



A comparative study of empirical and ensemble machine learning algorithms in predicting air over-pressure in open-pit coal mine

Hoang Nguyen^{1,2} · Xuan-Nam Bui^{1,2} · Quang-Hieu Tran¹ · Pham Van Hoa¹ · Dinh-An Nguyen¹ · Le Thi Thu Hoa¹ · Qui-Thao Le^{1,2} · Ngoc-Hoan Do^{1,3} · Tran Dinh Bao¹ · Hoang-Bac Bui^{4,5} · Hossein Moayed^{6,7}

Received: 9 April 2019 / Accepted: 26 December 2019
© Institute of Geophysics, Polish Academy of Sciences & Polish Academy of Sciences 2020

Abstract

This study aims to take into account the feasibility of three ensemble machine learning algorithms for predicting blast-induced air over-pressure (AOp) in open-pit mine, including gradient boosting machine (GBM), random forest (RF), and Cubist. An empirical technique was also applied to predict AOp and compared with those of the ensemble models. To employ this study, 146 events of blast were investigated with 80% of the total database (approximately 118 blasting events) being used for developing the models, whereas the rest (20% ~ 28 blasts) were used to validate the models' accuracy. RMSE, MAE, and R^2 were used as performance indices for evaluating the reliability of the models. The findings revealed that the ensemble models yielded more precise accuracy than those of the empirical model. Of the ensemble models, the Cubist model provided better performance than those of RF and GBM models with RMSE, MAE, and R^2 of 2.483, 0.976, and 0.956, respectively, whereas the RF and GBM models provided poorer accuracy with an RMSE of 2.579, 2.721; R^2 of 0.953, 0.950, and MAE of 1.103, 1.498, respectively. In contrast, the empirical model was interpreted as the poorest model with an RMSE of 4.448, R^2 of 0.872, and MAE of 3.719. In addition, other findings indicated that explosive charge capacity, spacing, stemming, monitoring distance, and air humidity were the most important inputs for the AOp predictive models using artificial intelligence.

Keywords Air over-pressure · Open-pit mine · Ensemble algorithm · Random forest · Gradient boosting machine · Cubist

Introduction

One of the most effective techniques for fragmenting rock in open-pit mines is blasting because of its advantages from technical and economical points of view. It can generate a large amount of rock for the subsequent operations (e.g., loading, transporting) with low cost (Jhanwar et al. 1999).

However, its ill side influences are not negligible, including air over-pressure (AOp), flyrock, ground vibration, dust, and fumes (Nguyen et al. 2018; Zhang et al. 2019; Shang et al. 2019) (Fig. 1). Of those, AOp is considered as a dangerous phenomenon, which is needed to control (Alal et al. 2018; Armaghani et al. 2015; Khandelwal and Kankar 2011;

✉ Xuan-Nam Bui
buixuannam@humg.edu.vn

¹ Department of Surface Mining, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang ward, Bac Tu Liem Dist., Hanoi, Vietnam

² Center for Mining, Electro-Mechanical Research, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Vietnam

³ Faculty of Mining, Saint-Petersburg Mining University, Saint Petersburg, Russia

⁴ Faculty of Geosciences and Geoengineering, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang ward, Bac Tu Liem Dist., Hanoi, Vietnam

⁵ Center for Excellence in Analysis and Experiment, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang ward, Bac Tu Liem Dist., Hanoi, Vietnam

⁶ Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam

⁷ Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Fig. 1 Illustration of the undesirable effects of blasting operations

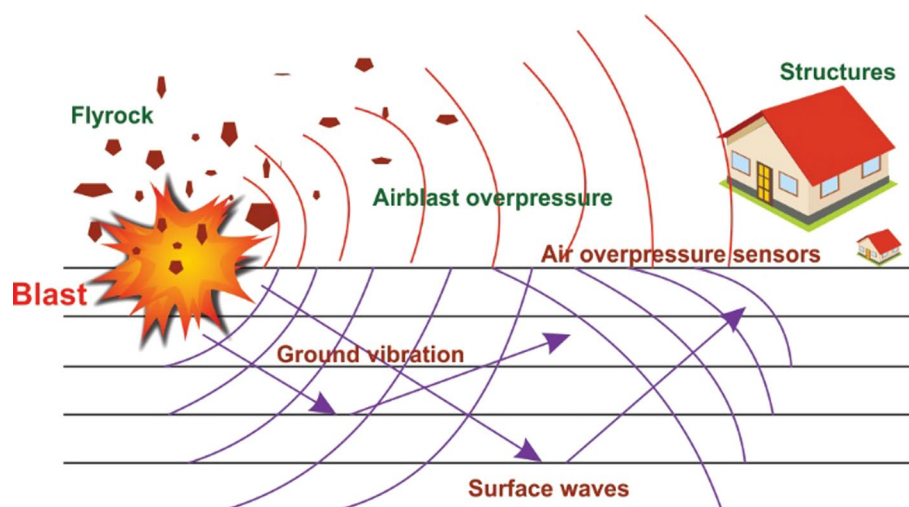


Table 1 Several empirical equations for predicting blast-induced AOp

No.	References	Empirical model
1	Siskind et al. (1980)	$AOp = k(SD)^{-\beta}$
2	Loder (1985)	$AOp = \frac{140 \sqrt[3]{\frac{W}{300}}}{D}$
3	McKenzie (1990)	$AOp = 165 - 24 \log (R/D^{1/3})$

k and β are the coefficients of the study site; SD denotes the scaled distance ($\text{m kg}^{0.33}$)

Khandelwal and Singh 2005; Nguyen and Bui 2018; Nguyen et al. 2017, 2018).

For predicting blast-induced AOp, several scholars proposed empirical equations, as listed in Table 1. Accordingly, the relationship between monitoring distance (D) and explosive charge per delay/maximum explosive charge capacity (W) was established through empirical equations.

Of the empirical equations in Table 1, the equation No.1 (USBM empirical equation) has been widely used to predict blast-induced AOp (Siskind et al. 1980; Hustrulid 1999; Walter 1990; Kuzu et al. 2009; Hasanipanah et al. 2016; Mahdiyar et al. 2018). However, the accuracy of empirical models was often not high due to some drawbacks of them, as discussed by Hasanipanah et al. (Hasanipanah et al. 2016), Mahdiyar et al. (Mahdiyar et al. 2018).

Recently, artificial intelligence (AI) became more appropriate and highly used in different fields, especially mining technology (Pierini et al. 2013; Rahmani and Farnood Ahmadi 2018; Montahaei and Oskooi 2014; Wiszniowski 2016; Naganna and Deka 2019; Piasecki et al. 2018; Nguyen et al. 2019a, b, c, d; Zhou et al. 2019; Asteris et al. 2016; Asteris and Nikoo 2019). In order to estimate blast-induced AOp, Hajihassani et al. (Hajihassani et al. 2014) trained an artificial neural network (ANN) by an evolutionary algorithm (Particle Swarm Optimization—PSO), namely

ANN-based PSO model, using 62 AOp datasets. Their results showed that the ANN-based PSO model performed properly in forecasting blast-caused AOp with the correlation coefficient (CC) of 0.94. In another study, Mohamad et al. (Mohamad et al. 2016) predicted blast-induced AOp by an ANN-based genetic algorithm (GA), abbreviated as GA-ANN, using 76 blasting events. Empirical and ANN models were also provided to predict AOp and compared them to those of the GA-ANN model. Their results interpreted that the GA-ANN model performed better than those of empirical and ANN models. Hasanipanah et al. (2016) used ANFIS, ANN, fuzzy system (FS) techniques, and an empirical equation for estimating blast-induced AOp. For developing these models, a group of 77 blasting events was used in their study. Their findings revealed that the ANFIS system was the most superior approach in forecasting AOp. Amiri et al. (2016) also introduced a new combination of k-nearest neighbors (KNN) and ANN models to predict AOp using 75 blasting events. Their results indicated that the KNN-ANN model predicted better than those of ANN and empirical models. Mahdiyar et al. (2018) also proposed three AI models to estimate AOp based on PSO algorithm and 80 blasting events. The results indicated that the PSO model estimated AOp very well with a promising result. Nguyen et al. (2019) also discovered a hybrid model based on clustering technique and backpropagation neural networks. In another study, Nguyen et al. (2018) performed a comparative study of MLP neural nets, BRNN, and HYFIS in estimating AOp. Their results showed that the MLP neural nets were the most superior model than those of the other models. They also developed another AI model based on ensemble of ANN and RF (i.e., ANNs-RF) for predicting AOp with an excellent result (Nguyen and Bui 2018). By the use of optimization algorithm, AminShokravi et al. (2018) demonstrated the potential of the PSO algorithm in predicting AOp with high accuracy. Bui et al. (2019) also evaluated the

performance of different AI techniques for estimating AOp in an open-pit coal mine, including RF, boosted regression trees, KNN, SVR, GP (Gaussian process), BART (Bayesian additive regression trees), and ANN. They claimed the feasibility of the mentioned AI techniques. ANN model was recommended as the best model in their study for estimating AOp. Zhou et al. (2019) also developed a novel AI model for forecasting AOp based on FS and firefly algorithm (FA), namely FS-FA model. A high prediction level was confirmed in their study for the proposed FS-FA model. Gao et al. (2019) also took into account the performance of the GA and group method of data handling (GMDH) for forecasting AOp. Eventually, their GA-GMDH model was proposed as a robust technique with an excellent agreement.

A review of the literature shows that blast-induced AOp predictive models were developed and proposed quite well. Nevertheless, they cannot apply and represent other locations/regions, whereas the effects of blast-induced AOp are different from country to country. In this study, blast-induced AOp was assessed and predicted by three ensemble machine learning algorithms, including RF, GBM (gradient boosting machine), and Cubist. An empirical model was also developed to predict and compare with those of ensemble models herein.

The rest of the present work is arranged as follows: “Study area and data used” section presents the study site and characteristics of the dataset; “Methods” section provides the principle of the approaches used; the preparation of the dataset is introduced in “Preparing the dataset” section; the development of the models is shown in “Establishing the AOp predictive models” section; some performance indices are presented in “Performance indices” section; and “Results and discussion” section reports the results and discussion.

Finally, “Conclusions and remarks” section presents our conclusions of this work.

Study area and data used

Study area

Herein, the Deo Nai open-pit coal mine, which is located in Quang Ninh Province, Vietnam, was selected as a special study area. It lies within latitudes $21^{\circ}00'00''\text{N}$ and $21^{\circ}02'00''\text{N}$ and between longitudes $107^{\circ}01'15''\text{E}$ and $107^{\circ}01'20''\text{E}$ (Fig. 2). The coal store is 42.5 Mt, and production capacity is 2.5 Mt/year; overburden is 20–30 Mt/year. (Vinacomin 2015). With a large amount of overburden per year and the hardness of rock being high (from 10 to 14 according to Protodiakonov’s classification (Bach et al. 2012)), blasting was selected as a proper technique for fragmenting rock in the mine. ANFO is the main explosive used in this mine, with the amount being up to 20 tons. The non-electric delay blasting method was applied to fragment rock with the diameter of borehole of 105 mm. The nearest distance from blasts to the residential area is about 400–500 m. Hence, the ill side effects of blasts are substantial.

Data collection and its characteristics

In this study, 146 events of blasting were investigated, with ten parameters being measured. Of the ten parameters, nine first variables were used as the inputs to predict the outcome of AOp, including powder factor (q), maximum explosive charge capacity (W), burden (B), length of stemming (T), spacing (S), number of rows per blast (N),

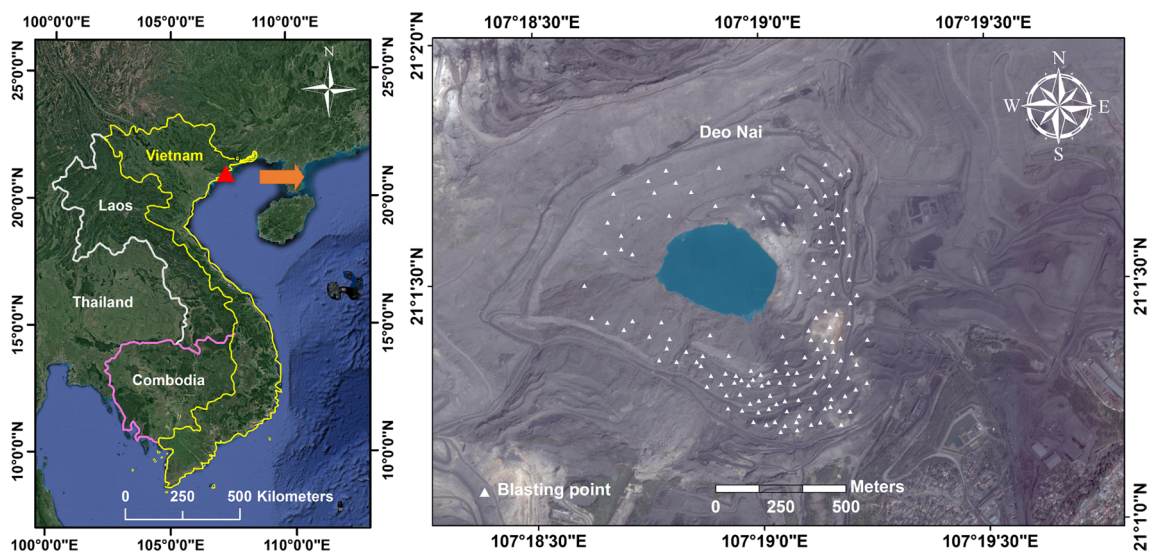


Fig. 2 Location of the study site

monitoring distance (D), bench height (H), and air humidity (RH) (Fig. 3). For monitoring blast-induced AOp, an instrument of Instanetel (Canada) was utilized with a microphone. According to the guideline of the producer, the microphone should be placed at the sensitive locations and straightforward with the direction of blasts (Fig. 4). Also, a handheld GPS was used to define D . RH was measured by Kanomax 2212 air quality meter (Japan). It is one of the most influential parameters for estimating AOp, which was recommended by Nguyen et al. (2018). The remaining inputs were extracted from the design of blasts. Table 2 shows the characteristics of inputs and output in this work.

Methods

Empirical

Empirical is one of the methods which is utilized to predict blast-produced AOp in open-cast mine. Of the empirical methods (as shown in Table 1), USBM empirical formula was widely applied to predict AOp in open-pit mines (Hajihassani et al. 2014; Armaghani et al. 2016). For example, Kuzu et al. (2009) used the empirical equation of the USBM to forecast AOp with a promising result. In the USBM

equation, the scaled distance was illustrated through W and D as follows:

$$SD = DW^{-0.33} \quad (1)$$

Subsequently, the USBM empirical equation can be computed according to Eq. 2:

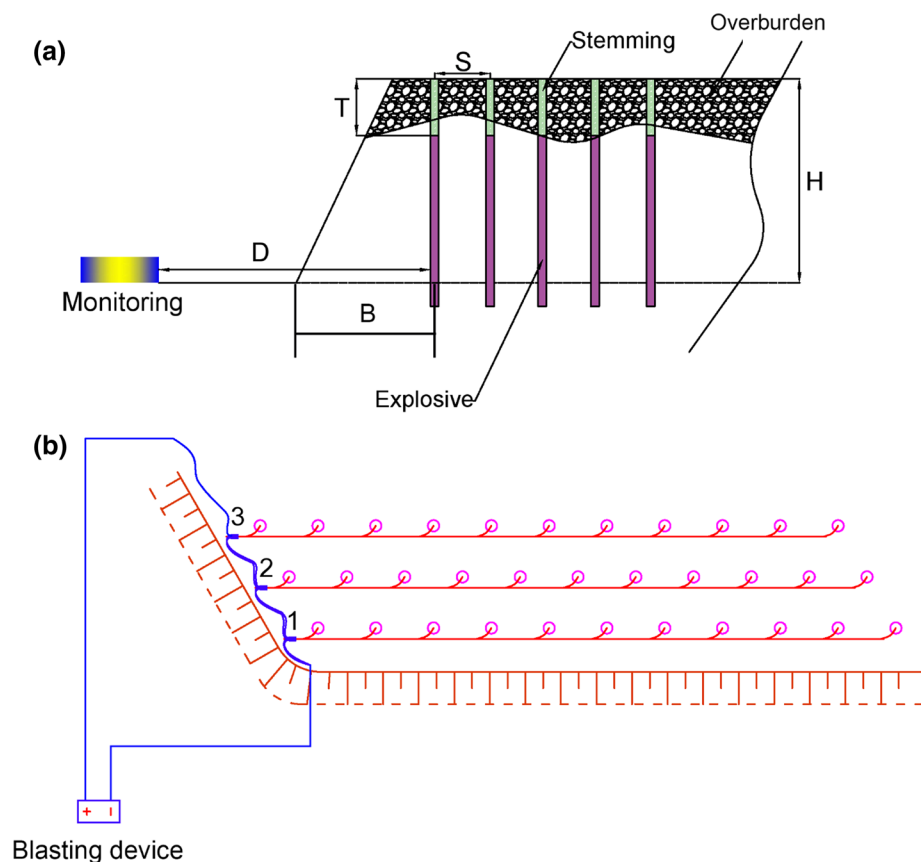
$$AOp = \gamma(SD)^{-\alpha} \quad (2)$$

where γ and α are the site factors.

Random forest

Decision tree (DT) is one of the branches of AI, and RF belongs to the DT branch, which was developed by Breiman (2001). As a robust DT model, RF can solve both classification and regression cases. Based on the different results of the trees, this method has been suggested as a suitable method for achieving predictive precision (Vigneau et al. 2018). In addition, this method used the results of the exclusive tree in the forest to present the best outcome. As a voter, each tree contributes its predictions for the final decision of RF (Gao et al. 2018). On the other hand, RF ensembles the predictions of the tree and making a final decision based on the obtained results. The key of

Fig. 3 Structure of the borehole and its parameters. **a** Parameters of blast design and **b** a combination plan of blasting



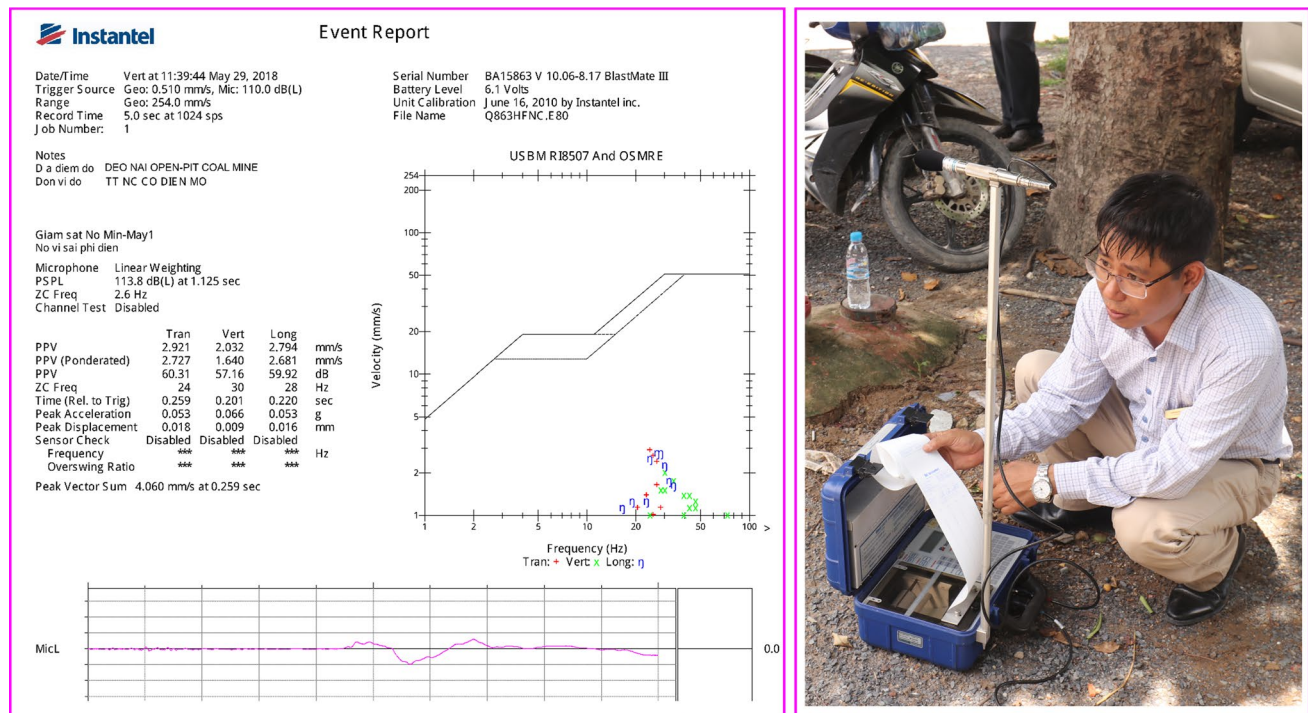


Fig. 4 Data collection for predicting AOP in this work

Table 2 Inputs, output, and their properties

Categories	<i>W</i>	<i>H</i>	<i>B</i>	<i>S</i>	<i>T</i>
Minimum	1376	13.00	7.500	7.400	6.200
Mean	13183	14.37	8.064	7.814	6.879
Maximum	24171	16.00	8.500	8.200	7.500
Standard deviation	4685.73	0.937	0.332	0.213	0.362
Categories	<i>q</i>	<i>N</i>	RH	<i>D</i>	AOp
Minimum	0.3500	2.000	76.00	180	92.26
Mean	0.4178	3.486	85.16	469	123.19
Maximum	0.4800	5.000	94.00	726	147.00
Standard deviation	0.035	1.216	4.817	159.896	11.912

the RF for regression is presented in three steps: (i) producing bootstrap instances as the tree number in the forest (n_{tree}) according to the database, (ii) expand a suitable regression tree for any bootstrap instance using random sampling of the estimators (m_{try}) (Dou et al. 2019). Of those variables, choose the most appropriate split and (iii) estimate recent perception using ensemble the estimated amounts of the trees (n_{tree}). For the regression issue (i.e., estimating AOp), the mean amount of the estimated values in the single tree is applied.

According to the training dataset, a prediction of the error rate may be calculated according to the two following steps:

- 1 At any iteration of bootstrap, estimate the non-information in the instance of bootstrap using the tree grown with the bootstrap instance, named “out-of-bag” (OOB).
- 2 Collect the OOB estimations and predict the error.

More details of the RF algorithm can be explained in (Nguyen and Bui 2018; Breiman 2001; Bui et al. 2019).

Gradient boosting machine

GBM is an ensemble approach that is suggested by Friedman (2002). It is an improved boosting algorithm and can be applied for regression, as well as classification problems (Friedman 2001). The boosting algorithm can be described according to the pseudocode in Fig. 5 (Friedman 2002).

Subsequently, Friedman (Friedman 1999) provided a particular algorithm based on the platform of boosting algorithm for various loss criteria like least squares:

$$\psi(y_{\text{AOp}}, T) = (y_{\text{AOp}} - T)^2 \quad (3)$$

Least absolute deviation:

Boosting algorithm

```

1   $F_0(x) = \arg \min_{\gamma} \sum_{i=1}^N \psi(y_i, \gamma)$ 
2  For  $m=1$  to  $M$  do:
3       $\tilde{y}_{im} = - \left[ \frac{\partial \psi(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{m-1}(x)}, i=1, N$ 
4       $\{R_{im}\}_1^L = L - \text{terminal node tree}(\{\tilde{y}_{im}, x_i\}_1^N)$ 
5       $\gamma_{im} = \arg \min_{\gamma} \sum_{x_i \in R_{im}} \psi(y_i, F_{m-1}(x_i) + \gamma)$ 
6       $F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_{im} 1(x \in R_{im})$ 
7  endFor.
```

Fig. 5 Pseudocode of the boosting algorithm

Gradient boosting machine

```

1   $F_0(x) = \arg \min_{\gamma} \sum_{i=1}^N \psi(y_i, \gamma)$ 
2  For  $m=1$  to  $M$  do:
3       $\{\pi(i)\}_1^N = \text{rand\_perm} \{i\}_1^N$ 
4       $\tilde{y}_{\pi(i)m} = - \left[ \frac{\partial \psi(y_{\pi(i)}, F(x_{\pi(i)}))}{\partial F(x_{\pi(i)})} \right]_{F(x)=F_{m-1}(x)}, i=1, \tilde{N}$ 
5       $\{R_{im}\}_1^L = L - \text{terminal node tree}(\{\tilde{y}_{\pi(i)m}, x_{\pi(i)}\}_1^{\tilde{N}})$ 
6       $\gamma_{im} = \arg \min_{\gamma} \sum_{x_{\pi(i)} \in R_{im}} \psi(y_{\pi(i)}, F_{m-1}(x_{\pi(i)}) + \gamma)$ 
7       $F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_{im} 1(x \in R_{im})$ 
8  endFor.
```

Fig. 6 Pseudocode of the GBM algorithm

$$\psi(y_{\text{AOp}}, T) = |y_{\text{AOp}} - T| \quad (4)$$

Huber M:

$$\psi(y_{\text{AOp}}, T) = (y_{\text{AOp}} - T)^2 1(|y_{\text{AOp}} - T| \leq \delta) + 2\delta(|y_{\text{AOp}} - T| - \delta/2) 1(|y_{\text{AOp}} - T| > \delta) \quad (5)$$

Let $\{y_{i,\text{AOp}}, x_{i,\text{AOp}}\}_1^N$ as the entire training information instance and $\{\pi(i)\}_1^N$ stands for random permutation for integers $\{1, \dots, N\}$. Then, a random subsample of size $\tilde{N} < N$ is predicted by $\{y_{\pi(i,\text{AOp})}, x_{\pi(i,\text{AOp})}\}_1^{\tilde{N}}$. The pseudocode of the GBM algorithm is described in Fig. 6 (Friedman 2002).

Cubist

Cubist algorithm (Rulequest 2016a, b) is one of the rule-based algorithms, which is utilized to make predictive models according to the input information analysis, whereas the See5/C5.0 method that is able to solve classification problems (Quinlan 2004), the Cubist can solve regression issues very well. The outcomes from the Cubist model are more priority than those of linear regression models. In addition, it is simpler than the ANN model (Rulequest 2016a, b).

The Cubist model is expanded based on Quinlan's M5 model tree (Quinlan 1992) with the capability to apply for thousands of input characteristics (Rulequest 2016a, b). In the Cubist model, the targets depend on the inputs, and it is computed based on the rule(s). A combination of different conditions with a linear function is conducted for these rules. The related linear function is used to estimate the output properly if a rule takes into consideration the whole requirements. The Cubist algorithm can perform multiple situations at the same time and then detect various distinct linear functions for estimating output. Therefore, Cubist can generate various models and mixes them based on the rules which are determined before. Developing multiple models with different rules and their combinations can assist Cubist model in attaining much higher levels of precision. More details of Cubist can be found in Refs. (Nguyen et al. 2019; Kuhn et al. 2012; Drzewiecki 2016; Kuhn et al. 2018; Bernat and Drzewiecki 2015).

Preparing the dataset

In this section, the AOp dataset is prepared as a geospatial database by the ArcGIS software; 146 records of blast were divided into two sections according to the recommendations of previous researchers (Nguyen et al. 2019a, b); 80% of the total datasets (approximate 118 events of blast) are selected by randomly and applied as the training dataset to build the AOp predictive models. The rest (28 records of the blast) were utilized as the testing dataset for evaluating the AOp models' performance. Summary of training and testing datasets is shown in Tables 3 and 4, respectively.

Table 3 Summary of the training dataset

Categories	<i>W</i>	<i>H</i>	<i>B</i>	<i>S</i>	<i>T</i>
Minimum	1376	13.00	7.500	7.400	6.200
Mean	13036	14.36	8.054	7.807	6.877
Maximum	24171	16.00	8.500	8.200	7.500
Standard deviation	4735.744	0.949	0.332	0.209	0.367
Categories	<i>q</i>	<i>N</i>	RH	<i>D</i>	AOp
Minimum	0.350	2.000	76.00	180	92.26
Mean	0.417	3.466	85.22	473.4	122.98
Maximum	0.480	5.000	94.00	726	147.00
Standard deviation	0.035	1.217	4.665	158.747	11.999

Table 4 Summary of the testing dataset

Categories	<i>W</i>	<i>H</i>	<i>B</i>	<i>S</i>	<i>T</i>
Minimum	1376	13.00	7.500	7.400	6.200
Mean	13183	14.37	8.064	7.814	6.879
Maximum	24171	16.00	8.500	8.200	7.500
Standard deviation	4497.849	0.903	0.330	0.233	0.349
Categories	<i>q</i>	<i>N</i>	RH	<i>D</i>	AOp
Minimum	0.3500	2.000	76.00	180	92.26
Mean	0.4178	3.486	85.16	469	123.19
Maximum	0.4800	5.000	94.00	726	147.00
Standard deviation	0.034	1.230	5.497	166.333	11.712

Establishing the AOp predictive models

For the empirical model, 118 blasting events (training dataset) were used to compute the site factors k and β . Microsoft Excel 2016 was used to define k and β by the use of a multivariate regression analysis technique. As a result, $k=208.26$ and $\beta=0.183$ are the optimal values of the USBM model

for predicting AOp. The USBM model (in this case) can be described as:

$$\text{AOp} = 208.026(SD)^{-0.183} \quad (6)$$

For the development of the ensemble models, the ten-fold cross-validation method, along with three repetitions, is utilized to avoid overfitting. Furthermore, the ensemble models used the same training as those used for the development of the USBM model. To develop the RF model, the number of trees was set equal to 2000 to meet the diversity of the forest (Nguyen et al. 2017). Then, the random predictor (m_{try}) was tuned to get the optimal performance of the RF model. Herein, m_{try} was set in the range of 1–50 as a trial and error procedure. Ultimately, an optimal value of m_{try} was determined for the RF model with $m_{\text{try}}=41$. Figure 7 shows the efficiency of the RF model for estimating AOp.

Unlike the RF model, the GBM model used four parameters to control the model's performance, such as the *number of trees*, *max tree depth*, *shrinkage*, and *n.minobsinnode*. A grid search method was also applied to define the optimal parameters for the GBM model. As a result, *number of trees*=500, *max tree depth*=4, *shrinkage*=0.1, and *n.minobsinnode*=5 were the best values for the GBM model in this case. GBM's performance is illustrated in Fig. 8.

To develop the Cubist model, *committees* and *neighbors* were used as the key parameters. The results indicated that the Cubist model reached optimal performance with *committees* of 80 and *neighbors* of 0, as shown in Fig. 9.

Performance indices

For evaluating the efficiency of the AOp predictive models, three performance indices were computed, including mean absolute error (MAE), coefficient of determination (R^2), and root mean square error (RMSE).

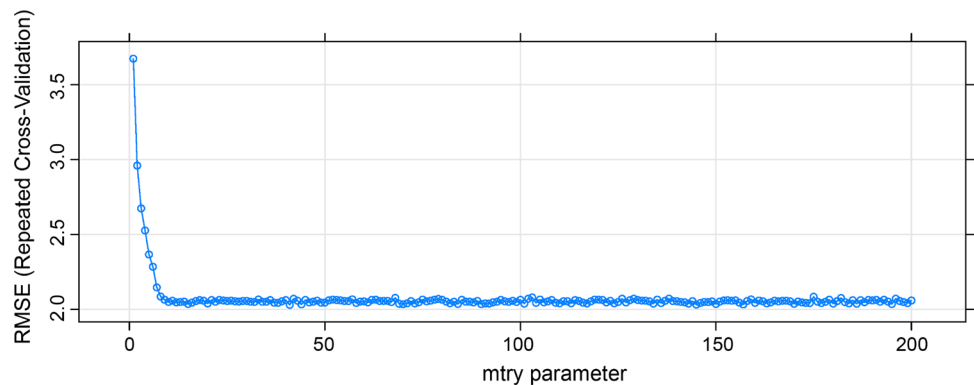
Fig. 7 RF modeling for prediction of AOp

Fig. 8 GBM modeling for prediction of AOp

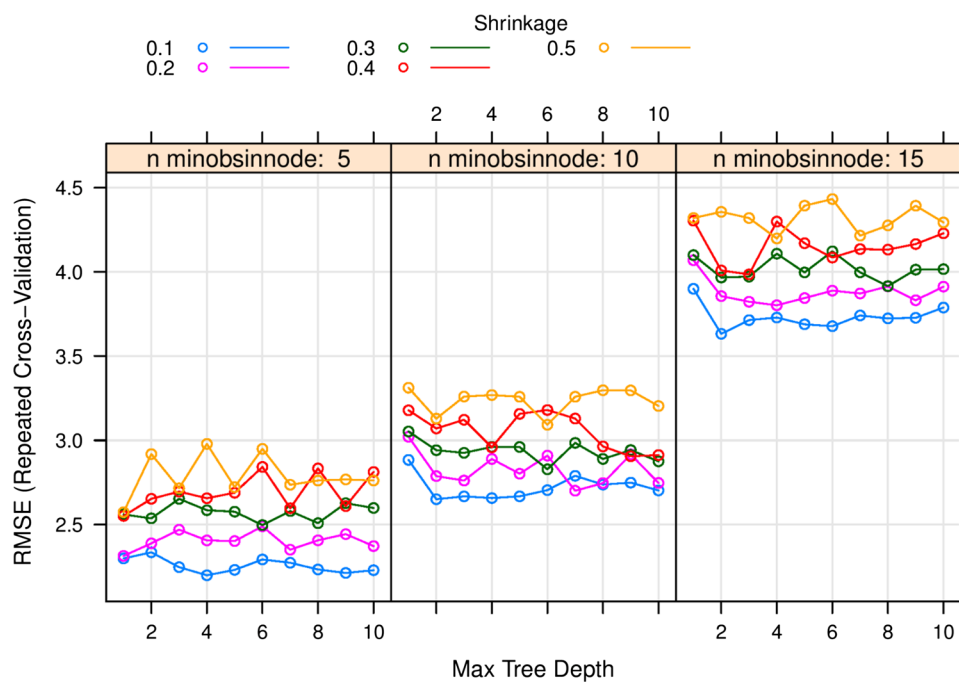


Fig. 9 Cubist modeling for prediction of AOp

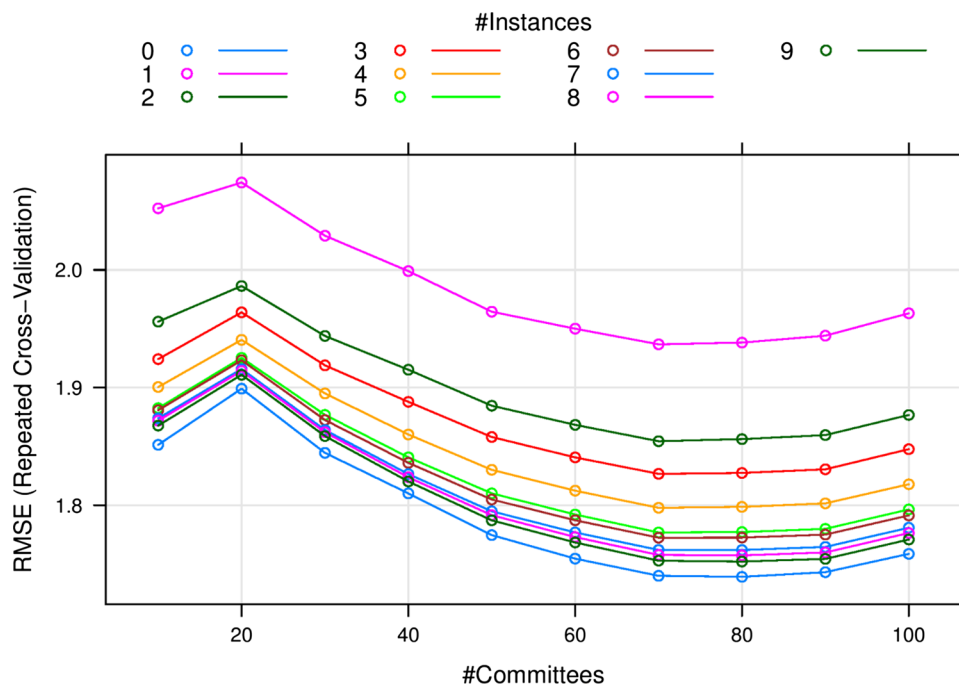
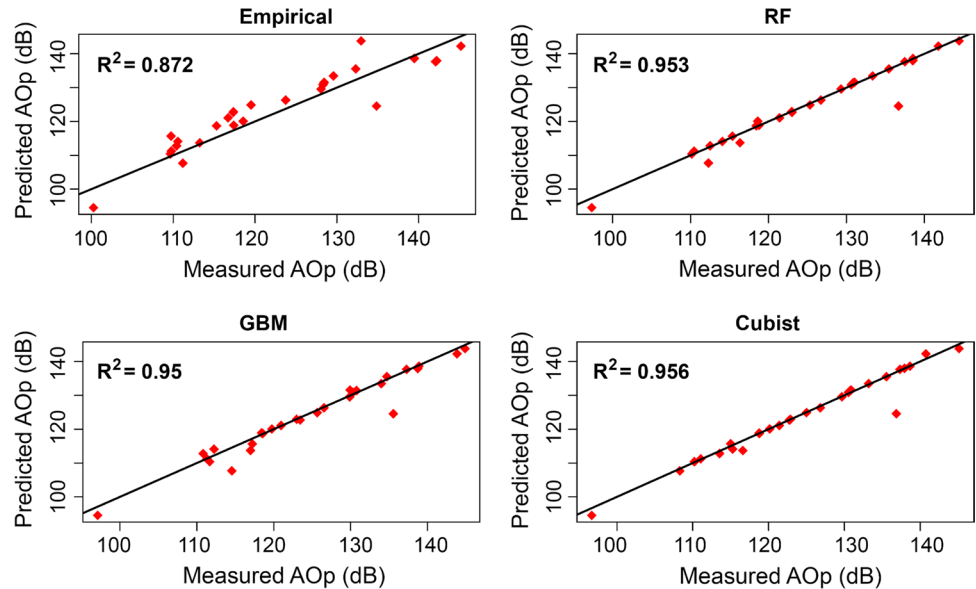


Table 5 Performance indices of the ensemble and empirical models

Method	Training dataset			Testing dataset		
	RMSE	R^2	MAE	RMSE	R^2	MAE
Empirical	4.838	0.871	4.101	4.448	0.872	3.719
RF	2.030	0.968	1.143	2.592	0.953	1.103
GBM	2.199	0.970	1.451	2.721	0.950	1.498
Cubist	1.739	0.969	0.980	2.483	0.956	0.976

Fig. 10 Relationship of measured and predicted AOp on the ensemble and empirical models

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{AOp} - \hat{y}_{AOp})^2} \quad (7)$$

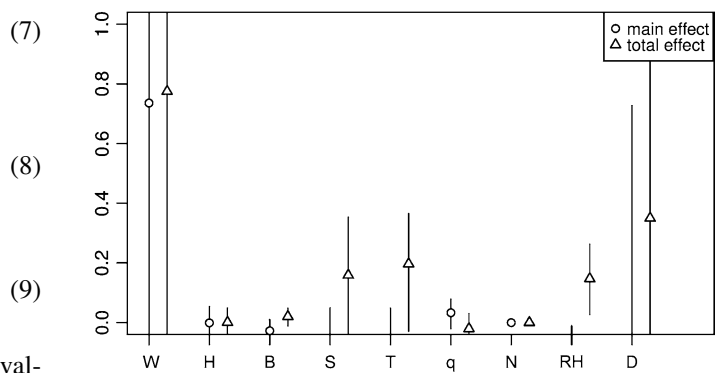
$$R^2 = 1 - \frac{\sum_i (y_{AOp} - \hat{y}_{AOp})^2}{\sum_i (y_{AOp} - \bar{y}_{AOp})^2} \quad (8)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{AOp} - \hat{y}_{AOp}| \quad (9)$$

n is the total number of observations. y_{AOp} is recorded values, \hat{y}_{AOp} is predicted values, and \bar{y}_{AOp} is the average of recorded values.

Results and discussion

Once the models were well established, their performance is evaluated and checked through the performance indices according to Eqs. (7–9). Table 5 shows the results, as well as the performance of the ensemble and empirical models on training/testing datasets.

**Fig. 11** Sensitivity analysis of the parameters

It can be easy to recognize that the ensemble models performed very well in this study. On the training dataset, the ensemble models obtained high performance with RMSE of 1.739–2.199; R^2 of 0.968–0.970; and MAE of 0.980–1.451. The similar results were also observed on the testing dataset for the ensemble models with RMSE of 2.483–2.721, R^2 of 0.950–0.956, and MAE of 0.976–1.498. In contrast to the ensemble models, the empirical model provided the poorest efficiency (i.e., RMSE = 4.838, 4.448;

$R^2 = 0.871, 0.872$; and $MAE = 4.101, 3.719$, on the training and testing datasets, respectively). Among three ensemble models (RF, GBM, Cubist), the Cubist model was the most dominant model with an RMSE of 2.483, R^2 of 0.956, and MAE of 0.976 on the testing database. Figure 10 shows the efficiency of the AOp predictive models in testing process.

Although the efficiency of the ensemble models is better than the empirical model in this study, however, the practical technique used only two input parameters (W and D) to estimate blast-induced AOp, whereas the ensemble models used nine input parameters for predicting the same objective. Therefore, a sensitivity analysis procedure was conducted to assess the effect of the inputs on the AOp predictive model (Tarantola et al. 2007; Saltelli et al. 2010). The results showed that W , S , T , RH , and D were the most influential parameters on the AOp predictive model, as illustrated in Fig. 11.

Conclusions and remarks

Based on the obtained results of this study, some conclusions and remarks are withdrawn as follows:

- Ensemble machine learning algorithms are good candidates for predicting blast-induced AOp than those of empirical methods, especially RF, GBM, and Cubist models. They should be considered to control the undesirable effects of blasting in practical engineering.
- Cubist is a robust ensemble AI model for predicting AOp in this study. Its accuracy can ensure safety for the surrounding environment. However, it should be reconsidered in other locations/areas.
- RF and GBM are also good AI techniques for predicting AOp. However, its performance seems not to satisfy. Therefore, they need to improve and further research.
- For predicting AOp, it is not only W and D , but also S , T , and RH are the important inputs for the development of the AOp predictive models. They should be carefully collected to ensure the accuracy level of the models.

Acknowledgements This research was supported by Hanoi University of Mining and Geology (HUMG), Hanoi, Vietnam; Duy Tan University, Da Nang, Vietnam; and the Center for Mining, Electro-Mechanical research of HUMG.

Compliance with ethical standards

Conflict of interest Authors declare that they have no conflict of interest

References

- Alel MNA, Upom MRA, Abdullah RA, Abidin MHZ (2018) Optimizing blasting's air overpressure prediction model using swarm intelligence. In: Journal of Physics: Conference Series, vol 1. IOP Publishing, p 012046
- AminShokravi A, Eskandar H, Derakhsh AM, Rad HN, Ghanadi A (2018) The potential application of particle swarm optimization algorithm for forecasting the air-overpressure induced by mine blasting. *Eng Comput* 34(2):277–285. <https://doi.org/10.1007/s00366-017-0539-5>
- Amiri M, Amnieh HB, Hasanipanah M, Khanli LM (2016) A new combination of artificial neural network and K-nearest neighbors models to predict blast-induced ground vibration and air-overpressure. *Eng Comput* 32(4):631–644
- Armaghani DJ, Hajiassani M, Marto A, Faradonbeh RS, Mohamad ET (2015) Prediction of blast-induced air overpressure: a hybrid AI-based predictive model. *Environ Monit Assess* 187(11):666
- Armaghani DJ, Hasanipanah M, Mohamad ET (2016) A combination of the ICA-ANN model to predict air-overpressure resulting from blasting. *Eng Comput* 32(1):155–171
- Asteris PG, Nikoo M (2019) Artificial bee colony-based neural network for the prediction of the fundamental period of infilled frame structures. *Neural Comput Appl*. <https://doi.org/10.1007/s00521-018-03965-1>
- Asteris P, Kolovos K, Douvika M, Roinos K (2016) Prediction of self-compacting concrete strength using artificial neural networks. *Eur J Environ Civ Eng* 20(sup1):s102–s122
- Bach NV, Nam BX, An ND, Hung TK (2012) Determination of blast-induced ground vibration for non-electric delay blasting (in Vietnamese). *J Sci Technol Hanoi Univ Min Geol* 38(02):25–28
- Bernat K, Drzewiecki W (2015) A study of selected textural features usefulness for impervious surface coverage estimation using Landsat images. In: Image and signal processing for remote sensing XXI, 2015. International society for optics and photonics, p 964327
- Breiman L (2001) Random for Mach Learn 45(1):5–32
- Bui XN, Nguyen H, Le HA, Bui HB, Do NH (2019a) Prediction of blast-induced air over-pressure in open-pit mine: assessment of different artificial intelligence techniques. *Nat Resour Res*. <https://doi.org/10.1007/s11053-019-09461-0>
- Bui X-N, Nguyen H, Le H-A, Bui H-B, Do N-H (2019b) Prediction of blast-induced air over-pressure in open-pit mine: assessment of different artificial intelligence techniques. *Nat Res Res*. <https://doi.org/10.1007/s11053-019-09461-0>
- Dou J, Yunus AP, Bui DT, Merghadi A, Sahana M, Zhu Z, Chen C-W, Khosravi K, Yang Y, Pham BT (2019) Assessment of advanced random forest and decision tree algorithms for modeling rainfall-induced landslide susceptibility in the Izu-Oshima Volcanic Island, Japan. *Sci Total Environ* 662:332–346
- Drzewiecki W (2016) Comparison of selected machine learning algorithms for sub-pixel imperviousness change assessment. In: Geodetic Congress (Geomatics), Baltic, 2016. IEEE, pp 67–72
- Friedman J (1999) Greedy function approximation: A stochastic boosting machine. Department of Statistics Stanford University
- Friedman JH (2001) Greedy function approximation: a gradient boosting machine. *Ann Stat* 29:1189–1232
- Friedman JH (2002) Stochastic gradient boosting. *Comput Stat Data Anal* 38(4):367–378
- Gao W, Guirao JL, Basavanagoud B, Wu J (2018) Partial multi-dividing ontology learning algorithm. *Inf Sci* 467:35–58
- Gao W, Alqahtani AS, Mubarakali A, Mavaluru D, Khalafi S (2019) Developing an innovative soft computing scheme for prediction of air overpressure resulting from mine blasting using GMDH

- optimized by GA. *Eng Comput.* <https://doi.org/10.1007/s00366-019-00720-5>
- Hajihassani M, Armaghani DJ, Sohaei H, Mohamad ET, Marto A (2014) Prediction of airblast-overpressure induced by blasting using a hybrid artificial neural network and particle swarm optimization. *Appl Acoust* 80:57–67
- Hasanipanah M, Armaghani DJ, Khamesi H, Amnieh HB, Ghoraba S (2016) Several non-linear models in estimating air-overpressure resulting from mine blasting. *Eng Comput* 32(3):441–455
- Hustrulid W (1999) *Blasting principles for open-pit blasting: theoretical foundations*. Balkema, Rotterdam
- Jhanwar J, Cakraborty A, Anireddy H, Jethwa J (1999) Application of air decks in production blasting to improve fragmentation and economics of an open pit mine. *Geotech Geol Eng* 17(1):37–57
- Khandelwal M, Kankar P (2011) Prediction of blast-induced air overpressure using support vector machine. *Arab J Geosci* 4(3–4):427–433
- Khandelwal M, Singh T (2005) Prediction of blast induced air overpressure in opencast mine. *Noise Vib Worldw* 36(2):7–16
- Kuhn M, Weston S, Keefer C, Coulter N (2012) Cubist models for regression. R package Vignette R package version 00 18
- Kuhn M, Weston S, Keefer C, Kuhn MM (2018) Package ‘Cubist’
- Kuzu C, Fisne A, Ercelebi S (2009) Operational and geological parameters in the assessing blast induced airblast-overpressure in quarries. *Appl Acoust* 70(3):404–411
- Loder B (1985) National Association of Australian State Road Authorities. In: *Australian Workshop for Senior ASEAN Transport Officials*, 1985, Canberra, 1987
- Mahdiyar A, Marto A, Mirhosseini SA (2018) Probabilistic air-overpressure simulation resulting from blasting operations. *Environ Earth Sci* 77(4):123
- McKenzie C (1990) Quarry blast monitoring: technical and environmental perspectives. *Quarry Manag* 17:23–24
- Mohamad ET, Armaghani DJ, Hasanipanah M, Murlidhar BR, Alel MNA (2016) Estimation of air-overpressure produced by blasting operation through a neuro-genetic technique. *Environ Earth Sci* 75(2):174
- Montahaei M, Oskooi B (2014) Magnetotelluric inversion for azimuthally anisotropic resistivities employing artificial neural networks. *Acta Geophys* 62(1):12–43. <https://doi.org/10.2478/s11600-013-0164-7>
- Naganna SR, Deka PC (2019) Artificial intelligence approaches for spatial modeling of streambed hydraulic conductivity. *Acta Geophys.* <https://doi.org/10.1007/s11600-019-00283-5>
- Nguyen H, Bui X-N (2018a) Feasibility of artificial neural network in predicting blast-induced air overpressure in open-pit mine. *J Min Ind* 01:60–66
- Nguyen H, Bui X-N (2018b) Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. *Nat Resour Res.* <https://doi.org/10.1007/s11053-018-9424-1>
- Nguyen H, Bui X-N, Tran Q-H (2017) Prediction of blast-induced air overpressure in Deo Nai open-pit coal mine using Random Forest algorithm. *J Min Ind* 06:47–53
- Nguyen H, Bui X-N, Tran Q-H, Le T-Q, Do N-H, Hoa LTT (2018a) Evaluating and predicting blast-induced ground vibration in opencast mine using ANN: a case study in Vietnam. *SN Appl Sci* 1(1):125. <https://doi.org/10.1007/s42452-018-0136-2>
- Nguyen H, Bui XN, Bui HB, Mai NL (2018b) A comparative study of artificial neural networks in predicting blast-induced air-blast overpressure at Deo Nai open-pit coal mine, Vietnam. *Neural Comput Appl.* <https://doi.org/10.1007/s00521-018-3717-5>
- Nguyen H, Bui X-N, Tran Q-H, Moayedi H (2019a) Predicting blast-induced peak particle velocity using BGAMs, ANN and SVM: a case study at the Nui Beo open-pit coal mine in Vietnam. *Environ Earth Sci* 78(15):479. <https://doi.org/10.1007/s12665-019-8491-x>
- Nguyen H, Bui X-N, Tran Q-H, Mai N-L (2019b) A new soft computing model for estimating and controlling blast-produced ground vibration based on hierarchical K-means clustering and cubist algorithms. *Appl Soft Comput* 77:376–386. <https://doi.org/10.1016/j.asoc.2019.01.042>
- Nguyen H, Moayedi H, Foong LK, Al Najjar HAH, Jusoh WAW, Rashid ASA, Jamali J (2019c) Optimizing ANN models with PSO for predicting short building seismic response. *Eng Comput.* <https://doi.org/10.1007/s00366-019-00733-0>
- Nguyen H, Drebenstedt C, Bui X-N, Bui DT (2019d) Prediction of blast-induced ground vibration in an open-pit mine by a novel hybrid model based on clustering and artificial neural network. *Nat Resour Res.* <https://doi.org/10.1007/s11053-019-09470-z>
- Nguyen H, Bui X-N, Bui H-B, Cuong DT (2019e) Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study. *Acta Geophys* 67(2):477–490. <https://doi.org/10.1007/s11600-019-00268-4>
- Nguyen H, Bui X-N, Moayedi H (2019f) A comparison of advanced computational models and experimental techniques in predicting blast-induced ground vibration in open-pit coal mine. *Acta Geophys.* <https://doi.org/10.1007/s11600-019-00304-3>
- Piasecki A, Jurasz J, Adamowski JF (2018) Forecasting surface water-level fluctuations of a small glacial lake in Poland using a wavelet-based artificial intelligence method. *Acta Geophys* 66(5):1093–1107. <https://doi.org/10.1007/s11600-018-0183-5>
- Pierini JO, Lavallo M, Telesca L, Gómez EA (2013) Investigating prediction performance of an artificial neural network and a numerical model of the tidal signal at Puerto Belgrano, Bahia Blanca Estuary (Argentina). *Acta Geophys* 61(6):1522–1537. <https://doi.org/10.2478/s11600-012-0093-x>
- Quinlan JR (1992) *Learning with continuous classes*. In: 5th Australian joint conference on artificial intelligence. Singapore, pp 343–348
- Quinlan R (2004) Data mining tools See5 and C5. 0
- Rahmani Y, Farnood Ahmadi F (2018) Application of InSAR in measuring Earth’s surface deformation caused by groundwater extraction and modeling its behavior using time series analysis by artificial neural networks. *Acta Geophys* 66(5):1171–1184. <https://doi.org/10.1007/s11600-018-0182-6>
- Rulequest (2016a) Data Mining with Cubist. <https://www.rulequest.com/cubist-win.html>. Accessed 26 Feb 2019
- Rulequest (2016b) Data Mining with Cubist. <https://www.rulequest.com/cubist-info.html> RuleQuest Research Pty Ltd., St. Ives, NSW, Australia
- Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010) Variance based sensitivity analysis of model output Design and estimator for the total sensitivity index. *Comput Phys Commun* 181(2):259–270
- Shang Y, Nguyen H, Bui X-N, Tran Q-H, Moayedi H (2019) A novel artificial intelligence approach to predict blast-induced ground vibration in open-pit mines based on the firefly algorithm and artificial neural network. *Nat Resour Res.* <https://doi.org/10.1007/s11053-019-09503-7>
- Siskind DE, Stachura VJ, Stagg MS, Kopp JW (1980) *Structure response and damage produced by airblast from surface mining*. Citeseer
- Tarantola S, Gatelli D, Kucherenko S, Mauntz W (2007) Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliab Eng Syst Saf* 92(7):957–960
- Vigneau E, Courcoux P, Symoneaux R, Guérin L, Villière A (2018) Random forests: a machine learning methodology to highlight the volatile organic compounds involved in olfactory perception. *Food Qual Prefer* 68:135–145
- Vinacomin (2015) Report on geological exploration of Coc Sau open pit coal mine, Quang Ninh, Vietnam (in Vietnamese-unpublished). VINACOMIN, Vietnam
- Walter E (1990) *Surface blast design*. Prentice Hall, New Jersey

- Wiszniowski J (2016) Applying the general regression neural network to ground motion prediction equations of induced events in the Legnica-Głogów copper district in Poland. *Acta Geophys* 64(6):2430–2448. <https://doi.org/10.1515/acgeo-2016-0104>
- Zhang X, Nguyen H, Bui X-N, Tran Q-H, Nguyen D-A, Bui DT, Moayedi H (2019) Novel soft computing model for predicting blast-induced ground vibration in open-pit mines based on particle swarm optimization and XGBoost. *Nat Resour Res*. <https://doi.org/10.1007/s11053-019-09492-7>
- Zhou J, Li E, Yang S, Wang M, Shi X, Yao S, Mitri HS (2019a) Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. *Saf Sci* 118:505–518
- Zhou J, Nekouie A, Arslan CA, Pham BT, Hasanipanah M (2019b) Novel approach for forecasting the blast-induced AOp using a hybrid fuzzy system and firefly algorithm. *Eng Comput*. <https://doi.org/10.1007/s00366-019-00725-0>