Research article

Prediction of blast-induced air overpressure using a hybrid machine learning model and gene expression programming (GEP): A case study from an iron ore mine

Mohammad Mirzehi Kalateh Kazemi[1], Zohreh Nabavi[1] and Manoj Khandelwal[2,*]

[1] Department of Mining Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran
[2] Institute of Innovation, Science and Sustainability, Federation University Australia, Ballarat, VIC 3350, Australia

* Corresponding: Email: m.khandelwal@federation.edu.au; Tel: 61353279821.

Abstract: Mine blasting can have a destructive effect on the environment. Among these effects, air overpressure (AOp) is a major concern. Therefore, a careful assessment of the AOp intensity should be conducted before any blasting operation in order to minimize the associated environmental detriment. Several empirical models have been established to predict and control AOp. However, the current empirical methods have many limitations, including low accuracy, poor generalizability, consideration only of linear relationships among influencing parameters, and investigation of only a few influencing parameters. Thus, the current research presents a hybrid model which combines an extreme gradient boosting algorithm (XGB) with grey wolf optimization (GWO) for accurately predicting AOp. Furthermore, an empirical model and gene expression programming (GEP) were used to assess the validity of the hybrid model (XGB-GWO). An analysis of 66 blastings with their corresponding AOp values and influential parameters was conducted to achieve the goals of this research. The efficiency of AOp prediction methods was evaluated in terms of mean absolute error (MAE), coefficient of determination (R^2), and root mean square error (RMSE). Based on the calculations, the XGB-GWO model has performed as well as the empirical and GEP models. Next, the most significant parameters for predicting AOp were determined using a sensitivity analysis. Based on

the analysis results, stemming length and rock quality designation (RQD) were identified as two variables with the greatest influence. This study showed that the proposed XGB-GWO method was robust and applicable for predicting AOp driven by blasting operations.

Abbreviations

Symbol	Explanation	Symbol	Explanation
MC	Maximum charge per delay	D	Distance from the blast point
RQD	Rock quality designation	B	Burden
S	Spacing	ST	Stemming length
H	Depth of the blast holes	RH	Air humidity
PF	Powder factor	NH	Number of holes
W	Air humidity	ICA	Imperialist competitive algorithm
BI	Blasting index	GP	Genetic programming
RMR	Rock mass rating	CA	Cubist algorithm
JA	Joint aperture	ANN	Artificial neural network
PSO	Particle swarm optimization	SFS	Stochastic fractal search
GA	Genetic algorithm	ANFIS	Adaptive neuro-fuzzy inference system
RF	Random forest algorithm	LM	Levenberg–Marquardt
CFNN	Cascaded forward neural network	PNN	Polynomial neural network
M5'	One of the decision tree algorithms	GBM	Gradient boosting machine
BRNN	Bayesian regularized neural networks	HYFIS	Hybrid neural fuzzy inference system
USBM	US Bureau of Mines		

1. Introduction

Drilling and blasting is the most common, most economical, fastest and cheapest method for fragmentation of rocks in mining operations [1]. Each explosion releases a huge amount of energy in the form of temperature and pressure. Meanwhile, moving and fragmenting take up a small amount of power. Various adverse environmental effects are caused by the remaining energy, such as flyrock, air overpressure, dust, ground vibration and noise [2–6]. Gases resulting from blasting contain unused energy that increases air pressure above normal. This phenomenon is known as air overpressure (AOp) [7]. As a destructive result of mine blasting, AOp can cause damage to nearby structures [8]. Consequently, accurate predictions of such destructive effects are necessary to increase blasting safety.

The intensity of AOp is affected by three groups of parameters, namely, blast design parameters, rock mass properties and explosive properties [9]. The second group consists of uncontrollable parameters, while the first and third groups include controllable parameters. Controllable parameters in group 1 include blasting type and blasting pattern design parameters. Furthermore, the characteristics of explosives (classified in group 3) like their type, power and density can be controlled

by engineers. Uncontrollable parameters in the second group consist of the rock mass's RQD and compressive, shear and tensile strengths. These mainly depend on the complex geological and geotechnical conditions of the rock mass.

Remennikov and Rose [10] have proposed solutions such as improving the structure of buildings and glass doors or using barriers as shields to reduce the effects of AOp. However, their efficacy was not significantly proven [11]. With similar objectives, experimental methods are also presented to calculate AOp. Also, Khandelwal and Kankar [12] and Armaghani et al. [13] showed that empirical equations have poor performance. In addition to low accuracy, conventional tools also have many limitations including studying specific areas, considering linear relationships between influence parameters and focusing on explosive charges per delay and monitoring distances [13].

Over the years, machine learning and artificial intelligence (AI) methods have been demonstrated to be acceptable and reliable for solving different engineering subjects, especially for prediction and optimization purposes [14–19]. For example, AI can be used to analyze the results of a blast, using machine learning algorithms to identify patterns, and then this data can be used to adjust the blast design for future operations [20,21]. Prediction of AOp has also been conducted using these methods. The artificial neural network (ANN) proposed by Khandelwal and Singh [22] was compared with regression methods to predict AOp. Results showed that ANN performed similarly to regression methods in predicting AOp with higher precision. Khandelwal and Kankar [12] for predicting AOp investigated and confirmed the efficacy of support vector machine (SVM). Nguyen et al. [23] evaluated and compared different ANN systems including ANN, BRNN and HYFIS to predict blast-induced AOp. Similarly, a heuristic algorithm, intended to optimize ANNs for predicting AOp, was used successfully by Armaghani et al. [24]. Also, Hasanipanah et al. [25] developed a hybrid SVR model that was optimized with PSO to predict AOp.

Several soft computing models were presented by Bui et al. [4] for predicting AOp, including SVM, ANN, boosted regression trees, k-nearest neighbors and RF. In another study, Zhou et al. [26] developed a hybrid model which combines a fuzzy system (FS) and the firefly algorithm (FFA). Their study demonstrated that FFA-FS could be used to predict AOp efficiently. Tran et al. [27] have investigated the effect of meteorology on AOp, using AOp prediction models. Their results show that meteorological conditions, especially wind speed and air humidity, have a noteworthy impact on blast-induced AOp. Zeng et al. [28] have developed an efficient method based on the Levenberg-Marquardt (LM) algorithm and cascaded forward neural network (CFNN) to predict AOp. In addition, the accuracy level of this technique has been tested using the generalized regression neural network (GRNN) and extreme learning machine (ELM). Table 1 summarizes some of the relevant previous studies conducted by various researchers.

Blasting operations have a range of effects on the environment, including air overpressure (AOp) [29–31]. Therefore, AOp prediction with high accuracy is essential to determining the safe zone around an operation site. In many studies, machine learning has been applied to evaluate and predict the adverse consequences of blasting. However, the studies have not addressed the evaluation and prediction of air overpressure via hybrid extreme gradient boosting (XGB). Thus, this article develops a hybrid model which combines extreme gradient boosting (XGB) with grey wolf optimization (GWO) to predict AOp in open pit iron mines accurately. Furthermore, gene expression programming (GEP) and an empirical model were used to assess the validity of the

hybrid model (XGB-GWO). Additionally, unlike other methods, GEP can create a simple mathematical expression that can be applied to different mining conditions, which is another advantage of this article.

Table 1. Some recent work with machine learning techniques for air overpressure prediction.

Research	Input	Method	R^2
Fang, et al. [32]	MC, PF, B, ST, S, D	GA-CA	0.968
Ye, et al. [3]	H, PF, MC, ST, S, B, RQD, NH, D	ANFIS-GA	0.945
		ANFIS-PSO	0.963
		ANFIS-SFS	0.987
Zeng, et al. [28]	PF, MC, ST, S, B, RQD, D	CFNN-LM	0.92
Murlidhar, et al. [34]	BI, JA, ST, PF, MC, D	GP, M5'	0.85
			0.88
Hajihassani, et al. [35]	H, PF, MC, ST, S, B, RQD, NH, D	PSO-ANN	0.94
Nguyen and Bui [36]	MC, D, P, B, S, ST	ANNs-RF	0.985
Harandizadeh and Armaghani [37]	B, S, ST, PF	ANFIS-PNN-GA	0.94
Nguyen, et al. [38]	MC, H, B, S, ST, PF, NH, RH, D	GBM	0.95
Nguyen, et al. [23]	MC, H, BS, ST, PF, NH, W, D	MLP	0.96
		BRNN	0.93
		HYFIS	0.80
AminShokravi, et al. [39]	MC, D, RMR	PSO	0.92
Armaghani, et al. [13]	MC, D	ICA-ANN	0.90
Khandelwal and Kankar [12]	MC, D	SVM	0.85

Note: Symbols are explained in the Abbreviations section.

2. Methodology

Machine learning models require parameters based on datasets. Many studies have shown that heuristic algorithms improve machine learning accuracy and stability [40,41]. Therefore, this study applied a hybrid predictive approach based on ideas from the extreme gradient boosting framework (XGB) as well as metaheuristic algorithms: namely, grey wolf optimization (GWO). Then, gene expression programming (GEP) and empirical models were employed to assess the validity of the optimized model (XGB-GWO). By using GWO we can find the optimal values of hyperparameters of the regression model. Therefore, the intelligent optimization algorithm by the adjustment of three key parameters of the XGB model (n_estimator, maximum_depth and learning_rate) achieves higher accuracy. N_estimator, maximum_depth and learning_rate represent the number of trees, the maximum depth of a tree and the shrinkage coefficient of the tree, respectively.

2.1. Extreme gradient boosting (XGB)

Chen et al. [42] proposed a method based on gradient boosting [43–46]. Many engineering fields have implemented XGBoost for classifying and predicting problems [47]. It has performed well

because of the advantages of parallel processing, regularization and efficient tree pruning. XGBoost can solve a wide variety of data science problems quickly and accurately with parallel tree boosting. The core of this algorithm is optimizing the objective function [48].

In each iteration, XGBoost calibrates the previous predictor using the residual. Loss function optimization (LOF) is involved in this process. However, regularization is applied to the objective function during calibration to reduce overfitting. Using this description, Equation (1) describes the objective function as two parts: regularization and training loss.

$$Obj(\Theta) = \Omega\,(\Theta)\, +\, L(\Theta), \tag{1}$$

In Equation (1), Θ and Ω represent the parameters trained from the data and related to regularization, respectively. Regularization controls the complexity of the model in order to avoid overfitting [49]. L denotes the training loss function, which measures the model's fit to the training data. Complexity can be defined in a variety of ways. Equation (2), however, is often used to calculate the complexity of each tree.

$$\Omega\,(\mathcal{F}) = \,\gamma T + \frac{1}{2}\,\lambda\,\sum_{j=1}^{T}\omega_j^2, \tag{2}$$

where γ denotes each leaf's complexity, T represents the number of DT leaves, ω is the vector of scores on the leaves and λ scales the penalty. Next, XGBoost combines loss function (LOF) general gradient boosting with the second-order Taylor expansion. When the mean squared error is taken into account, equation (3) can obtain the objective function.

$$J^{(t)} \,\approx\, \gamma T + \sum_{j=1}^{T}\left[\left(\sum_{i\in i_j}g_i\right)\omega_j + 0.5\left(\sum_{i\in i_j}h_i + \lambda\,\right)\omega_j^2\right], \tag{3}$$

The MSE loss function's first and second derivatives are g_i and h_i, respectively. Also, the q function is used to assign data points to leaves. As a final step, we calculate the XGBoost objective function by referring to Equation (4).

$$Obj = \sum_{j=1}^{T}\left[\,G_j\omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2\right] + \gamma T, \tag{4}$$

Here ω_j are independent of each other. The definitions of the two terms G_j and H_j are given in Equation (5).

$$G_j = \sum_{i\in i_j}g_i\,,H_j = \sum_{i\in i_j}h_i, \tag{5}$$

In general, finding the minimum of a quadratic function can be applied to optimizing the objective function. In order to evaluate the performance of the model after splitting a particular node in DT, an objective function is used. If the model performs better than before, this division will be accepted. Otherwise, the division will come to a halt. Due to the inclusion of regularization phenomena, XGBoost can prevent over-installation more effectively [50].

2.2. Grey wolf optimization (GWO)

GWO is a metaheuristic optimization algorithm proposed by Mirjalili et al. [51] which reflects grey wolf family social systems in nature. Social hierarchy is very strict in the grey wolf community. Usually, wolf levels are divided into four categories: α, β, δ and ω. α is primarily responsible for making overall decisions at the first level, and the β wolf assists the α wolf at the second level. At the third level, δ must follow the decisions of α and β wolves. Furthermore, the lowest rank for wolves is ω. Also, this group must follow wolves with a higher rank. Mathematical models can be developed, and optimal solutions can be found based on wolves' hunting process and social hierarchy.

α represents the optimal solution, β denotes the second, and δ shows the third optimal solution. All other applicant solutions are denoted by ω. In the wolf pack, D indicates the distance between the prey and the individual (Equations (6) and (7)).

$$D = |CX_P(t) - X|, \tag{6}$$

$$C = 2r_1, \tag{7}$$

where C, D and $X_P(t)$ are the step length coefficient, the current number of iterations and prey location, respectively. Also, r_1 and X are a random number ranging in (0,1) and the location of a grey wolf, respectively.

A pack of individuals tries to shorten the distance between their prey and themselves by constantly updating the following Equations (8) and (9):

$$X(t - 1) = X_P(t) - A.D, \tag{8}$$

$$A = 1ar_2 - a, \tag{9}$$

where r_2 and A are a random number ranging in (0,1) and the convergence influence factor, respectively. A decreases linearly with the number of iterations from 2 to 0.

Since α, β and δ possess high levels of intelligence, they can carry more information about where prey can be found. This will enable them to lead the group to the hunting grounds gradually. Gradually, the pack makes its way to the hunting grounds. As a result, three optimal solutions will now be available to ignore or find other solutions, and gradually they will find the global optimal solution based on the three optimal solutions, which are outlined in Equations (10) to (12).

$$D_\alpha = |C_1 X_\alpha(t) - X|, \tag{10}$$

$$D_\beta = |C_1 X_\beta(t) - X|, \tag{11}$$

$$D_\delta = |C_1 X_\delta(t) - X|. \tag{12}$$

A decision will be made as to which individuals to include in the remaining pack relying on a joint decision by α, β and δ. It is now necessary to move the position shown in Equation (13).

$$X(t-1) = \frac{X_1 + X_2 + X_3}{3},\tag{13}$$

It is possible to summarize the grey wolf optimization algorithm in this way: during the optimization process, it constantly updates the solution area for the location search problem and finds the ideal solution at the end (Figure 1).
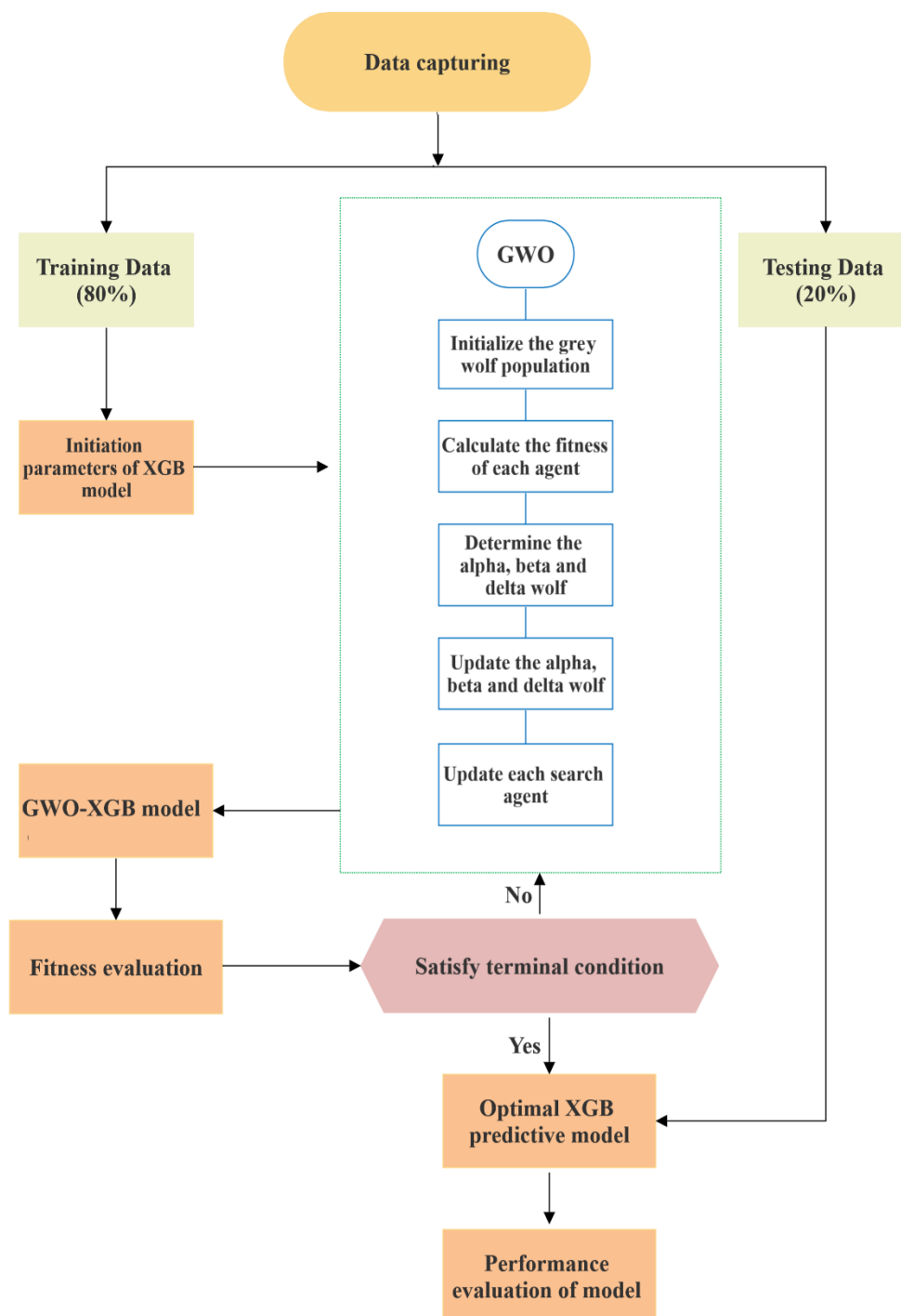

Figure 1. Structure of hybrid method based on XGB and GWO for predicting air overpressure.

2.3. *Gene expression programming (GEP)*

The GEP method was introduced by Ferreira [52] and combines genetic programming (GP) and genetic algorithms (GA). As an AI developmental approach, GEP has corrected some of GA's and GP's weaknesses, such as tree systems. Both of these methods differ in the following ways: the concept of individuals or answers. In the GEP algorithm, individuals are defined as binary fixed-length chromosomes. Answers from the GP method can include tree systems of varying sizes. Because GEP combines GA and GP, it produces chromosomes of fixed length and tree systems with different sizes and shapes, identified as an expression tree (ET). The GEP algorithm's structure consists of several elements, including function, terminal, operator, fitness, and stop criteria [52].

Chromosomes are composed of two parts: a head and a tail, making them a fixed length. Functions and terminals are included in the head part, and terminals are also included in the end part. As a result of the complexity of the problems, a specific equation cannot be used to calculate the head part's length. This term is defined as the input of the GEP method. Trial and error are one of the solutions [53]. Furthermore, the tail part's length can be calculated using Equation (14).

$$t = h(n_{max} - 1) + 1, \tag{14}$$

where n_{max}, h and t represent the number of arguments of the functions and the length of the head and tail.

Solitary chromosomes in the initial population are evaluated according to a fitness function developed for gap problems. A number of genetic operators are used to adapt the considered chromosomes. Depending on the problem conditions, each chromosome may contain functions, constants and terminals [54,55]. The following is a general description of the GEP algorithm:

Step 1: Depending on the problem's conditions (size) being studied, a certain number of chromosomes should be determined (randomly).

Step 2: Expression trees and mathematical equations are used to select the initial population chromosomes.

Step 3: The chromosomes are fitted according to the overall fitness function (RMSE or R^2). Alternative methods, such as the roulette wheel method, are used if the stopping criterion is not met.

Step 4: During this step, the GEP algorithm's genetic operators, known as the core, must be linked to the rest of the chromosomes.

Step 5: Lastly, the process of creating the next generation begins, and the process is repeated until new structures are created.

In order to decode the programs on the chromosomes, Karva (K-Expression) was invented to express the codes. Inversions, mutations, triple recombination operators and triple transposition operators have been introduced so far as genetic operators that are used for chromosome modification [54]. The structure of the GEP method is illustrated in Figure 2.
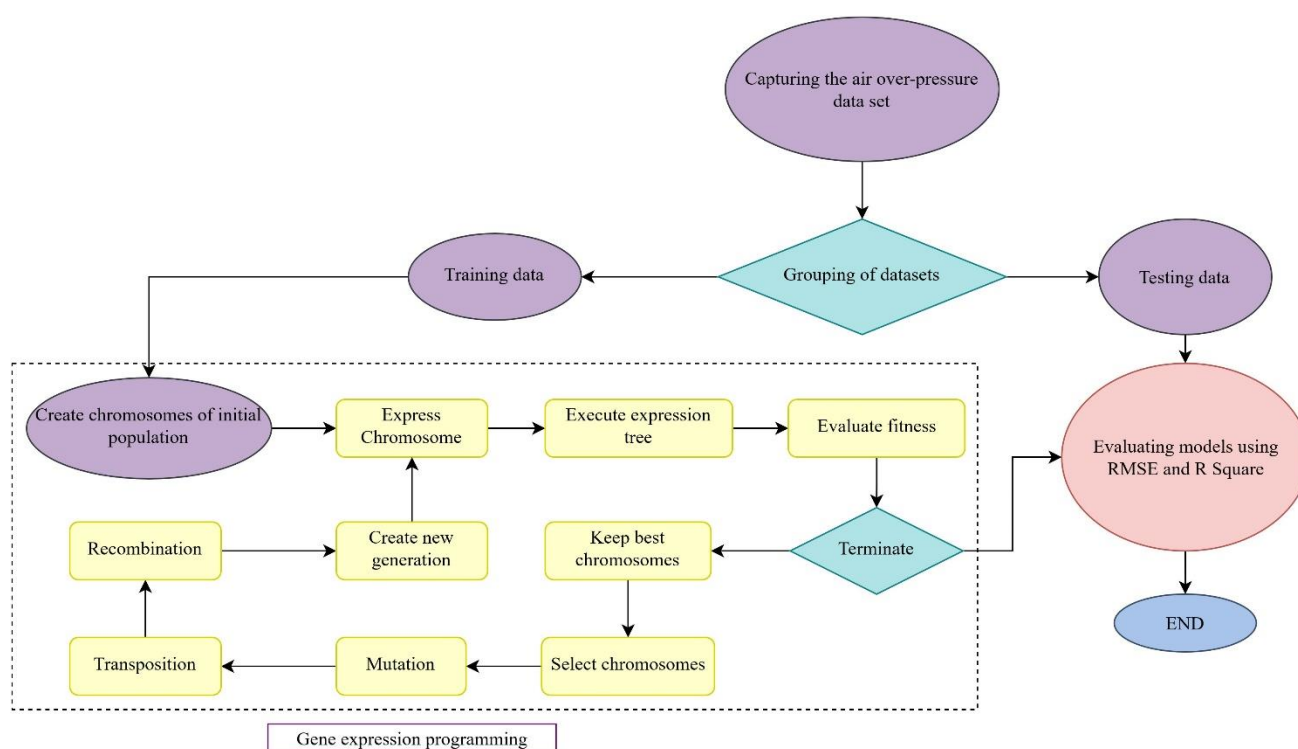
Figure 2. Structure of the GEP method for prediction of air overpressure.

2.4. Empirical

In open-pit mines, the blast-induced AOp has been estimated using empirical methods. The technique proposed by the United States Bureau of Mines (USBM) is the most widely used empirical equation for predicting AOp among all available methods. Kuzu et al. [56] calculated the scaled-distance (SD) factor for AOp predictions. USBM is an empirical method for predicting AOp based on site factors, the maximum charge per delay (MC) and monitoring distance (D) in open-pit mines. SD values determine the relationship between MC and D [56]:

$$SD = DMC^{-0.33} \tag{15}$$

The equation recommended by USBM has been widely adopted by researchers and is expressed in the following manner [57]:

$$AOp = \gamma(SD)^{-\alpha} \tag{16}$$

By regression analysis, AOp in decibels (dB) is calculated, and γ and α are site factors.

2.5. Model verification and evaluation

Evaluation and verification of models are essential steps during model development. In order to determine whether a model is of high quality and whether the results produced are adequate for the goals pursued, it is necessary to test it. In this study, training and testing sets are used to train and verify predictive models. Predicted and actual values are compared using relevant evaluation indicators.

In this study, hybrid models will be evaluated to ensure their reliability [58]. R^2, RMSE, and MAE are evaluation indicators.

As defined in Equation 17, between actual and predicted values, R^2 represents the square of the correlation. Further, the RMSE represents the standard deviation between predicted and actual values (Equation 18). Equation 19 defines mean absolute error as the measure of error between paired observations describing the same phenomenon [59].

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{17}$$

$$RMSE = \sqrt{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2 / N} \tag{18}$$

$$MAE = \frac{\sum_{i=1}^{n} \hat{y}_i - y_i}{n} \tag{19}$$

3. Case study and data collection

This study analyzes the case of Chadormalu, a large open-pit iron mine located 120 kilometers northeast of Yazd, Iran. Figure 3 depicts the location of the Chadormalu mine in Iran in relation to Tehran (the capital of Iran) and Yazd. It is estimated that there are about 400 million tons of mineable ore reserves. Magnetite and hematite are the main components of the deposit, according to mineralogical studies. Mine blasting is primarily carried out with ANFO explosives.

Figure 3. The location of the Chadormalu mine in Iran relative to Tehran (the capital) and Yazd city.

In the current study, input and output parameters were used in developing the model. The influential input parameters used for air overpressure (AOp) prediction are given in Table 2. Burden (B), number of holes (NH), hole depth (H), spacing (S), powder factor (PF), distance (D), rock quality

designation (RQD), stemming (ST) and the maximum charge per delay (MC) have been employed as the output parameters of the models to predict air overpressure (AOp).

Data were collected as follows: D was determined by handheld GPS, and AOp values from blasting operations were recorded with a sound level meter (SLM). Using blast design, the remaining parameters were collected. Table 2 shows the influential input and output parameters' details, symbols and statistical descriptions.

For model development, 66 data pairs were used to create a database related to air overpressure (AOp). From the organized database, 20% of the data set was selected to test the model to ensure consistency.

Table 2. Input and output data with details, symbols and statistical descriptions.

Type of data	Variable	Symbol	Minimum	Maximum	Mean	Std. Deviation
Input	Burden (m)	B	1.5	1.7	2.29	0.6
	Hole depth (m)	H	10	24	15.41	3.95
	Spacing (m)	S	2.8	4	3.37	0.4
	Powder factor (kg/m^3)	PF	0.34	0.66	0.487	0.13
	Distance (m)	D	300	600	522.73	132
	Rock quality designation (%)	RQD	60	89	74	9.51
	Stemming (m)	ST	1.7	2.5	2.14	0.24
	Number of holes	NH	12	59	35.14	12.4
	Maximum charge per delay (kg)	MC	60	160	89.58	31.84
Output	Air overpressure (dB)	AOp	89.1	122.6	107.82	8.77

4. Results and discussion

Predicting air overpressure requires the preparation of a database. This database was divided into training and test sets using the most common division ratio of 80%–20% [60]. The air overpressure prediction model was evaluated using several performance indicators, including R^2, RMSE and MAE. All predictive models use the same training and test data sets.

4.1. Hybrid model result (XGB-GWO)

In order to avoid complexity in XGBoost modeling, three stopping criteria were considered, namely, n estimators, learning rate and maximum depth. If significant values are assessed for each parameter, overfitting can occur. Therefore, XGB parameters have been optimally determined using grey wolf optimization (GWO). Figure 1 illustrates the method used to develop models based on the XGB. It was time to set the XGB model's parameters. The optimization algorithm's parameters are

shown in Table 3. Additionally, Table 3 shows the optimal parameters of the model based on the optimization process.

Table 3. The optimization algorithm's parameter and optimal parameters for the hybrid model.

Algorithm	Parameters	Value	Optimal parameters
GWO	Convergence parameter	[2,0]	learning_rate = 0.06
			maximum_depth = 1
			n_estimator = 204

Based on the training data set, Figure 4 shows the correlation between the actual values and the predicted values. The training data are largely scattered near the closest fit line for the intelligent model, showing that training effects remain relatively favorable. XGB-based optimization techniques have been proposed that demonstrate high training effects with R^2 values of 0.96. After the hybrid intelligent model has been trained, it is verified and evaluated using the testing data set. Figure 4 shows that the test data set is mostly distributed near the perfectly fitted line. Because the predicted AOp values and the actual AOp values are within the correlation, they can be classified as correlated. Hybrid models achieve high levels of prediction accuracy, with R^2 values of 0.98.
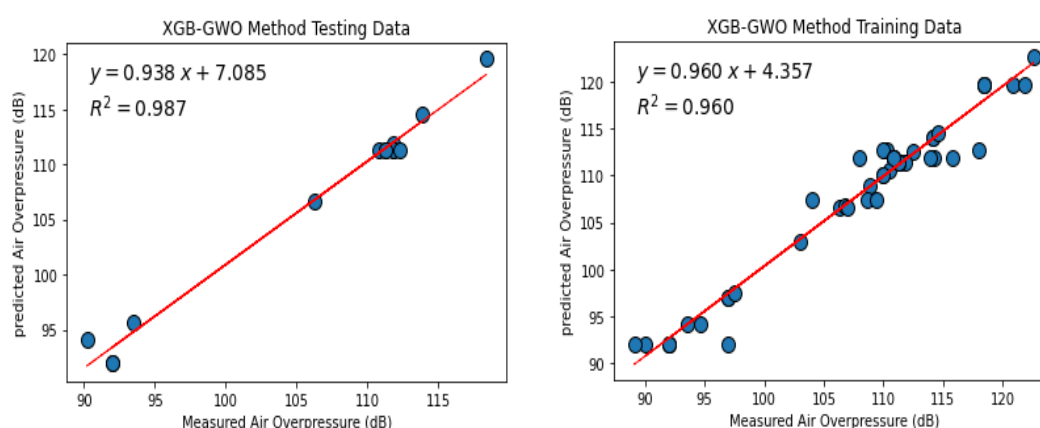

Figure 4. Actual values and predicted values on the testing and training dataset by XGB-GWO Method.

4.2. GEP model result

The flowchart shown in Figure 2 illustrates the gene expression programming (GEP) modeling process. The GEP was designed using the same testing and training datasets as the previous sections. The final relationship between the initial data and air overpressure was analyzed and determined by GeneXproTools (version 5.0).

To build an efficient model, it is imperative to consider the fitting parameters. The number of chromosomes determines how long it takes for the model to run, which is crucial to the model's performance. For a suitable architecture, the size of the head and the number of genes must also be considered. Each component's complexity and the number of related equations are determined by the

head size and the number of genes. To achieve the most suitable GEP model, a trial and error mechanism was applied (Table 4).

Table 4. Constructed 10 GEP models for predicting air overpressure.

GEP parameters	Value									
	GEP model number									
	1	2	3	4	5	6	7	8	9	10
Fitness function	MAE	MAE	RMSE	MAE	MAE	RMSE	RMSE	RMSE	MAE	RMSE
No. of chromosomes	40	40	40	30	40	40	40	40	40	40
Head size	8	8	8	8	8	8	8	8	8	8
No. of genes	3	3	3	3	4	4	4	5	5	4
Linking function	+	×	×	×	×	×	+	+	+	1

By using R^2, RMSE and MAE indices, for the testing and training data sets, the performance prediction of GEP models was evaluated. Of the ten models stated in Table 4, the five which had the most accurate predicting of air overpressure were selected (Table 5). Therefore, based on the results in Table 5, model No. 6 was the most accurate model among all models made by the GEP method. Figure 5 illustrates the scatter plot of the predicted and measured air overpressure in the selected GEP model by training and test data.

Table 5. Performance indices of GEP models.

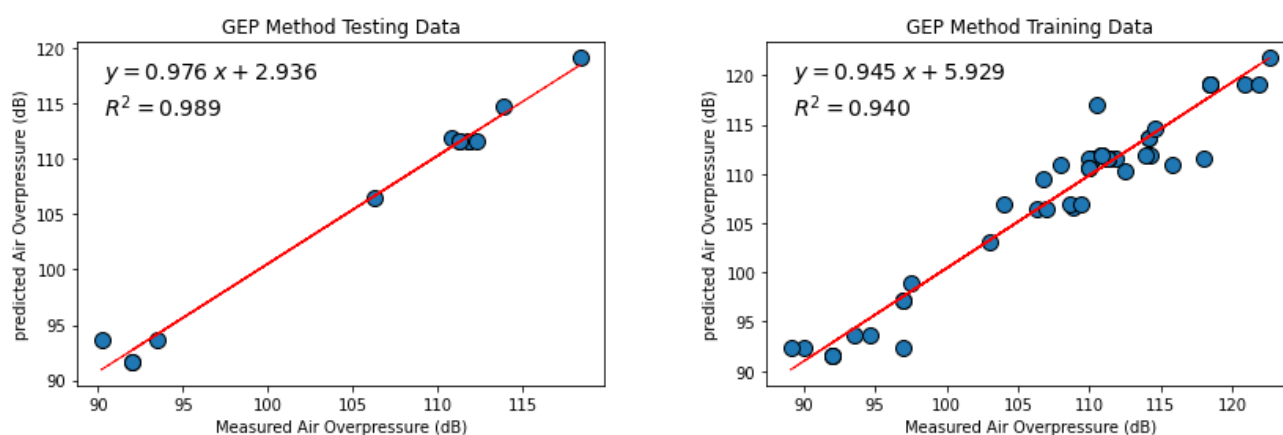| Method	Model no. Table (6)	Training			Testing			Rank
		R^2	RMSE	MAE	R^2	RMSE	MAE	
GEP	3	0.93	2.26	1.83	0.95	2.13	1.4	4
	6	0.94	2.08	1.42	0.99	1.04	0.63	1
	7	0.94	2.06	1.47	0.98	1.64	1.06	2
	9	0.92	2.56	1.68	0.98	1.35	0.84	5
	10	0.93	2.25	1.57	0.98	1.36	0.91	3

Figure 5. Measured and predicted air overpressure on the training and test dataset by selected GEP Method.

The number of genes and head size in model No. 6 are 8 and 4, respectively. The expression tree (sub-ETs of each gene) of this model is illustrated in Figure 6. In addition to the expression trees, functions were linked together to construct a large and complex ET. Expressions (20) to (23) are mathematical expressions of each of the genes. Finally, the developed GEP equation for predicting air overpressure is shown in Equation (24).

$$Gene1: MC + [Exp\ (6.26 - S) + 52.272] \tag{20}$$

$$Gene2: H + \left[\frac{1}{\left((6.11 * ST) - (H - 7.78)\right) - (B + 1.81)}\right] \tag{21}$$

$$Gene3: PF - \left[Ln\ (\ Ln\ ((-6.86 * NH)^2 - \ MC^2))\right] \tag{22}$$

$$Gene4: \left[\frac{\left(\left(\left(\frac{H - 6.89}{MC}\right) * S^2\right) - S\right)}{MC}\right] \tag{23}$$

$$
\begin{aligned}
Air\ &overpressure \\
= &\left[MC + [Exp\ (6.26 - S) + 52.272]\right] \\
* &\left[H + \left[\frac{1}{\left((6.11 * ST) - (H - 7.78)\right) - (B + 1.81)}\right]\right] \\
* &\left[PF - \left[Ln\ (\ Ln\ ((-6.86 * NH)^2 - \ MC^2))\right]\right] \\
* &\left[\frac{\left(\left(\left(\frac{H - 6.89}{MC}\right) * S^2\right) - S\right)}{MC}\right]
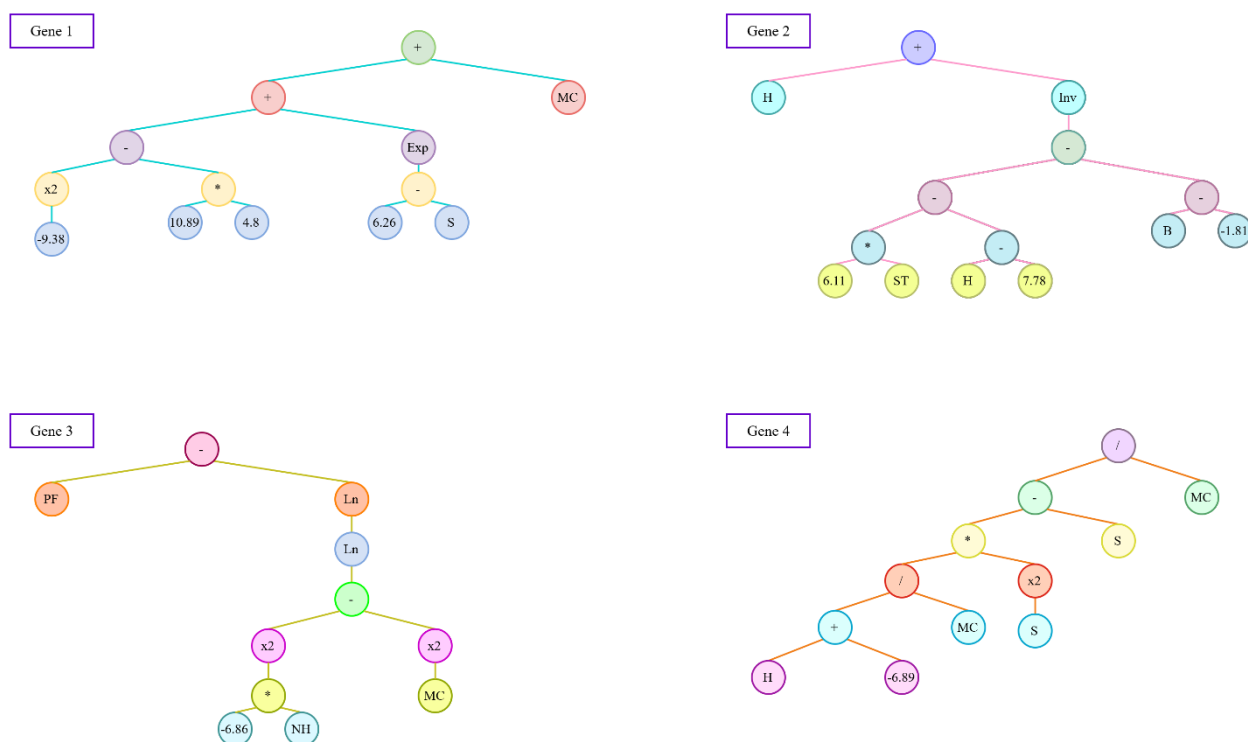\end{aligned} \tag{24}
$$

Figure 6. Sub-ETs of each gene for the best GEP model with addition as a linking function.

4.3. Empirical model result

In the empirical model, site factors γ and α were computed from 52 blasting events (training data set). A multivariate regression analysis was performed using Microsoft Excel 2019. Therefore, the optimal values of γ and α for the USBM model for predicting AOp are 173 and 0.1. Using the USBM model in this case, we can describe it as follows:

$$AOp = 173(SD)^{0.1} \tag{25}$$

Also, Figure 7 illustrates the scatter plot of the predicted and measured air overpressure in the empirical (USBM) model by training and test data.

4.4. Comparison of models and validation performance

The efficiency of the predictive models is evaluated in this section. In this study, R^2, RMSE and MAE (Equations 15 to 17) were employed to assess predictive models' performance. Regarding Table 6, the above statistical criteria and performance comparison for techniques were determined for the testing and training data sets. Based on these results, the correctness rate and performance of the hybrid XGBoost (XGB-GWO) method are preferable to the GEP and empirical models.
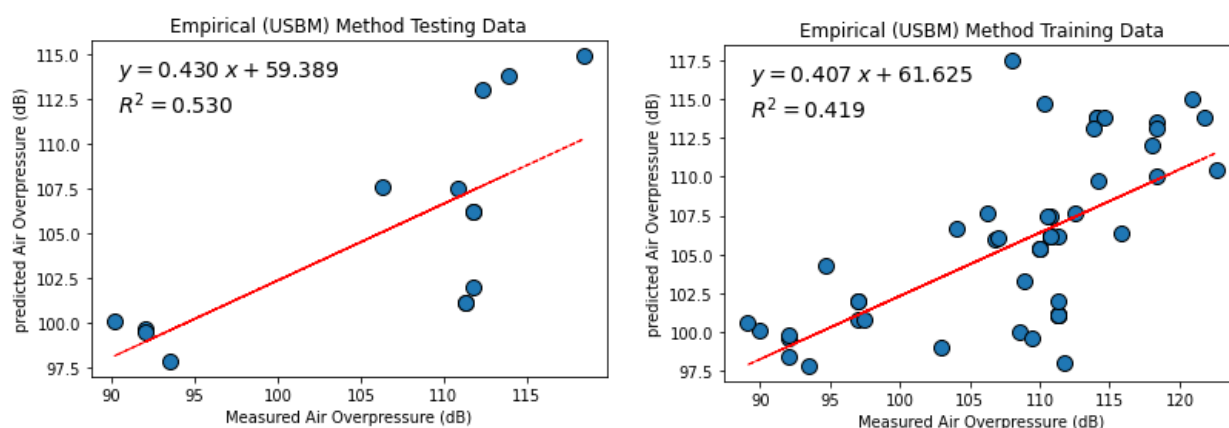
Figure 7. Measured and predicted air overpressure on the training and test dataset by Empirical Method.

Table 6. Statistics of predictive models and performance comparison.

Method	Training Data			Testing Data		
	R^2	RMSE	MAE	R^2	RMSE	MAE
XGB- GWO	0.96	1.87	1.02	0.983	1.42	0.69
GEP	0.94	2.06	1.47	0.989	1.04	0.63
Empirical	0.419	6.98	6.11	0.53	6.68	5.92

In the next step, the selected AOp prediction models' accuracies are compared, as shown in Figure 8. The accomplishment of the models in predicting the air overpressure in the testing and training data set is shown in Figure 8. Also, based on Figure 8, the GWO-XGB technique gives the most reliable and steady results in AOp prediction among the GEP and empirical models.
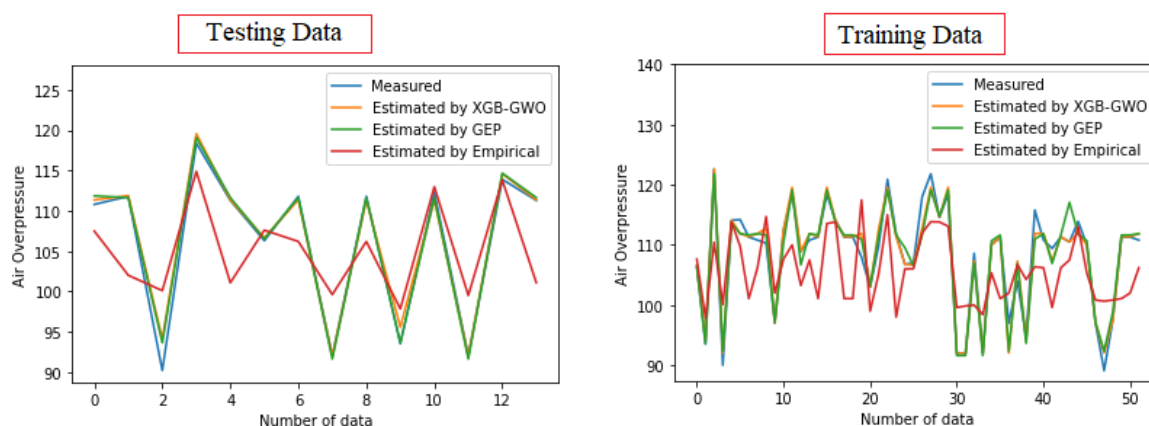

Figure 8. Predictive models with prediction values on training and testing datasets.

As shown in Figure 9, the box plots show the distribution functions corresponding to the measured and predicted AOp values. Based on an exhaustive analysis of Figure 9, the XGB-GWO approach yielded the most promising performance relative to the GEP model due to its similar probability distribution to observational results.
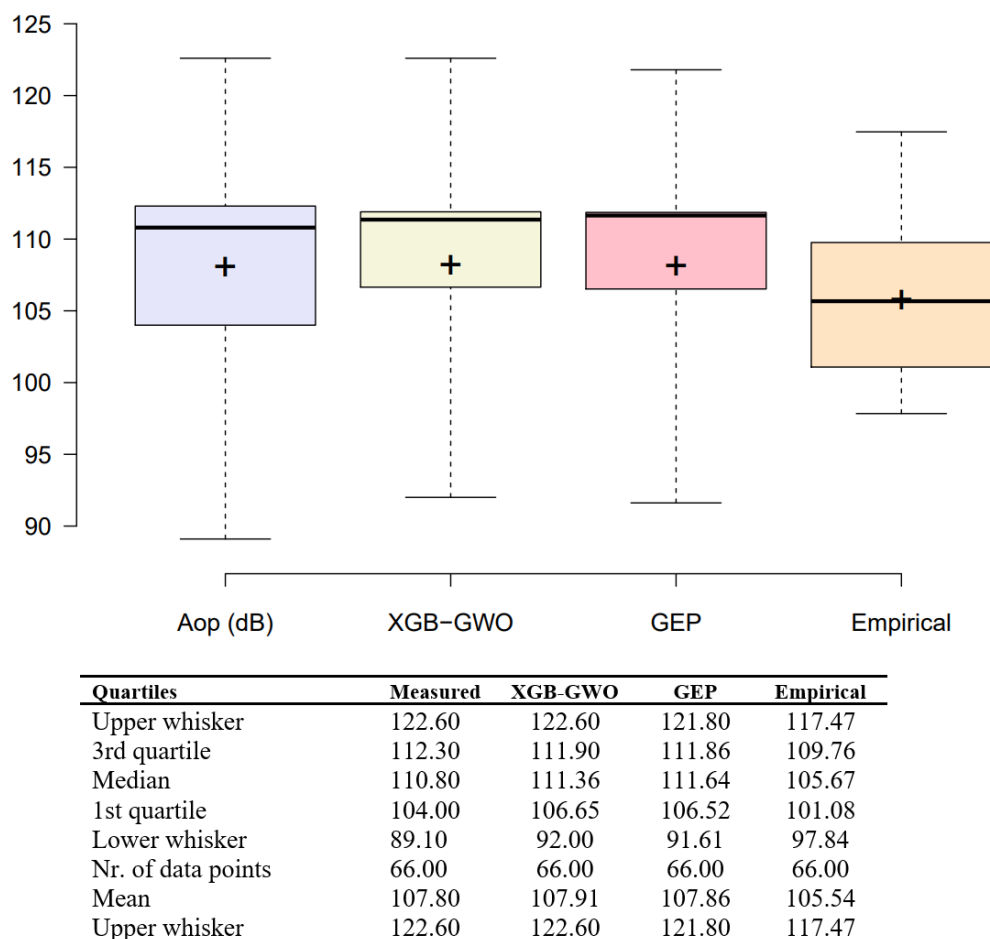

Quartiles	Measured	XGB-GWO	GEP	Empirical
Upper whisker	122.60	122.60	121.80	117.47
3rd quartile	112.30	111.90	111.86	109.76
Median	110.80	111.36	111.64	105.67
1st quartile	104.00	106.65	106.52	101.08
Lower whisker	89.10	92.00	91.61	97.84
Nr. of data points	66.00	66.00	66.00	66.00
Mean	107.80	107.91	107.86	105.54
Upper whisker	122.60	122.60	121.80	117.47

Figure 9. The box plot of predictive models (distribution function).

A Taylor diagram illustrating the performance of the predictive models is presented in this subsection. Taylor diagrams are used to assess the accuracy of models by showing them in two dimensions [61]. Indicators of the relationship between the actual and predicted observations are R, RMSE and standard deviation. Each model is denoted by a term in the Taylor diagram. In an ideal model, the position of the point should coincide with the reference point. Figure 10 illustrates the predictive models developed in this study. While both models are highly accurate at predicting air overpressure, Figure 9 shows that the hybrid XGB-GWO model is better.
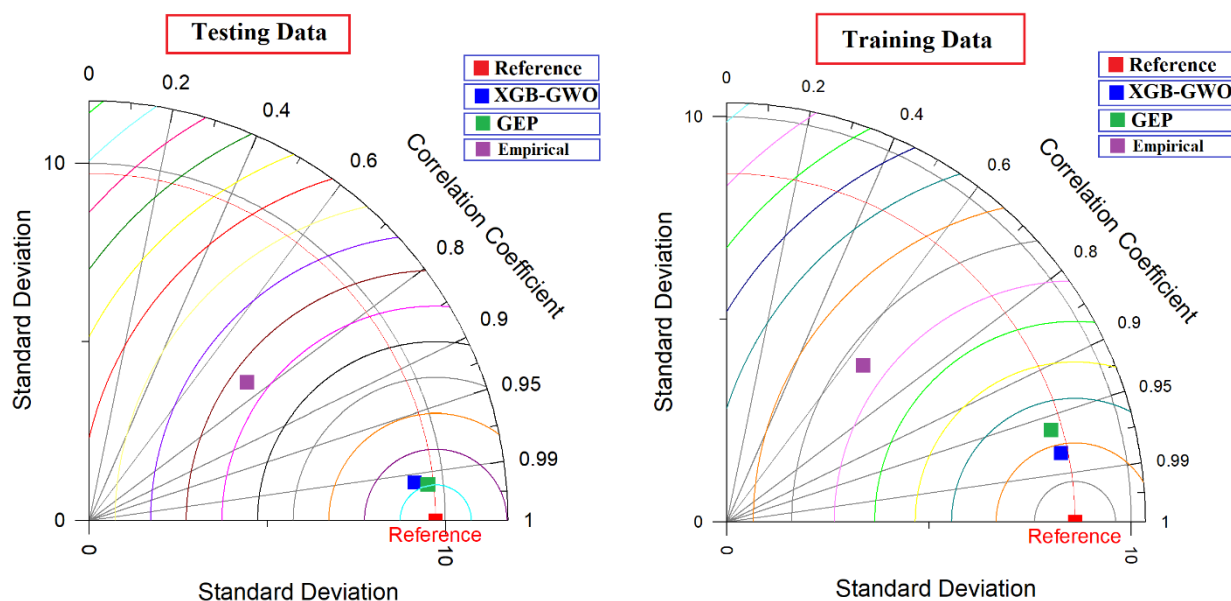
Figure 10. Performance comparison of predictive models with testing and training data in Taylor diagrams.

4.5. Decay of air overpressure

Understanding the decay of overpressures during the explosion of an explosive charge in an open pit mine is crucial in ensuring the safety of personnel and equipment. The pressure waves generated during an explosion can cause significant damage to structures and equipment, as well as pose a risk to the health and safety of personnel. By examining the decay of overpressures, it is possible to determine the safe distance for personnel and equipment from the explosion site, as well as the appropriate level of protective measures required. This information can also inform the design of blast patterns and other explosives-related procedures in order to minimize the risk of harm. Therefore, studying the decay of overpressures during explosions is an important aspect of ensuring the safe and effective use of explosives in open pit mines [62].

In this case, only two distances have been measured, and it is necessary to use empirical and GEP methods to estimate the decay of overpressures at different distances. The resulting plot can provide valuable insights into the behavior of air overpressure as it decays over distance, allowing us to identify any unexpected trends or anomalies that may require further investigation. In this regard, Figure 11 plots the measured, empirical and GEP predicted AOp values for different distances from the blasting site.
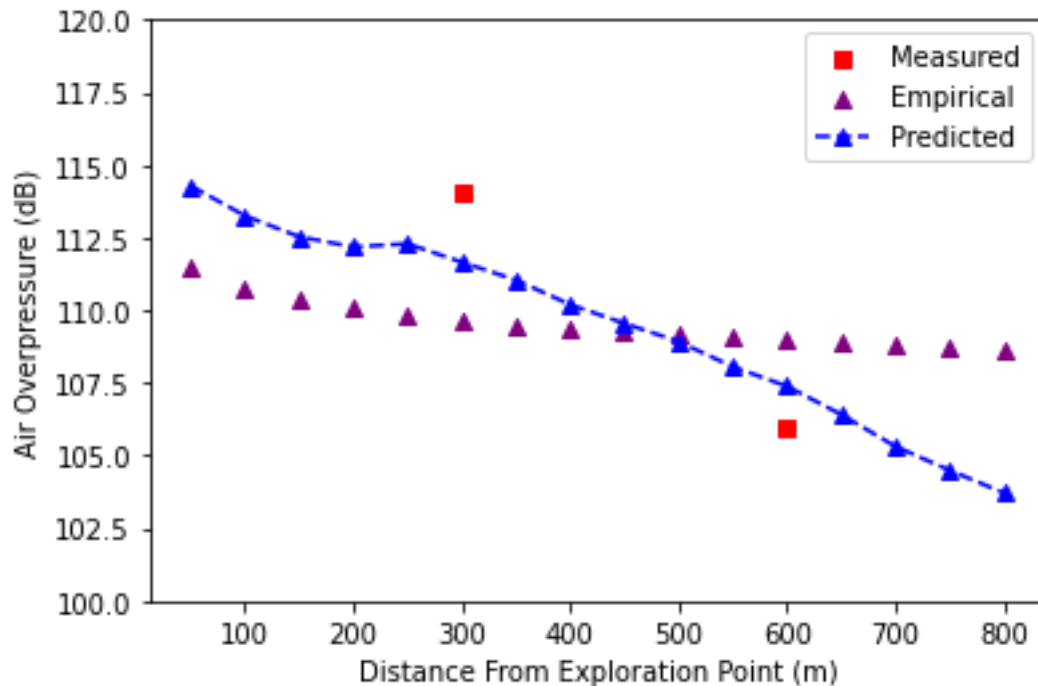
Figure 11. Comparison of measured, empirical, and GEP predicted AOp values at different distances from the blasting site.

Analyzing the plot, we can see that the measured AOp values are highest at the closest distance to the explosion site (300 m) and decrease as the distance increases. Both the empirical and GEP predicted AOp values follow similar trends, with the GEP method predicting AOp values first that are generally higher than the empirical values. The plot also shows that the GEP method generally provides a reasonable estimate of the AOp values at different distances from the explosion site, although it tends to underestimate the AOp values at shorter distances.

5. Sensitivity analysis

During the final modeling stage, the output of the model is analyzed for its sensitivity to the input parameters. Analyzing the sensitivity of input parameters can provide insights into how they affect model output (objective function). For determining sensitivity analysis, the cosine amplitude method (CAM) is one method [63–65]. Equation (26) describes the CAM method in terms of an n-dimensional space.

$$X = \{X_1, X_2, X_3 \dots, X_n\} \tag{26}$$

The m-dimensional length vector X_i, which is part of the array X, represents a variable in the given context.

$$X_i = \{X_{i1}, X_{i2}, X_{i3} \dots, X_{im}\} \tag{27}$$

This implies that the different dimensions of X_{ij} are interrelated with those of X_{jk}, and the degree of correlation between them can be expressed mathematically as shown in the following equation.

$$R_{ij} = \frac{\sum_{k=1}^{n} X_{ik} X_{jk}}{\sqrt{\sum_{k=1}^{m} X_{ik}^2 \sum_{k=1}^{m} X_{jk}^2}} \qquad (28)$$

Input parameters have a greater influence on output as R_{ij} approaches one. It has been shown that input parameters have a significant effect on the output when this parameter is above 0.9 [63]. In Figure 12, the regression prediction parameters are sensitivity analyzed. According to Figure 10, stemming length and RQD had the most significant effect on the air overpressure among the input elements.
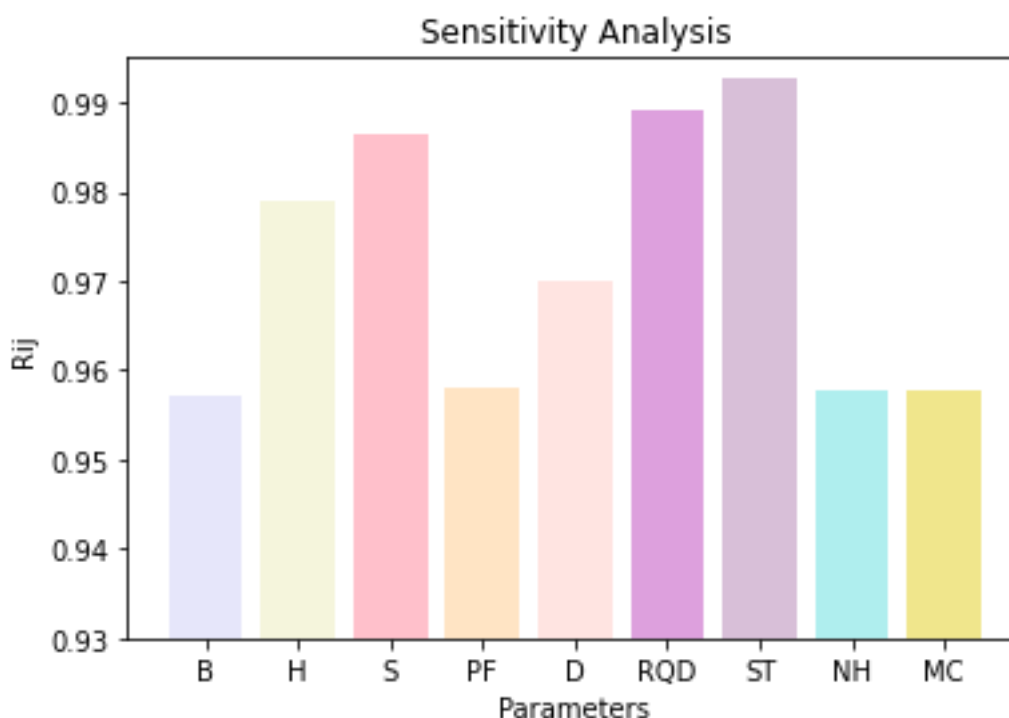

Figure 12. The sensitivity analysis and effect of input data on air overpressure.

6. Conclusions

Blasting is a commonly used method in open-pit mines for breaking down rocks. However, it can result in adverse effects, such as air over-pressure (AOp), ground vibration, flyrock, backbreak, and dust, which can have a negative impact on the surrounding environment. To address this issue, it is essential to predict and control the effects of blasting. In this study, an efficient and practical hybrid machine learning model (XGB-GWO) was proposed for predicting AOp values, and its performance was evaluated using gene expression programming (GEP) and an empirical model. The accuracy of the predictive models was assessed using standard criteria such as R^2, RMSE and MAE. Additionally, a sensitivity analysis was conducted using the cosine amplitude method (CAM) to determine the intensity of input parameters at AOp. Overall, this study highlights the significance of developing effective models to predict the impact of blasting and minimize its adverse effects on the environment. In conclusion, the hybrid XGB models proposed in this study demonstrate considerable potential in predicting AOp and can aid XGB in adjusting hyperparameters. The performance of the predictive

models for test data falls within the following ranges: XGB-GWO (MAE: 0.69; RMSE: 1.42; R2: 0.983), GEP (MAE: 0.63; RMSE: 1.04; R2: 0.989) and empirical (MAE: 5.92; RMSE: 6.68; R2: 0.53). It is worth noting that the XGB-GWO hybrid model outperforms the other models in terms of overall performance. Furthermore, a sensitivity analysis technique called cosine amplitude was used to determine the importance of input variables, which revealed that the stemming length and RQD had the most significant impact on the penetration rate. This study demonstrates that the proposed XGB-GWO model is robust and effective in predicting blast-induced AOp, indicating its potential for practical applications in the mining industry.

Conflict of interest

The authors declare no conflict of interest.

References

1. Nguyen H (2019) Support vector regression approach with different kernel functions for predicting blast induced ground vibration: a case study in an open-pit coal mine of Vietnam. *SN Appl Sci* 1: 283. https://doi.org/10.1007/s42452-019-0295-9

2. Yang H, Hasanipanah M, Tahir MM, et al. (2020) Intelligent prediction of blasting-induced ground vibration using ANFIS optimized by GA and PSO. *Nat Resour Res* 29: 739–750. https://doi.org/10.1007/s11053-019-09515-3

3. Li G, Kumar D, Samui P, et al. (2020) Developing a new computational intelligence approach for approximating the blast-induced ground vibration. *Appl Sci* 10: 434. https://doi.org/10.3390/app10020434

4. Bui XN, Nguyen H, Le HA, et al. (2020) Prediction of blast-induced air over-pressure in open-pit mine: assessment of different artificial intelligence techniques. *Nat Resour Res* 29: 571–591. https://doi.org/10.1007/s11053-019-09461-0

5. Trigueros E, Cánovas M, Muñoz JM, et al. (2017) A methodology based on geomechanical and geophysical techniques to avoid ornamental stone damage caused by blast-induced ground vibrations. *Int J Rock Mech Min Sci* 93: 196–200. https://doi.org/10.1016/j.ijrmms.2016.12.013

6. Khandelwal M, Saadat M (2015) A Dimensional Analysis Approach to Study Blast-Induced Ground Vibration. *Rock Mech Rock Eng* 48: 727–735.

7. Persson PA, Holmberg R, Lee J (2018) *Rock blasting and explosives engineering*. CRC press.

8. Chen L, Asteris PG, Tsoukalas MZ, et al. (2022) Forecast of Airblast Vibrations Induced by Blasting Using Support Vector Regression Optimized by the Grasshopper Optimization (SVR-GO) Technique. *Appl Sci* 12: 9805. https://doi.org/10.3390/app12199805

9. Segarra P, Domingo JF, López LM, et al. (2010) Prediction of near field overpressure from quarry blasting. *Appl Acoustics* 71: 1169–1176. https://doi.org/10.1016/j.apacoust.2010.07.008

10. Remennikov AM, Rose TA (2007) Predicting the effectiveness of blast wall barriers using neural networks. *Int J Impact Eng* 34: 1907–1923. https://doi.org/10.1016/j.ijimpeng.2006.11.003

11. Jahed Armaghani D, Hasanipanah M, Mahdiyar A, et al. (2018) Airblast prediction through a hybrid genetic algorithm-ANN model. *Neural Comput Applic* 29: 619–629. https://doi.org/10.1007/s00521-016-2598-8

12. Khandelwal M, Kankar PK (2011) Prediction of blast-induced air overpressure using support vector machine. *Arab J Geosci* 4: 427–433.

13. Jahed Armaghani D, Hajihassani M, Marto A, et al. (2015) Prediction of blast-induced air overpressure: a hybrid AI-based predictive model. *Environ Monit Assess* 187: 666. https://doi.org/10.1007/s10661-015-4895-6

14. Zhou J, Li X, Mitri HS (2018) Evaluation method of rockburst: state-of-the-art literature review. *Tunn Undergr Sp Tech* 81: 632–659. https://doi.org/10.1016/j.tust.2018.08.029

15. Liao X, Khandelwal M, Yang H, et al. (2020) Effects of a proper feature selection on prediction and optimization of drilling rate using intelligent techniques. *Eng Comput* 36: 499–510. https://doi.org/10.1007/s00366-019-00711-6

16. Jamei M, Ahmadianfar I, Chu X, et al. (2020) Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: New approach. *J Hydrol* 589: 125335. https://doi.org/10.1016/j.jhydrol.2020.125335

17. Dai Y, Khandelwal M, Qiu Y, et al. (2022) A hybrid metaheuristic approach using random forest and particle swarm optimization to study and evaluate backbreak in open-pit blasting. *Neural Comput Applic* 34: 6273–6288. https://doi.org/10.1007/s00521-021-06776-z

18. Sharma M, Agrawal H, Choudhary BS (2022) Multivariate regression and genetic programming for prediction of backbreak in open-pit blasting. *Neural Comput Applic* 34: 2103–2114. https://doi.org/10.1007/s00521-021-06553-y

19. Lawal AI, Kwon S, Hammed OS, et al. (2021) Blast-induced ground vibration prediction in granite quarries: An application of gene expression programming, ANFIS, and sine cosine algorithm optimized ANN. *Int J Min Sci Technol* 31: 265–277. https://doi.org/10.1016/j.ijmst.2021.01.007

20. Abd Elwahab A, Topal E, Jang HD (2023) Review of machine learning application in mine blasting. *Arab J Geosci* 16: 133. https://doi.org/10.1007/s12517-023-11237-z

21. Dumakor-Dupey NK, Arya S, Jha A (2021) Advances in blast-induced impact prediction—A review of machine learning applications. *Minerals* 11: 601. https://doi.org/10.3390/min11060601

22. Khandelwal M, Singh TN (2005) Prediction of blast induced air overpressure in opencast mine. *Noise Vib Worldwide* 36: 7–16. https://doi.org/10.1260/0957456053499095

23. Nguyen H, Bui XN, Bui HB, et al. (2020) A comparative study of artificial neural networks in predicting blast-induced air-blast overpressure at Deo Nai open-pit coal mine, Vietnam. *Neural Comput Applic* 32: 3939–3955. https://doi.org/10.1007/s00521-018-3717-5

24. Jahed Armaghani D, Hasanipanah M, Tonnizam Mohamad E (2016) A combination of the ICA-ANN model to predict air-overpressure resulting from blasting. *Eng Comput* 32: 155–171. https://doi.org/10.1007/s00366-015-0408-z

25. Hasanipanah M, Shahnazar A, Bakhshandeh Amnieh H, et al. (2017) Prediction of air-overpressure caused by mine blasting using a new hybrid PSO–SVR model. *Eng Comput* 33: 23–31. https://doi.org/10.1007/s00366-016-0453-2

26. Zhou J, Nekouie A, Arslan CA, et al. (2020) Novel approach for forecasting the blast-induced AOp using a hybrid fuzzy system and firefly algorithm. *Eng Comput* 36: 703–712. https://doi.org/10.1007/s00366-019-00725-0

27. Tran QH, Nguyen H, Bui XN, et al. (2021) Evaluating the Effect of Meteorological Conditions on Blast-Induced Air Over-Pressure in Open Pit Coal Mines, *Proceedings of the International Conference on Innovations for Sustainable and Responsible Mining 2021*, Springer, Cham. 170–186. https://doi.org/10.1007/978-3-030-60839-2_9

28. Zeng J, Jamei M, Nait Amar M, et al. (2022) A novel solution for simulating air overpressure resulting from blasting using an efficient cascaded forward neural network. *Eng Comput* 38: 2069–2081. https://doi.org/10.1007/s00366-021-01381-z

29. Akande JM, Aladejare AE, Lawal AI (2014) Evaluation of the environmental impacts of blasting in Okorusu fluorspar mine, Namibia. *Int J Eng Technol* 4: 101–108.

30. Bakhtavar E, Sadiq R, Hewage K (2021) Optimization of blasting-associated costs in surface mines using risk-based probabilistic integer programming and firefly algorithm. *Nat Resour Res* 30: 4789–4806. https://doi.org/10.1007/s11053-021-09935-0

31. Khademian A, Bagherpour R (2017) Environmentally sustainable mining through proper selection of explosives in blasting operation. *Environ Earth Sci* 76: 166. https://doi.org/10.1007/s12665-017-6483-2

32. Fang Q, Nguyen H, Bui XN, et al. (2020) Estimation of blast-induced air overpressure in quarry mines using cubist-based genetic algorithm. *Nat Resour Res* 29: 593–607. https://doi.org/10.1007/s11053-019-09575-5

33. Ye J, Dalle J, Nezami R, et al. (2022) Stochastic fractal search-tuned ANFIS model to predict blast-induced air overpressure. *Eng Comput* 38: 497–511. https://doi.org/10.1007/s00366-020-01085-w

34. Ramesh Murlidhar B, Yazdani Bejarbaneh B, Jahed Armaghani D, et al. (2021) Application of tree-based predictive models to forecast air overpressure induced by mine blasting. *Nat Resour Res* 30: 1865–1887. https://doi.org/10.1007/s11053-020-09770-9

35. Hajihassani M, Armaghani DJ, Sohaei H, et al. (2014) Prediction of airblast-overpressure induced by blasting using a hybrid artificial neural network and particle swarm optimization. *Appl Acoust* 80: 57–67. https://doi.org/10.1016/j.apacoust.2014.01.005

36. Nguyen H, Bui XN (2019) Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. *Nat Resour Res* 28: 893–907. https://doi.org/10.1007/s11053-018-9424-1

37. Harandizadeh H, Armaghani DJ (2021) Prediction of air-overpressure induced by blasting using an ANFIS-PNN model optimized by GA. *Appl Soft Comput* 99: 106904. https://doi.org/10.1016/j.asoc.2020.106904

38. Nguyen H, Bui XN, Tran QH, et al. (2020) A comparative study of empirical and ensemble machine learning algorithms in predicting air over-pressure in open-pit coal mine. *Acta Geophys* 68: 325–336. https://doi.org/10.1007/s11600-019-00396-x

39. AminShokravi A, Eskandar H, Derakhsh AM, et al. (2018) The potential application of particle swarm optimization algorithm for forecasting the air-overpressure induced by mine blasting. *Eng Comput* 34: 277–285. https://doi.org/10.1007/s00366-017-0539-5

40. Fan J, Wu L, Ma X, et al. (2020) Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. *Renewable Energy* 145: 2034–2045. https://doi.org/10.1016/j.renene.2019.07.104

41. Ghorbani MA, Deo RC, Yaseen ZM, et al. (2018) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theor Appl Climatol* 133: 1119–1131. https://doi.org/10.1007/s00704-017-2244-0

42. Chen T, He T, Benesty M, et al. (2015) Xgboost: extreme gradient boosting. *R Package Version 0.4–2* 1: 1–4.

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

44. Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data An* 38: 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

45. Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Ann Statist* 28: 337–407. https://doi.org/10.1214/aos/1016218223

46. Hastie T, Tibshirani R, Friedman J (2001) The elements of statistical learning, *Springer series in statistics*, New York, NY, USA.

47. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794. https://doi.org/10.1145/2939672.2939785

48. Zhou J, Li E, Wang M, et al. (2019) Feasibility of stochastic gradient boosting approach for evaluating seismic liquefaction potential based on SPT and CPT case histories. *J Perform Constr Fac* 33: 04019024.

49. Gao W, Wang W, Dimitrov D, et al. (2018) Nano properties analysis via fourth multiplicative ABC indicator calculating. *Arab J Chem* 11: 793–801. https://doi.org/10.1016/j.arabjc.2017.12.024

50. Nabavi Z, Mirzehi M, Dehghani H, et al. (2023) A hybrid model for back-break prediction using XGBoost machine learning and metaheuristic algorithms in Chadormalu iron mine. *J Min Environ* 9. doi: 10.22044/jme.2023.12796.2323.

51. Mirjalili S, Mirjalili SM, Lewis A. (2014). Grey wolf optimizer. *Adv Eng Softw* 69: 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007

52. Ferreira C (2001) Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst* 13: 87–129. https://doi.org/10.48550/arXiv.cs/0102027

53. Faradonbeh RS, Armaghani DJ, Amnieh HB, et al. (2018) Prediction and minimization of blast-induced flyrock using gene expression programming and firefly algorithm. *Neural Comput Applic* 29: 269–281. https://doi.org/10.1007/s00521-016-2537-8

54. Zhou J, Li C, Koopialipoor M, et al. (2021) Development of a new methodology for estimating the amount of PPV in surface mines based on prediction and probabilistic models (GEP-MC). *Int J Min Reclam Environ* 35: 48–68. https://doi.org/10.1080/17480930.2020.1734151

55. Faradonbeh RS, Hasanipanah M, Amnieh HB, et al. (2018) Development of GP and GEP models to estimate an environmental issue induced by blasting operation. *Environ Monit Assess* 190: 351. https://doi.org/10.1007/s10661-018-6719-y

56. Kuzu C, Fisne A, Ercelebi SG (2009) Operational and geological parameters in the assessing blast induced airblast-overpressure in quarries. *Appl Acoust* 70: 404–411. https://doi.org/10.1016/j.apacoust.2008.06.004

57. Hasanipanah M, Jahed Armaghani D, Khamesi H, et al. (2016) Several non-linear models in estimating air-overpressure resulting from mine blasting. *Eng Comput* 32: 441–455. https://doi.org/10.1007/s00366-015-0425-y

58. Zhou J, Qiu Y, Armaghani DJ, et al. (2021) Predicting TBM penetration rate in hard rock condition: a comparative study among six XGB-based metaheuristic techniques. *Geosci Front* 12: 101091. https://doi.org/10.1016/j.gsf.2020.09.020

59. Koopialipoor M, Fahimifar A, Ghaleini EN, et al. (2020) Development of a new hybrid ANN for solving a geotechnical problem related to tunnel boring machine performance. *Eng Comput* 36: 345–357. https://doi.org/10.1007/s00366-019-00701-8

60. Bunkley N (2008) Joseph Juran, Pioneer in Quality Control, Dies. *NYT*.

61. Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106: 7183–7192. https://doi.org/10.1029/2000JD900719

62. Chi M, Jiang H, Lan X, et al. (2021) Study on overpressure propagation law of vapor cloud explosion under different building layouts. *ACS Omega* 6: 34003–34020. https://doi.org/10.1021/acsomega.1c05332

63. Sirjani AK, Sereshki F, Ataei M, et al. (2022) Prediction of Backbreak in the Blasting Operations using Artificial Neural Network (ANN) Model and Statistical Models (Case study: Gol-e-Gohar Iron Ore Mine No. 1). *Arch Min Sci* 67: 107–121. https://doi.org/10.24425/ams.2022.140705

64. Nikafshan Rad H, Bakhshayeshi I, Wan Jusoh WA, et al. (2020) Prediction of flyrock in mine blasting: a new computational intelligence approach. *Nat Resour Res* 29: 609–623. https://doi.org/10.1007/s11053-019-09464-x

65. Khandelwal M, Saadat M (2015) A Dimensional Analysis Approach to Study Blast-Induced Ground Vibration. *Rock Mech Rock Eng* 48: 727–735. https://doi.org/10.1007/s00603-014-0604-y