ISSN 2616-5767 Vol.8, Issue 1: 1-7, DOI: 10.25236/AJETS.2025.080101

Prediction and Interpretability Analysis of Mine Blasting Vibration PPV Based on the BOA-XGBoost Model

Zhang Wentao[1,a], Wang Haibo[1,b,*]

[1]*School of Civil Engineering and Architecture, Anhui University of Science and Technology, Huainan, China*
[a]*2250319620@qq.com*, [b]*wanghb_aust@163.com*
[*]*Corresponding author*

Abstract: *Accurately predicting the Peak Particle Velocity (PPV) of blasting vibrations is crucial for controlling the hazards of blasting vibrations. To improve the accuracy of PPV prediction, a model based on Bayesian Optimization Algorithm (BOA) and Extreme Gradient Boosting (XGBoost) is proposed, under feature selection conditions. First, correlation analysis combined with variance analysis is used to filter the initial features. Then, the Bayesian optimization algorithm is applied to fine-tune the hyperparameters of XGBoost, and the optimal hyperparameters are input into the prediction model for training, testing, and evaluation. Finally, the SHAP method is used for interpretability analysis of the model. The results show that optimizing the XGBoost hyperparameters through Bayesian optimization can alleviate overfitting caused by improper hyperparameter selection, improving the model's prediction accuracy and generalization ability. Compared to five other models, BOA-XGBoost demonstrates higher prediction accuracy and stronger nonlinear fitting performance. The importance ranking of features influencing the PPV of blasting vibrations is as follows:* $D > Q_{max} > T > Pf > L > N$. *The blast center distance and number of holes have a negative impact on PPV, while the maximum charge per blast section, stemming length, and powder factor have a positive impact.*

Keywords: *Blasting vibration PPV prediction; interpretability analysis; Extreme gradient boosting algorithm; bayesian optimization algorithm; SHAP method*

1. Introduction

Blasting, as an efficient rock-breaking method, is widely used in open-pit mining operations. However, the vibration effects generated during blasting can threaten the stability of surrounding structures[1]. Accurately predicting the Peak Particle Velocity (PPV) of blasting vibrations before the operation, optimizing blasting parameters, and implementing reliable safety measures can reduce the adverse impacts of blasting vibrations, which is of significant importance for protecting the auxiliary facilities of the mining site.

In engineering practice, blasting technicians typically use regression analysis to establish empirical formulas based on charge weight and distance to describe the relationship between blasting vibrations and these factors. Based on these empirical formulas, the vibrations caused by blasting can be predicted. Traditional empirical formulas include the Sadovski formula, Indian Bureau of Standards formula, and U.S. Bureau of Mines formula[2].However, in practical applications, due to the fact that peak particle velocity is influenced by multiple factors, traditional empirical formulas only consider a limited number of influencing factors. With the development of artificial intelligence, machine learning has been increasingly applied to engineering data analysis, providing new ideas for blasting vibration prediction. Hu Xiaobing[3]developed a blasting vibration prediction program based on BP neural networks and MATLAB. Yue Zhonwen[4] combined Particle Swarm Optimization (PSO) with Least Squares Support Vector Machine (LSSVM) to construct the PSO-LSSVM blasting vibration prediction model. Wang XinYu et al [5] combined the whale optimization algorithm with the support vector machine algorithm to establish a hybrid WOA-SVM model for predicting blasting vibration, and achieved better prediction results. Machine learning models are capable of identifying the underlying relationships between input features and target variables, and they demonstrate excellent performance in predictive analysis. However, although the above models can make accurate predictions, the modeling results are difficult to interpret and cannot reveal the underlying physical processes from a

ISSN 2616-5767 Vol.8, Issue 1: 1-7, DOI: 10.25236/AJETS.2025.080101

mechanistic perspective, which undermines the credibility of the model.

In summary, this study first introduces correlation analysis and variance analysis to select the initial features, then applies Bayesian optimization Algorithm to fine-tune the hyperparameters of the Extreme Gradient Boosting (XGBoost) algorithm, establishing the BOA-XGBoost blasting vibration prediction model. Using actual blasting vibration monitoring data, the model is trained, tested, and evaluated. Finally, the SHAP method is used for interpretability analysis of the model, with the aim of providing a reference for the scientific and precise control of blasting vibrations.

2. Basic Theory

2.1. XGBoost

Extreme Gradient Boosting (XGBoost)[6] is an ensemble learning algorithm, particularly suited for classification and regression tasks in supervised learning, and is an important tool in machine learning.

The XGBoost model can be expressed as the sum of multiple trees:

$$\hat{y}_i = \sum_{t=1}^{t} f_t(x_i) \tag{1}$$

where, x_i is input features, f_t is the t-th tree, $\hat{y}_i$ is the predicted value of the target variable.

The overall objective function of XGBoost is as follows:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{2}$$

where, $l(\bullet)$ represents the model's loss function, n denotes the number of samples, $\Omega(f_t)$ is the regularization term.

In the optimization process of XGBoost, the loss function used for constructing decision trees is often approximated using a second-order Taylor expansion. The Taylor expansion of the loss function in the equation at $y_i, \hat{y}_i^{(t-1)} + f_t(x_i)$ can be approximately expressed as:

$$L^{(t)} \approx \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^{2}(x_i) \right] + \Omega(f_t) \tag{3}$$

where, g_i and h_i represent the first-order and second-order derivatives of the loss function at $f_t(x_i)$ respectively.

Through this expansion, the model can be gradually optimized, allowing each newly added tree to effectively reduce the loss.

2.2. Bayesian Optimization Algorithm

The Bayesian Optimization Algorithm[7] is a global hyperparameter optimization algorithm that uses Bayes' theorem in its optimization process. The core idea is to use conditional probability to transform the prior probability model into a posterior probability distribution, thereby actively selecting the next sample point. This iterative process significantly reduces the number of iterations, shortens the search time, and enhances the generalization ability of the model.

2.3. Feature Selection

Feature selection aims to identify the most contributive features from the raw data to reduce model complexity, improve training efficiency, and ultimately enhance the model's generalization ability. This study primarily employs a feature selection method that combines Spearman's correlation analysis and variance analysis (ANOVA)[8,9].

ISSN 2616-5767 Vol.8, Issue 1: 1-7, DOI: 10.25236/AJETS.2025.080101

2.4. Performance metrics

The model is evaluated using three performance metrics: coefficient of determination (R^2),root mean squared error(*RMSE*), and mean absolute error (*MAE*)[10]. R^2represents the linear relationship between the observed and predicted values. *RMSE* is used to represent the dispersion of the results, while *MAE* indicates the accuracy of the results.

2.5. SHAP method

Shapley Additive Explanations (SHAP) [11]is an innovative explanation tool based on cooperative game theory, which can deeply reveal the reasons behind model predictions through visualizations, providing important support for the reliability of the predictions.

2.6. Model Framework

The technical framework of BOA-XGBoost prediction model constructed in this study is shown in Fig. 1, firstly, feature selection is carried out, and then 80% of the data in the database is randomly selected for training, and the remaining 20% of the data is used for testing. The XGBoost hyperparameters are optimized using the Bayesian optimization algorithm, the optimal hyperparameters are substituted into the model for training and testing, the model is evaluated by three indicators, and finally the model is analyzed for interpretability.

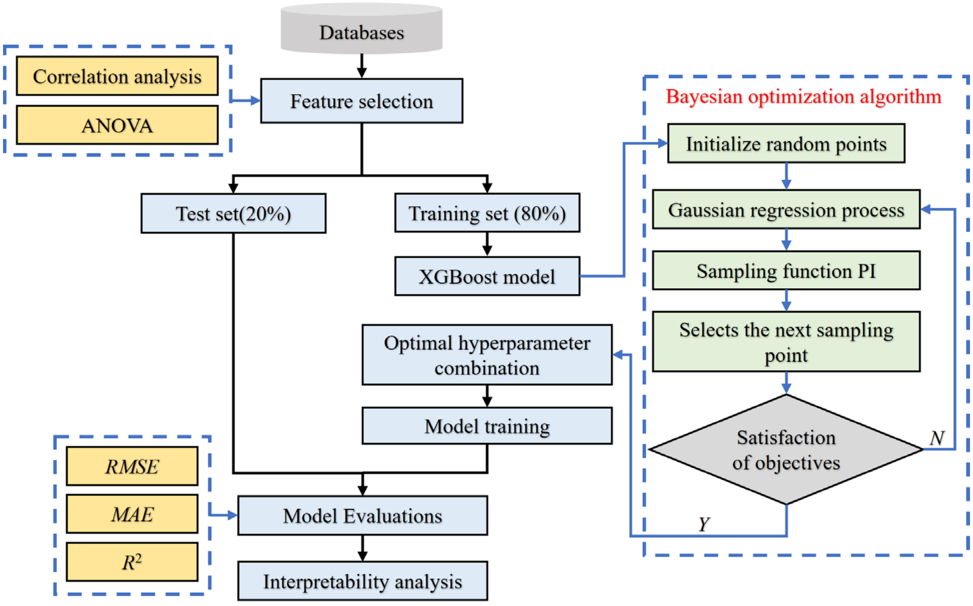

Figure 1: BOA-XGBoost model framework

3. Dataset Preparation

The dataset in this study comes from the Dahuangshan Building Materials Mine in Dinghai District, Zhoushan City. The site uses medium-deep-hole bench blasting technology for ore extraction. To facilitate ore transportation and processing, the company built a sand and gravel processing system at the foot of the mountain on the northern side of the blasting area. The extracted ore is transported to the processing plant for crushing and screening. As the blasting operations continue, the relative position between the bench and the existing slope dynamically changes. When the blasting bench approaches the slope, the dynamic load effect caused by the blasting increases, further raising the potential risk of slope instability and collapse, which poses a severe threat to the safety of the factory building below. Multiple blasting vibration monitoring sessions were conducted using the NUBOX-8016 vibration meter based on the blasting operation conditions. After removing invalid data, a total of 60 data sets were collected. An overview of the mining area and blasting vibration tests is shown in Figure 2.

ISSN 2616-5767 Vol.8, Issue 1: 1-7, DOI: 10.25236/AJETS.2025.080101

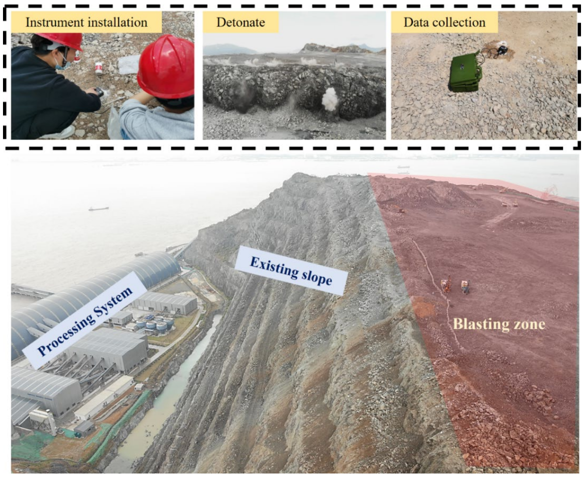

Figure 2: Blasting Vibration Test in the Mining Area

There are many factors that affect the intensity of blasting vibrations. In this study, seven initial features were selected: the maximum charge per blast section (Q_{max}), blasting center distance (D), number of holes (N), hole depth (L), charge weight (W_{total}), the powder factor (Pf), and stemming length (T). The correlation between the features is shown in Figure 3. It can be observed that there are varying degrees of correlation between the features, indicating an issue of information redundancy in the initial features. ANOVA was used to calculate the F-statistic to measure the relationship between the features and the target variable PPV, and the results are shown in Figure 4. From Figures 3 and 4, it can be seen that charge weight has a relatively significant correlation with other features but has a minimal impact on the target variable. Therefore, the charge weight feature was removed.

Figure 3: Correlation Heatmap

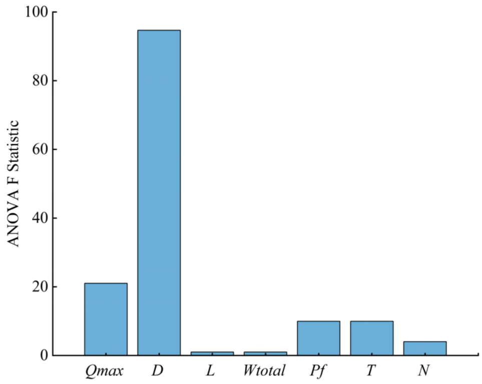

Figure 4: F-statistics for Each Feature

4. Model Training, Prediction and Evaluation

4.1. Hyperparameter Optimization

The hyperparameters of the XGBoost algorithm selected in this study and their predefined search Space are shown in Table 1.

Table 1: Predefined search space for XGBoost algorithm hyperparameters

Hyperparameter	Meaning	Search Space
max_depth	Decision tree depth	1~20
n_estimators	Number of iterations	1~300
learning rate	Learning rate	0.01~0.3
subsample	Sample sampling rate	0.5~1
colsample_bytree	Queue sampling rate	0.5~1
gamma	Control whether splitting is allowed	0~5

The Bayesian algorithm built into the Optuna library was used for hyperparameter optimization, with the Mean Squared Error (*MSE*) as the optimization objective and 500 iterations. Figure 5

ISSN 2616-5767 Vol.8, Issue 1: 1-7, DOI: 10.25236/AJETS.2025.080101

illustrates the hyperparameter optimization process.

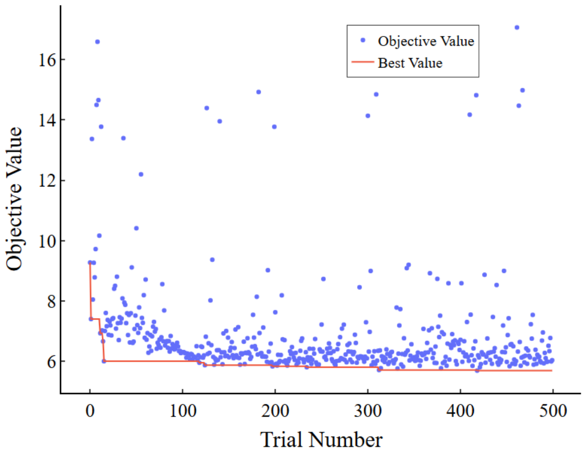

Figure 5: Optimization history

In the figure, the x-axis represents the trial number, and the y-axis shows the objective function value obtained from each trial. It can be observed that the optimization process quickly reduces the objective value in the early stage (less than 15 trials) and gradually approaches convergence, demonstrating the efficiency of the Bayesian algorithm.The optimal hyperparameters obtained through Bayesian optimization are: max_depth = 9, n_estimators = 140, learning rate = 0.054, subsample = 0.89, colsample_bytree = 0.935, gamma = 4.74.

4.2. Model Training and Testing

The optimal hyperparameters are incorporated into the XGBoost model for training and testing. Meanwhile, to verify the superiority of the prediction model developed in this study, unoptimized XGBoost, BPNN, SVR models, as well as the Sadowski formula and multiple linear regression models, are used to predict the PPV. The training and testing results of the different models are shown in Figure 6.

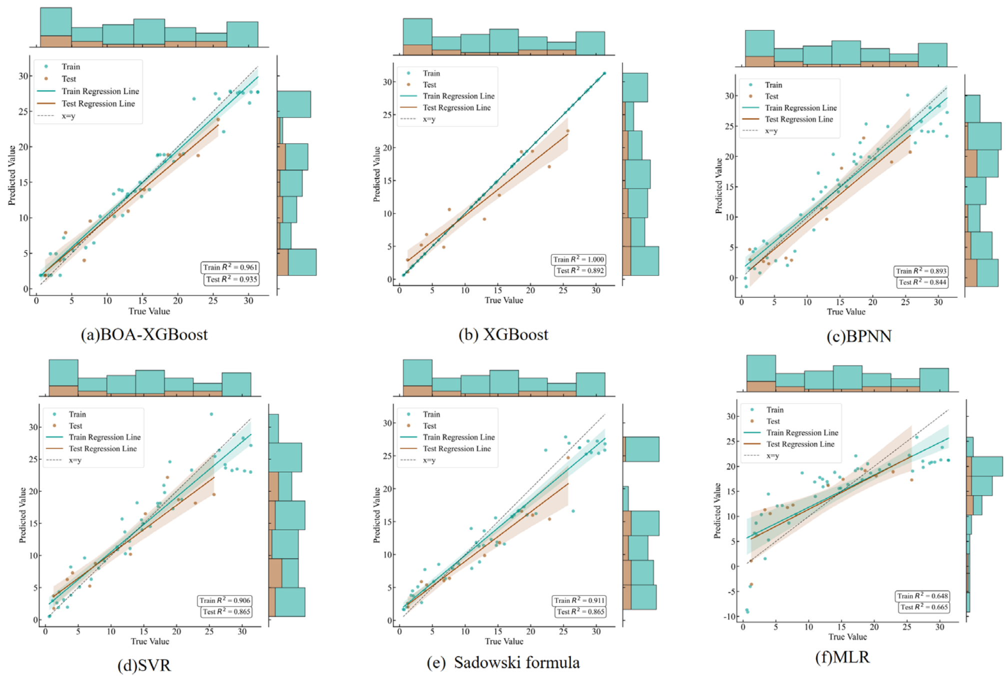

Figure 6: Model Training and Testing Results

ISSN 2616-5767 Vol.8, Issue 1: 1-7, DOI: 10.25236/AJETS.2025.080101

4.3. Model Evaluation

A statistical summary of the R^2, *RMSE*, and *MAE* results for all models is presented in Table 2. From Table 2, it can be observed that the XGBoost model outperforms the BP neural network, SVR, empirical formulas, and MLR models in terms of predictive performance. The standalone XGBoost model exhibits much better performance on the training set compared to the test set, showing a clear overfitting issue. After Bayesian optimization, the performance of the XGBoost model on the training set decreases, while its predictive ability on the test set improves, leading to a significant enhancement in the model's overall stability and generalization ability.

Table 2: Comparison of performance metrics for different models

Model	Training Set			Test Set		
	R^2	*RMSE*	*MAE*	R^2	*RMSE*	*MAE*
BOA-XGBoost	0.961	1.907	1.516	0.935	2.143	1.781
XGBoost	1	0.003	0.002	0.892	2.758	2.341
BPNN	0.893	3.151	2.566	0.844	3.312	2.991
SVR	0.906	2.950	2.048	0.865	3.085	2.655
Sadowski formula	0.911	2.876	2.285	0.865	3.078	2.437
MLR	0.648	5.706	4.812	0.665	4.852	4.132

5. Interpretability Analysis

The SHAP algorithm package in Python was used to perform interpretability analysis on the BO-XGBoost model. The feature importance bar chart and summary plot are shown in Figure 7. The feature importance bar chart displays the contribution of each input feature to the model's prediction. The summary plot, combining the advantages of feature importance and feature effect plots, intuitively shows the distribution of SHAP values for each feature. The scatter points range in color from blue to red, indicating the change of feature values from low to high. Each point represents the SHAP value of a sample, and the horizontal axis shows the size of the SHAP value, reflecting the strength of the feature's impact on the prediction result. The further the point is from zero, the more significant the impact of the feature on the model output. Positive SHAP values indicate a positive impact, while negative SHAP values indicate a negative impact.

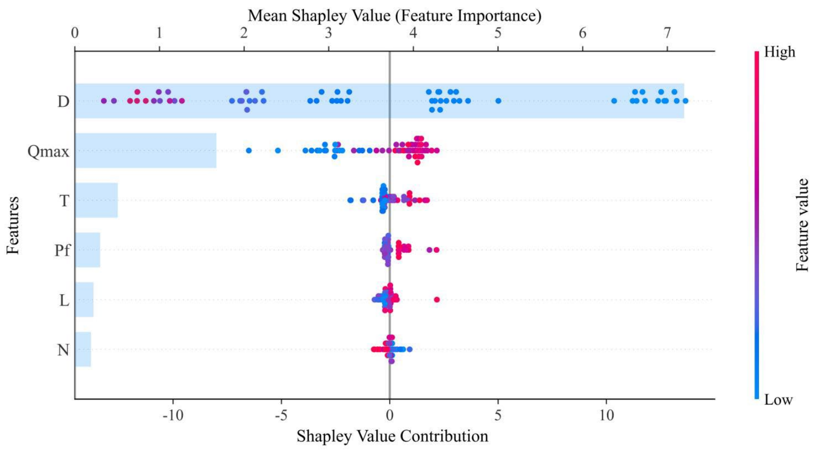

Figure 7: Summary chart

From Figure 7, it can be seen that the importance of the input features in influencing the model output is ranked as follows: $D > Q_{max} > T > Pf > L > N$. The blasting center distance (D) is the most important feature influencing the model's prediction, with a significant inverse relationship with the predicted Peak Particle Velocity (PPV). As the blasting center distance increases, the corresponding SHAP value decreases, and the predicted PPV also decreases. This relationship is highly consistent with the actual physical law, as the vibration energy gradually dissipates with the increase in distance from the blast. As a specific manifestation of the explosive energy, the PPV decreases as the blast center distance increases.

The importance of the maximum charge per blast section(Q_{max}), the powder factor (Pf)and stemming length (T) on the model output decreases sequentially. The common characteristic of these features is that as their values increase, the corresponding SHAP value increases, leading to a higher predicted PPV. This can be explained by the fact that an increase in $Q_{\max}$ and Pf represents a higher initial explosive energy, which leads to an increase in PPV. An appropriate stemming length can prevent the explosive gases from escaping prematurely, allowing the borehole to maintain high pressure for a longer period, which significantly enhances the pressure from the expanding gases and prolongs their effect on the borehole wall.

The impact of hole depth (L) on the model output is unclear, with the sample scatter points mainly concentrated around zero. The number of holes (N) has the least impact on the model's prediction. As the number of holes increases, the SHAP value predicted by the model decreases, and the predicted PPV also decreases. This is because a staggered firing pattern is used in the field blasting design, with each hole representing a separate blasting section. The increase in the number of holes implies an increase in the number of detonation segments, which to some extent acts as an interference damping, weakening the peak vibration velocity.

6. Conclusions

The Bayesian optimization algorithm was used to optimize the hyperparameters of XGBoost, improving the overfitting issue caused by improper hyperparameter selection, and enhancing the model's prediction accuracy and generalization ability.Compared with the other five models, BOA-XGBoost has higher prediction accuracy and stronger nonlinear fitting performance, indicating the effectiveness and feasibility of this model in blasting vibration PPV prediction.The importance of each feature in influencing the blasting vibration PPV is ranked as follows: $D > Q_{\max} > T > Pf > L > N$. The blasting center distance and the number of holes have an inverse effect on PPV, while maximum charge weight per blast section,stemming length, and the powder factor have a positive effect on PPV.

References

[1] Gou Y, Shi X, Huo X, et al. Motion parameter estimation and measured data correction derived from blast-induced vibration: New insights[J]. Measurement, 2018, 135.

[2] HE L, YIN L, ZHONG D W, et al. Research review on blast vibration intensity, waveform and spectrum:prediction and active Control[J]. Blasting, 2024, 41(03):189-204+262.

[3] HU X B, CHEN Z Y, WEI G P, et al. Blasting Vibration Prediction System Based on BP Neural Network[J]. Mining Research and Development, 2020, 40(09):154-158.

[4] Yue Z W, Wu Y X, Wei Z, et al. Prediction of blasting vibration effect in open-pit mine based on PSO-LSSVM model[J]. Engineering Blasting, 2020, 26(6):1-8.

[5] WANG X Y, CAO P F, XIAO Y Q, et al. Blasting vibration prediction based on WOA-SVM model[J]. Mining and Metallurgical Engineering, 2023, 43(4):48-51.

[6] Chen T, Guestrin C. Xgboost:A scalable tree boos-ting system[C]//Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, 2016:785-794.

[7] Snoek J, Larochelle H, Adams R P. Practical Bayesian Optimization of Machine Learning Algorithms [J]. Advances in neural information processing systems, 2012, 4.

[8] Zhou S, Z. -X. Z, Luo X. Huang Y. Yu Z. Yang X. Predicting dynamic compressive strength of frozen-thawed rocks by characteristic impedance and data-driven methods[J]. Journal of rock mechanics and geotechnical engineering, 2024, 16(7):2591-2606.

[9] Kang J Y, Zhang S X, Zhang Q P, et al. Transformer fault diagnosis method based on ANOVA and BO-SVM[J]. High Voltage Technology, 2023, 49(05):1882-1891.

[10] Yu Z, Shi X, Zhou J, et al. A new multikernel relevance vector machine based on the HPSOGWO algorithm for predicting and controlling blast-induced ground vibration[J]. Engineering with Computers, 2020.

[11] Chelgani S, H. Nasiri, M. Alidokht. Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A “conscious-lab” development[J]. International Journal of Mining Science and Technology, 2021, 31(06):1135-1144.