determine the penalty weights for misclassified samples based on their ordering relationships, and to better classify these misclassified samples by adjusting the parameters of the model.

3.3 Model building

The model established in this paper is data-driven, and its development process can be divided into three parts: data preprocessing, splitting, and hyperparameters tuning. Firstly, the data are normalized to eliminate the impact of scale differences. Secondly, the dataset is divided into a partition according to 6:2:2 as training set (321), validation set (108), and test set (108), where the validation set is used in the process of hyperparameter tuning. Finally since the modeling process involves the augmented data, the tuning strategy for each model is borrowed from the method in (Rodrigues et al., 2023), in which the parameters of the classifiers are set by the tuning of the balanced dataset.

The establishment process of the VAE-classifier can be divided into three parts. First, we trained a VAE network with the training set on the parameters of the VAE network, and we mainly adjusted the number of neurons in its connection layer and latent layer, and then we determined it by observing its loss function curves. Figure 9 shows its loss function curve. Then we use VAE to augment data to make the original data balanced. Finally, the balanced dataset is used to train a classifier model, where the parameters of the classifier model are obtained by a Bayesian optimizer.

In order to fully express the advantages of the model, not only advanced ensemble learning methods, but also

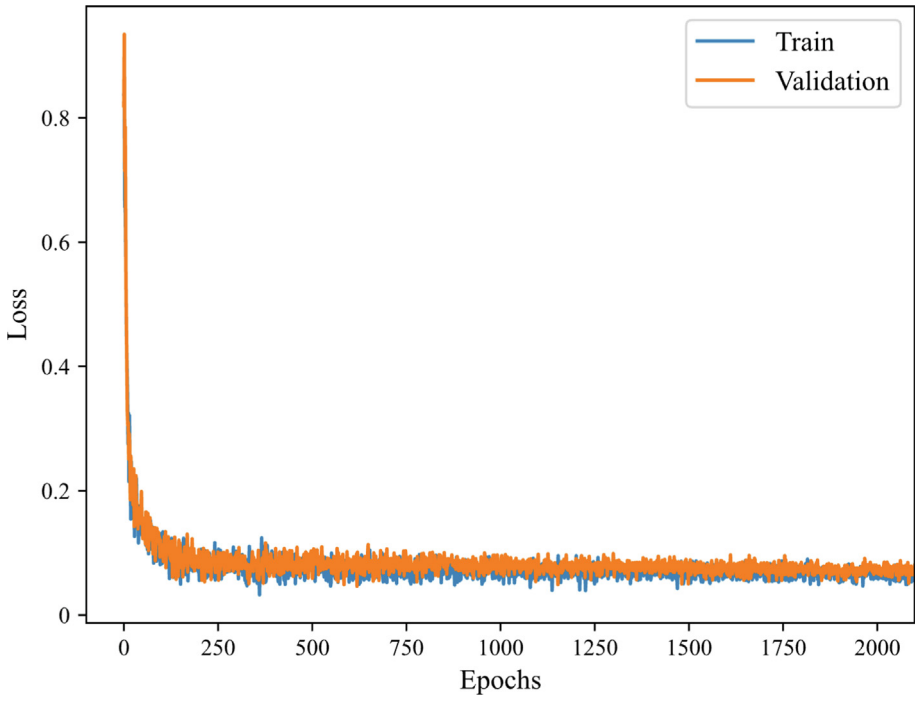

Fig. 9. Loss curve of the VAE training process.

the commonly used linear regression classifiers are used for the model comparison. In comparison with different data generation methods, we also use three sampling-based data generation methods, SMOTE, SMOTE-ENN, and SMOTE-Tomek, and model the corresponding balanced datasets using NGBoost. In summary, we have built a total of 15 models, such as VAE-NGBoost, VAE-CatBoost, VAE-XGBoost, VAE-LightGBM, VAE-RF, VAE-LR, SMOTE-NGBoost, SMOTE-ENN-NGBoost, SMOTE-Tomek-NGBoost, etc.

4 Model evaluation

Considering the data imbalance, we used the evaluation indicators in the confusion matrix that are specialized for imbalanced data (Susan & Kumar, 2020; Tao et al., 2024). In addition to that, we also use 5-fold stratified cross-validation to verify the generalization performance of the model.

4.1 Confusion matrix

Confusion matrix can help us visualize the performance of a classification model directly and thoroughly. Considering the redundancy of the images, we only give a representative comparison of the confusion matrix plots of the best in model evaluation. Figure 10 is a schematic representation of the confusion matrix for the VAE-NGBoost and NGBoost. The values around this figure can be calculated by following classification evaluation metrics from bottom left to top right, Precision (P), negative predictive rate (NPR), Accuracy (ACC), Specificity (S), and Recall (R). These evaluation indicators can be calculated by Eqs. (18)–(22).

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}}, \tag{18}$$

$$\mathrm{NPR} = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{19}$$

$$\mathrm{ACC} = \frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FP}+\mathrm{TN}+\mathrm{FN}}, \tag{20}$$

$$S = \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}}, \tag{21}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}}, \tag{22}$$

$$\text{F1-score} = \frac{2PR}{P+R}, \tag{23}$$

$$\text{G-mean} = \sqrt{R \times S}. \tag{24}$$

For imbalanced datasets, many researchers have explored the selection of their evaluation indexes, where ACC, F1-score, and G-mean are widely used. F1-score and G-mean can be calculated by Eqs. (23) and (24). In Table 1, the confusion matrix information and three indexes of those models are presented.

Figure 11 shows the three metrics of these 15 models from the data generation method and classifier perspective, where VAE-NGBoost performs best. From the standpoint of model improvement by VAE, the VAE model, compared with the original model, got a significant improvement in both its G-mean and F1-score, even though the ACC did not have a significant improvement, which sug-

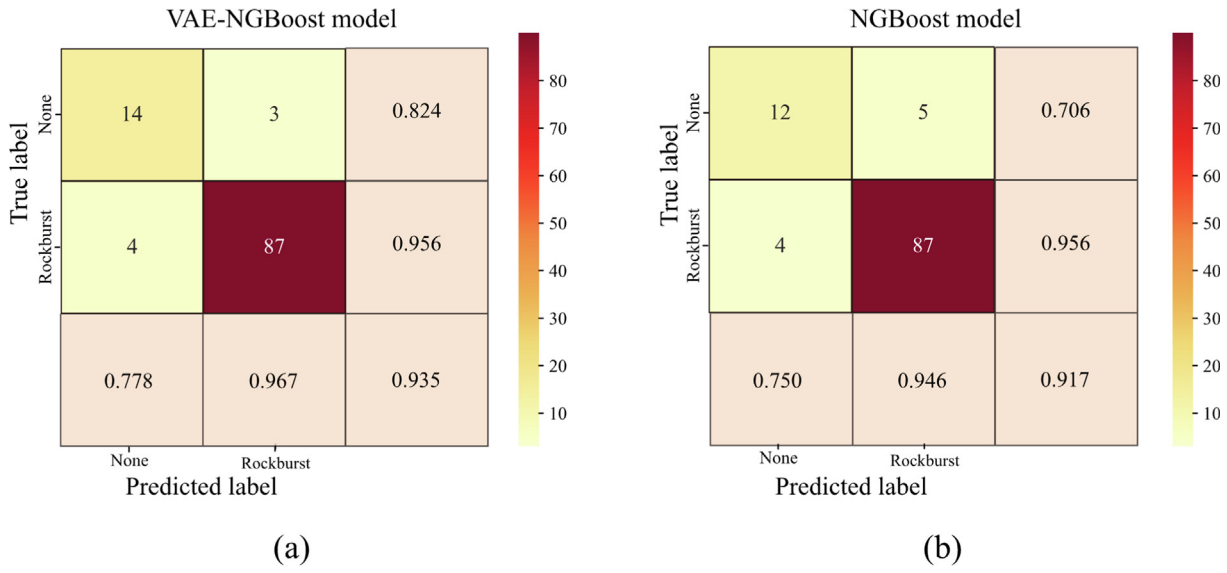

Fig. 10. Confusion matrix for (a) VAE-NGBoost, and (b) NGBoost.

Table 1
Confusion matrix for different machine learning models.

Model	True label	Predicted label		ACC	G-mean	F1-score
		None	Rock burst			
VAE-NGBoost	None	14	3	0.935	0.887	0.800
	Rock burst	4	87			
VAE-CatBoost	None	13	4	0.935	0.860	0.788
	Rock burst	3	88			
VAE-XGBoost	None	14	3	0.917	0.877	0.757
	Rock burst	6	85			
VAE-LightGBM	None	12	5	0.917	0.821	0.727
	Rock burst	4	87			
VAE-RF	None	12	5	0.917	0.821	0.727
	Rock burst	4	87			
VAE-LR	None	16	1	0.787	0.845	0.582
	Rock burst	22	69			
NGBoost	None	12	5	0.917	0.821	0.727
	Rock burst	4	87			
CatBoost	None	11	6	0.917	0.791	0.710
	Rock burst	3	88			
XGBoost	None	11	6	0.907	0.787	0.688
	Rock burst	4	87			
LightGBM	None	10	7	0.907	0.754	0.667
	Rock burst	3	88			
RF	None	11	6	0.917	0.791	0.710
	Rock burst	3	88			
LR	None	1	16	0.852	0.243	0.111
	Rock burst	0	91			
SMOTE-NGBoost	None	14	3	0.889	0.861	0.700
	Rock burst	9	82			
SMOTEENN-NGBoost	None	13	4	0.833	0.804	0.591
	Rock burst	14	77			
SMOTETomek-NGBoost	None	15	2	0.889	0.886	0.714
	Rock burst	10	81			

gests that the VAE augmented data can improve the model's ability to categorize the minor categories. Compared to other methods of augmenting data, the G-mean of these models are all improved compared to the original model in terms of improving the performance of NGBoost. However, the ACC of the models built by these three methods is severely reduced, and the ACC of the SMOTE-ENN model has the most severe reduction.

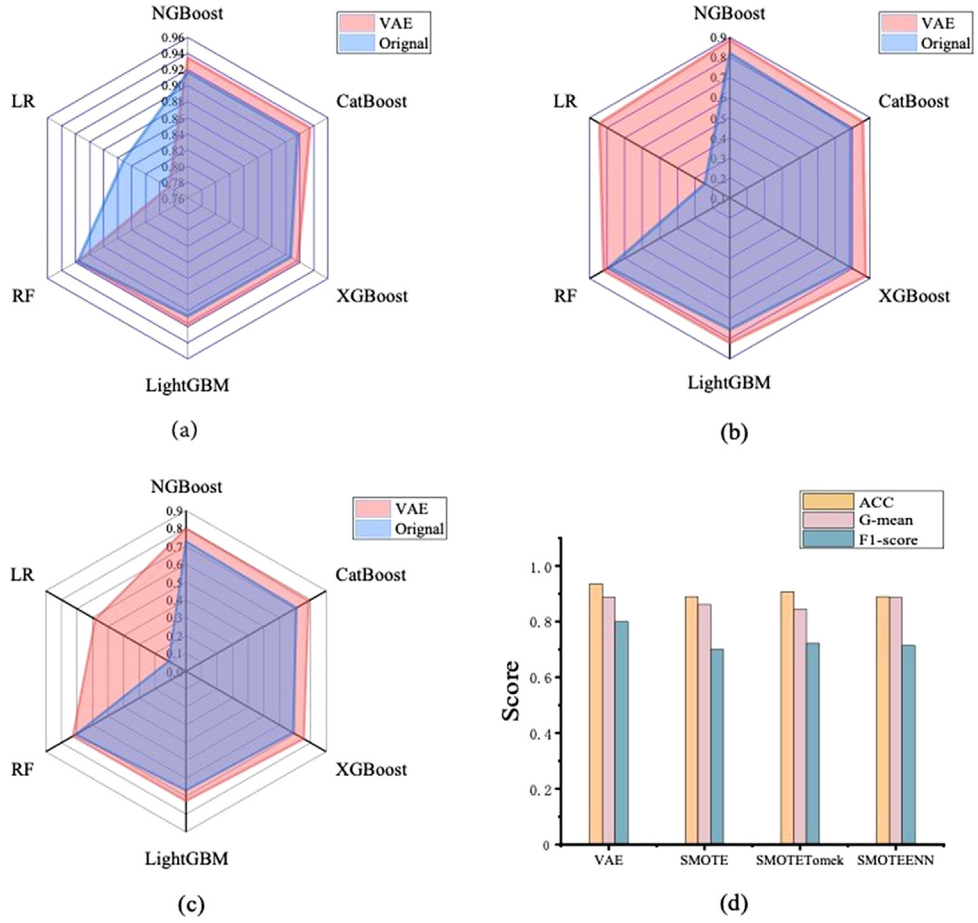

Fig. 11. Comparison of 15 models from different perspectives.(a) ACC score, (b) G-mean score,(c) F1-score, and (d) generative method perspective.

4.2 5-fold stratified cross-validation

In the validation of model generalization performance, we use ACC and G-mean of 5-fold stratified cross-validation to measure. Stratified cross-validation is a commonly used method for evaluating models of imbalanced datasets, which ensures that the distribution of categories in each point is similar to the distribution of categories in the entire training set, thus better assessing the performance of the model. Compared with the general stratified cross-validation, since this paper uses the method of generating data, in order to prevent the "spoofing" brought about by augmented data, we need to make the split dataset process before the generative process (Tholke et al., 2023). Tables 2 and 3 are the ACC and G-mean scores of stratified 5-fold cross-validation for the 15 models, respectively.

As you can see in both tables, the VAE-NGBoost has a significant increase in performance in all folds, except for a small decrease in performance in folds 3 and 4 compared with NGBoost. This can be due to the quality of the data. VAE generates a small amount of noise in the process of generating the data. When the selected data in a certain fold are of good quality, the noise brought by VAE may cause a marginal fluctuation in the model performance. However, in the fair evaluation, overall VAE-NGBoost outperforms NGBoost in cross-validation. To be specific, in the third and fourth folds, especially the fourth fold, when the data quality is already relatively sound, the noise is comparatively large, which will bring a negative impact on the model evaluation and reduce the performance of the model.

From all the models, the VAE-NGBoost model has the highest average value in both ACC and G-mean. In addition to this, VAE-RF also performs very well, with similar ACC and G-mean as VAE-NGBoost. From the angle of VAE, VAE-classifier is also generally rated higher than the other original models in terms of stratified 5-fold cross validation. From the angle of different generative methods compared with the best performance model VAE-NGBoost, the other data augmentation methods show a significant drop in ACC but a significant rise in recognition of minority classes compared to the original NGBoost, and there is no doubt that the performance is far worse than that of VAE.

In summary, the comprehensive comparison of these models demonstrates that the classical logistic regression model performs very poorly, with its G-mean only around

Table 2
ACC in 5-fold stratified cross-validation.

Model	ACC in 5-fold stratified cross-validation					Averaged ACC
	1	2	3	4	5	
VAE-NGBoost	0.930	0.930	0.895	0.907	0.941	0.921
VAE-CatBoost	0.942	0.930	0.895	0.907	0.918	0.918
VAE-XGBoost	0.953	0.919	0.907	0.860	0.953	0.918
VAE-LightGBM	0.965	0.895	0.884	0.884	0.941	0.914
VAE-RF	0.965	0.919	0.907	0.895	0.918	0.921
VAE-LR	0.814	0.767	0.733	0.767	0.741	0.765
NGBoost	0.884	0.907	0.919	0.919	0.929	0.911
CatBoost	0.919	0.907	0.930	0.907	0.929	0.918
XGBoost	0.953	0.884	0.907	0.884	0.941	0.914
LightGBM	0.942	0.907	0.919	0.860	0.918	0.909
RF	0.919	0.919	0.919	0.884	0.906	0.909
LR	0.884	0.895	0.884	0.849	0.882	0.879
SMOTE-NGBoost	0.884	0.826	0.872	0.791	0.882	0.851
SMOTEENN-NGBoost	0.826	0.756	0.802	0.802	0.906	0.818
SMOTETomek-NGBoost	0.895	0.837	0.872	0.791	0.882	0.856

Table 3
G-mean in 5-fold stratified cross-validation.

Model	G-mean in 5-fold stratified cross-validation					Averaged G-mean
	1	2	3	4	5	
VAE-NGBoost	0.925	0.925	0.763	0.675	0.811	0.819
VAE-CatBoost	0.894	0.805	0.803	0.768	0.702	0.795
VAE-XGBoost	0.938	0.842	0.846	0.474	0.816	0.783
VAE-LightGBM	0.907	0.788	0.797	0.665	0.764	0.784
VAE-RF	0.944	0.842	0.768	0.763	0.753	0.814
VAE-LR	0.885	0.760	0.716	0.735	0.709	0.761
NGBoost	0.688	0.641	0.774	0.815	0.759	0.735
CatBoost	0.753	0.748	0.779	0.809	0.759	0.770
XGBoost	0.860	0.688	0.768	0.665	0.764	0.749
LightGBM	0.764	0.748	0.774	0.656	0.645	0.717
RF	0.702	0.753	0.729	0.713	0.641	0.708
LR	0.408	0.500	0.480	0.539	0.408	0.467
SMOTE-NGBoost	0.897	0.753	0.861	0.626	0.782	0.784
SMOTEENN-NGBoost	0.862	0.787	0.849	0.631	0.794	0.785
SMOTETomek-NGBoost	0.830	0.799	0.861	0.711	0.737	0.788

0.5, influenced by the decision boundary bias generated by the imbalanced data. Moreover, in the ensemble learning model, compared to RF, the ensemble model based on the gradient boosting algorithm gradually optimizes the performance of the model through the continuous learning of the residuals, which leads to the model's performance being better. Relative to other data augmented methods, the VAE augmented data more closely match the original samples and with relatively less sample noise, which results in a higher performance of the model built in this way. Combining the advantages of VAE and ensemble learning, the VAE-NGBoost built in this paper performs significantly better than the other models.

5 Model explanation

Model explanation is the key to understanding and trusting the decision-making mechanism of the complex ML models. Currently, it seems that even though complex models can exhibit high performance, we still do consider using more transparent and simple models, such as linear regression. Therefore, model explanation is crucial for black-box models, which not only helps us adjust the decision-making process of ML models but also improves the models' reliability, robustness, and user acceptance. In this section, the model explanation is divided into two main parts: global analysis and local analysis. The following section focuses on these two parts to explain the VAE-NGBoost model, which has the best performance of the above models.

5.1 Global explanatory analysis

Global explanatory analysis of ML models refers to integrating and analyzing a large amount of data to reveal the relationship between individual features in the data and the model output as a whole. The combined use of multiple methods can provide more comprehensive and accurate

explanation and analysis results, helping us to better understand and apply ML models. Therefore, we use two methods, sensitivity analysis, and SHAP analysis, to interpret the VAE-NGBoost model.

5.1.1 Sobol sensitivity

Sensitivity analysis is a method to investigate how the output of a model responds to changes in its input. B. Liu et al. (2020), Liu et al. (2023c), and Zhou et al. (2020) used this method to gain insight into the relative importance of different input parameters to the model or system output. It serves not only to assess the stability of model outputs in case input data are inaccurate or fluctuating but also for feature selection, allowing us to identify and exclude features that have minimal impact on the output. Various methods have been proposed in the field of sensitivity analysis, with Sobol's method being the most widely employed in engineering (Yong et al., 2008).

Sobol sensitivity analysis is based on the principle of total variance decomposition, which decomposes the variance of the outputs into the contributions of individual input variables and their interactions. Specifically, Sobol sensitivity analysis assesses the significance of input variables by calculating the variance of their main and interaction effects, usually using the first-order, second-order, and total-order sensitivity indices for the analysis (Liu et al., 2020). First-order sensitivity reflects the direct effect of changes in a single feature on the model output, second-order sensitivity reflects the effect of interactions between two features on the model output, and total-order sensitivity reflects the degree to which a single feature has an overall effect on the model output, both directly and indirectly through interactions with other features. In the following, define the first-order sensitivity as $S1_i$, the second-order sensitivity as $S2_{i,t}$, and the total sensitivity as ST_i, where i, t are the i th and t th feature. From the introduction in Saltelli et al. (2008), the formula can be expressed by Eqs. (25)–(27), where $\mathrm{Var}(Y)$ is the variance of all numbers in Y.

In this section, we will employ the $S1_i$ and ST_i to explain the VAE-NGBoost model, where $S1_i$ and ST_i are calculated by Salib (Herman & Usher, 2017), which is a python library specifically designed for sensitivity analysis.

$$S1_i = \frac{\mathrm{Var}_{X_i}(E_{X_{\sim i}}(Y|X_i))}{\mathrm{Var}(Y)} \tag{25}$$

$$S2_{i,t} = \frac{\mathrm{Var}_{X_i}\left(E_{X_{\sim i,t}}(Y|X_i, X_t)\right)}{\mathrm{Var}(Y)} \tag{26}$$

$$ST_i = \frac{E(\mathrm{Var}_{X_i}(E_{X_{\sim i}}(Y|X_i)))}{\mathrm{Var}(Y)} \tag{27}$$

Figure 12 demonstrates the first-order and total-order sensitivity values of each feature. The first-order sensitivity of the features suggests that the feature that primarily affects rock burst predictive model is the $\tau_{\max}$, whose first-order sensitivity index is 0.58. The first-order sensitivities of the other features are all less than 0.1, which suggests that a change in the values of these few features will only have a small direct effect on the model output. A small first-order sensitivity does not mean that the feature is

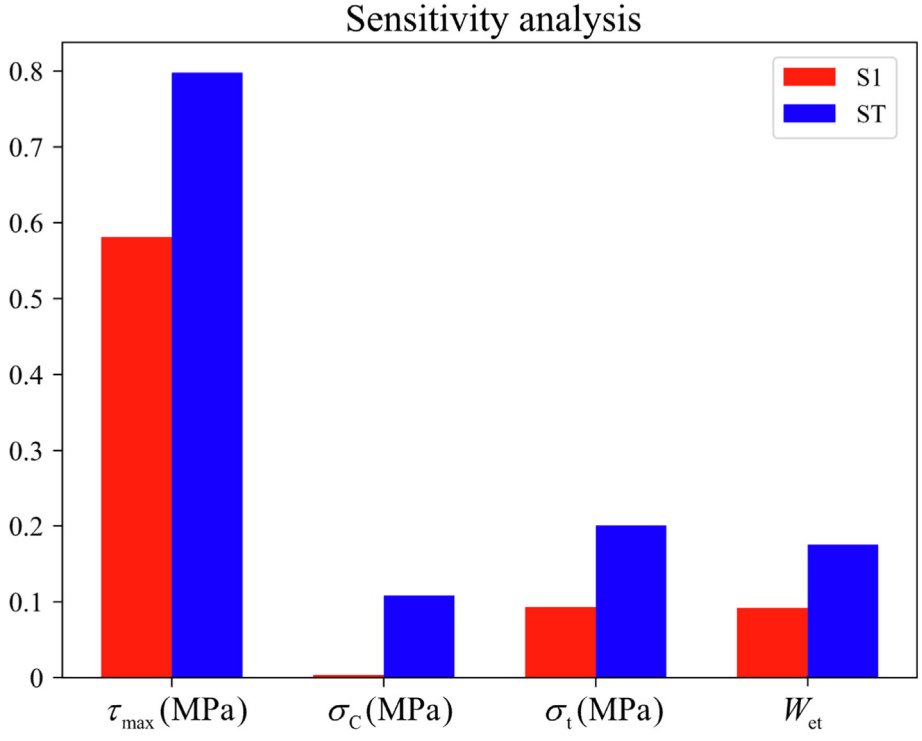

Fig. 12. First-order and total-order sensitive indices of four features.

worthless; given the nonlinear relationship between these features, the total-order sensitivity can often be more responsive to the impact of the feature.

The feature affecting the rock burst predictive model most is τ_{max}, whose total-order sensitivity value is 0.80. The total-order sensitivities of the other features, although relatively low, have a significant increase compared to their first order sensitivities, implying that the interactions existing between the individual features have a significant impact on the model output. This reveals that these four chosen features are all valuable to the model performance, which is consistent with the literature collected above. In the study of rock burst, researchers also tend to use these four indicators and their combination of values such as C, B_1 and B_2 to predict rock burst.

5.1.2 Tree SHAP

Tree SHAP is a method for explaining the prediction results of ML models. Lin et al. (2023) and Liu et al. (2024) used SHAP interpretability analysis to carefully interpret the model developed, and the consistency of their results with published experimental results improved the reliability of the model. It is based on the concept of SHAP values and is used to determine how much each feature contributes to the prediction results. In contrast to sensitivity analysis, Tree SHAP not only measures the importance of each feature, but also roughly indicates the positive or negative correlation between the feature and the model output by the positive or negative SHAP value. The Tree SHAP method is often applicable to models based on tree structures, so we also use Tree SHAP for the VAE-NGBoost explanation presented in this paper.

Figure 13(a) and (b) shows the feature importance of the four features and the distribution of the SHAP value, respectively. From Fig. 13(a), we can see that the importance of each feature is consistent with the distribution of the total-order of sensitivity in the sensitivity analysis, which suggests that the most important feature for the model is τ_{max}, both in terms of variance contribution and marginal contribution. From Fig. 13(b), we can see that the feature τ_{max} and the occurrence of rock burst is positively correlated, i.e., the more the rock is subjected to the maximum tangential stress, the more likely the rock burst will occur. The correlation between the remaining three features and the occurrence of rockbursts is ambiguous, i.e., these three features tend to influence the model predictions in conjunction with other features. This is consistent with the conclusions drawn in the sensitivity analysis.

5.2 Local explanatory analysis

Local explanatory analysis is the process of explaining model predictions at the individual level. Local explanatory analysis can provide individuals with personalized recommendations and decision support. It is very meaningful in engineering examples, where we can provide individuals with customized recommendations and decision support based on their specific situation by understanding the drivers of the predicted results of the data, provided that the model performs well. In this subsection, we will carry out the local explanation of the model through Anchor and Tree SHAP. For data selection, we used one rockburst and one non-rockburst sample from the test as example templates, and the details of these two samples are shown in Table 4.

5.2.1 Anchor

Anchor is a method for explaining model predictions and has the advantage of providing intuitive and explainable results. ElShawi et al. (2021) compared multiple

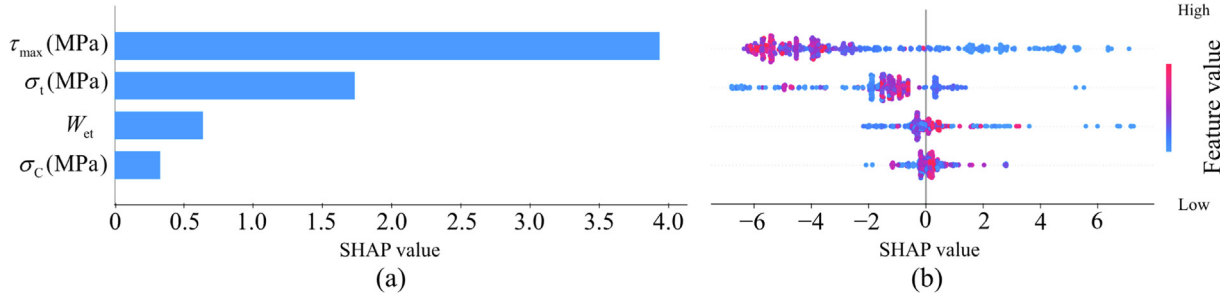

Fig. 13. Feature importance and SHAP values of the features. (a) Mean(|SHAP value|)(average impact on model output magnitude), and (b) SHAP value (impact on model output).

Table 4
Details of two samples.

Location of the data in dataset	Features				Situation	Data source
	τ_{max}(MPa)	σ_C(MPa)	σ_t(MPa)	W_{et}		
306	29.7	116	2.7	3.7	Rockburst	Daxiangling tunnel YK61 + 445 (Afraei et al.,2019)
316	15.97	114.07	11.96	2.4	Non-rockburst	Biotite granite porphyry(Zhou et al., 2016)

Table 5
VAE-NGBoost prediction rules generated by Anchor.

Samples	Predict outcome	Rule of features				Precision	Coverage
		τ_{max}(MPa)	σ_C(MPa)	σ_t(Mpa)	W_{et}		
306	Rock burst	>26.8	≤125.7	≤5.65	>3.1	0.95	0.07
316	None	≤26.8	≤125.7	>5.65		0.95	0.39

machine learning interpretability methods on multiple datasets and showed that anchor points achieved the highest performance on the trust metric. The Anchor algorithm is based on the paper by Ribeiro et al. (Ribeiro et al. (2018). Its computation is based on a heuristic search algorithm, which searches in a continuous loop to find the optimal explanation that satisfies a confidence threshold. Compared to local interpretable model-agnostic explanations (LIME), Anchor illustrates the scope of "explanation" through coverage, which solves the problem of overconfidence in the explanation that LIME tends to cause. In addition, Anchor generates an explanation as a set of feature constraints, which is intuitively explainable and allows us to better understand the decision-making process of the model.

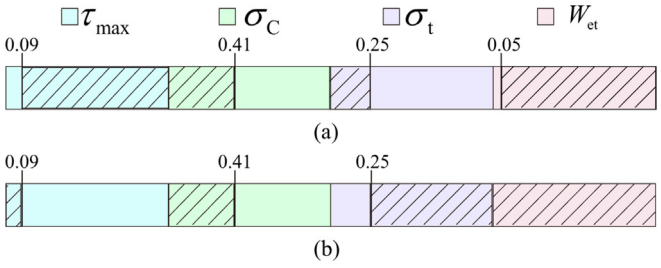

Fig. 14. (a) Rules of the 306th sample, and (b) rule of the 316th sample generated by Anchor.

In Table 5, you can find the results of Anchor's explanation of the VAE-CatBoost prediction about these two samples. As can be seen from Table 5, the model's prediction outcome of the 306th sample is rock burst due to the rule of $\tau_{max} > 26.8$, $\sigma_C \leq 125.7$, $\sigma_t \leq 5.65$ and $W_{et} > 3.1$. A comparison of the two samples shows that different values of τ_{max} and σ_t can affect the prediction results when σ_C is under the same rule, which is also in line with the results of the global sensitivity analysis and SHAP analysis mentioned above, where the sensitivity as well as the importance of τ_{max} and σ_t are in the first and second place. To make it easier to understand the size of the scope of the rules generated by Anchor, Fig. 14 describes its limited scope in percentage terms.

5.2.2 Tree SHAP

In addition to global interpretability, Tree SHAP also provides local explanation for each datum. Compared to Anchor, Tree SHAP not only quantifies the size of the contribution to be made to each feature, but also indicates the relevance of that contribution to the output results. Figure 15 shows the explanation of the model output results by Tree SHAP.

As can be seen in Fig. 15(a), τ_{max}, σ_t and W_{et} have a negative effect on the model's prediction of 'rockburst', while σ_C has a positive effect. Since the sum of the base value and the shap value of each feature is negative, the model predicts that rockburst will occure in this space, and the

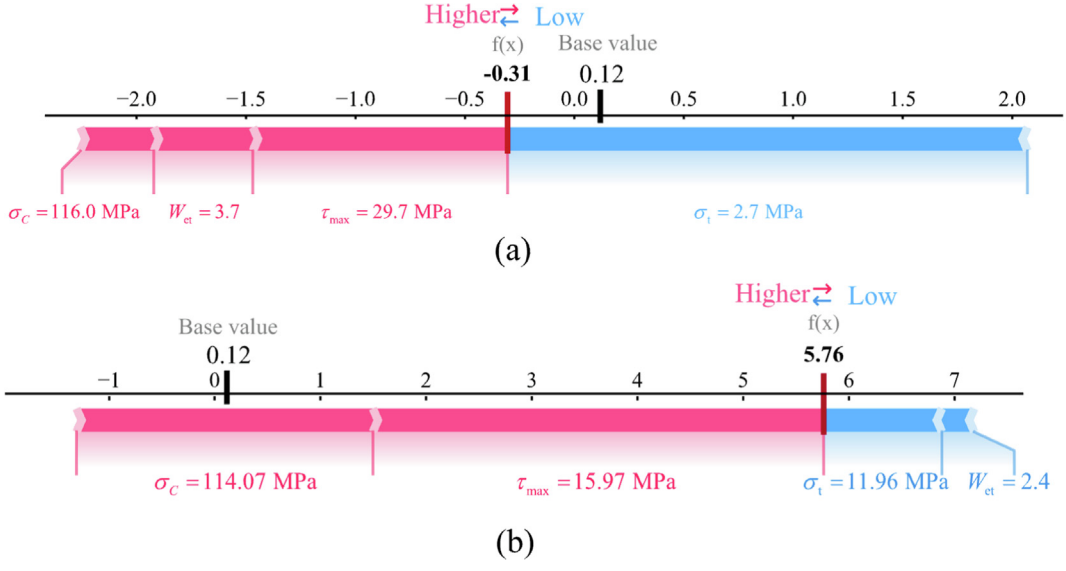

Fig. 15. Local SHAP value analysis. (a) The 306th sample, and (b) the 316th sample.

same analytical principle applies to Fig. 15(b). Considering the results of the first order sensitivity analysis of the data, the model is explained mainly in terms of the direct effect of τ_{max}, σ_t and the interaction σ_C and W_{et} with other features. From the comparison of these two sets of data, the positive SHAP value produced by τ_{max} is significantly higher due to the decrease in τ_{max}, i.e., a decrease in σ_t reduces the likelihood of rock burst. This is consistent with the conclusions in (Xu et al., 2022), in which it is clearly stated that the direct cause of rock burst is the disturbance and superimposition of external loads on the surrounding rock, i.e., an increase in τ_{max} increases the likelihood of rock burst. Due to the elevation of σ_t, the negative SHAP value produced by σ_t is also elevated, i.e., the elevation of σ_t also reduces the likelihood of rock burst, which is consistent with the results in Du et al. (2006) and Guo et al. (2022). In that literature, it is clearly stated that the greater the rock brittleness indices, the more prone to rock burst, and when σ_C is constant, the greater the σ_t, the greater the brittleness indices. For the other features σ_C and W_{et}, there is a large difference in their SHAP values, even though the difference in their values is not large. From this, we can also see that there is a strong interaction between the features, where a change in one feature affects the contribution of the other feature to the model output.

In summary, we have used three methods, Sobol sensitive analysis, Tree SHAP and Anchor, for global and local explanation of VAE-NGBoost. From the results, these three explanatory methods can corroborate each other, and their results are consistent with the theoretical basis in the previous literature, which improves the utility of the black model and further increases our trust in the model.

6 Conclusions

In this paper, we proposed a multilevel explainable ensemble learning model equipped with a variational autoencoder for rock burst analysis that can handle imbalanced engineering data. We collect real-world recorded data of 537 rock bursts from opening source literature for model training and testing, and analyze the correlation between data structure and features. For parameter settings, we used Bayesian optimization and the algorithm-specific parameter tuning strategies of imbalanced datasets to find the optimal hyperparameter settings for all ML models in the pipeline. Finally, for the model explanation, we formulated the multilevel XAI model with Sobol sensitivity analysis and Tree SHAP for a global explanation of high-performance ensemble learning models. In order to have a more detailed understanding of the model's output explanation for each instance, we also analyzed the local explanation for both instances using Anchor and Tree SHAP. The data-driven model in this study involves rock mechanics, engineering geology, deep learning, and explainable machine learning to provide a powerful method for predicting rock burst occurrences and understanding their mechanisms.

(1) Compared to the re-sampling based imbalanced data processing method, the use of VAE to generate a few class samples better solves the problem of imbalanced data, and the model built using this dataset significantly outperforms the original dataset as well as other balanced datasets.
(2) The gradient boosting based machine learning model performs better on this dataset than the standard ensemble learning model RF and the single linear regression model. Combining the advantages of VAE and the ensemble learning model based on the gradient boosting algorithm, the VAE-NGBoost constructed in this paper shows powerful performance: ACC=0.935, G-mean = 0.887, F1-score = 0.800 of test data, and the average ACC and G-mean scores of 5-fold stratified cross-validation are 0.921 and 0.819, respectively.
(3) Multilevel XAI not only gives us a macroscopic view of the impact of each feature on the model, but also allows us to specifically explain the role that each feature plays on the model output in each instance. Its results can inform future feature collection to further improve the performance of the model.
(4) The results from the different perspectives of the explainable model show that the factors affecting the VAE-NGBoost in this paper are broadly consistent with the records of published literature and the experiences of the engineering practice. This consistency gives us sufficient reliability in using VAE-NGBoost for predicting rock burst.

Due to the VAE, the model is more suitable for rockburst analysis when the data are currently imbalanced, and further optimization of the model will be required when more and more balanced rockburst data are collected in the future. In the actual prediction, we need to collect the same features, and then use VAE-NGBoost for prediction. Finally, the effectiveness of the model needs to be determined by the local interpretation analysis of XAI and engineering experience.

While the results of this study are valuable, there are some limitations to consider. The applicability of VAE could not be clarified before the model was built, and it is clear from the 5-fold cross-validation that, in some situations, the model works better without VAE despite the imbalanced data. The ML model is a data-driven model that simplifies the complexity of the field situation.

In the future, we will look for ways in which data quality can be clarified and combined with VAE as a condition for using VAE. Collect more relevant features to discover the complexity patterns in engineering practices.

In summary, the main contributions of this paper are the establishment of an extensive rock burst database and the revelation of the best way to expand the data and the

ensemble learning models, especially the VAE-NGBoost model for predicting rock burst. In addition, we investigate the framework of explainable methods suitable for ensemble classifiers, which improves the transparency and reliability of complex black-box models.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

CRediT authorship contribution statement

Shan Lin: Writing – review & editing, Supervision, Software, Methodology, Investigation, Funding acquisition, Conceptualization. **Zenglong Liang:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Data curation, Conceptualization. **Miao Dong:** Writing – review & editing, Visualization, Software. **Hongwei Guo:** Writing – review & editing, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Hong Zheng:** Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Grant Nos. 42107214 and 52130905).

References

Afraei, S., Shahriar, K., & Madani, S. H. (2019). Developing intelligent classification models for rock burst prediction after recognizing significant predictor variables, Section 1: Literature review and data preprocessing procedure. *Tunnelling and Underground Space Technology, 83*, 324–353.

Ahmad, M., Katman, H. Y., Al-Mansob, R. A., Ahmad, F., Safdar, M., & Alguno, A. C. (2022). Prediction of Rockburst Intensity Grade in Deep Underground Excavation Using Adaptive Boosting Classifier. *Complexity, 2022*, 615620.

Bijlsma, S., Bobeldijk, L., Verheij, E. R., Ramaker, R., Kochhar, S., Macdonald, I. A., et al. (2006). Large-scale human metabolomics studies: A strategy for data (pre-) processing and validation. *Analytical Chemistry, 78*(2), 567–574.

Cai, M. F., Wang, J. A., & Wang, S. H. (2001). Analysis on energy distribution and prediction of rock burst during deep mining excavation in linglong gold mine. *Chinese Journal of Rock Mechanics and Engineering, 20*(1), 38–42 (in Chinese).

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Chen, T. Q., Guestrin, C., & Assoc Comp, M. (2016). *XGBoost: A Scalable Tree Boosting System.* Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Aug 13-17, San Francisco, CA.

Dong, X. B., Yu, Z. W., Cao, W. M., Shi, Y. F., & Ma, Q. L. (2020). A survey on ensemble learning. *Frontiers of Computer Science, 14*(2), 241–258.

Du, Z. J., Xu, M. G., Liu, Z. P., & Wu, X. (2006). Laboratory integratedevaluation method for engineering wall rock rock-burst. *Gold* (11), 26-30.(in Chinese).

Duan, T., Avati, A., Ding, D. Y., Thai, K. K., Basu, S., Ng, A., et al. (2019). NGBoost: Natural Gradient Boosting for Probabilistic Prediction. Paper presented at the 25th Americas Conference on Information Systems of the Association-for-Information-Systems (AMCIS), Aug 15–17, Cancun, MEXICO.

ElShawi, R., Sherif, Y., Al-Mallah, M., & Sakr, S. (2021). Interpretability in healthcare: A comparative study of local machine learning interpretability techniques. *Computational Intelligence, 37*(4), 1633–1650.

Faradonbeh, R. S., Taheri, A., & Karakus, M. (2022). The propensity of the over-stressed rock masses to different failure mechanisms based on a hybrid probabilistic approach. *Tunnelling and Underground Space Technology, 119*, 104214.

Feng, G. L., Feng, X. T., Chen, B. R., & Xiao, Y. X. (2015). Microseismic sequences associated with rockbursts in the tunnels of the Jinping II hydropower station. *International Journal of Rock Mechanics and Mining Sciences, 80*, 89–100.

Gong, F. Q., & Li, X. B. (2007). A distance discriminant analysis method for prediction of possibility and classification of rockburst and its application. *Chinese Journal of Rock Mechanics and Engineering, 26*(5), 1012–1018 (in Chinese).

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). Generative Adversarial Networks. arXiv:1406.2661. Retrieved from https://ui.adsabs.harvard.edu/abs/2014arXiv1406.2661G. doi:10.48550/arXiv.1406.2661.

Guo, J., Guo, J. W., Zhang, Q. L., & Huang, M. J. (2022). Research on Rockburst Classification Prediction Based on BP-SVM Model. *Ieee Access, 10*, 50427–50447.

Hao, J., Shi, K. B., Wang, X. L., Bai, X. J., & Chen, G. M. (2016). Application of cloud model to rating of rockburst based on rough set of FCM algorithm. *Rock and Soil Mechanics, 37*(3), 859–866 (in Chinese).

He, M. C., & Wang, Q. (2023). Rock dynamics in deep mining. *International Journal of Mining Science and Technology, 33*(9), 1065–1082.

He, M. C., Xia, H. M., Jia, X. N., Gong, W. L., Zhao, F., & Liang, K. Y. (2012). Studies on classification, criteria and control of rockbursts. *Journal of Rock Mechanics and Geotechnical Engineering, 4*(2), 97–114.

Herman, J., & Usher, W. (2017). SALib: An open-source Python library for Sensitivity Analysis. *Journal of Open Source Software, 2*(9), 97.

Jia, Y. P. (2014). *Study on prediction method and theorial model of rockburst.* (Publication No.2 page-130) [Doctoral dissertation, Zhejiang University]. CNKI. (in Chinese).

Jiang, L. F. (2008). *Study on prediction and prevention of rockburst in anlu tunnel.* (Publication No.12 page-88) [Master's thesis, Southwest Jiaotong University]. (in Chinese).

Ke, G. L., Meng, Q., Finley, T., Wang, T. F., Chen, W., Ma, W. D., et al. (2017). *LightGBM: A Highly Efficient Gradient Boosting Decision Tree.* Paper presented at the 31st Annual Conference on Neural Information Processing Systems (NIPS), Dec 04-09, Long Beach, CA.

Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S. H., et al. (2021). A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. *IEEE Access, 9*, 109960–109975.

Kingma, D. P., & Welling, M. (2013). Auto-Encoding Variational Bayes. arXiv:1312.6114. Retrieved from https://ui.adsabs.harvard.edu/abs/2013arXiv1312.6114K. https://doi.org/10.48550/arXiv.1312.6114.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. [Review]. *Nature, 521*(7553), 436–444.

Lee, J., & Jeong, S. (2016). Experimental Study of Estimating the Subgrade Reaction Modulus on Jointed Rock Foundations. *Rock Mechanics and Rock Engineering, 49*(6), 2055–2064.

Li, D. Y., Liu, Z. D., Armaghani, D. J., Xiao, P., & Zhou, J. (2022). Novel ensemble intelligence methodologies for rockburst assessment in complex and variable environments. [Article]. *Scientific Reports, 12*(1), 1–23.

Li, T. Z., Li, Y. X., & Yang, X. L. (2017). Rock burst prediction based on genetic algorithms and extreme learning machine. *Journal of Central South University, 24*(9), 2105–2113.

Li, X. H., Wang, X. F., Kang, Y., & He, Z. (2005). Artificial neural network for prediction of rockburst in deep-buried long tunnel. In J. Wang, X. Liao & Z. Yi (Eds.), *Advances in Neural Networks - Isnn 2005, Pt 3, Proceedings* (Vol. 3498, pp. 983-986).

Liang, W. Z., Sari, Y. A., Zhao, G. Y., McKinnon, S. D., & Wu, H. (2021). Probability Estimates of Short-Term Rockburst Risk with Ensemble Classifiers. *Rock Mechanics and Rock Engineering, 54*(4), 1799–1814.

Lin, S., Liang, Z. L., Zhao, S. X., Dong, M., Guo, H. W., & Zheng, H. (2023). A comprehensive evaluation of ensemble machine learning in geotechnical stability analysis and explainability. [Article; Early Access]. *International Journal of Mechanics and Materials in Design*.

Liu, B. K., & Lu, W. Z. (2022). Surrogate models in machine learning for computational stochastic multi-scale modelling in composite materials design. *International Journal of Hydromechatronics, 5*(4), 336–365.

Liu, B. K., Lu, W. Z., Olofsson, T., Zhuang, X. Y., & Rabczuk, T. (2024). Stochastic interpretable machine learning based multiscale modeling in thermal conductivity of Polymeric graphene-enhanced composites. *Composite Structures, 327*, 117601.

Liu, B. K., Penaka, S. R., Lu, W. Z., Feng, K. L., Rebbling, A., & Olofsson, T. (2023a). Data-driven quantitative analysis of an integrated open digital ecosystems platform for user-centric energy retrofits: A case study in northern Sweden. *Technology in Society, 75*, 102347.

Liu, B. K., Vu-Bac, N., & Rabczuk, T. (2021). A stochastic multiscale method for the prediction of the thermal conductivity of Polymer nanocomposites through hybrid machine learning algorithms. *Composite Structures, 273*, 114269.

Liu, B. K., Vu-Bac, N., Zhuang, X. Y., Fu, X. L., & Rabczuk, T. (2022a). Stochastic full-range multiscale modeling of thermal conductivity of Polymeric carbon nanotubes composites: A machine learning approach. *Composite Structures, 289*, 115393.

Liu, B. K., Vu-Bac, N., Zhuang, X. Y., Fu, X. L., & Rabczuk, T. (2022b). Stochastic integrated machine learning based multiscale approach for the prediction of the thermal conductivity in carbon nanotube reinforced polymeric composites. *Composites Science and Technology, 224*, 109425.

Liu, B. K., Vu-Bac, N., Zhuang, X. Y., Lu, W. Z., Fu, X. L., & Rabczuk, T. (2023b). Al-DeMat: A web-based expert system platform for computationally expensive models in materials design. *Advances in Engineering Software, 176*, 103398.

Liu, B. K., Wang, Y. Z., Rabczuk, T., Olofsson, T., & Lu, W. Z. (2023c). Multi-scale modeling in thermal conductivity of Polyurethane incorporated with Phase Change Materials using Physics-Informed Neural Networks. arXiv:2307.16785. Retrieved from https://ui.adsabs.harvard.edu/abs/2023arXiv230716785L. https://doi.org/10.48550/arXiv.2307.16785.

Liu, D. Y., & Liu, G. S. (2019). *A Transformer-Based Variational Autoencoder for Sentence Generation.* Paper presented at the International Joint Conference on Neural Networks (IJCNN), Jul 14-19, Budapest, HUNGARY.

Liu, B. K., Vu-Bac, N., Zhuang, X. Y., & Rabczuk, T. (2020). Stochastic multiscale modeling of heat conductivity of Polymeric clay nanocomposites. *Mechanics of Materials, 142*, 103280.

Liu, R., Ye, Y. C., Zhang, G. Q., Yao, N., Chen, H., & Wang, Q. H. (2019). Grading Prediction Model of Rockburst Based on Rough Set-Multidimensional Normal Cloud. *Metal Mine* (3), 48-55. (in Chinese).

Liu, H. X., Zhao, G. Y., Xiao, P., & Yin, Y. T. (2023). Ensemble Tree Model for Long-Term Rockburst Prediction in Incomplete Datasets. *Minerals, 13*(1), 103.

Liu, Y. P., Jiang, H. K., Wang, Y. F., Wu, Z. H., & Liu, S. W. (2022). A conditional variational autoencoding generative adversarial networks with self-modulation for rolling bearing fault diagnosis. *Measurement, 192*, 110888.

Micci-Barreca, D. (2001). A preprocessing scheme for high-cardinality categorical attributes in classification and prediction problems. *ACM SIGKDD Explorations Newsletter, 3*(1), 27–32.

Mienye, I. D., & Sun, Y. X. (2023). A Deep Learning Ensemble With Data Resampling for Credit Card Fraud Detection. *IEEE Access, 11*, 30628–30638.

Mirza, B., Haroon, D., Khan, B., Padhani, A., & Syed, T. Q. (2021). Deep Generative Models to Counter Class Imbalance: A Model-Metric Mapping With Proportion Calibration Methodology. *IEEE Access, 9*, 55879–55897.

Pan, Y. S., & Wang, A. W. (2023). Disturbance response instability theory of rock bursts in coal mines and its application. *Geohazard Mechanics, 1*(1), 1–17.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). *CatBoost: unbiased boosting with categorical features.* Paper presented at the 32nd Conference on Neural Information Processing Systems (NIPS), Dec 02-08, Montreal, CANADA.

Pu, Y. Y., Apel, D. B., & Xu, H. W. (2019). Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier. *Tunnelling and Underground Space Technology, 90*, 12–18.

Puh, M., & Brkic, L. (2019). *Detecting Credit Card Fraud Using Selected Machine Learning Algorithms.* Paper presented at the 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), May 20-24, Opatija, CROATIA.

Ribeiro, M. T., Singh, S., Guestrin, C., & Aaai. (2018). *Anchors: High-Precision Model-Agnostic Explanations.* Paper presented at the 32nd AAAI Conference on Artificial Intelligence / 30th Innovative Applications of Artificial Intelligence Conference / 8th AAAI Symposium on Educational Advances in Artificial Intelligence, Feb 02-07, New Orleans, LA.

Rodrigues, A. D. P., Luna, A. S., & Pinto, L. (2023). An evaluation strategy to select and discard sampling preprocessing methods for imbalanced datasets: A focus on classification models. *Chemometrics and Intelligent Laboratory Systems, 240*, 104933.

Saltelli, Ratto, A., Andres, M., & Campol, T. (2008). Global Sensitivity Analysis. The Primer.

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks, 61*, 85–117.

Shi, X. Z., Zhou, J., Dong, L., Hu, H. Y., Wang, H. Y., & Chen, S. R. (2010). Application of unascertained measurement model to prediction of classification of rockburst intensity. *Chinese Journal of Rock Mechanics and Engineering, 29*(S1), 2720–2726 (in Chinese).

Susan, S., & Kumar, A. (2020). The balancing trick: Optimized sampling of imbalanced datasets—A brief survey of the recent State of the Art. *Engineering Reports, 3*(4), 12298.

Tao, S. T., Peng, P., Li, Y. F., Sun, H. Y., Li, Q., & Wang, H. W. (2024). Supervised contrastive representation learning with tree-structured parzen estimator Bayesian optimization for imbalanced tabular data. *Expert Systems with Applications, 237*, 121294.

Tasci, E., Zhuge, Y., Camphausen, K., & Krauze, A. V. (2022). Bias and class imbalance in oncologic data-towards inclusive and transferrable ai in large scale oncology data sets. *Cancers, 14*(12), 2897.

Tholke, P., Mantilla-Ramos, Y. J., Abdelhedi, H., Maschke, C., Dehgan, A., Harel, Y., et al. (2023). Class imbalance should not throw you off balance: Choosing the right classifiers and performance metrics for brain decoding with imbalanced data. *Neuroimage, 277*, 120253.

Tian, R. (2021). *Research and Application of Rockburst Intensity Classification Prediction Model Based on Machine Learning Algorithm.* (Publication No.5 page-143) [Doctoral dissertation, Inner Mongolia University Of Science & Technology]. (in Chinese).

Topuz, B., & Alp, N. C. (2023). Machine learning in architecture. *Automation in Construction, 154*, 105012.

Wagner, H. (2019). Deep mining: A rock engineering challenge. *Rock Mechanics and Rock Engineering, 52*(5), 1417–1446.

Wang, C. L., Chen, Z., Liao, Z. F., Hou, X. L., Li, H. T., Wang, A. W., et al. (2020). Experimental investigation on predicting precursory changes in entropy for dominant frequency of rockburst. *Journal of Central South University, 27*(10), 2834–2848.

Wang, J., Liu, P., Ma, L., & He, M. C. (2022). A rockburst proneness evaluation method based on ultidimensional cloud model improved by control variable method and rockburst database. *Lithosphere, 2021*.

Wang, J. C., Ma, H. J., & Yan, X. H. (2023a). rockburst intensity classification prediction based on multi-model ensemble learning algorithms. *Mathematics, 11*(4), 838.

Wang, Y. T., Wei, Y. X., & Wang, H. (2023b). A class imbalanced wafer defect classification framework based on variational autoencoder generative adversarial network. *Measurement Science and Technology, 34*(2), 024008.

Waqar, M. F., Guo, S. F., & Qi, S. W. (2023). A comprehensive review of mechanisms, predictive techniques, and control strategies of rockburst. *Applied Sciences-Basel, 13*(6), 3950.

Xia, B. W. (2007). *Study on prediction and forecast of geologic disaster in highway tunned construction.* (Publication No.1 page-94) [Master's thesis, Chongqing University]. (in Chinese).

Xia, Y. Y., Zhang, C., Wang, C. X., Liu, H. J., Sang, X. X., Liu, R., et al. (2023). Prediction of bending strength of glass fiber reinforced methacrylate-based pipeline UV-CIPP rehabilitation materials based on machine learning. *Tunnelling and Underground Space Technology, 140*, 105319.

Xing, Y., Kulatilake, P., & Sandbak, L. A. (2018). Effect of rock mass and discontinuity mechanical properties and delayed rock supporting on tunnel stability in an underground mine. *Engineering Geology, 238*, 62–75.

Xu, L. M., Lu, K. X., Pan, Y. S., & Qin, Z. J. (2022). Study on rock burst characteristics of coal mine roadway in China. *Energy Sources Part a-Recovery Utilization and Environmental Effects, 44*(2), 3016–3035.

Xu, M. G., Du, Z. J., Yao, G. H., & Liu, Z. P. (2008). Rockburst prediction of chengchao iron mine during deep mining. *Chinese Journal of Rock Mechanics and Engineering, S1*, 2921–2928.

Xue, Y. G., Bai, C. H., Kong, F. M., Qiu, D. H., Li, L. P., Su, M. X., et al. (2020a). A two-step comprehensive evaluation model for rockburst prediction based on multiple empirical criteria. *Engineering Geology, 268*, 105515.

Xue, Y. G., Bai, C. H., Qiu, D. H., Kong, F. M., & Li, Z. Q. (2020b). Predicting rockburst with database using particle swarm optimization and extreme learning machine. *Tunnelling and Underground Space Technology, 98*, 103287.

Xue, Y. G., Li, Z. Q., Li, S. C., Qiu, D. H., Tao, Y. F., Wang, L., et al. (2019). Prediction of rock burst in underground caverns based on rough set and extensible comprehensive evaluation. *Bulletin of Engineering Geology and the Environment, 78*(1), 417–429.

Yao, J. M., & He, F. L. (2008). *Countermeasure research on preventing rock burst with hard roof by energy mechanism.* Paper presented at the International Young Scholars Symposium on Rock Mechanics, Apr 28-May 02, Beijing, Peoples R China.

Yong, T., Reed, P. M., Wagener, T., & Werkhoven, K. V. (2008). Comparison of parameter sensitivity analysis methods for lumped watershed model. *Paper presented at the World Environmental & Water Resources Congress*.

Zhang, K., Schölkopf, B., Muandet, K., & Wang, Z. (2013). *Domain adaptation under target and conditional shift*. Paper presented at the Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28.

Zhang, L. W., Zhang, D. Y., & Qiu, D. H. (2010). Application of extension evaluation method in rockburst prediction based on rough set theory. *Journal of China Coal Society, 35*(9), 1461–1465 (in Chinese).

Zhang, L. W., Zhang, X. Y., Wu, J., Zhao, D. K., & Fu, H. (2020). Rockburst prediction model based on comprehensive weight and extension methods and its engineering application. *Bulletin of Engineering Geology and the Environment, 79*(9), 4891–4903.

Zhang, Y., & Liu, Q. (2022). On IoT intrusion detection based on data augmentation for enhancing learning on unbalanced samples. *Future Generation Computer Systems-the International Journal of Escience, 133*, 213–227.

Zhang, Y., Su, G. S., & Yan, L. B. (2011). Method of identifying rockburst grades based on gaussian process machine learning. *Chinese Journal of Underground Space and Engineering, 7*(02), 392–397 (in Chinese).

Zhao, H. B. (2005a). Classification of rockburst using support vector machine. *Rock and Soil Mechanics*(04), 642-644. (in Chinese).

Zhao, H. B. (2005b). *Rockburst prediction using evolutionary support vector machine.* Paper presented at the Asia Pacific Symposium on Safety 2005, Nov 02-04, Shaoxing, Peoples R China.

Zhao, H. B., & Chen, B. R. (2020). Data-Driven Model for Rockburst Prediction. *Mathematical Problems in Engineering, 2020*, 5735496.

Zhao, J., Song, Y., Wang, L., Guo, H., Marigentti, F., & Liu, X. (2023). Forecasting the eddy current loss of a large turbo generator using hybrid ensemble Gaussian process regression. *Engineering Applications of Artificial Intelligence, 121*, 106022.

Zhao, Y., Nasrullah, Z., & Li, Z. (2019). PyOD: A Python Toolbox for Scalable Outlier Detection. *Journal of Machine Learning Research, 20*, 96.

Zhou, H., Chen, S. K., Li, H. R., Liu, T., & Wang, H. L. (2021). Rockburst prediction for hard rock and deep-lying long tunnels based on the entropy weight ideal point method and geostress field inversion: A case study of the Sangzhuling Tunnel. *Bulletin of Engineering Geology and the Environment, 80*(5), 3885–3902.

Zhou, H., Chen, S. K., Zhang, G. Z., Wang, H. L., He, H. D., & Feng, J. (2020a). Efficiency coefficient method and ground stress field inversion for rockburst prediction in deep and long tunnel. *Journal of Engineering Geology, 28*(06), 1386–1396 (in Chinese).

Zhou, J., Li, X. B., & Mitri, H. S. (2016). Classification of Rockburst in Underground Projects: Comparison of Ten Supervised Learning Methods. *Journal of Computing in Civil Engineering, 30*(5), 4016003.

Zhou, S., Jia, Y., & Wang, C. (2020b). Global Sensitivity Analysis for the Polymeric Microcapsules in Self-Healing Cementitious Composites. *Polymers, 12*(12), 2990.

Zhu, H. L., Wu, X., Luo, Y. L., Jia, Y., Wang, C., Fang, Z., et al. (2023). Prediction of Early Compressive Strength of Ultrahigh-Performance Concrete Using Machine Learning Methods. *International Journal of Computational Methods, 20*(8), 2141023.

Zhuang, X. Y., & Zhou, S. (2019). The Prediction of Self-Healing Capacity of Bacteria-Based Concrete Using Machine Learning Approaches. *Cmc-Computers Materials & Continua, 59*(1), 57–77.