Acta Geophysica (2019) 67:477–490
https://doi.org/10.1007/s11600-019-00268-4

RESEARCH ARTICLE - SOLID EARTH SCIENCES

# Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: a case study

Hoang Nguyen[1,2] · Xuan-Nam Bui[1,2] · Hoang-Bac Bui[3,4] · Dao Trong Cuong[5]

Received: 26 February 2018 / Accepted: 19 February 2019 / Published online: 25 February 2019

**Abstract**

Ground vibration is one of the most undesirable effects induced by blasting operations in open-pit mines, and it can cause damage to surrounding structures. Therefore, predicting ground vibration is important to reduce the environmental effects of mine blasting. In this study, an eXtreme gradient boosting (XGBoost) model was developed to predict peak particle velocity (PPV) induced by blasting in Deo Nai open-pit coal mine in Vietnam. Three models, namely, support vector machine (SVM), random forest (RF), and *k*-nearest neighbor (KNN), were also applied for comparison with XGBoost. To employ these models, 146 datasets from 146 blasting events in Deo Nai mine were used. Performance of the predictive models was evaluated using root-mean-squared error (RMSE) and coefficient of determination ($R^2$). The results indicated that the developed XGBoost model with RMSE = 1.554, $R^2$ = 0.955 on training datasets, and RMSE = 1.742, $R^2$ = 0.952 on testing datasets exhibited higher performance than the SVM, RF, and KNN models. Thus, XGBoost is a robust algorithm for building a PPV predictive model. The proposed algorithm can be applied to other open-pit coal mines with conditions similar to those in Deo Nai.

**Keywords** eXtreme gradient boosting · XGBoost · Ground vibration · Peak particle velocity

## Introduction

Blasting is one of the highly effective methods in open-cast mining when used to move rocks and overburden. However, only 20–30% of explosion energy is used for rock fragmentation (Chen and Huang 2001; Coursen 1995; Gad et al. 2005; Gao et al. 2018e). The remaining energy is wasted and generates undesirable effects such as ground vibration, air-blast overpressure (AOp), fly rock, and back break (Ak and Konuk 2008; Bui et al. 2019; Chen and Huang 2001; Ghasemi et al. 2016; Hajihassani et al. 2014; Hasanipanah et al. 2017a; Monjezi et al. 2011a; Nguyen and Bui 2018b; Nguyen et al. 2018a). Among these effects, PPV is one of the most undesirable effects because it may be harmful to humans and structures. To reduce the adverse effects of blasting operations, many researchers have proposed empirical equations to predict PPV; among these researchers are the United States Bureau of Mines (Duvall and Fogelson1962; Ambraseys and Hendron 1968; Davies et al.1964; Standard 1973; Roy 1991). However, influencing parameters are numerous, and the relationship among them is complicated. Thus, the empirical methods may not be entirely suitable for predicting PPV in open-cast mines (Ghasemi et al. 2013; Hajihassani et al. 2015; Hasanipanah et al. 2015; Monjezi et al. 2011b, 2013; Nguyen and Bui 2018a; Nguyen et al. 2018b, 2019; Saadat et al. 2014).

Nowadays, artificial intelligence (AI) is well known as a robust tool for solving the real-life problems (Alnaqi et al. 2019; Gao et al. 2018a, c; Moayedi and Nazir 2018; Moayedi et al. 2019; Moayedi and Rezaei 2017). Many researchers

✉ Hoang Nguyen
nguyenhoang@humg.edu.vn

[1] Department of Surface Mining, Mining Faculty, 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 Geosciences and Geoengineering, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Vietnam

[4] 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

[5] Ministry of Industry and Trade, Hanoi, Vietnam

have studied and applied AI in predicting blast-induced issues, especially blast-produced PPV. Longjun et al. (2011) applied two benchmark algorithms for estimating PPV, including support vector machine (SVM) and random forest (RF); two other parameters with 93 explosions were used as training datasets, and 15 observations among 93 views were selected as testing datasets. Their study indicated that the SVM and RF models performed well in estimating blast-induced PPV. The SVM model was introduced as a superior model in their study. Hasanipanah et al. (2017b) also developed a Classification and regression tree (CART) model to predict PPV at Miduk copper mine (Iran) using 86 blasting events. Multiple regression (MR) and various empirical techniques were also considered to predict PPV and compared with the CART model. As a result, the CART model was exhibited better performance than the other models with RMSE = 0.17 and $R^2 = 0.95$ in their study. In another work, Chandar et al. (2017) estimated blast-induced PPV using ANN model; 168 blasting operations were collected in dolomite, coal mine, and limestone (Malaysia) for their aim. The results indicated that the ANN model, with $R^2 = 0.878$ for the three mines, is the best among the approaches used in their study. Metaheuristics algorithm was also considered and used to predict PPV by Faradonbeh and Monjezi (2017), i.e., gene expression programming (GEP); 115 blasting operations were used for their study. Accordingly, a formula based on the GEP was developed to estimate PPV as the first step in their study. Then, it was compared with several nonlinear and general equation models as the second step as well. Their results designated that the GEP model was better than the other models in forecasting blast-induced PPV. Similar works can be found at those references (Faradonbeh et al. 2016; Hasanipanah et al. 2017c; Sheykhi et al. 2018; Taheri et al. 2017).

In this study, an XGBoost model was developed to predict blast-induced PPV in Deo Nai open-pit coal mine (Vietnam). Three other models were also produced, including SVM, RF, and KNN for comparison with the constructed XGBoost model.

This paper is organized as follows. Section “two” describes the site study and the data used. Section “three” provides an overview of the algorithms used in this study. Section “four” reports the results and discussion. Section “five” shows the validation of the constructed models. Finally, Section “six” presents our conclusions.

## Site study and data used

### Study area

With the total area up to ~ 6 $\text{Km}^2$, the Deo Nai open-pit coal mine was a large open-cast coal mine in Vietnam (Fig. 1). It is located in Quang Ninh province, Vietnam, with the proven reserve is 42.5 Mt, and productivity is 2.5 Mt/year. The study area has a complex geological structure, includes many different phases and faults. Conglomerate, siltstone, sandstone, claystone, and argillic rock were included in the overburden of this mine (Vinacomin 2015). The hardness of these rocks (f) in the range of 11–12 according to Protodiakonov’s classification (Protodiakonov et al. 1964); specific weight ($\gamma$) in the range of 2.62–2.65 $\text{t/m}^3$. Therefore, blasting operations for rock fragmentation in this mine is a high-performance method.

However, the Deo Nai open-pit coal mine is located near residential areas (Fig. 1), which have a distance of approximately 400 m from the blasting sites. Moreover, the capacity of burden must explode significantly in a blast of up to more 20 tons, and the adverse effects (especially PPV) of the blasting operation to the surrounding environment are substantial. Thus, we have selected this area as a case study to consider and predict PPV caused by blasting operations with the aim of controlling the undesirable effects on the environment and residential areas.

### Data collection

To conduct this study, 146 blasting events were collected with nine parameters, such as the number of borehole rows per blast ($N$), charge per delay ($Q$), powder factor ($q$), length of stemming ($T$), burden ($B$), monitoring distance ($D$), spacing ($S$), bench height ($H$), and time interval between blasts ($\Delta t$) which were considered as nine input parameters to predict the outcome, i.e., PPV. Table 1 shows a brief of the datasets used in this study.

For monitoring PPV, the Blastmate III instrument (Instantel, Canada) was used with the specifications that are shown in Table 2. In this study, PPV values were recorded in the range of 2.140 to 33.600 mm/s. A GPS device was used to determine D. The remaining parameters were extracted from blast patterns.

## Preview of XGBoost, SVM, RF, and KNN

### eXtreme gradient boosting (XGBoost)

XGBoost is an improved algorithm based on the gradient boosting decision proposed by (Friedman et al. 2000, 2001; Friedman 2001, 2002). XGBoost, which was created and developed by Chen and He (2015), can construct boosted trees efficiently, operate in parallel, and solve both classification and regression problems. The core of the algorithm is the optimization of the value of the objective function. It implements machine learning algorithms under the gradient boosting framework. XGBoost can

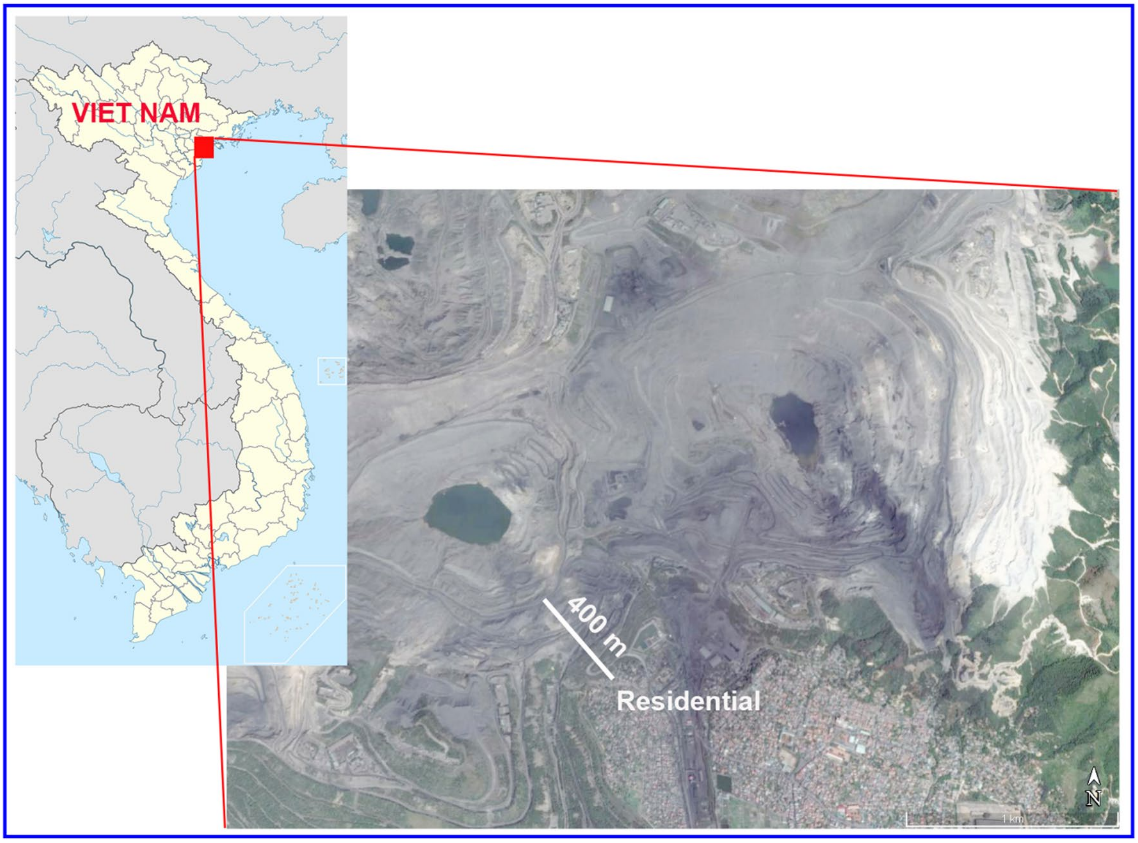


**Fig. 1** Location of the study area

**Table 1** Blasting events recorded for this study

| $Q$ (Kg) | $H$ (m) | $B$ (m) | $S$ (m) | $T$ (m) |
|---|---|---|---|---|
| Minimum: 3007 | Minimum: 13.00 | Minimum: 7.500 | Minimum: 7.40 | Minimum: 6.20 |
| Mean: 13,324 | Mean: 14.38 | Mean: 8.071 | Mean: 7.82 | Mean: 6.87 |
| Maximum: 24,171 | Maximum: 16.0 | Maximum: 8.50 | Maximum: 8.20 | Maximum: 7.50 |
| $q$ (kg/m$^3$) | $N$ (borehole) | $D$ (m) | $\Delta t$ (ms) | PPV (mm/s) |
| Minimum: 0.3500 | Minimum: 2.000 | Minimum: 180.0 | Minimum: 17.00 | Minimum: 2.140 |
| Mean: 0.4184 | Mean: 3.525 | Mean: 465.8 | Mean: 27.64 | Mean: 13.322 |
| Maximum: 0.480 | Maximum: 5.00 | Maximum: 726 | Maximum: 42.00 | Maximum: 33.600 |

**Table 2** Basic parameters of the PPV monitoring instrument Source: http://www.instantel.com

| Features | Parameters |
|---|---|
| Range | 0.127–254 mm/s |
| Accuracy | ± 0.5 mm/s or 5% (0.02 in/s) |
| Transducer density | 2.13 g/cc (133 lbs/ft3) |

solve many data science problems in a fast and accurate way with parallel tree boosting such as gradient boosting decision tree and gradient boosting machine.

An objective function usually consists of two parts (training loss and regularization):

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

where $L$ is the training loss function and $\Omega$ is the regularization term. The training loss is used to measure the model performance on training data. The regularization term aims to control the complexity of the model such as overfitting (Gao et al. 2018d). Various ways are conducted to define complexity. However, the complexity of each tree is often computed as the following equation:

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

where $T$ is the number of leaves and $\omega$ is the vector of scores on leaves.

The structure score of XGBoost is the objective function defined as follows:

$$\text{Obj} = \sum_{j=1}^{T} \left[ G_j\omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2 \right] + \gamma T, \tag{3}$$

where $\omega_j$ are independent of each other. The form $G_j\omega_j + \frac{1}{2}(H_j + \lambda)\omega_j^2$ is quadratic and the best $\omega_j$ for a given structure $q(x)$.

## Support vector machines (SVM)

SVM is a machine learning method based on statistical theory and developed by (Cortes and Vapnik 1995). This method continues to be applied to high-performing algorithms with slight tuning. Similar to CART, SVM can also be used to solve classification and regression problems. According to Cortes and Vapnik (1995), SVM was used for classification analysis. SVR, a version of SVM for regression analysis, was proposed by Drucker et al. (1997).

In SVM, fitting data $\{x_i, y_i\}$, $(i = 1, 2, \ldots, n)$, $x_i \in R^n$, $y_i \in R$ with a function $f(x) = w \cdot x + b$ is a problem. Thus, according to SVM theory, the fitting problem function is expressed as follows:

$$f(x) = w \cdot x + b = \sum_{i=1}^{k} (a_i - a_i^*)K(xx_i) + b \tag{4}$$

where $a_i$, $a_i^*$, and $b$ are obtained by solving subsequent second optimization problems. Usually, a small fraction of $a_i$, $a_i^*$ is not zero; this fraction is called support vector.

Max:

$$w(a, a^*) = -\frac{1}{2} \sum_{i,j=1}^{k} (a_i - a_i^*)(a_j - a_j^*)K(x_ix_j) + \sum_{i=1}^{k} y_i(a_i - a_i^*) - \varepsilon \sum_{i=1}^{k} (a_i + a_i^*), \tag{5}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{k} (a_i - a_i^*) = 0 \\ 0 \le a_i, a_i^* \le C, (i = 1, 2, \ldots, k) \end{cases}, \tag{6}$$

where $C$ is a penalty factor that shows the penalty degree to samples of excessive error $\varepsilon$; $K(x_ix_j)$ is kernel function, which solves calculation problems of high dimension skillfully by introducing kernel functions. These functions are mainly of the following types:

1. Linear kernel

$$K(x, y) = x \cdot y, \tag{7}$$

2. Polynomial kernel

$$K(x, y) = [(x \cdot y) + 1]^d; \quad d = (1, 2, \ldots), \tag{8}$$

3. Radial original kernel function

$$K(x, y) = \exp\left[\frac{-\|x - y\|^2}{\sigma^2}\right], \tag{9}$$

4. Two-layer neural kernel

$$K(x, y) = \tanh\left[a(x \cdot y) - \delta\right]. \tag{10}$$

In this study, the SVM method with a polynomial kernel function is used to develop the SVM model for anticipating PPV.

## Random forest (RF)

RF is one of the decision tree algorithms and introduced by Breiman (2001) for the first time. It is well known as a robust non-parametric statistical technique for both regression and classification problems. On the other hand, RF was introduced as an ensemble method based on the results from different trees to achieve predictive accuracy (Vigneau et al. 2018). For each new observation, RF combines the predicted values from the individual tree in the forest to give the best result. In the forest, each tree roles as a voter for the final decision of the RF (Gao et al. 2018b). The core of the RF model for regression can be described as three steps follow:

*Step 1* Create bootstrap samples as the number of the tree in the forest ($n_{\text{tree}}$) based on the dataset.
*Step 2* Develop an unpruned regression tree for each bootstrap sample by random sampling of the predictors ($m_{\text{try}}$). Among those variables, select the best split.
*Step 3* Predict new observation by ensemble the predicted values of the trees ($n_{\text{tree}}$). For the regression problem as well as predicting blast-induced PPV, the average value of the predicted values by the individual tree in the forest used.

Based on the training dataset, an estimate of the error rate can be obtained by the following:

- At each bootstrap iteration, predict the data not in the bootstrap sample using the tree grown with the bootstrap sample, called “out-of-bag” (OOB).
- Aggregate the OOB predictions and calculate the error rate.

The implementation of the RF algorithm for predicting blast-induced PPV in this study is shown in Fig. 2. More details of the RF algorithm can be found at those references (Breiman 2001; Bui et al. 2019; Nguyen and Bui 2018b).

### *k*-nearest neighbor (KNN)

KNN is known as a favorite technique for solving regression and classification problems in machine learning and introduced by Altman (1992). Based on the closest neighbors (*k* neighbors), the KNN algorithm determines the testing point and classify them. On the other hand, the KNN algorithm does not learn anything from training data. It only remembers the weights of neighbors in the functional space. When it comes to forecasting a new observation, it searches similar results and calculates the distance to those neighbors. Therefore, KNN is classified as “lazy learning” algorithms (Fig. 3).

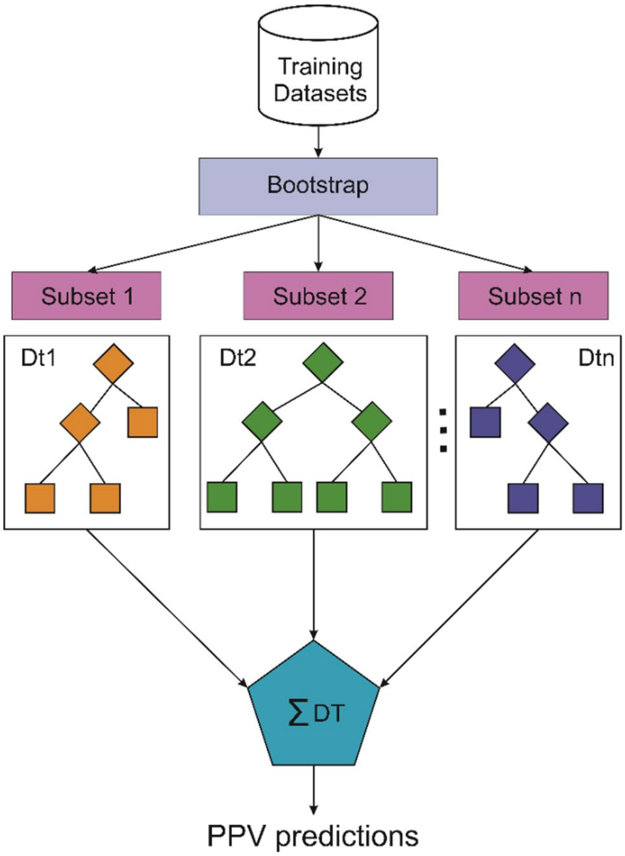


**Fig. 2** Workflow of RF in predicting blast-induced PPV

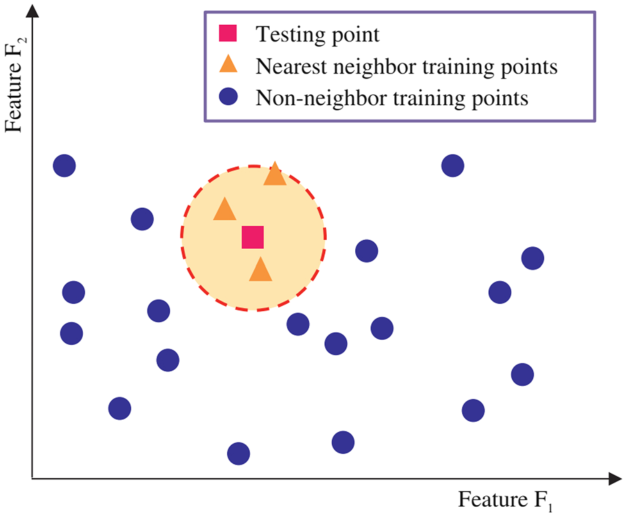

Testing point
Nearest neighbor training points
Non-neighbor training points
Feature F2
Feature F1

**Fig. 3** Illustration of KNN algorithm for two-dimensional feature space (Hu et al. 2014)

For regression problems as well as predicting blast-induced PPV, the KNN algorithm uses a weighted average of the *k*-nearest neighbors, computed by their distance inversely. The KNN for regression can be worked as four steps follow:

*Step 1* Determine the distance from the query sample to the labeled samples.

$$d(x_{tr}, x_t) = \sqrt{\sum_{n=1}^{N} w_n (x_{tr,n} - x_{t,n})^2} \tag{11}$$

where $N$ is the number of features; $x_{tr,n}$ and $x_{t,n}$ denote the $n$th feature values of the training ($x_{tr}$) and testing ($x_t$) points, respectively; $w_n$ is the weight of the $n$th feature and lies interval [0,1].
*Step 2* Order the labeled examples by increasing distance.
*Step* 3 Based on RMSE (Eq. 12), define the optimal number of neighbors. Cross-validation can be used for this task.
*Step 4* Calculate the average distance inversely with *k*-nearest neighbors.

## Results and discussion

In this study, the datasets are divided into two sections: training and testing. Of the total datasets, 80% (approximately 118 blasting events) are used for the training process, and the rest (28 observations) are used for the testing process. The training dataset is used for the development of the mentioned models. The testing dataset is used to assess the performance of the constructed models.

To evaluate the performance of the constructed models, two criteria statistical include determination coefficient ($R^2$) and root-mean-square error (RMSE) are used with RMSE provide an idea of how wrong all predictions are (0 is perfect), and $R^2$ provides an idea of how well the model fits the data (1 is perfect, 0 is worst). In this study, RMSE and $R^2$ were computed using the following equations:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{12}$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{13}$$

where $n$ denotes for the number of data, $y_i$ and $\hat{y}_i$ denotes the measured and predicted values, respectively; $\bar{y}$ is the mean of the measured values.

Additionally, the Box–Cox transform and 10-fold cross-validation methods are used to avoid overfitting/underfitting.

## XGBoost

In XGBoost, two stopping criteria, namely, *maximum tree depth* and *nrounds,* were considered to prevent complexity in modeling. Selecting the significant values for *maximum tree depth* and the *nrounds* causes excessive growth of the tree and an overfitting problem. Therefore, the *maximum tree depth* is set in the 1–3 range, and *nrounds* is set as 50, 100, and 150.

To achieve an optimum combination of these two parameters, a trial-and-error procedure was conducted with the range of two settings proposed. The performance indices, which include RMSE and $R^2$, were calculated to evaluate the XGBoost models on both the training and testing datasets (Table 3).

Based on Table 3, nine XGBoost models were developed and evaluated. The results of the XGBoost models in Table 3 are very close to each other, which causes difficulty in selecting the best model. Thus, a simple procedure with the ranking method proposed by Zorlu et al. (2008) is applied in Table 4. The XGBoost models in Table 4 are ranked and evaluated through ranking indicators. The results of the overall grade for XGBoost models 1–9 are summarized in Table 5.

According to Table 5, model 1 with the total rank value of 35 reached the highest value among all the constructed XGBoost models. On other words, the XGBoost model No. 1 performed better than the other XGBoost models in this study.

## Support vector machine (SVM)

In SVM, the kernel function with polynomial kernel was used to develop the SVM models. Two stopping criteria, namely *degree* and *cost*, were considered to prevent complexity in modeling. Also, the *scale* parameter was held constant at a value of 0.1. In this study, we select the range of 1–3 for the *degree* and set the *cost* as 0.25, 0.5, and 1.

To achieve an optimum combination of these two parameters, a trial-and-error procedure was also conducted similarly to that for the XGBoost method with the range of the two SVM parameters. The performance indices, namely, RMSE and $R^2$, were calculated to evaluate the SVM models on both the training and testing datasets (Table 6).

Table 6 shows some low-performance models such as nos. 1, 4, 7, 2, 5. However, some models exhibit high performances that are almost similar. Thus, a simple ranking method should be applied to determine the best SVM model among the developed ones, as shown in Table 7. Table 8 indicates the total rank of the SVM models 1–9.

According to Table 8, model 6 with a total rank of 32 achieved the best performance among all the developed

**Table 3** Performance indicators of the XGBoost models

| Technique | Model no. | *n*rounds | max_depth | Training process | | Testing process | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | $R^2$ | RMSE | $R^2$ |
| XGBoost | 1 | 150 | 1 | 1.554 | 0.955 | 1.742 | 0.952 |
| | 2 | 150 | 2 | 1.636 | 0.951 | 1.783 | 0.951 |
| | 3 | 150 | 3 | 1.711 | 0.945 | 1.840 | 0.948 |
| | 4 | 100 | 1 | 1.568 | 0.954 | 1.745 | 0.951 |
| | 5 | 100 | 2 | 1.609 | 0.953 | 1.765 | 0.950 |
| | 6 | 100 | 3 | 1.662 | 0.949 | 1.853 | 0.947 |
| | 7 | 50 | 1 | 1.725 | 0.951 | 1.947 | 0.946 |
| | 8 | 50 | 2 | 1.618 | 0.953 | 1.748 | 0.952 |
| | 9 | 50 | 3 | 1.617 | 0.952 | 1.808 | 0.949 |

**Table 4** The ranking of the XGBoost models based on their performance

| Technique | Model | RMSE | $R^2$ | Rank for RMSE | Rank for $R^2$ | Total rank |
|---|---|---|---|---|---|---|
| XGBoost | Training 1 | 1.554 | 0.955 | 9 | 9 | 18 |
| | Training 2 | 1.636 | 0.951 | 4 | 4 | 8 |
| | Training 3 | 1.711 | 0.945 | 2 | 1 | 3 |
| | Training 4 | 1.568 | 0.954 | 8 | 8 | 16 |
| | Training 5 | 1.609 | 0.953 | 7 | 7 | 14 |
| | Training 6 | 1.662 | 0.949 | 3 | 2 | 5 |
| | Training 7 | 1.725 | 0.951 | 1 | 3 | 4 |
| | Training 8 | 1.618 | 0.953 | 5 | 6 | 11 |
| | Training 9 | 1.617 | 0.952 | 6 | 5 | 11 |
| | Testing 1 | 1.742 | 0.952 | 9 | 8 | 17 |
| | Testing 2 | 1.783 | 0.951 | 5 | 6 | 11 |
| | Testing 3 | 1.840 | 0.948 | 3 | 3 | 6 |
| | Testing 4 | 1.745 | 0.951 | 8 | 7 | 15 |
| | Testing 5 | 1.765 | 0.950 | 6 | 5 | 11 |
| | Testing 6 | 1.853 | 0.947 | 2 | 2 | 4 |
| | Testing 7 | 1.947 | 0.946 | 1 | 1 | 2 |
| | Testing 8 | 1.748 | 0.952 | 7 | 9 | 16 |
| | Testing 9 | 1.808 | 0.949 | 4 | 4 | 8 |

**Table 5** Total rank of XGBoost models

| Technique | Model no. | Total rank |
|---|---|---|
| XGBoost | **1** | **35** |
| | 2 | 19 |
| | 3 | 9 |
| | 4 | 31 |
| | 5 | 25 |
| | 6 | 9 |
| | 7 | 6 |
| | 8 | 27 |
| | 9 | 19 |

The best model was shown in bold

SVM models. Thus, we conclude that model 6 is the best SVM model with the SVM method. Note that, the same training and testing datasets were applied for the development of the SVM models as those used for the XGBoost models.

## Random forest (RF)

With the RF technique, two stopping criteria called $n_{\text{tree}}$ and $m_{\text{try}}$ were considered to prevent complexity and reduce the running time of the model. A trial-and-error procedure with $n_{\text{tree}}$ is discussed in the range of 50–150, whereas $m_{\text{try}}$ set as 5, 7, and 9 is implemented in Table 9. Likewise to the development of the XGBoost and SVM models, the same training and testing datasets were applied for the development of the RF models in this study.

**Table 6** Performance indices of SVM models

| Method | Model no. | Degree | Cost | Training data | | Testing data | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | $R^2$ | RMSE | $R^2$ |
| SVM | 1 | 1 | 0.25 | 2.763 | 0.877 | 3.875 | 0.805 |
| | 2 | 2 | 0.25 | 2.213 | 0.917 | 2.991 | 0.869 |
| | 3 | 3 | 0.25 | 2.074 | 0.923 | 2.580 | 0.926 |
| | 4 | 1 | 0.5 | 2.725 | 0.880 | 3.763 | 0.808 |
| | 5 | 2 | 0.5 | 2.185 | 0.918 | 2.753 | 0.890 |
| | 6 | 3 | 0.5 | 2.036 | 0.921 | 2.566 | 0.934 |
| | 7 | 1 | 1 | 2.664 | 0.883 | 3.908 | 0.783 |
| | 8 | 2 | 1 | 2.186 | 0.916 | 2.469 | 0.916 |
| | 9 | 3 | 1 | 2.152 | 0.911 | 2.440 | 0.944 |

**Table 7** Performance indices of SVM models with the rank

| Method | Model | RMSE | $R^2$ | Rank for RMSE | Rank for $R^2$ | Total rank |
|---|---|---|---|---|---|---|
| SVM | Training 1 | 2.763 | 0.877 | 1 | 1 | 2 |
| | Training 2 | 2.213 | 0.917 | 4 | 6 | 10 |
| | Training 3 | 2.074 | 0.923 | 8 | 9 | 17 |
| | Training 4 | 2.725 | 0.880 | 2 | 2 | 4 |
| | Training 5 | 2.185 | 0.918 | 6 | 7 | 13 |
| | Training 6 | 2.036 | 0.921 | 9 | 8 | 17 |
| | Training 7 | 2.664 | 0.883 | 3 | 3 | 6 |
| | Training 8 | 2.186 | 0.916 | 5 | 5 | 10 |
| | Training 9 | 2.152 | 0.911 | 7 | 4 | 11 |
| | Testing 1 | 3.875 | 0.805 | 2 | 2 | 4 |
| | Testing 2 | 2.991 | 0.869 | 4 | 4 | 8 |
| | Testing 3 | 2.580 | 0.926 | 6 | 7 | 13 |
| | Testing 4 | 3.763 | 0.808 | 3 | 3 | 6 |
| | Testing 5 | 2.753 | 0.890 | 5 | 5 | 10 |
| | Testing 6 | 2.566 | 0.934 | 7 | 8 | 15 |
| | Testing 7 | 3.908 | 0.783 | 1 | 1 | 2 |
| | Testing 8 | 2.469 | 0.916 | 8 | 6 | 14 |
| | Testing 9 | 2.440 | 0.944 | 9 | 9 | 18 |

**Table 8** Total rank of SVM models

| Technique | Model no. | Total rank |
|---|---|---|
| SVM | 1 | 6 |
| | 2 | 18 |
| | 3 | 30 |
| | 4 | 10 |
| | 5 | 23 |
| | **6** | **32** |
| | 7 | 8 |
| | 8 | 24 |
| | 9 | 29 |

The best model was shown in bold

Based on Table 9, all of the nine constructed RF models are suitable for estimating blast-produced PPV in this study. Some of the RF models, such as models 5–9, provide higher performance than others. However, the results of the models are nearly similar. Thus, concluding which model is the best for the RF technique is difficult. A ranking technique was used to identify the best model for the RF technique, as reported in Table 10. Additionally, a total ranking of the RF models is computed in Table 11.

According to Tables 10 and 11, RF model 7 with a total ranking value of 30 reached the highest value among all the developed RF models. Thus, we can conclude that RF model 7 with $n_{\text{tree}} = 150$ and $m_{\text{try}} = 9$ is the superior model in the RF technique for anticipating blast-produced PPV in this study.

**Table 9** The RF models performance for predicting blast-induced PPV

| Technique | Model no. | $n_{\text{tree}}$ | $m_{\text{try}}$ | Training process | | Testing process | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | $R^2$ | RMSE | $R^2$ |
| RF | 1 | 150 | 5 | 1.816 | 0.944 | 2.837 | 0.885 |
| | 2 | 100 | 5 | 1.819 | 0.944 | 2.975 | 0.874 |
| | 3 | 50 | 5 | 1.803 | 0.944 | 3.126 | 0.859 |
| | 4 | 150 | 7 | 1.736 | 0.948 | 2.031 | 0.939 |
| | 5 | 100 | 7 | 1.737 | 0.948 | 1.945 | 0.943 |
| | 6 | 50 | 7 | 1.745 | 0.948 | 1.925 | 0.944 |
| | 7 | 150 | 9 | 1.719 | 0.946 | 1.811 | 0.947 |
| | 8 | 100 | 9 | 1.723 | 0.946 | 1.804 | 0.948 |
| | 9 | 50 | 9 | 1.717 | 0.946 | 1.886 | 0.944 |

**Table 10** The RF models with their rank through performance indicators

| Technique | Model | RMSE | $R^2$ | Rank for RMSE | Rank for $R^2$ | Total rank |
|---|---|---|---|---|---|---|
| RF | Training 1 | 1.816 | 0.944 | 2 | 1 | 3 |
| | Training 2 | 1.819 | 0.944 | 1 | 3 | 4 |
| | Training 3 | 1.803 | 0.944 | 3 | 2 | 5 |
| | Training 4 | 1.736 | 0.948 | 6 | 8 | 14 |
| | Training 5 | 1.737 | 0.948 | 5 | 9 | 14 |
| | Training 6 | 1.745 | 0.948 | 4 | 7 | 11 |
| | Training 7 | 1.719 | 0.946 | 8 | 6 | 14 |
| | Training 8 | 1.723 | 0.946 | 7 | 4 | 11 |
| | Training 9 | 1.717 | 0.946 | 9 | 5 | 14 |
| | Testing 1 | 2.837 | 0.885 | 3 | 3 | 6 |
| | Testing 2 | 2.975 | 0.874 | 2 | 2 | 4 |
| | Testing 3 | 3.126 | 0.859 | 1 | 1 | 2 |
| | Testing 4 | 2.031 | 0.939 | 4 | 4 | 8 |
| | Testing 5 | 1.945 | 0.943 | 5 | 5 | 10 |
| | Testing 6 | 1.925 | 0.944 | 6 | 7 | 13 |
| | Testing 7 | 1.811 | 0.947 | 8 | 8 | 16 |
| | Testing 8 | 1.804 | 0.948 | 9 | 9 | 18 |
| | Testing 9 | 1.886 | 0.944 | 7 | 6 | 13 |

**Table 11** Total ranking of RF models

| Technique | Model no. | Total rank |
|---|---|---|
| RF | 1 | 9 |
| | 2 | 8 |
| | 3 | 7 |
| | 4 | 22 |
| | 5 | 24 |
| | 6 | 24 |
| | **7** | **30** |
| | 8 | 29 |
| | 9 | 27 |

The best model was shown in bold

## *k*-nearest neighbor (KNN)

In this study, nine KNN models were developed with the *k neighbors* set in a range of 3–11 through training datasets. The performance of the KNN models was evaluated using the testing dataset as the second step in the development of the KNN models. Note that the same datasets were used for the development of the KNN models as those used for the development of the models above. The performance indices of the KNN models are shown in Table 12.

As shown in Table 12, the results of the constructed KNN models are close to one another. Thus, determining which model is the most optimal among the built KNN models is difficult. A simple ranking method similar to the previous sections was applied to the KNN technique. The performance indices of the KNN models with their rank were

**Table 12** The KNN models performance in this study

| Technique | Model no. | *k* | Training process | | Testing process | |
|---|---|---|---|---|---|---|
| | | | RMSE | $R^2$ | RMSE | $R^2$ |
| KNN | 1 | 3 | 2.878 | 0.853 | 3.708 | 0.791 |
| | 2 | 4 | 2.864 | 0.861 | 3.581 | 0.809 |
| | 3 | 5 | 2.796 | 0.876 | 3.161 | 0.859 |
| | 4 | 6 | 2.857 | 0.878 | 3.326 | 0.851 |
| | 5 | 7 | 2.985 | 0.867 | 3.203 | 0.870 |
| | 6 | 8 | 3.008 | 0.867 | 3.278 | 0.874 |
| | 7 | 9 | 3.006 | 0.866 | 3.368 | 0.861 |
| | 8 | 10 | 3.061 | 0.859 | 3.466 | 0.850 |
| | 9 | 11 | 3.017 | 0.864 | 3.434 | 0.863 |

**Table 13** Performance of the KNN models with the rank

| Technique | Model | RMSE | $R^2$ | Rank for RMSE | Rank for $R^2$ | Total rank |
|---|---|---|---|---|---|---|
| KNN | Training 1 | 2.878 | 0.853 | 6 | 1 | 7 |
| | Training 2 | 2.864 | 0.861 | 7 | 3 | 10 |
| | Training 3 | 2.796 | 0.876 | 9 | 8 | 17 |
| | Training 4 | 2.857 | 0.878 | 8 | 9 | 17 |
| | Training 5 | 2.985 | 0.867 | 5 | 7 | 12 |
| | Training 6 | 3.008 | 0.867 | 3 | 6 | 9 |
| | Training 7 | 3.006 | 0.866 | 4 | 5 | 9 |
| | Training 8 | 3.061 | 0.859 | 1 | 2 | 3 |
| | Training 9 | 3.017 | 0.864 | 2 | 4 | 6 |
| | Testing 1 | 3.708 | 0.791 | 1 | 1 | 2 |
| | Testing 2 | 3.581 | 0.809 | 2 | 2 | 4 |
| | Testing 3 | 3.161 | 0.859 | 9 | 5 | 14 |
| | Testing 4 | 3.326 | 0.851 | 6 | 4 | 10 |
| | Testing 5 | 3.203 | 0.870 | 8 | 8 | 16 |
| | Testing 6 | 3.278 | 0.874 | 7 | 9 | 16 |
| | Testing 7 | 3.368 | 0.861 | 5 | 6 | 11 |
| | Testing 8 | 3.466 | 0.850 | 3 | 3 | 6 |
| | Testing 9 | 3.434 | 0.863 | 4 | 7 | 11 |

**Table 14** Total rank of KNN models

| Technique | Model no. | Total rank |
|---|---|---|
| KNN | 1 | 9 |
| | 2 | 14 |
| | **3** | **31** |
| | 4 | 27 |
| | 5 | 28 |
| | 6 | 25 |
| | 7 | 20 |
| | 8 | 9 |
| | 9 | 17 |

The best model was shown in bold

**Table 15** Statistical values for selected predictive models

| Model | Training data | | Testing data | |
|---|---|---|---|---|
| | RMSE | $R^2$ | RMSE | $R^2$ |
| XGBoost | 1.554 | 0.955 | 1.742 | 0.952 |
| SVM | 2.036 | 0.921 | 2.566 | 0.934 |
| RF | 1.719 | 0.946 | 1.811 | 0.947 |
| KNN | 2.796 | 0.876 | 3.161 | 0.859 |

calculated and the results are presented in Table 13. Additionally, Table 14 shows the total rank of KNN models.

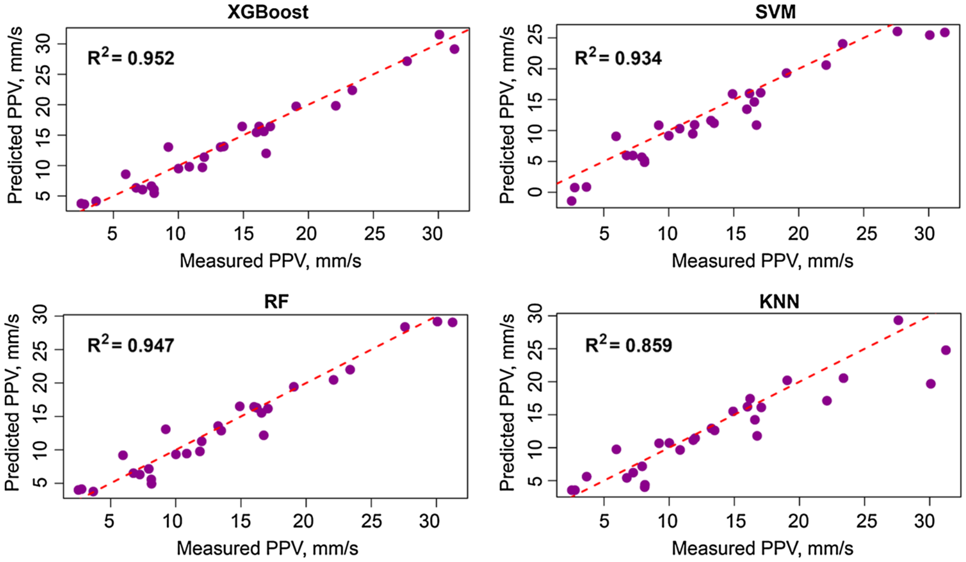
According to Tables 13 and 14, nine KNN models were ranked with the value of total rank in the range of 9–31. As shown in the tables, KNN model 3 with an entire rank value of 31 achieved the highest value among the developed KNN models.

## Validation performance of models

In this study, two statistical criteria, namely, $R^2$ and RMSE, were employed to measure the performance of the selected predictive models and computed using Eqs. (12–13). After the optimal models for each technique were selected, the values of the aforementioned statistical criteria for all models were calculated for both the training and testing datasets, as indicated in Table 15. According to these results, the accuracy level of the XGBoost technique is better than those of the SVM, RF, and KNN models. Figure 4 demonstrates the performance of the models in forecasting blast-induced PPV on the testing dataset.

Figure 5 presents a useful way to consider the spread of the estimated accuracies for the various methods and how they relate among the XGBoost, SVM, RF, and KNN techniques. According to Fig. 5, the KNN technique has the lowest accuracy level with several outliers, whereas the XGBoost technique exhibits the highest accuracy level without outliers. The RF technique can also provide an approximation of the XGBoost performance. However, a closer look shows that the developed XGBoost model offers higher performance than the RF model. Furthermore, the RF technique appears to have outliers, whereas the established XGBoost model has none. Additionally, the accuracy of the selected PPV predictive models was also

**Fig. 4** Measured versus predicted values on the testing dataset



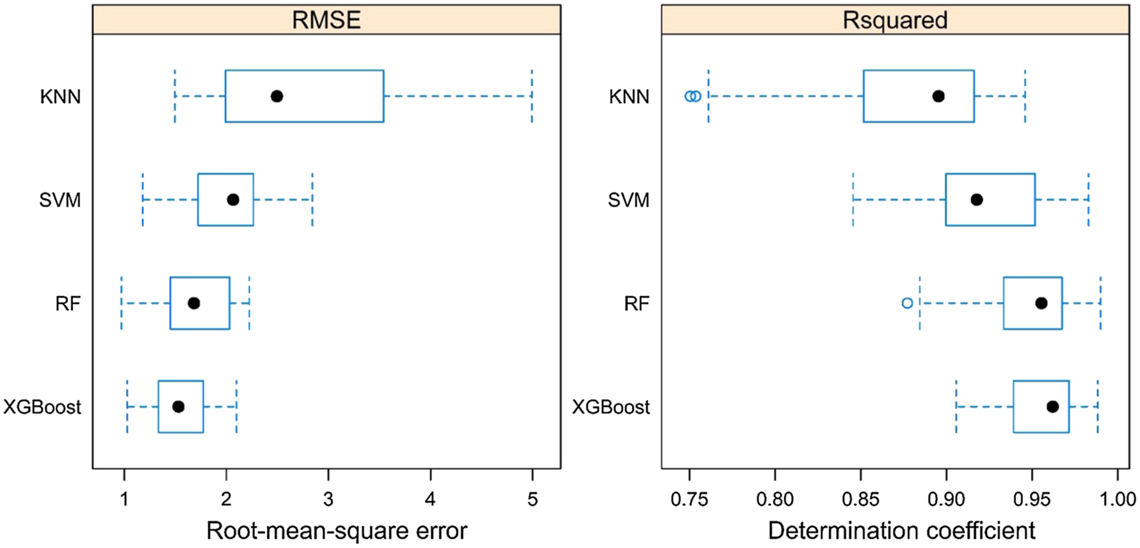


**Fig. 5** Comparison of machine learning algorithms in box and whisker plots

compared and shown in Fig. 6. According to Fig. 6, among the developed models, the XGBoost technique yields the most reliable results in forecasting blast-produced PPV.

Considering the input variables in this study, it shows that the number of input variables is high (9 input variables). Therefore, an analysis procedure of sensitivity was performed to find out which input variable(s) is/are the most influential parameters on blast-induced PPV as shown in Fig. 7. As a result, $Q$ (charge) and D (distance) are the most influential factors on blast-induced PPV in this study. They should be used in practical engineering to control blast-induced PPV. The other input parameters were also effected on blast-induced PPV but not much.

## Conclusions and recommendations

In practice, an accurate and efficient estimation of PPV is essential to reduce the environmental effects of blasting operations, especially near residential areas. This study developed the XGBoost, SVM, RF, and KNN models to predict PPV caused by blasting operations in the Deo Nai open-pit coal mine in Vietnam. Nine input parameters ($Q$, $H$, $B$, $S$, $T$, $q$, $N$, $D$, and $\Delta t$) were used to predict PPV from 146 blasting events at the mine. For modeling purposes, all datasets were divided into training and testing sets, with 80% (118 observations) of the entire dataset used for training and 20% (28 representations) for testing. The performance

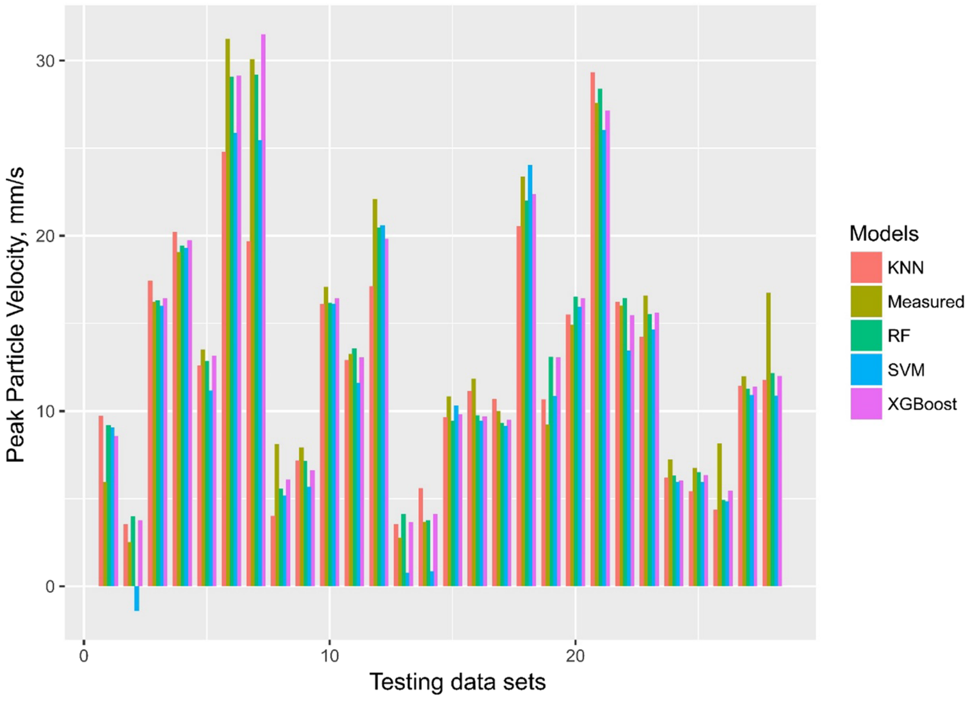


**Fig. 6** Prediction values of selected predictive models on testing datasets

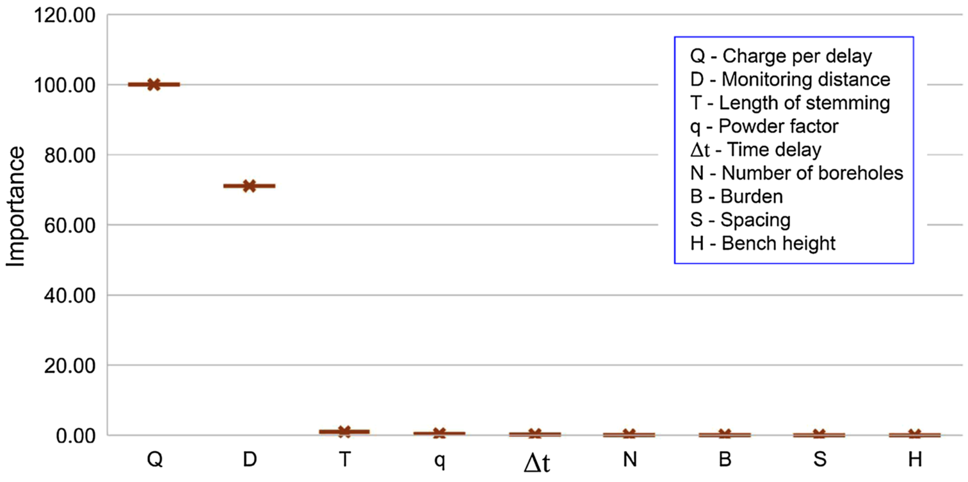


**Fig. 7** Sensitivity analysis of independent variables for the PPV predictive model

of the predictive models was evaluated based on two criteria, namely, $R^2$ and RMSE, using the training and testing datasets. Based on the results of this study, RMSE values of 1.554 and 1.742 were obtained for the XGBoost model on the training and testing datasets, respectively. These values are the smallest among the RMSE values of the constructed models, which shows that the XGBoost model can be introduced as a new approach to solve environmental problems caused by blasting. Furthermore, $R^2$ values of 0.955 and 0.952, respectively, for the training and testing datasets of the XGBoost technique indicate that the capability of the proposed technique is slightly higher than that of the other developed models for PPV prediction.

Although XGBoost was a robust model for predicting blast-induced PPV in this study, it is still needed to be further studied for improving the accuracy level as well as the computational time. Also, a hybrid model based on XGBoost and another algorithm are also a good idea for future works.

**Acknowledgements** We would like to thank the Hanoi University of Mining and Geology (HUMG), Vietnam; Ministry of Education and Training of Vietnam (MOET); The Center for Mining, Electro-Mechanical research of HUMG.

## Compliance with ethical standards

**Conflict of interest** On behalf of all authors, I hereby attest that no conflict of interest exists in financial relationships, intellectual property, or any point related to publishing ethics.

## References

Ak H, Konuk A (2008) The effect of discontinuity frequency on ground vibrations produced from bench blasting: a case study. Soil Dynamics and Earthquake Engineering 28:686–694

Alnaqi AA, Moayedi H, Shahsavar A, Nguyen TK (2019) Prediction of energetic performance of a building integrated photovoltaic/thermal system thorough artificial neural network and hybrid particle swarm optimization models. Energy Convers Manag 183:137–148

Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 46:175–185

Ambraseys NR, Hendron AJ (1968) Dynamic behavior of rock masses. In: Stagg KG, Zienkiewicz OC (eds) Rock mechanics in engineering practices. Wiley, New York, pp 203–207

Breiman L (2001) Random forests. Mach Learn 45:5–32

Bui X-N, Nguyen H, Le H-A, Bui H-B, Do N-H (2019) Prediction of blast-induced air over-pressure in open-pit mine: assessment of different artificial intelligence. Tech Natl Resour Res. https://doi.org/10.1007/s11053-019-09461-0

Chandar KR, Sastry V, Hegde C (2017) A critical comparison of regression models and artificial neural networks to predict ground vibrations. Geotech Geol Eng 35:573–583

Chen T, He T (2015) Xgboost: extreme gradient boosting R package version 04-2

Chen G, Huang SL (2001) Analysis of ground vibrations caused by open pit production blasts–a case study. Fragblast 5(1–2):91–107

Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20:273–297

Coursen DL (1995) Method of reducing ground vibration from delay blasting. Google Patents

Davies B, Farmer I, Attewell P (1964) Ground vibration from shallow sub-surface blasts. Engineer 217(5644):553–559

Drucker H, Burges CJ, Kaufman L, Smola AJ, Vapnik V (1997) Support vector regression machines. In: Advances in neural information processing systems, pp 155–161

Duvall WI, Fogelson DE (1962) Review of criteria for estimating damage to residences from blasting vibrations. US Department of the Interior, Bureau of Mines

Faradonbeh RS, Monjezi M (2017) Prediction and minimization of blast-induced ground vibration using two robust metaheuristic algorithms. Eng Comput 33(4):835–851. https://doi.org/10.1007/s00366-017-0501-6

Faradonbeh RS, Armaghani DJ, Majid MA, Tahir MM, Murlidhar BR, Monjezi M, Wong H (2016) Prediction of ground vibration due to quarry blasting based on gene expression programming: a new model for peak particle velocity prediction. Int J Environ Sci Technol 13:1453–1464

Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

Friedman JH (2002) Stochastic gradient boosting. Comput Stat Data Anal 38:367–378

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). Ann Stat 28:337–407

Friedman J, Hastie T, Tibshirani R (2001) The elements of statistical learning vol 1. vol 10. Springer series in statistics New York, NY, USA

Gad EF, Wilson JL, Moore AJ, Richards AB (2005) Effects of mine blasting on residential structures Journal of performance of constructed facilities 19:222–228

Gao W, Dimitrov D, Abdo H (2018a) Tight independent set neighborhood union condition for fractional critical deleted graphs and ID deleted graphs. Discrete Continuous Dyn Syst S 123–144. https://doi.org/10.3934/dcdss.2019045

Gao W, Guirao JL, Basavanagoud B, Wu J (2018b) Partial multidividing ontology learning algorithm. Inf Sci 467:35–58

Gao W, Guirao JLG, Abdel-Aty M, Xi W (2018c) An independent set degree condition for fractional critical deleted graphs. Discrete Continuous Dyn Syst S 12:877–886. https://doi.org/10.3934/dcdss.2019058

Gao W, Wang W, Dimitrov D, Wang Y (2018d) Nano properties analysis via fourth multiplicative ABC indicator calculating. Arab J Chem 11:793–801

Gao W, Wu H, Siddiqui MK, Baig AQ (2018e) Study of biological networks using graph theory. Saudi J Biol Sci 25:1212–1219

Ghasemi E, Ataei M, Hashemolhosseini H (2013) Development of a fuzzy model for predicting ground vibration caused by rock blasting in surface mining. J Vib Control 19:755–770

Ghasemi E, Kalhori H, Bagherpour R (2016) A new hybrid ANFIS–PSO model for prediction of peak particle velocity due to bench blasting. Eng Comput 32:607–614

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

Hajihassani M, Armaghani DJ, Marto A, Mohamad ET (2015) Ground vibration prediction in quarry blasting through an artificial neural network optimized by imperialist competitive algorithm. Bull Eng Geol Environ 74:873–886

Hasanipanah M, Monjezi M, Shahnazar A, Armaghani DJ, Farazmand A (2015) Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement 75:289–297

Hasanipanah M, Armaghani DJ, Amnieh HB, Majid MZA, Tahir MM (2017a) Application of PSO to develop a powerful equation for prediction of flyrock due to blasting. Neural Comput Appl 28:1043–1050

Hasanipanah M, Faradonbeh RS, Amnieh HB, Armaghani DJ, Monjezi M (2017b) Forecasting blast-induced ground vibration developing a CART model. Eng Comput 33:307–316

Hasanipanah M, Golzar SB, Larki IA, Maryaki MY, Ghahremanians T (2017c) Estimation of blast-induced ground vibration through a soft computing framework. Eng Comput 33(4):951–959. https://doi.org/10.1007/s00366-017-0508-z

Hu C, Jain G, Zhang P, Schmidt C, Gomadam P, Gorka T (2014) Data-driven method based on particle swarm optimization and *k*-nearest neighbor regression for estimating capacity of lithium-ion battery. Appl Energy 129:49–55

Longjun D, Xibing L, Ming X, Qiyue L (2011) Comparisons of random forest and support vector machine for predicting blasting vibration characteristic parameters. Procedia Eng 26:1772–1781

Moayedi H, Nazir R (2018) Malaysian experiences of peat stabilization, state of the art. Geotech Geol Eng 36:1–11

Moayedi H, Rezaei A (2017) An artificial neural network approach for under-reamed piles subjected to uplift forces in dry sand. Neural Comput Appl. https://doi.org/10.1007/s00521-017-2990-z

Moayedi H, Raftari M, Sharifi A, Jusoh WAW, Rashid ASA (2019) Optimization of ANFIS with GA and PSO estimating α ratio in driven piles. Eng Comput. https://doi.org/10.1007/s00366-018-00694-w

Monjezi M, Bahrami A, Varjani AY, Sayadi AR (2011a) Prediction and controlling of flyrock in blasting operation using artificial neural network. Arab J Geosci 4:421–425

Monjezi M, Ghafurikalajahi M, Bahrami A (2011b) Prediction of blast-induced ground vibration using artificial neural networks. Tunn Undergr Space Technol 26:46–50

Monjezi M, Hasanipanah M, Khandelwal M (2013) Evaluation and prediction of blast-induced ground vibration at Shur River Dam, Iran, by artificial neural network. Neural Comput Appl 22:1637–1643

Nguyen H, Bui X-N (2018a) A comparison of artificial neural network and empirical technique for predicting blast-induced ground vibration in open-pit mine. In: Mining sciences and technology—XXVI, Mong Cai, Hanoi, Vietnam. Industry and trade of the socialist Republic of Vietnam, pp 177–182

Nguyen H, Bui X-N (2018b) Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. Natl Resour Res. https://doi.org/10.1007/s11053-018-9424-1

Nguyen H, Bui X-N, Bui H-B, Mai N-L (2018a) 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, Le T-Q, Do N-H, Hoa LTT (2018b) Evaluating and predicting blast-induced ground vibration in open-cast mine using ANN: a case study in Vietnam SN. Appl Sci 1:125. https://doi.org/10.1007/s42452-018-0136-2

Nguyen H, Bui X-N, Tran Q-H, Mai N-L (2019) 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. https://doi.org/10.1016/j.asoc.2019.01.042

Protodiakonov M, Koifman M, Chirkov S, Kuntish M, Tedder R (1964) Rock strength passports and methods for their determination. Nauka, Moscow

Roy PP (1991) Prediction and control of ground vibration due to blasting. Colliery Guard 239:215–219

Saadat M, Khandelwal M, Monjezi M (2014) An ANN-based approach to predict blast-induced ground vibration of Gol-E-Gohar iron ore mine. Iran J Rock Mech Geotech Eng 6:67–76

Sheykhi H, Bagherpour R, Ghasemi E, Kalhori H (2018) Forecasting ground vibration due to rock blasting: a hybrid intelligent approach using support vector regression and fuzzy C-means clustering. Eng Comput 34(2):357–365. https://doi.org/10.1007/s00366-017-0546-6

Standard I (1973) Criteria for safety and design of structures subjected to under ground blast ISI, IS-6922

Taheri K, Hasanipanah M, Golzar SB, Majid MZA (2017) A hybrid artificial bee colony algorithm-artificial neural network for forecasting the blast-produced ground vibration. Eng Comput 33:689–700

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 Vietnamse-unpublished). VINACOMIN, Vietnam

Zorlu K, Gokceoglu C, Ocakoglu F, Nefeslioglu H, Acikalin S (2008) Prediction of uniaxial compressive strength of sandstones using petrography-based models. Eng Geol 96:141–158