Multiscale and Multidisciplinary Modeling, Experiments and Design (2025) 8:138
https://doi.org/10.1007/s41939-024-00725-8

ORIGINAL PAPER

Estimation of powder factor in mine blasting: feasibility of tree-based predictive models

Danial Jahed Armaghani[1] · Mohammad Hayati[2] · Ehsan Momeni[2] · Mohammad Bagher Dowlatshahi[2] · Panagiotis G. Asteris[3]

Received: 16 April 2024 / Accepted: 28 December 2024

Abstract
Drilling and blasting is a process frequently used in rock-surface and deep excavation. For a proper drilling plan, accurate prediction of the amount of explosive material is essential to reduce the environmental effects associated with blasting operations. This study introduces a series of tree-based models, namely extreme gradient boosting machine (XGBoost), gradient boosting machine (GBM), adaptive boosting machine (AdaBoost), and random forest (RF), for predicting powder factor (PF) values obtained from blasting operations. The predictive models were constructed based on geomechanical characteristics at the blasting site, blasting pattern parameters, and rock material properties. These tree-based models were designed and tuned to minimize system error or maximize accuracy in predicting PF. Subsequently, the best model from each category was evaluated using various statistical metrics. It was found that the XGBoost model outperformed the other implemented techniques and exhibited outstanding potential in establishing the relationship between PF and input variables in the training set. Among the input parameters, hole diameter received the highest significance rating for predicting the system output, while the point load index had the least impact on the PF values.

Keywords Powder factor · Drilling and blasting · Geomechanical characteristics · Blasting pattern parameters · Tree-based models

1 Introduction

The blasting technique is a common method in surface and deep rock excavation. The blasting pattern parameters and their effective designs are always crucial in any rock excavation project. Optimal selection of the powder factor is significant for tunnel blasting from the economic point of view (Gokhale 2010). Additionally, the powder factor is an essential factor in the assessment of blast efficiency and its proper determination can affect the blasting results, which is fragmentation. The powder factor is defined as the ratio of the weight of the explosive to the volume of rock that must be blasted (Mohamed et al. 2015). An increase or decrease in the powder factor directly affects the rock fragmentation rate. An excessive amount of powder factor can lead to smaller sizes of fragmentation, which may not be of interest. On the other hand, a lower amount of powder factor can lead to larger rock fragments, which may require a secondary blast (Singh et al. 2016). Some researchers reported well-established experimental models for powder factor estimation (Kuznetsov 1973; Roy 2005). In addition, many researchers reported that powder factor is related to rock mass properties and/or rock quality index (Chakraborty et al. 1997; Adesida 2022).

Several researchers highlighted the importance of powder factor in relevant artificial intelligence (AI)-based studies.

✉ Panagiotis G. Asteris
panagiotisasteris@gmail.com

Danial Jahed Armaghani
danial.jahedarmaghani@uts.edu.au

Mohammad Hayati
hayati.m@lu.ac.ir

Ehsan Momeni
Momeni.E@lu.ac.ir

Mohammad Bagher Dowlatshahi
dowlatshahi.mb@lu.ac.ir

[1] School of Civil and Environmental Engineering, University of Technology Sydney, Sydney, NSW, Australia

[2] Faculty of Engineering, Lorestan University, Khorramabad, Iran

[3] Computational Mechanics Laboratory, School of Pedagogical and Technological Education, Heraklion, 14121 Athens, Greece

Published online: 15 January 2025

Rezaei et al. (2011) developed a fuzzy inference system for flyrock distance prediction. Powder factor, rock density, hole depth, to name a few were among the inputs of their intelligent predictive model. It is worth mentioning that they used nearly 490 sets of data for their model construction. Using 102 sets of data Zhou et al. (2020) developed a predictive model of particle peak velocity. They considered powder factor as one of the important input parameters. Lawal and Idris (2020) proposed an artificial neural network (ANN)-based predictive model of ground vibration. They identified powder factor as well as maximum charge per delay and distance from blasting as influential parameters.

Ghasemi et al. (2014) introduced an intelligent model for flyrock distance estimation. Powder factor, stemming, hole diameter, were among their proposed model. It should be mentioned that the coefficient of determination value of their model was 0.83. Nevertheless, in the aforementioned studies, powder factor was utilized as the influential input parameters for different types of AI-based predictive models. It is well-established that proper selection of input parameters can lead to a better AI-based predictive model (Armaghani et al. 2014). Miguel-García and Gómez-González (2019) implemented regression technique for powder factor estimation. They showed that powder factor is related to the point load index test, porosity and density. As expected, their results showed that powder factor is higher with rocks of low porosity, high density and high point load index test. Kahryman et al. (1998) showed that the powder factor can be estimated using comminution concept. Chakraborty et al. (1997) conducted a study on the prediction of powder factor in mixed-face condition. They showed that powder factor can be estimated using mean rock mass quality as well as mean tunnel rock blasting index. The correlation coefficient values for the former and the latter models were 0.82 and 0.9, respectively. Leu et al. (1998) investigated the applicability of ANN in forecasting the values of powder factor based on rock properties. The required data for his ANN-based model were taken from a railway tunnel constructed on the eastern coast of Taiwan. They used 328 sets of data for ANN model construction. According to their study, some parameters such as rock quality designation (RQD), rock mass rating (RMR), rock strength and tunnel orientation were set as model inputs. Their results (root mean square error, RMSE of 0.02983) showed that ANN is a feasible tool in predicting the powder factor values. Saemi and Gilani (2006) developed an ANN-based predictive model for powder factor values. Overall, 12 input parameters were used for their model construction. The input parameters of their model were related to the geological conditions and rock strength parameters as well as RMR. The coefficient of determination (R^2) value of 0.935 for testing data showed the workability of their constructed ANN-based predictive model of powder factor values. Jong and Lee (2004) showed the importance of geological condition on the powder factor for tunnel blasting.

As discussed, powder factor as one of the blasting pattern parameters plays a significant role in controlling and reducing risk associated with blasting operations. Therefore, it is important to propose a practical solution for predicting this parameter using AI techniques. These techniques have a proven capacity in solving civil and mining problems (Armaghani et al. 2022, 2021a, b; Momeni et al. 2021, 2020, 2014; Abdi et al. 2020; Ghorbani and Yagiz 2024; Bunawan et al. 2018; Huat et al. 2024; Fattahi et al. 2024; Ghanizadeh et al. 2022; Asteris et al. 2021; He et al. 2024; Momeni and Abdi 2022) and more specifically in predicting and minimizing blasting environmental issues (Hasanipanah et al. 2015; Zhou et al. 2024; Ding et al. 2024; Rezaei et al. 2023). These techniques with capacity of handling large datasets, identifying complex patterns, and providing accurate predictions are considered ideal for addressing the challenges associated with blasting operations.

Since there are only a few published papers on powder factor prediction using AI, it is challenging to determine the best category of AI techniques in this area. However, the aforementioned blasting-related studies can shed some light on identification of the proper AI method. Among the various AI techniques such as neural networks and vector regression models, to name a few, high accuracy has been achieved using tree-based approaches. These models have been widely utilized in technological sciences and engineering due to their ability to: (i) eliminate subjective uncertainty, (ii) handle large datasets with greater modeling speed, (iii) identify the most influential factors on output, and (iv) visualize nonlinear data (Chang and Chen 2005; Zhang et al. 2020a; Baker et al. 2006).

In addressing blasting-related environmental issues, tree-based models such as random forest (RF), AdaBoost, eXtreme gradient boosting (XGB), and CatBoost have also been successfully employed to solve problems related to ground vibration, back-break, flyrock, and air-overpressure (Han et al. 2020; Zhang et al. 2020; Yu et al. 2021; Barkhordari et al. 2022; Yari et al. 2023; Ye et al. 2021). Additionally, these techniques have successfully been proposed to address other issues in civil and geotechnical engineering (Bardhan et al. 2024; Asteris et al. 2024; Fakharian et al. 2024; Wang et al. 2021; Naderpour et al. 2024; Mohammed et al. 2021; Pham et al. 2020; He et al. 2021). Therefore, the authors decided to implement the mentioned tree-based models for predicting powder factor using an extensive experimental database.

The next section explores the basic concepts of the tree-based techniques utilized in this study. Section III describes the dataset in more detail, covering input sources and system outputs. Sections IV and V focus on the modeling approach and present the results and discussion. Finally, Sects. VI, VII

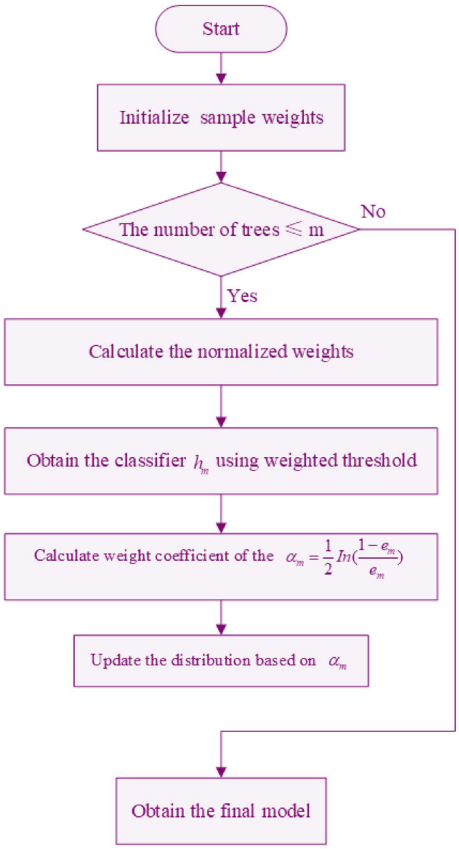

Fig. 1 RF general flowchart

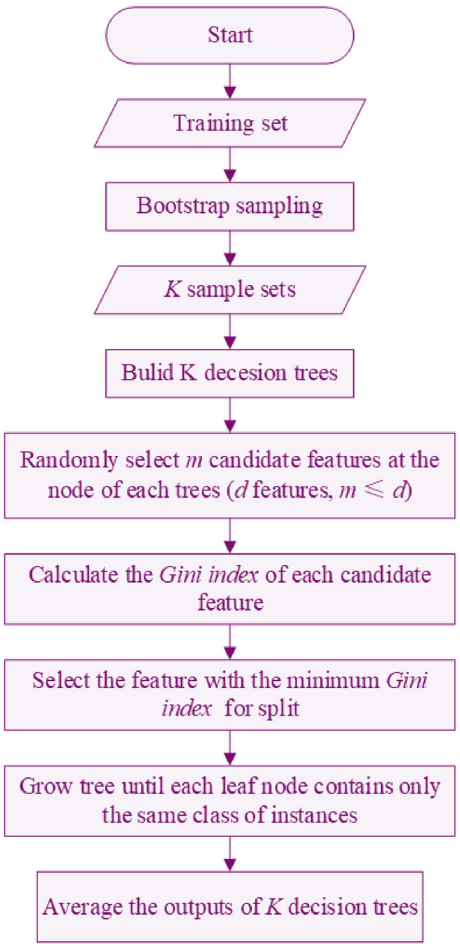

Fig. 2 AdaBoos general flowchart

and VIII provide an interpretation of the model, highlight its limitation and conclude with some remarks on proposing a practical tool for predicting the powder factor.

2 Principles of the used models

2.1 Random forest (RF)

Number equations consecutively with equation numbers in RF is an ensemble model that trains samples to solve regression problems using multiple regression trees (Breiman 2001; Asteris et al. 2022). The flowchart in Fig. 1 illustrates the process of developing an RF model for prediction purposes. The unpredictability inherent in RF is mostly manifested in two ways: (1) given a training set containing N samples, a subset of size N is picked using bootstrap sampling from all the training samples for the purpose of creating an individual tree. (2) A random subset of all characteristics is used to find the best split point for each node. RF performs well in machine learning tasks. It is capable of processing multidimensional data and boasts a quick training time, excellent anti-interference, and generalization capabilities. However, RF is prone to overfitting in several high-noise regression tasks. Also, when given data with many different qualities, attributes with a larger value division have a big impact on how well RF works.

2.2 Adaptive boosting machine (AdaBoost)

Using training data, Adaboost creates a succession of weak learners, which are then linearly combined into a strong model by modifying the weight distribution of samples (Freund and Schapire 1997). The flowchart for developing the AdaBoost model is shown in Fig. 2. When samples with substantial mistakes are used in the training process, the weights of those samples are improved, and the sample with the updated weight is used in the training process again. As a result, in the succeeding training phase, the next learner pays close attention to the training sample that contains big mistakes. When faced with noise, AdaBoost, on the other hand, concentrates on studying these complex data, resulting in a significant rise in the weight of the complex samples and, ultimately, the deterioration of the model.

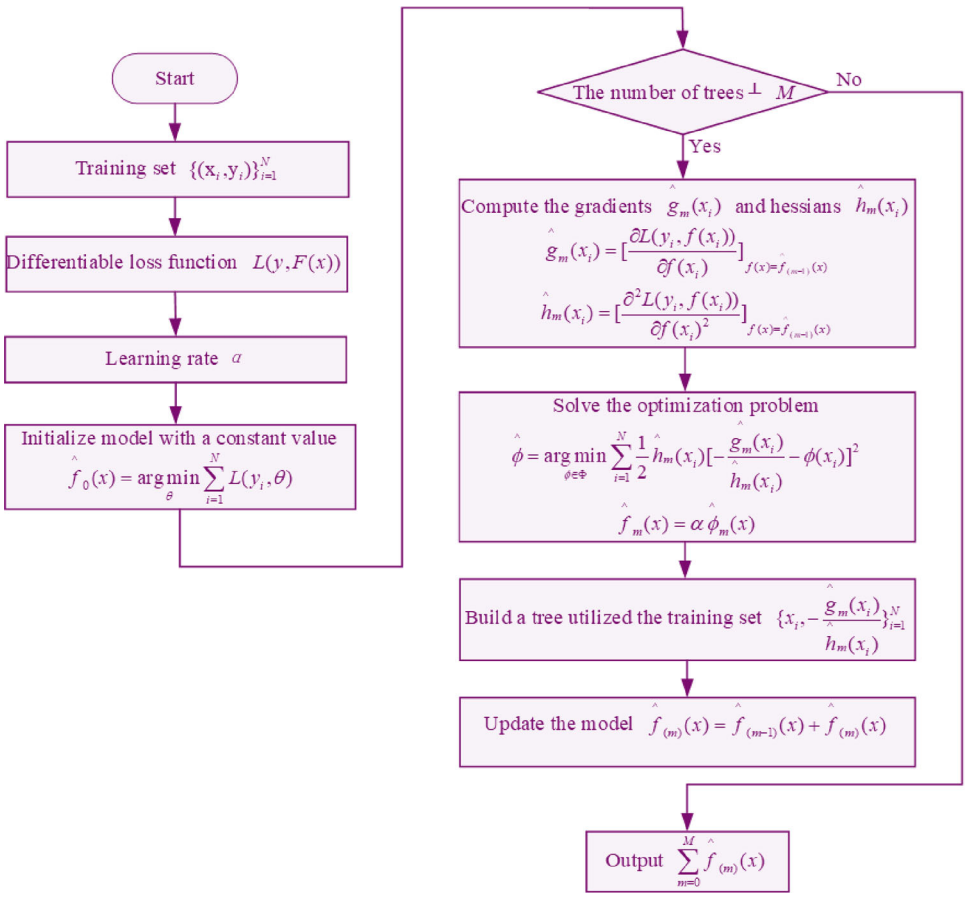

Fig. 3 Step-by-step flowchart of GBM in prediction problems

2.3 Gradient boosting machine (GBM)

The GBM algorithm (Friedman 2001) is an enhancement to the boosting method. When working with squared or exponential loss functions, the optimization method for boosting may be quite straightforward. On the other hand, it is difficult to optimize other loss functions while using boosting. The optimization method is straightforward when the loss function is a squared loss function or an exponential loss function. However, the optimization of generic loss functions is not straightforward. To solve this limitation, Freidman (2001) created the GBM, which uses the steepest descent as the approximation approach in order to minimize error. The value of the negative gradient of the loss function in the present model (Eq. 1) is used as an approximation to the value of the residual of the boosting tree in the regression problem to a regression tree in the current model.

The flowchart for GBM is depicted in Fig. 3. Initially, GBM calculates the constant value that minimizes the loss function, which is a single-rooted tree with no other nodes in the tree. A further calculation is made to determine how much of the negative gradient of the loss function is present in the current model, which is then utilized to estimate the residual. It is the residual in the case of a squared loss function, and it is an approximation of the residual in the case of a generic loss function. Later, the regression leaf node area (R_mj) is calculated in order to fit the approximations of residuals to the regression leaf node region. After that, the loss function is made as small as possible, the regression tree is changed, and the final model is made.

GBM is capable of dealing with a wide range of data types in a flexible manner and achieving high accuracy of prediction with just a minimal number of hyper-parameter adjustments. However, since the base learner in GBM has a serial connection, it is impossible to train data in parallel with the base learner. Furthermore, when dealing with high-dimensional sparse data sets, GBM has low performance.

$$r_{mi} = -\left(\frac{\partial L(y_i,\ f(x_i))}{\partial f(x_i)}\right)_{f(x)=f_{t-1}(x)} \quad (m = 1, 2, .., M,\ \ i = 1, 2, ..N) \tag{1}$$

$$f_0(x) = \arg\min_{c} \sum_{i=1}^{N} L(y_i,\ c) \tag{2}$$

$$f_M(x) = \sum_{m=1}^{M} \sum_{j=1}^{J} c_{mj} I(x \in R_{mj}) \tag{3}$$

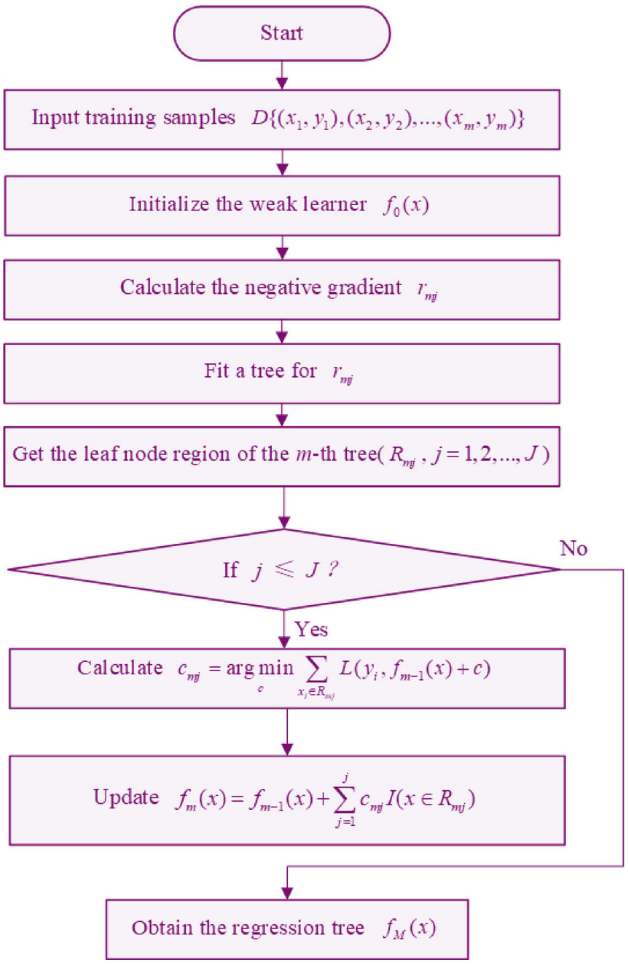

Fig. 4 General flowchart of XGBoost for proposing predictive models

In Eq. 1, $L(\cdot)$ represents the loss function and $f(x)$ represents the tree. In Eq. 2, c depicts the mean value of samples. In Eq. 3, $I(\cdot)$ equal to one indicates that the condition in the parenthesis is true, whereas 0 indicates that it is not.

2.4 Extreme gradient boosting machine (XGBoost)

XGBoost can do convex optimisation and uses second-order Taylor expansion on the loss function (Chen and Guestrin 2016). Meanwhile, distributed computing may help speed up the modeling process. Figure 4 shows the XGBosot flowchart. XGBoost beats GBM in the following ways: One for linear learners: XGBoost (2) XGBoost's goal function is a second-order Talor expansion of cost. Three changes are made to decrease over-fitting and calculation. Iteratively allocating learning rate reduces the weight of the tree and improves learning space.

3 Case study and the implemented database

The database which was utilized in this study was from the Seimare Dam. The dam is located on the Seimare river in Badreh county in Ilam province in Iran. The Seimare River is 417 km long and was formed by the joining of the Gharehsu and Gamasiab rivers. It is a concrete double-arc dam with the coordinates 33°17′ N, 47°12′ E. The dam is under construction in the southwestern part of the folded Zagros on the northern flank of the Ravandi anticline in the Kafenil valley. The bedrock is almost limestone (Shahbazan-Asmari formation). Figure 5 shows the location of the case study, and the profile of the water tunnel at Seimare Dam is tabulated in Table 1.

Overall, the results of 567 recorded drilling and tunnel blasting operations (including the parameters in Table 2) together with the relevant powder factors were collected and investigated. In order to select the acceptable blasting patterns among them, three basic conditions were determined as follows: (a) minimum progress, which is defined as 85% of the bore length at each blasting stage; (b) the back break at each stage in the permitted range, which is determined based on the tunnel cross-section and type of rock in the area and is an average of 200 mm and; (c) a lack of drilling fraction at each blasting stage. Out of 567 series of collected data, 294 series satisfied these conditions and were used for model developmental. The following parameters were chosen to predict the powder factor for blasting due to their influence on the powder factor as discussed earlier: RMR, hole diameter (HD), point load index (PLI) which represents the rock strength at small scale, and joint set slope (JSS). Table 2 shows a summary of the implemented data in this study.

The scatter distributions between any two variables are displayed in Fig. 6. As can be seen in these figures, there are positive and negative relationships between inputs and PF. HD and RMR have the highest correlations with model output (i.e., PF), which is in line with the other studies in the literature (Adesida 2022). In addition, a reverse relationship has been observed between JSS and PLI with PF. In order to have a better understanding regarding the used database in this research, 50 data samples of the entire data are presented in Appendix 1.

4 Modeling and results

The entire dataset, consisting of 294 samples, was divided into two portions: a training set, which comprised 80 percent, and a testing set, which comprised the remaining 20 percent. This division was chosen based on recommendations from the literature (Yari et al. 2023; Ye et al. 2021), aiming to identify the most effective combinations for model

Fig. 5 Location of Seimare Dam in Ilam, Iran

Table 1 Profile of Seimare Dam water tunnel

Main tunnel length	1290 m
Total drilling volume	190,900 m^3
Total concrete volume	34,729 m^3
Water tunnel number	1 course
Position	Left side of the river
Total length	1476 m
Internal diameter	11 m

development and evaluation. Additionally, we explored other combinations such as (70–30%), (90–10%), and (75–25%) on the same database; however, we found that the 80–20% split yielded the best results. The training components were used to construct ensemble models based on trees, while the testing components were employed to assess the models' capabilities. In this research, tree-based prediction models were constructed using Scikit-learn (Pedregosa et al. 2011), an open-source Python library.

One of the most critical factors influencing the performance of a tree-based model is the selection of optimal hyperparameters, such as the number of trees or boosting iterations. In this study, we utilized four ensemble tree models for PF prediction. These four ensemble tree models exhibit strong robustness and generalization abilities, often performing well even with default hyperparameters. We employed the Python library Scikit-learn to develop these four ensemble tree models. Table 3 presents the hyperparameters used for modeling tree-based models. In RF, the nodes were not considered pure until further extension of the tree. Each boosting model has a unique configuration for its hyperparameters (as detailed in Table 3). This variability in hyperparameter settings arose from the distinct training principles underlying each of the various models.

The input parameters (RMR, HD, PLI, and JSS) were standardized to ensure comparability, and outlier detection was conducted to improve data quality. The dataset was divided into 80% for training and 20% for testing based on literature-recommended practices, and the split was validated as optimal through experimentation with other ratios. Hyperparameters for each model were tuned using grid search to identify optimal configurations, balancing computational efficiency and accuracy. Cross-validation was employed to reduce overfitting and improve generalization. Model-specific techniques, such as loss function minimization in GBM and the application of second-order Taylor expansion in XGBoost, were implemented to enhance performance. Computational resources, including a high-performance system with parallel processing capabilities, were utilized to manage the intensive training processes efficiently.

Once the hyperparameters had been determined, we inputted the training samples into their respective models. We constructed four tree-based models to capture the extensive nonlinear interactions between PF and the input parameters. Ultimately, we successfully obtained four PF

Table 2 Specifications of inputs and output (powder factor) parameters

Parameter	Symbol	Unit	Category	Min	Max	Mean
Point load index test	PLI	MPa	Input	1.82	3.63	2.52
Joint set slope	JSS	deg	Input	25	85	60.796
Hole diameter	HD	mm	Input	32	45	38.27
Rock mass rating	RMR	–	Input	35	72	60.282
Powder factor	PF	kg/m^3	Output	1.94	3.47	2.622

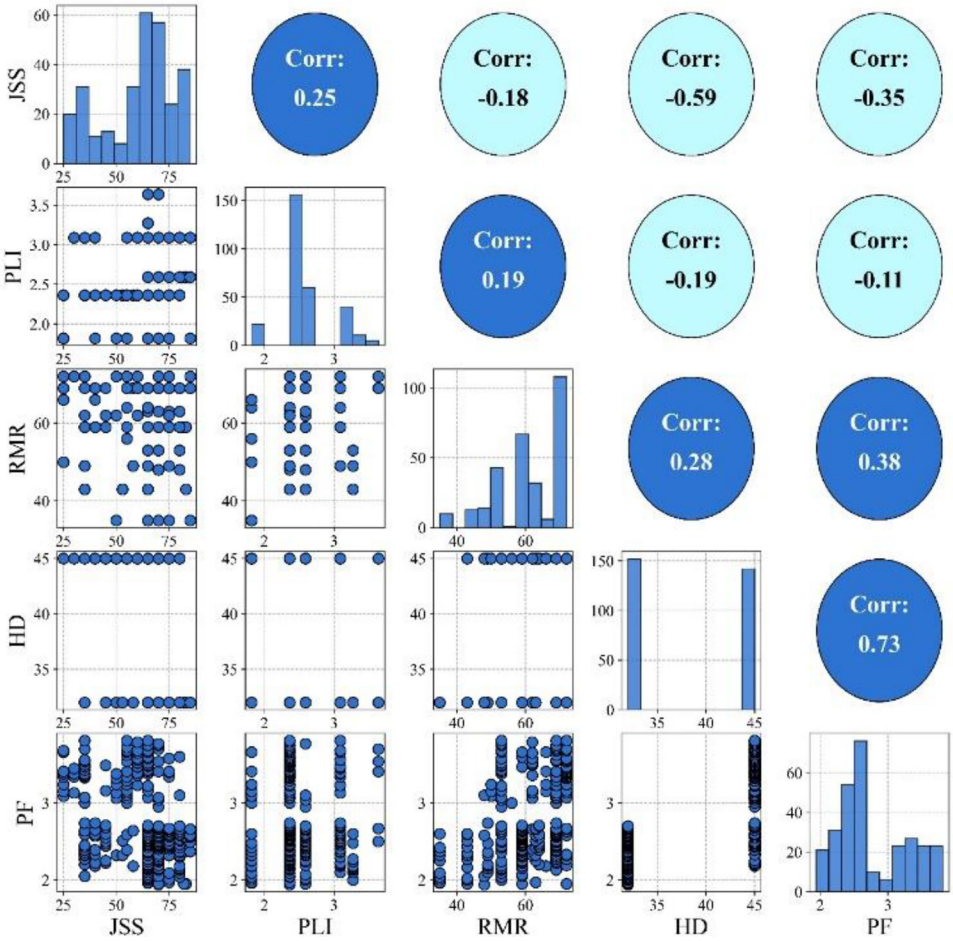

Fig. 6 Scatter plot of input parameters and PF together with correlation coefficients between variables

Table 3 The used hyper-parameters values in the developed tree-based models

Model	Parameters	Value
RF	Number of trees	100
	The minimum sample number of internal nodes for splitting	2
	The minimum sample number of leaf nodes	1
AdaBoost	The maximum number of trees	50
	Learning rate	1.0
	Loss function	Linear
GBM	The number of boosting iteration	100
	The minimum sample number of internal nodes for splitting	2
	The minimum sample number of leaf nodes	1
	Learning rate	0.1
	Maximum depth in each tree	3
XGBoost	The number of boosting iteration	100
	Learning rate	0.30
	Maximum depth in each tree	6

assessment models. We assessed the models' capabilities using the remaining testing set. To evaluate the performance of the tree-based models in regression tasks, we calculated R^2, mean absolute error (MAE), RMSE, variance accounted for (VAF), and the A-20 index. For the computation of MAE, RMSE, VAF, and the A-20 index, you can use the following formulas (Khotbehsara et al. 2014; Zeng et al. 2022):

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(m_i - y_i)^2}{\sum_{i=1}^{N}(m_i - \overline{m})^2} \tag{4}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|m_i - y_i| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - y_i)^2} \tag{6}$$

$$VAF = \left(1 - \frac{var(m_i - y_i)}{var(m_i)}\right) \times 100 \tag{7}$$

$$A - 20 = \frac{m20}{N} \tag{8}$$

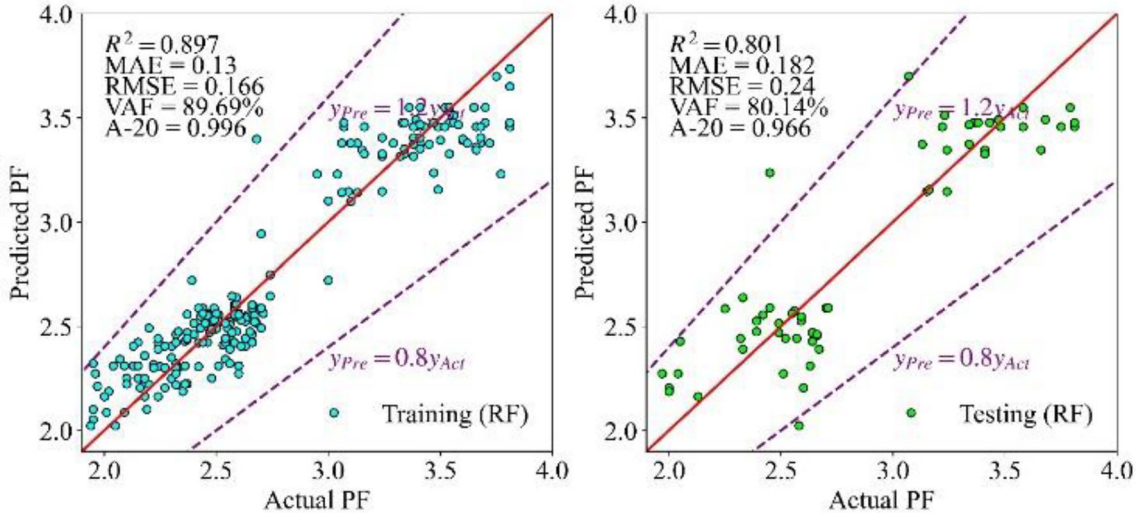

Fig. 7 Modeling results of RF to forecast PF

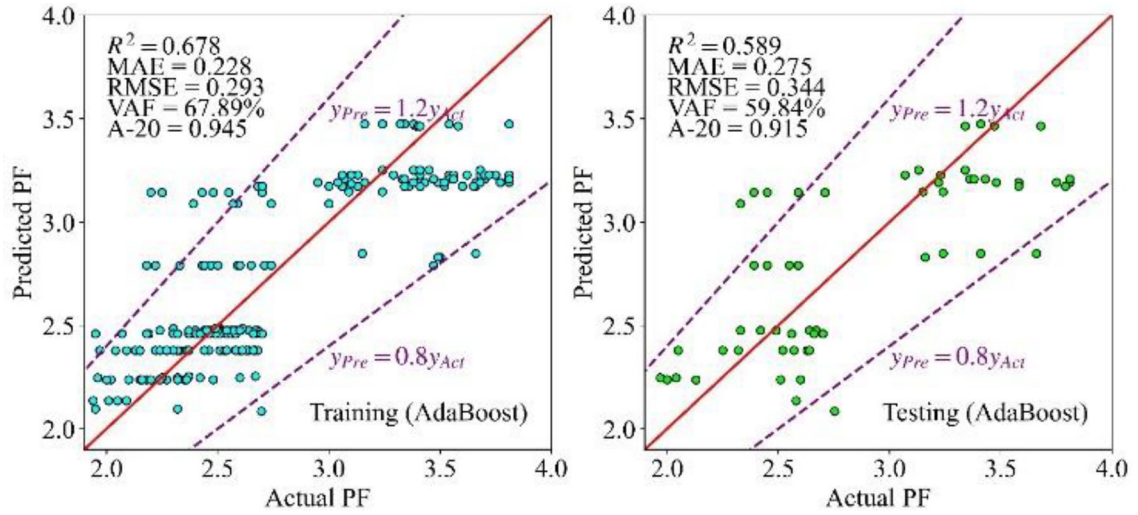

Fig. 8 Modeling results of AdaBoost to forecast PF

In these questions, m_i is the actual values of PF, y_i is the predicted values of PF, m- is the mean of actual PF values, and N is the number of data samples. In addition, is used to calculate the variance of the value in the bracket, and m20 is the number of samples for which the predicted PF values is the range of 0.8–1.2 times the actual PF values. R^2 equals 1, MAE equals 0, RMSE equals 0, VAF equals 100 (percent), and A-20 equals 1 in the situation in which the projected value precisely matches the actual value.

The results of the training and testing for four different tree-based models are depicted in Figs. 7, 8, 9 and 10. In these figures, the red line signifies that the estimated value equals the actual value, and points that align with it reflect a flawless forecast. The A-20 index is calculated by dividing the total number of points by the number of points inside the two purple lines with dots. Although both the training and testing stages of all tree-based models produced favorable results for all four models, AdaBoost did not achieve a high level of accuracy. Further explanations concerning these models and their predictive capabilities for PF will be provided in the next section.

5 Discussion

The evaluation indicators were discussed in depth in the previous part of the report. The Taylor diagram was developed as part of this research project so that the tree-based models could be compared to one another. In 2001, Taylor (2005) introduced the Taylor diagram, which was first used to examine the capacity of multiple weather models to replicate each other. As can be seen in Eq. 9, the correlation coefficient,

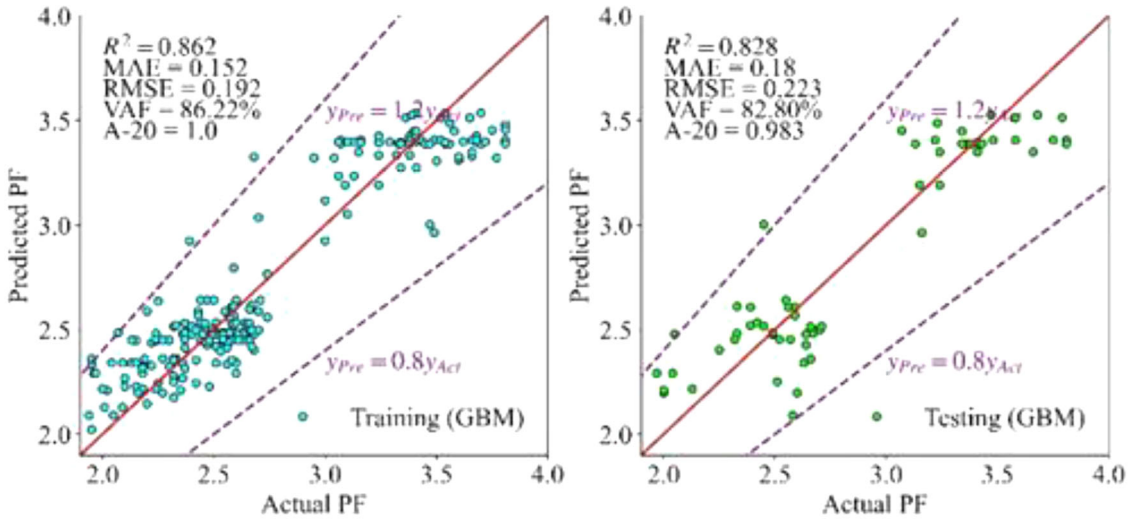

Fig. 9 Modeling results of GBM to forecast PF

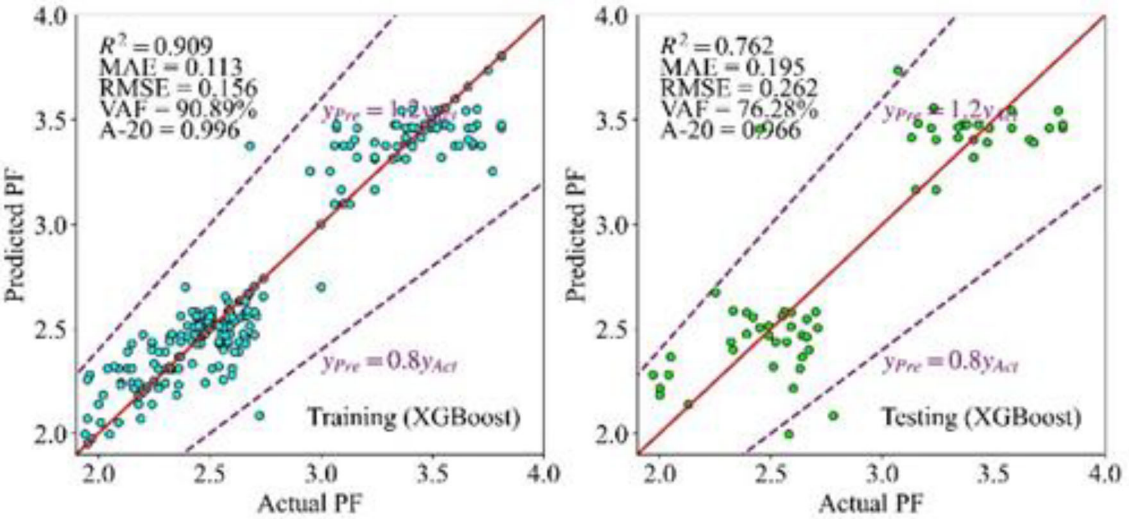

Fig. 10 Modeling results of XGBoost to forecast PF

centered RMSE, and standard deviation all have a cosine connection with one another. Ingeniously incorporating the correlation coefficient, centered RMSE, and standard deviation into a polar diagram is where the Taylor diagram gets its inspiration from.

$$E\prime^2 = \sigma_p^2 + \sigma_a^2 - 2\sigma_p\sigma_a R \tag{9}$$

In Eq. 9, is the centered RMSE between the predicted and actual variables, and are the variances of estimated and measured variables respectively, and is the correlation coefficient between estimated and measured variables.

The Taylor diagrams of the training outcomes are shown in Fig. 11, and the testing results are displayed in Fig. 12. The standard deviation is shown by the distance between the point representing the model and the origin point, and the correlation coefficient is represented by the ticks on the clockwise arc around the point representing the model. The real PF is represented by the point labelled "Reference" with the start shape on the x-axis, and the distance from each of the other points to the point labelled "Reference" reflects the centered RMSE (i.e., the red dotted line).

When working with Taylor diagrams, the placement of points on the graph may be used as a criterion for determining the capabilities of the relevant model. The models that are represented by points that are located closer to the 'Reference' point have capabilities that are more desirable. In accordance with this guiding concept, the XGBoost model exhibits the highest possible performance in the training set, but the GBM model is the most effective one in terms of its performance in the testing set. In the outcomes of the training

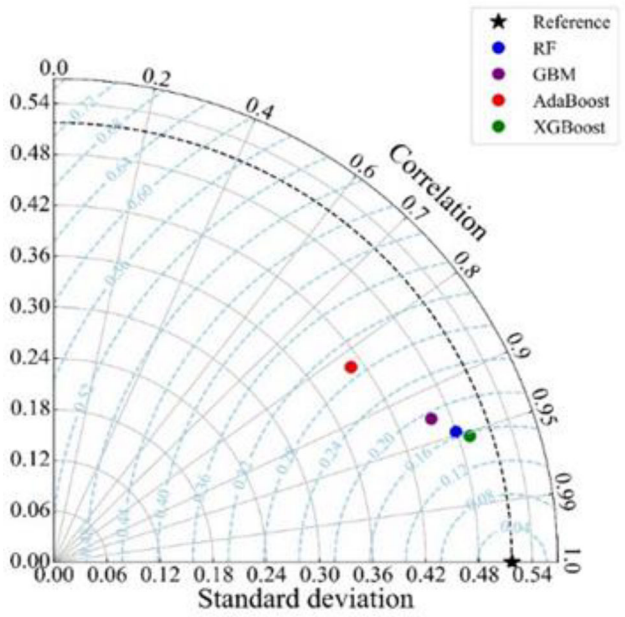

Fig. 11 The Taylor diagram of training results for four tree-based models

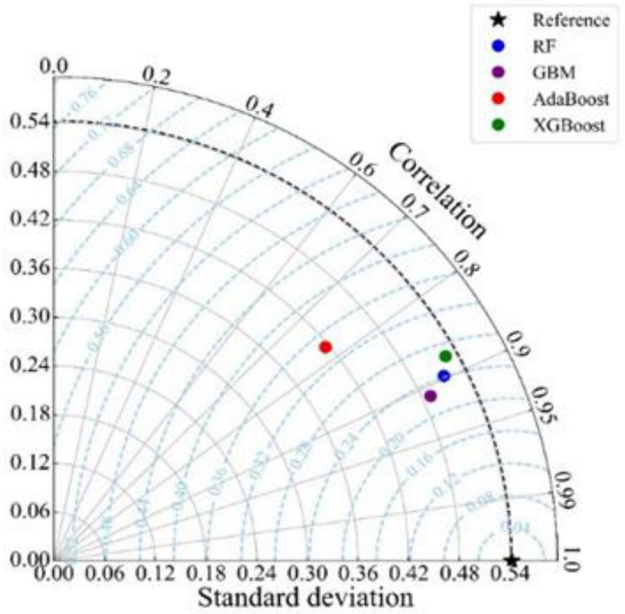

Fig. 12 The Taylor diagram of testing results for four tree-based models

and testing, there were four points that represented four tree-based models that were located within the range of the black dotted line. This indicated that the standard deviation of their projected PF values was less than the actual PF value. It was found that RF, GBM and XGBoost have very similar performances in terms of both training and testing, hence, there is a need to implement a system for selecting the best (more accurate) model in predicting PF. These tree-based models

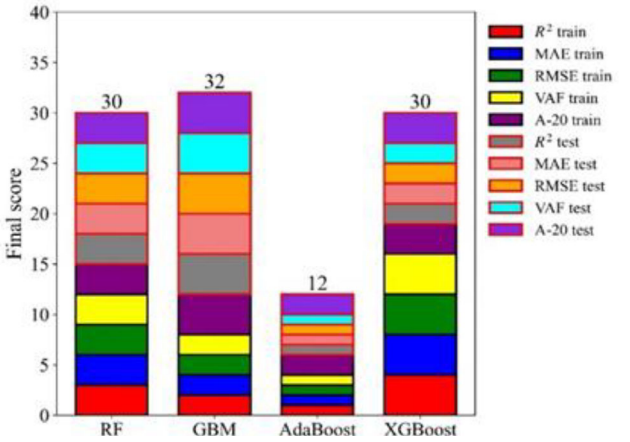

Fig. 13 The ranking results for GBM, RF, XGBoost, and AdaBoost in predicting PF

were graded using a ranking method that was developed by Zorlu et al. (2008), and the same system was utilized to rate the models. The higher score on the training set represents the greatest learning capacity, and the greater capability on the testing set implies the generalization ability and practical application of the model. The ranking mechanism takes into account both the performance of the model in the training set and in the testing set, and it is able to represent the full scope of the model's capabilities.

Figure 13 and Table 4 display the ranking results for the tree-based models in predicting PF. The four models were graded according to their performance in training and testing datasets using identical metrics. The ratings varied from 1 to 4, and a higher number is connected with better competence. The total score in the associated dataset is the sum of the scores of the model in five indicators. The final score is the sum of the overall scores of the model on the training and testing datasets. All models were compared and ranked according to the final score. Consequently, considering all performance indices and datasets, the model ranking was: GBM > RF = XGBoost > AdaBoost. Therefore, this study introduces the GBM model as the best tree-based technique for predicting PF.

The GBM model consistently achieved the highest scores across the evaluation metrics, demonstrating exceptional performance in both the training and testing phases. This superior performance highlights its strong generalization ability and robustness in handling nonlinear relationships between input parameters and the PF. The RF and XGBoost models followed closely, sharing comparable performance in most metrics, particularly in terms of R^2 and A-20 index values. Although AdaBoost showed relatively lower accuracy and generalization capabilities, it still provided reasonable predictions within acceptable limits.

Table 4 The ranking results for four tree-based models

Model	Dataset	R^2		MAE		RMSE		VAF (%)		A-20 index		Total score	Final score
		Value	Score	Value	Score	Value	Score	Value	Score	Value	Score		
XGBoost	Train	0.909	4	0.113	4	0.156	4	90.89	4	0.996	3	19	30
	Test	0.762	2	0.195	2	0.262	2	76.28	2	0.966	3	11	
GBM	Train	0.862	2	0.152	2	0.192	2	86.22	2	1	4	12	32
	Test	0.828	4	0.180	4	0.223	4	82.80	4	0.983	4	20	
AdaBoost	Train	0.678	1	0.228	1	0.293	1	67.89	1	0.945	2	6	12
	Test	0.589	1	0.275	1	0.344	1	59.84	1	0.915	2	6	
RF	Train	0.897	3	0.130	3	0.166	3	89.69	3	0.996	3	15	30
	Test	0.801	3	0.182	3	0.240	3	80.14	3	0.966	3	15	

The comparative analysis underscores the capability of ensemble learning techniques in predictive modeling, with GBM emerging as the most effective tree-based method for PF prediction. Its combination of high prediction accuracy, efficient use of input data, and adaptability to complex data structures positions it as a valuable tool for practical applications in blasting design and optimization. Therefore, this study introduces the GBM model as the best tree-based technique for predicting PF, offering a reliable solution for minimizing risks and improving blasting efficiency.

6 Model interpretation

The permutation importance technique was developed in order to assess the relative significance of the input parameters so that tree-based models could be explained. The significance of a given input parameter may be evaluated by calculating the degree to which the score, like, suffers when the parameter is eliminated (Pedregosa et al. 2011). The relevance score of the input parameters for each of the four models is shown in Fig. 14. As it can be seen, we have different descriptions for each model. HD shows the highest influence for all models, however, XGBoost introduces the highest importance (i.e., 0.88) among the others. After HD, RMR receives the highest importance in the PF data, followed by JSS and PII variables. These types of analyses can help designers to identify the most effective parameters, and through them they can have an optimum design and the best blasting pattern parameters in order to minimize the associated risks.

7 Limitations and future directions

The number of data samples used for proposing machine learning solutions in the geotechnical and mining areas is always a concern. In the current study, 294 data samples were utilized, which is significantly higher than the average number of data samples used in blasting environmental studies, typically around 100 data samples. Therefore, the proposed models are considered to be more generalized compared to those found in the existing literature. However, it is worth noting that the number of collected data samples can be increased to 400 or even 500, allowing for a broader coverage of input and output parameters. Furthermore, future studies can employ separate validation data to ensure the accurate assessment of the proposed models.

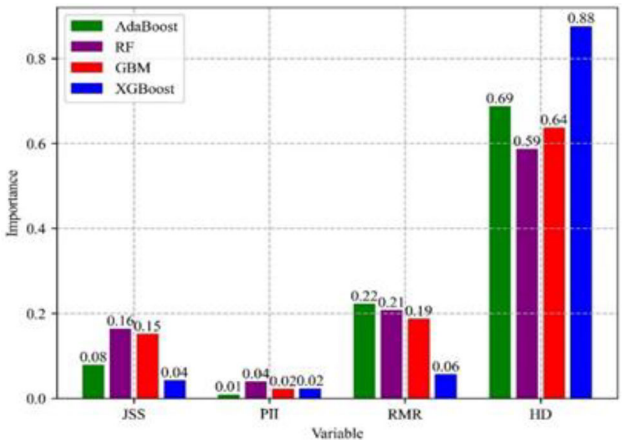

Fig. 14 The importance score of three input variables used in this study

Future studies should consider conducting a sensitivity analysis to quantify the impact of individual input parameters on the predicted powder factor and to further validate the robustness of the developed models. To enhance the practicality and appeal of ML techniques, future research can be planned and conducted using theory-guided machine learning (TGML). TGML is a unique paradigm aimed at enhancing the utility of ML models in advancing scientific discovery by harnessing a wealth of scientific data. Its fundamental objective is to make scientific consistency an essential

element of generalizable model development. The aim is to leverage well-established theories or empirical equations published in the field of blasting design to construct a comprehensive database capable of accommodating a wide range of influential factors. Through the application of distinctive and advanced ML techniques, this collected information can then be employed to create a highly accurate and generalizable model.

8 Conclusions

This study examined the use and capacity of tree-based techniques in predicting PF in mine blasting operations. The following conclusions can be drawn from the results of this research study:

1. Among the models tested, GBM demonstrated the best overall performance with R^2 values of 0.862 (training) and 0.828 (testing), along with the lowest RMSE and MAE values in the testing phase. Its ability to generalize across datasets makes it the most reliable for PF prediction.
2. The HD was the most critical factor influencing PF predictions, followed by RMR. JSS and PLI showed less importance but still contributed to the model's accuracy. These findings align with existing literature and provide actionable insights for optimizing blasting designs and future projects.
3. The comprehensive evaluation using five metrics (R^2, MAE, RMSE, VAF, and A-20 Index) ensured a rigorous comparison of models, providing confidence in the selection of GBM as the most effective technique.

Appendix 1 50 data samples used in this study

Data no.	JSS (°)	PLI (MPa)	RMR	HD (mm)	PF kg/m^3
1	70	2.36	48	32	1.94
2	65	2.59	59	32	2.07
3	75	1.82	35	32	2.09
4	80	2.36	53	32	2.09
5	65	3.09	49	32	2.1
6	80	3.09	59	32	2.18
7	40	1.82	66	45	2.18
8	70	2.59	72	32	2.19
9	35	3.09	49	32	2.2
10	65	2.36	64	45	2.2
11	70	2.36	69	32	2.27
12	35	2.36	72	32	2.27
13	80	3.09	59	32	2.3
14	70	2.36	69	32	2.3
15	45	2.36	59	45	2.44
16	70	3.09	59	32	2.45
17	35	2.36	59	45	2.45
18	65	3.09	64	45	2.45
19	75	2.59	59	32	2.46
20	85	3.09	69	32	2.49
21	65	3.64	69	32	2.5
22	45	2.36	59	45	2.5
23	53	2.36	43	32	2.51
24	58	2.36	69	32	2.64
25	40	2.36	59	45	2.65
26	70	2.59	53	32	2.66
27	70	2.59	53	32	2.66
28	60	3.09	72	45	3.39
29	75	2.36	72	45	3.4
30	65	2.36	53	45	3.41
31	25	1.82	66	45	3.41
32	35	3.09	69	45	2.59
33	40	2.36	69	45	2.59
34	65	1.82	35	32	2.6
35	65	3.27	43	32	2.6
36	55	2.36	72	45	3.24
37	55	2.36	69	45	3.29
38	50	2.36	72	45	3.32
39	35	2.36	72	45	3.33
40	25	2.36	69	45	3.34
41	30	3.09	72	45	3.34
42	35	2.36	72	45	3.7
43	55	2.36	69	45	3.72
44	65	2.36	53	45	3.75
45	70	2.36	69	45	3.75
46	70	2.59	53	45	3.77
47	85	3.09	69	32	2.57

Data no.	JSS (°)	PLI (MPa)	RMR	HD (mm)	PF kg/m^3
48	40	3.09	69	45	2.57
49	70	2.36	48	32	2.58
50	70	2.36	59	32	2.58

Author contributions D.J.A., M.H., E.M., M.B.D. and P.G.A. Conceptualization. D.J.A., M.H. and E.M. methodology. D.J.A., M.H. and E.M. software. D.J.A., M.H., E.M. and P.G.A. validation. D.J.A., M.H. and E.M. formal analysis. M.H. and E.M. resources. M.H. and E.M. data curation. D.J.A., M.H., E.M., M.B.D. and P.G.A. writing—original draft preparation. D.J.A., M.H., E.M., M.B.D. and P.G.A. writing—review and editing. D.J.A., M.H. and E.M. visualization. P.G.A. supervision. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by HEAL-Link Greece.

Data availability statement The data will be available upon a reasonable request. No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abdi Y, Momeni E, Khabir RR (2020) A reliable PSO-based ANN approach for predicting unconfined compressive strength of sandstones. Open Constr Build Technol J 14(1):237–249

Adesida PA (2022) Powder factor prediction in blasting operation using rock geo-mechanical properties and geometric parameters. Int J Min Geo-Eng 56(1):25–32

Armaghani DJ, Hajihassani M, Mohamad ET, Marto A, Noorani SA (2014) Blasting-induced flyrock and ground vibration prediction through an expert artificial neural network based on particle swarm optimization. Arab J Geosci 7(12):5383–5396. https://doi.org/10.1007/s12517-013-1174-0

Armaghani DJ, Harandizadeh H, Momeni E (2021a) Load carrying capacity assessment of thin-walled foundations: an ANFIS–PNN model optimized by genetic algorithm. Eng Comput. https://doi.org/10.1007/s00366-021-01380-0

Armaghani DJ, Harandizadeh H, Momeni E, Maizir H, Zhou J (2021b) An optimized system of GMDH-ANFIS predictive model by ICA for estimating pile bearing capacity. Artif Intell Rev. https://doi.org/10.1007/s10462-021-10065-5

Armaghani DJ, Harandizadeh H, Momeni E, Maizir H, Zhou J (2022) An optimized system of GMDH-ANFIS predictive model by ICA for estimating pile bearing capacity. Artif Intell Rev 55(3):2313–2350. https://doi.org/10.1007/s10462-021-10065-5

Asteris PG et al (2022) Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques. Constr Build Mater 322:126500. https://doi.org/10.1016/j.conbuildmat.2022.126500

Asteris PG et al (2022) Slope stability classification under seismic conditions using several tree-based intelligent techniques. Appl Sci 12(3):1753

Asteris PG, Gavriilaki E, Kampaktsis PN, Gandomi AH, Armaghani DJ, Tsoukalas MZ, Gkaliagkousi E (2024) Revealing the nature of cardiovascular disease using DERGA, a novel data ensemble refinement greedy algorithm. Int J Cardiol 412:132339

Baker C, Lawrence R, Montagne C, Patten D (2006) Mapping wetlands and riparian areas using Landsat ETM+ imagery and decision-tree-based models. Wetlands 26(2):465–474

Bardhan A, Ozcan NT, Asteris PG, Gokceoglu C (2024) Hybrid ensemble paradigms for estimating tunnel boring machine penetration rate for the 10-km long Bahce-Nurdagi twin tunnels. Eng Appl Artif Intell 136:108997

Barkhordari MS, Armaghani DJ, Fakharian P (2022) Ensemble machine learning models for prediction of flyrock due to quarry blasting. Int J Environ Sci Technol 19(9):8661–8676

Breiman L (2001) Random forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324

Bunawan AR, Momeni E, Armaghani DJ, Rashid ASA (2018) Experimental and intelligent techniques to estimate bearing capacity of cohesive soft soils reinforced with soil-cement columns. Measurement 124:529–538

Chakraborty AK, Jethwa JL, Dhar BB (1997) Predicting powder factor in mixed-face condition: development of a correlation based on investigations in a tunnel through basaltic flows. Eng Geol 47(1–2):31–41

Chang L-Y, Chen W-C (2005) Data mining of tree-based models to analyze freeway accident frequency. J Saf Res 36(4):365–375

Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: Proceedings of the 22Nd ACM sigkdd international conference on knowledge discovery and data mining, vol. 10, no. 2939672.2939785. ACM, New York, NY, USA, pp 785–794

de Miguel-García E, Gómez-González JF (2019) A new methodology to estimate the powder factor of explosives considering the different lithologies of volcanic lands: a case study from the island of Tenerife, Spain. Tunn Undergr Space Technol 91:103023

Ding X, Hasanipanah M, Ulrikh DV (2024) Hybrid metaheuristic optimization algorithms with least-squares support vector machine and boosted regression tree models for prediction of air-blast due to mine blasting. Nat Resour Res 33(3):1349–1363

Fakharian P, Nouri Y, Ghanizadeh AR, Jahanshahi FS, Naderpour H, Kheyroddin A (2024) Bond strength prediction of externally bonded reinforcement on groove method (EBROG) using MARS-POA. Compos Struct 349:118532

Fattahi H et al (2024) Optimizing pile bearing capacity prediction: insights from dynamic testing and smart algorithms in geotechnical engineering. Measurement 230:114563

Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. J Comput Syst Sci 55(1):119–139

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

Ghanizadeh AR, Ghanizadeh A, Asteris PG, Fakharian P, Armaghani DJ (2023) Developing bearing capacity model for geogrid-reinforced stone columns improved soft clay utilizing MARS-EBS hybrid

method. Transport Geotech 38:100906. https://doi.org/10.1016/j.trgeo.2022.100906
Ghasemi E, Amini H, Ataei M, Khalokakaei R (2014) Application of artificial intelligence techniques for predicting the flyrock distance caused by blasting operation. Arab J Geosci 7(1):193–202. https://doi.org/10.1007/s12517-012-0703-6
Ghorbani E, Yagiz S (2024) Estimating the penetration rate of tunnel boring machines via gradient boosting algorithms. Eng Appl Artif Intell 136:108985
Gokhale BV (2010) Rotary drilling and blasting in large surface mines. CRC Press, Boca Raton
Han H, Jahed Armaghani D, Tarinejad R, Zhou J, Tahir MM (2020) Random forest and bayesian network techniques for probabilistic prediction of flyrock induced by blasting in quarry sites. Nat Resour Res 29:655–667
Hasanipanah M, Monjezi M, Shahnazar A, Jahed Armaghani D, Farazmand A (2015) Feasibility of indirect determination of blast induced ground vibration based on support vector machine. Measurement (London) 75:289–297. https://doi.org/10.1016/j.measurement.2015.07.019
He Z, Armaghani DJ, Masoumnezhad M, Khandelwal M, Zhou J, Murlidhar BR (2021) A combination of expert-based system and advanced decision-tree algorithms to predict air-overpressure resulting from quarry blasting. Nat Resour Res 30(2):1889–1903
He B, Armaghani DJ, Lai SH, He X, Asteris PG, Sheng D (2024) A deep dive into tunnel blasting studies between 2000 and 2023—a systematic review. Tunn Undergr Space Technol 147:105727
Huat CY, Armaghani DJ, Lai SH, Motaghedi H, Asteris PG, Fakharian P (2024) Analyzing surface settlement factors in single and twin tunnels: a review study. J Eng Res. https://doi.org/10.1016/j.jer.2024.05.009
Jong YH, Lee CI (2004) Influence of geological conditions on the powder factor for tunnel blasting. Int J Rock Mech Min Sci. https://doi.org/10.1016/j.ijrmms.2004.03.095
Kahryman A, Sül ÖL, Demyrcy A (1998) Estimating powder factor from comminution concept. Miner Resour Eng 7(02):69–77
Khotbehsara MM, Zadshir M, Miyandehi BM, Mohseni E, Rahmannia S, Fathi S (2014) Rheological, mechanical and durability properties of self-compacting mortar containing nano-TiO2 and fly ash. J Am Sci 10(11):222–228
Kuznetsov VM (1973) The mean diameter of the fragments formed by blasting rock. Sov Min 9(2):144–148
Lawal AI, Idris MA (2020) An artificial neural network-based mathematical model for the prediction of blast-induced ground vibrations. Int J Environ Stud 77(2):318–334. https://doi.org/10.1080/00207233.2019.1662186
Leu S-S, Lin S-F, Chen C-K, Wang S-W (1998) Analysis of powder factors for tunnel blasting using neural networks. Fragblast 2(4):433–448
Mohamed F, Hafsaoui A, Talhi K, Menacer K (2015) Study of the powder factor in surface bench blasting. Procedia Earth Planet Sci 15:892–899
Mohammed AS, Asteris PG, Koopialipoor M, Alexakis DE, Lemonis ME, Armaghani DJ (2021) Stacking ensemble tree models to predict energy performance in residential buildings. Sustainability 13(15):8298
Momeni E, Nazir R, Armaghani DJ, Maizir H (2014) Prediction of pile bearing capacity using a hybrid genetic algorithm-based ANN. Measurement 57:122–131
Momeni E, Dowlatshahi MB, Omidinasab F, Maizir H, Armaghani DJ (2020) Gaussian process regression technique to estimate the pile bearing capacity. Arab J Sci Eng 45(10):8255–8267. https://doi.org/10.1007/s13369-020-04683-4
Momeni E, Poormoosavian M, Tehrani HS, Fakher A (2021) Reliability analysis and risk assessment of deep excavations using random-set finite element method and event tree technique. Transport Geotech 29:100560
Momeni E, Abdi Y (2022) Application of group method of data handling (GMDH) technique in predicting UCS of limestones. Iran J Eng Geol 15(2):173–176
Naderpour H, SoltaniMatin A, Kheyroddin A, Fakharian P, Ezami N (2024) Optimizing seismic performance of tuned mass dampers at various levels in reinforced concrete buildings. Buildings 14(8):2443
Pedregosa F et al (2011) Scikit-learn: machine learning in Python. J Mach LeArn Res 12:2825–2830
Pham BT et al (2020) A novel approach for classification of soils based on laboratory tests using Adaboost, tree and ANN modeling. Transport Geotech. https://doi.org/10.1016/j.trgeo.2020.100508
Rezaei M, Monjezi M, Yazdian Varjani A (2011) Development of a fuzzy model to predict flyrock in surface mining. Saf Sci 49(2):298–305. https://doi.org/10.1016/j.ssci.2010.09.004
Rezaei M, Monjezi M, Matinpoor F, Bolbanabad SM, Habibi H (2023) Simulation of induced flyrock due to open-pit blasting using the PCA-CART hybrid modeling. Simul Model Pract Theory 129:102844
Roy PP (2005) Rock blasting: effects and operations. CRC Press, Boca Raton
Saemi M, Gilani SO (2006) Determining of optimal powder factor in tunnel blasting using neural network systems. Iran J Min Eng 1(1):49–55
Singh PK, Roy MP, Paswan RK, Sarim MD, Kumar S, Jha RR (2016) Rock fragmentation control in opencast blasting. J Rock Mech Geotech Eng 8(2):225–237
Taylor KE (2005) Taylor diagram primer, working paper, January, pp 1–4
Wang S, Zhou J, Li C, Armaghani DJ, Li X, Mitri HS (2021) Rockburst prediction in hard rock mines developing bagging and boosting tree-based ensemble techniques. J Cent South Univ 28(2):527–542
Yari M, Armaghani DJ, Maraveas C, Ejlali AN, Mohamad ET, Asteris PG (2023) Several tree-based solutions for predicting flyrock distance due to mine blasting. Appl Sci 13(3):1345
Ye J, Koopialipoor M, Zhou J, Armaghani DJ, He X (2021) A novel combination of tree-based modeling and Monte Carlo simulation for assessing risk levels of flyrock induced by mine blasting. Nat Resour Res 30:225–243
Yu Q, Monjezi M, Mohammed AS, Dehghani H, Armaghani DJ, Ulrikh DV (2021) Optimized support vector machines combined with evolutionary random forest for prediction of back-break caused by blasting operation. Sustainability 13(22):12797
Zhang W et al (2020a) State-of-the-art review of soft computing applications in underground excavations. Geosci Front 11(4):1095–1106. https://doi.org/10.1016/j.gsf.2019.12.003
Zhang H, Zhou J, Jahed Armaghani D, Tahir MM, Pham BT, Van Huynh V (2020b) A combination of feature selection and random forest techniques to solve a problem related to blast-induced ground vibration. Appl Sci 10(3):869
Zhou J, Asteris PG, Armaghani DJ, Pham BT (2020) Prediction of ground vibration induced by blasting operations through the use of the Bayesian Network and random forest models. Soil Dyn Earthq Eng 139:106390. https://doi.org/10.1016/j.soildyn.2020.106390
Zhou J et al (2024) Advanced machine learning methods for prediction of blast-induced flyrock using hybrid SVR methods. CMES Comput Model Eng Sci 140(2):1595–1617
Zorlu K, Gokceoglu C, Ocakoglu F, Nefeslioglu HA, Acikalin S (2008) Prediction of uniaxial compressive strength of sandstones using petrography-based models. Eng Geol 96(3–4):141–158. https://doi.org/10.1016/j.enggeo.2007.10.009

Zeng J, Roy B, Kumar D et al (2022) Proposing several hybrid PSO-extreme learning machine techniques to predict TBM performance. Eng Comput 38(Suppl 5):3811–3827